BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
Department: “Information Processes and Decision Support Systems”

Kristel Gjergj Vula

Transformer-Based Detection and Epidemic-Model Analysis of Fake-News Dynamics on Social Networks

Dissertation Thesis

for acquiring Educational and Scientific Degree “DOCTOR”

In Professional field 4.6. “Informatics and Computer Science”

in Doctoral Program: “Informatics”

Scientific Supervisor: Prof. Dr. Vassil Georgiev Guliashki

Sofia, 2026

Table of Contents

Dedication

I would like to dedicate my thesis to my beloved husband and my precious daughter.

Together, they have been my strength, my inspiration, and my greatest blessing.

Acknowledgments

I consider myself truly fortunate to have been surrounded by great and supportive people along the way of my PhD journey.

I would like to express my deepest gratitude to my scientific supervisor, Prof. Dr. Vassil Guliashki, for his professional guidance, valuable recommendations, and continuous support throughout the preparation of this dissertation. His expertise, constructive criticism, and encouragement have significantly contributed to the development of this research work. I am sincerely thankful for his trust, patience, and academic mentorship during the entire period of my doctoral studies.

I would also like to express my sincere gratitude to the Bulgarian Academy of Sciences and to the Institute of Information and Communication Technologies for providing me with the opportunity and academic environment to pursue my doctoral studies. I am deeply thankful to all the professors, researchers, administrative staff, and colleagues who were kind, supportive, and helpful throughout this journey.

A very special thanks goes to my dear friend Aida, who encouraged me to continue my doctoral studies at IICT-BAS and believed in me from the very beginning. Her support, advice, and confidence in my abilities were a constant source of motivation during this important stage of my academic and personal life. I am truly grateful for her friendship and for the role she played in inspiring me to take this step.

A special appreciation for their support is for my dearest colleagues, and friends from Alexander Moisiu University of Durres, for long talks about this challenge, support in my daily workdays, and their friendship.

Lastly, I would like to send my deepest appreciation to my lifelong love, Shpend, whose persistent emotional support has made this day possible, and our daughter, Solin, who has been part of this journey from before her birth.

(Page intentionally left blank)

Dissertation Structure

Vula, K. Transformer-Based Detection and Epidemic-Model Analysis of Fake-News Dynamics on Social Networks. Scientific supervisor: Prof. Dr. Vassil Georgiev Guliashki. Department Information Processes and Decision Support Systems. Doctoral Program: "Informatics". Sofia, 2026. The dissertation with 210 pages, consists of an introduction, 3 chapters, a concluding section summarizing the results obtained and future research directions, a separate thesis contributions section, a list of 5 publications on the dissertation, declaration of originality of the thesis, and a bibliography of 191 references.

Keywords

Fake-news detection; Transformer models; Temporal networks; SIR/SEIZ epidemic modelling; Approximate Bayesian Computation; Intervention analysis.

Roadmap of the dissertation

The thesis is structured in the following parts:

Introduction

The Introduction presents the motivation for writing the PhD thesis, the social and technological relevance of fake news, and the need to analyse false information through Transformer-based detection, temporal diffusion modelling, epidemic-style propagation and ABC/ABC-SMC uncertainty-aware inference.

Chapter 1

This chapter presents a detailed and analytical overview of fake-news definitions, classical and Transformer-based detection methods, ensemble and hybrid approaches, temporal-network diffusion, SIR/SEIZ epidemic models, ABC inference, and the research gaps that motivate the proposed thesis.

Chapter 2

This chapter provides detailed information on the proposed Transformer-ABC-Epidemic framework, including dataset selection, preprocessing, binary label harmonization, temporal cascade construction, SIR/SEIZ modelling, ABC/ABC-SMC posterior inference, forecasting, and the baseline posterior diffusion-risk functional.

Chapter 3

This chapter outlines the experimental implementation and validation of the developed framework using content-based fake-news datasets and diffusion-oriented social-network cascades, reporting detection results, temporal cascade analysis, SIR/SEIZ comparison, ABC/ABC-SMC calibration, forecasting, and intervention scenarios

Conclusion

The conclusion summarizes the results obtained from the developed framework and experimental analysis, together with the study limitations, future research directions, thesis contributions, related publications, citations and supplementary materials.

List of Abbreviations

Abbreviation	Meaning
ABC	Approximate Bayesian Computation
ABC-SMC	Approximate Bayesian Computation - Sequential Monte Carlo
BERT	Bidirectional Encoder Representations from Transformers
BigBird	Sparse-attention Transformer model for long sequences
CoAID	COVID-19 Healthcare Misinformation Dataset
FRN	Fake and Real News dataset
GNN	Graph Neural Network
ISOT	ISOT Fake News Dataset
LLM	Large Language Model
mBERT	Multilingual BERT
NLP	Natural Language Processing
RoBERTa	Robustly Optimized BERT Pretraining Approach
SBERT	Sentence-BERT
SEIZ	Susceptible-Exposed-Infected-Skeptic model
SIR	Susceptible-Infected-Recovered model
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
XLM-R	XLM-RoBERTa multilingual Transformer model

Introduction

The rapid expansion of digital communication has transformed the production, circulation, and interpretation of information. News, political claims, health-related messages, crisis updates, and personal opinions are now generated and shared continuously through online platforms and social networks. This environment has increased access to information, but it has also created favourable conditions for the rapid dissemination of false, misleading, or manipulative content. Fake news is therefore not only a problem of incorrect textual content; it is also a problem of visibility, repetition, interaction, amplification, and public trust.

The social impact of fake news has become especially visible in domains where information strongly influences collective behaviour, including public health, elections, financial decisions, geopolitical conflicts, emergency communication, and social cohesion. During the COVID-19 pandemic, misleading claims circulated together with verified public-health information, generating confusion and weakening the ability of citizens to distinguish reliable evidence from rumours, speculation, and deliberate disinformation. Similar patterns have been observed in political and social-media contexts, where false claims may spread faster than corrections and may persist even after fact-checking or official clarification [2], [3], [4], [53], [78].

The computational study of fake news has traditionally focused on content-based detection. In this formulation, a news item, claim, post, or article is represented as textual input and classified as fake or real using machine-learning or deep-learning models. Classical approaches rely on lexical, statistical, stylistic, and source-related features, such as Bag-of-Words, TF-IDF, n-grams, readability indicators, and supervised classifiers including Logistic Regression, Naive Bayes, Support Vector Machines, Random Forests, and boosting methods. These methods remain useful as interpretable baselines, but they often have limited capacity to capture contextual meaning, implicit contradiction, sarcasm, framing, or domain shift.

Transformer-based models have changed the landscape of fake-news detection by providing contextual language representations that can model semantic relationships across an input sequence. Architectures such as BERT, RoBERTa, mBERT, XLM-RoBERTa, SBERT, and BigBird can represent short social-media posts, longer article-style inputs, multilingual text, and semantic similarity relations more effectively than sparse lexical models [6]-[12]. In the context of fake-news detection, these models can estimate a probability that a given item belongs to the fake-news class and can therefore provide a content-risk signal for subsequent analysis. However, even strong Transformer classification performance does not automatically explain how the content spreads through a social system.

This limitation is central to the present dissertation. A fake-news item with high textual risk may remain socially insignificant if it is not shared widely, while another item with moderate textual uncertainty may become harmful if it reaches influential users, crosses communities, or triggers repeated engagement. Therefore, fake-news analysis should not be reduced to classification accuracy alone. The societal risk of fake news depends on the interaction between the credibility of the content and the temporal dynamics of its diffusion. In other words, the key research challenge is not only to decide whether a message is fake, but also to understand whether, how, and under what uncertainty it propagates through social networks.

Temporal-network analysis provides a necessary bridge between content detection and diffusion modelling. Unlike static graph representations, which aggregate interactions over an observation window, temporal networks preserve the chronological order of interactions. This is essential because information can only travel along time-respecting paths. Retweets, replies, reposts, comments, shares, quotes, and other timestamped interactions can be used to reconstruct directed temporal cascades associated with a specific fake-news claim. From these cascades, observable quantities such as new active spreaders, cumulative active spreaders, cascade size, time to peak, inter-event delays, user influence, and community crossing can be extracted [52]-[55].

Epidemic-style models offer an interpretable mathematical language for representing fake-news diffusion. Classical models such as SIR describe movement from susceptible users to active spreaders and then to inactive, corrected, or removed states. Rumour and misinformation models extend this logic by considering exposure, hesitation, skepticism, correction, and loss of interest [31]-[35], [67], [73]. In this dissertation, SIR is retained as a transparent baseline, while SEIZ is used as the main diffusion model because it includes exposed and skeptic or non-spreading states. This makes SEIZ more suitable for fake-news cascades in which users may observe content, delay sharing, reject it, encounter correction, or stop contributing to the cascade.

Nevertheless, epidemic analogies must be used carefully. Online information sharing is not identical to biological infection. A repost does not prove belief, silence does not prove rejection, and platform data usually do not record every exposure, private message, deleted post, algorithmic recommendation, or decision not to engage. Social-network cascades are therefore partial observations of a hidden diffusion process. A deterministic fitted curve may appear precise while ignoring substantial uncertainty about transmission, recovery, exposure, skepticism, and correction-related parameters. This motivates the use of uncertainty-aware inference rather than relying only on point estimates.

Approximate Bayesian Computation (ABC) and ABC-SMC are suitable for this setting because they support likelihood-free inference for models that can be simulated but whose exact likelihood is unavailable or intractable [19]-[25], [40]. In fake-news diffusion, a candidate SIR or SEIZ simulator can generate synthetic cascades under sampled parameter values, and the simulated trajectories can be compared with observed temporal-cascade summaries. Parameters that generate sufficiently similar simulations are retained as posterior samples. ABC-SMC improves this process by progressively reducing tolerance levels and concentrating particles in parameter regions that better reproduce the observed spreading pattern.

The dissertation therefore treats fake-news analysis as an integrated content-to-cascade problem. The content layer estimates whether a textual item is likely to be fake. The temporal-network layer reconstructs how the item spreads through directed timestamped interactions. The epidemic layer interprets the spreading trajectory through SIR and SEIZ mechanisms. The ABC/ABC-SMC layer estimates uncertainty in the epidemic parameters and supports posterior predictive simulation. Finally, posterior forecasts and counterfactual intervention scenarios provide a cautious basis for analysing how visibility reduction, debunking, or early combined mitigation could affect the future trajectory of a cascade.

A further methodological difficulty concerns dataset suitability. Content-based datasets usually contain labelled texts suitable for supervised classification, but they often lack user-level temporal interactions. Diffusion-oriented datasets provide timestamps, retweets, replies, or cascade traces, but they may not contain full article text or standard fake/real labels. This dissertation therefore separates content-based evidence from diffusion-oriented evidence. This separation prevents the methodological error of treating classification datasets as if they were temporal-network datasets, or treating cascade records as if they were ordinary independent text-classification samples.

The experimental scope of the dissertation reflects this evidence-layer distinction. Content-based datasets are used to evaluate Transformer and baseline detection performance across short-text, long-document, healthcare-related, English, and low-resource Albanian settings. Diffusion-oriented datasets are used to reconstruct fake-news cascades, extract time-indexed spreading curves, compare SIR and SEIZ model fits, estimate posterior uncertainty through ABC and ABC-SMC, and evaluate short-horizon forecasting and intervention-oriented scenarios. In this way, the dissertation preserves the interpretability of each empirical layer while also connecting them into one coherent framework.

The inclusion of low-resource Albanian fake-news detection also strengthens the dissertation's scope. Many fake-news detection studies focus on English or high-resource languages, while smaller languages may have fewer labelled datasets, less developed fact-checking infrastructure, and greater variation in orthography, morphology, and local political vocabulary. Multilingual Transformer models such as mBERT and XLM-RoBERTa make it possible to evaluate fake-news detection in this context, while synthetic or validation-oriented temporal examples show how a diffusion pipeline could be extended when real temporal data become available.

The proposed research is motivated by the need for a framework that is both technically integrated and scientifically cautious. It is integrated because it combines Transformer-based detection, temporal-network representation, epidemic modelling, likelihood-free Bayesian inference, posterior prediction, and intervention-oriented scenario analysis. It is cautious because it does not infer private belief directly, does not treat model probabilities as absolute fact-checking truth, and does not claim causal platform-policy effects from observational cascades. Instead, it provides a reproducible and auditable methodology for analysing how fake-news content becomes socially active and how uncertain its future diffusion may be.

The need for such an integrated approach is reinforced by the way fake-news datasets are created and used. Many benchmark corpora are valuable for classification because they provide labels and textual content, but they often encode source-specific, topic-specific, or style-specific patterns that can inflate performance if they are not interpreted cautiously. Conversely, temporal diffusion datasets provide information about user activity and spreading behaviour, but they may contain incomplete hydration, missing content, partial claim labels, or platform-specific constraints. A high-quality dissertation framework must therefore make the evidential role of each dataset explicit and avoid transferring conclusions from one layer of analysis to another without methodological justification.

The dissertation also addresses the tension between predictive performance and interpretability. Transformer models can achieve strong classification scores, but their outputs become more useful for decision support when they are connected to transparent downstream quantities. Temporal cascades provide observable evidence of spread; SIR and SEIZ models provide interpretable state transitions; ABC and ABC-SMC provide posterior uncertainty rather than a single fitted parameter vector. This layered design makes it possible to examine which part of the evidence comes from text, which part comes from temporal interaction, which part comes from the epidemic simulator, and which part comes from posterior predictive uncertainty.

Another important aspect of the research is the distinction between modelling and causal interpretation. The intervention scenarios in this dissertation are not presented as direct measurements of real platform moderation policies. Instead, they are controlled counterfactual simulations based on calibrated SEIZ dynamics. This distinction is important for scientific validity. It allows the dissertation to discuss debunking, visibility reduction, and early combined mitigation while avoiding unsupported claims about what would necessarily happen on a real platform. The scenario analysis is therefore used as a structured decision-support illustration, not as a substitute for randomized platform experiments or causal policy evaluation.

The dissertation is also positioned within the author's research trajectory on fake-news diffusion, temporal networks, BigSEIZ modelling, SIR/SEIZ comparison, and BigBird-assisted SEIZ with ABC [87], [108], [86], [88], [89]. These works motivate the integrated direction developed here, but the dissertation extends them by consolidating the detection, diffusion, posterior inference, forecasting, and intervention components into one coherent research framework. The resulting contribution is methodological as well as empirical: it defines how multiple forms of evidence can be connected, evaluated, and interpreted under explicit assumptions.

Chapter 1. Literature Review and Theoretical Foundations

1.1 Introduction

Fake-news detection has become a central research problem in natural language processing, machine learning, data mining, social-network analysis, and computational social science. The increasing dependence on online platforms for political, health-related, social, and economic information has created an environment in which misleading information can be produced and distributed at high speed. In this context, fake-news detection is not limited to identifying false statements; it also involves analyzing how language, source credibility, user behavior, platform mechanisms, and social reactions contribute to the spread and perceived reliability of information [1], [2], [3].

The literature on fake-news detection is commonly organized into several methodological families. Earlier work used shallow statistical classifiers and handcrafted features, while recent work relies increasingly on deep contextual language models, graph-based representations and integrated content-diffusion frameworks. In the present chapter, Sections 1.2-1.4 focus on conceptual definitions and detection methods, while Sections 1.5-1.7 extend the discussion toward temporal networks, epidemic modelling and Approximate Bayesian Computation for diffusion inference.

A typical fake-news detection system follows a pipeline that begins with the collection of news articles, claims, tweets, posts, or comments. The text is then cleaned, normalized, and represented numerically using either handcrafted features, such as term frequency-inverse document frequency (TF-IDF), or learned contextual embedding's, such as those generated by BERT or RoBERTa. A classifier then predicts the credibility label of the content. Depending on the dataset, the task may be binary, where the system distinguishes fake from real news, or multiclass, where the system predicts more nuanced labels such as true, mostly true, half true, barely true, false, or doubtful [1], [4].

The importance of this literature review lies in establishing the theoretical and methodological foundation for the proposed thesis framework. Classical machine learning provides interpretable baselines; Transformer models provide contextual semantic understanding; temporal networks represent time-ordered interactions; and epidemic models with ABC support uncertainty-aware diffusion analysis. These components motivate the integrated framework developed in Chapter 2, where the operational bridge is the calibrated fake-news probability score produced by the selected Transformer detection layer.

At the author level, the review is grounded in Zhou and Zafarani's survey of fake-news theories, Shu et al.'s data-mining perspective on social-media fake-news detection, Patwa et al.'s COVID-19 fake-news dataset work, and the Transformer-based language-modelling contributions of Devlin et al., Liu et al., Pires et al., Reimers and Gurevych, Conneau et al., and Zaheer et al. [2], [3], [30], [6]-[12]. For the diffusion component, the chapter draws on Kermack and McKendrick, Daley and Kendall, Jin et al., Raponi et al., Kucharski, Del Vicario et al., Cinelli et al., Bakshy et al., Scholtes et al., Rosvall et al., Lambiotte et al., Kauk et al., and Murayama et al., whose work links epidemic modelling, temporal networks, polarization, echo chambers and online false-news spread [31], [32], [33], [34], [35], [36], [37], [13]-[15], [38], [16]-[18], [39]. For uncertainty-aware inference, the review uses the ABC and simulation-based inference literature of Beaumont et al., Marjoram et al., Sisson et al., Toni et al., Sunnåker et al., Kypraios et al., Minter and Retkute, Marin et al., Cranmer et al., Papamakarios and Murray, Gomez-Rodriguez et al., and Farajtabar et al. [19]-[25], [40], [26]-[29]. This author-centred synthesis avoids treating publication venues as evidence in themselves and instead emphasizes the specific scholarly contributions that support the dissertation framework.

1.2 Defining Fake News

1.2.1 Fake News: Definitions and Characteristics

The definition of fake news is a central theoretical issue because the term is used differently in journalism, political communication, information science, and machine-learning research. In a broad sense, fake news may refer to any false or misleading information that appears in a news-like form. In a narrower scientific sense, it is often restricted to intentionally and verifiably false information presented as legitimate news and designed to mislead readers [2], [4]. This distinction is important because not all false information is intentionally deceptive, and not all deceptive information is presented in a conventional news format.

This dissertation adopts a working definition that is suitable for both content-based detection and diffusion analysis: fake news is false, misleading, partially false, or manipulatively framed information distributed through online platforms or social networks in a form that resembles factual reporting. The definition includes news articles, social-media posts, claims, health-related misinformation, political narratives, and doubtful content whose truth status may be incomplete or disputed. This formulation is intentionally broader than a strict binary definition because false and misleading online information often exists along a credibility spectrum rather than as a perfectly separable fake/real dichotomy.

A further distinction concerns intentional and unintentional misinformation. Disinformation implies deliberate deception, whereas misinformation may be shared without a conscious intention to mislead. This difference is conceptually important, but it is difficult to encode in datasets because annotators usually observe the text, source, and context rather than the actual psychological intention of the author. Consequently, many computational studies treat intentionality as a latent or contextual variable instead of a directly observed label [3], [41].

Fake-news detection may be formulated as a binary or multiclass task. Binary classification simplifies the problem into fake and real labels and is useful for benchmark experiments, probability scoring, and integration with downstream diffusion models. Multiclass classification is more expressive because it can represent intermediate truthfulness categories such as partially false, unverified, doubtful, or misleading. However, multiclass annotation usually increases ambiguity and reduces inter-annotator agreement. In this thesis, binary fake/real probability is used as the main experimental detection output, while multiclass and doubtful-news categories remain essential for theoretical interpretation and for understanding the limits of binary labelling.

1.2.2 Related Concepts

Fake news is closely related to several concepts that overlap but are not identical. Confusing these concepts can lead to inconsistent dataset construction, weak annotation protocols, and unreliable evaluation. The conceptual boundary is especially important when models are trained on one dataset and then evaluated in another domain, because labels such as rumor, satire, disinformation, doubtful news, and clickbait may encode different annotation assumptions.

Table 1.1 summarizes the main concepts related to fake news and explains their relevance to automated detection. The table is retained because it clarifies the terminology used throughout the thesis and supports the later distinction between content-level credibility classification and network-level diffusion modelling.

Table 1.1. Main concepts related to fake news and their relevance to detection.

Concept	Meaning	Relevance to fake-news detection
Misinformation	False or inaccurate information that may be shared without deliberate intent to deceive.	Important when the intention of the source is unknown, unavailable, or not directly observable in the dataset [41].
Disinformation	False or misleading information produced or disseminated deliberately to deceive.	Closest to the narrow definition of fake news and highly relevant to intentional manipulation [2], [3].

Malinformation	True or partly true information used in a misleading context or with harmful intent.	Requires contextual analysis because the problem may lie in framing, timing, or selective disclosure [41].
Rumors	Unverified information that circulates before its truth status is confirmed or rejected.	Relevant to temporal verification, evolving labels, and diffusion before confirmation [111], [53].
Propaganda	Strategically framed communication designed to influence beliefs, attitudes, or behaviour.	May overlap with fake news when selective facts, emotional framing, or fabricated claims are used to influence audiences.
Satire / parody	Content that intentionally imitates news style for humour, criticism, or exaggeration.	Can be misclassified as fake news unless genre, intention, and publication context are considered [111].
Clickbait	Sensational or misleading headlines designed to attract attention and increase engagement.	May be false, partly true, or simply manipulative; useful for detecting engagement-driven misinformation.
Doubtful news	Content whose truthfulness is uncertain, mixed, weakly supported, or insufficiently verified.	Important for uncertainty-aware and delayed-decision classification rather than forced binary labelling.
Infodemic	A large-scale mixture of accurate, inaccurate, and misleading information, especially during crises.	Highly relevant to public-health misinformation and COVID-19 fake-news research [30].

The concept of doubtful news is particularly important for this dissertation. Doubtful content does not always satisfy the strict definition of fake news, but it may still spread rapidly and influence public perception before its truth status is settled. In a Transformer-based system, doubtful news may correspond to low confidence, contradictory evidence, disagreement among fact-checking sources, or conflicting signals between textual content and source credibility. Therefore, doubtful news should not be treated simply as an error category; it represents a real epistemic condition in online information environments.

This conceptual boundary is also consistent with deception-oriented and rumour-verification studies, including tree-structured rumor detection and conversational-thread analysis, which distinguish fabricated news, satire, hoaxes, propaganda, unverified rumours and evidence-based fact-checking tasks instead of treating every doubtful item as the same binary object [103], [105], [111], [160]-[162], [173], [177]-[178].

1.2.3 Challenges in Defining Fake News

The definition of fake news is difficult because truthfulness is not always stable, objective, or immediately verifiable. Some claims depend on context, timing, source reliability, or the availability of later evidence. A claim that is unverified at the time of publication may later become confirmed, while a claim that appears accurate may become misleading when new evidence emerges. This problem of evolving truth is particularly visible during crises, elections, wars, and public-health emergencies, when information changes rapidly and verification is incomplete.

Annotation ambiguity is one of the most serious methodological problems in fake-news detection. Annotators may disagree because they interpret sarcasm, satire, political framing, missing context, or partial truth differently. Dataset labels may also reflect the criteria of a specific fact-checking organization, the political context of a claim, or the level of evidence available at a particular moment. As a result, models trained on such labels may learn the annotation policy of a dataset rather than a general definition of falsity. This limitation is especially important when binary labels are used for content that is actually mixed, doubtful, or context-dependent.

Subjectivity and polarization further complicate the problem. In polarized communities, users may interpret the same claim according to ideological alignment rather than evidence. However, a detection model must not equate disagreement with falsity. The label should be grounded in verifiable evidence, transparent annotation criteria, and source documentation rather than audience preference or political identity [4], [3].

Fact-checking also has limitations. It may be delayed, incomplete, domain-specific, or inconsistent across organizations. Some claims are too vague to verify directly, while others combine true and false elements in a single narrative. Fact-checking databases may focus on high-visibility claims and underrepresent low-resource languages, local misinformation, or private-platform circulation. Consequently, fact-checking should be treated as an important but imperfect source of ground truth, and model outputs should be interpreted probabilistically rather than as absolute truth judgments.

1.3 Classical Machine Learning Approaches for Fake-News Detection

1.3.1 Text Representation Methods

Classical machine-learning approaches represent the first major family of computational methods for fake-news detection. They transform textual content into numerical feature vectors and then apply statistical classifiers. In the context of this dissertation, these methods are retained as interpretable baselines and as historical foundations for understanding why Transformer-based models are needed.

The most common text representations include Bag-of-Words, TF-IDF, word and character n-grams, and stylometric features. Bag-of-Words represents a document through the frequency or presence of tokens without modelling word order. N-grams partially address this limitation by preserving short local sequences of words or characters. Stylometric features capture writing style through punctuation, readability, sentence length, lexical diversity, sentiment, or part-of-speech patterns. These representations are useful because they are transparent and computationally efficient, but they remain limited when fake-news interpretation depends on context, implication, sarcasm, or narrative framing.

TF-IDF is retained as the principal classical representation because it is widely used in fake-news detection and provides a strong baseline for sparse textual classification. It gives higher importance to terms that are frequent in one document but relatively rare across the corpus:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d)\log\left(\frac{N}{\text{DF}(t)}\right) \tag{1.1}$$

In Equation (1.1), TF(t,d) is the frequency of term t in document d, DF(t) is the number of documents containing the term, and N is the number of documents in the corpus. The equation is useful for identifying discriminative lexical cues, but it does not capture semantic equivalence between different wordings of the same claim.

1.3.2 Classical Classification Models

Classical fake-news classifiers include Logistic Regression, Naive Bayes, Support Vector Machines, Random Forest, Gradient Boosting and XGBoost. In this thesis these models are treated as background and interpretable baselines rather than as the central methodological contribution. Their inclusion is supported by established text-classification and fake-news literature: Support Vector Machines are grounded in Joachims' text-classification work [42], Random Forests in Breiman's random-forest model [43], and XGBoost in the scalable gradient-boosting system of Chen and Guestrin [44]. Comparative fake-news studies and surveys also report Logistic Regression, Naive Bayes, SVM, Random Forest and boosting methods as classical baselines for fake-news detection [3], [45], [46].

Table 1.2 summarizes the main classical models and their relevance to fake-news detection. The table is kept in compact form in order to support comparison with the Transformer-based detection methods discussed later in the chapter.

Table 1.2. Classical machine learning approaches for fake-news detection.

Approach	**Typical features**	**Representative literature**	**Strengths**	**Limitations**
Logistic Regression	TF-IDF, n-grams, metadata	Shu et al. [3]; Agarwal et al. [45]; Zhang et al. [46]	Simple, interpretable, and strong as a linear baseline.	Limited ability to model nonlinear semantic relations.
Naive Bayes	Word counts, binary terms, TF-IDF	Shu et al. [3]; Agarwal et al. [45]; Zhang et al. [46]	Fast, scalable, and effective for small text datasets.	Relies on a strong conditional-independence assumption.
Support Vector Machine	TF-IDF, word and character n-grams	Joachims [42]; Agarwal et al. [45]; Zhang et al. [46]	Effective in high-dimensional sparse spaces.	Requires careful kernel and parameter selection for large datasets.
Random Forest	Stylometric, lexical, readability, and metadata features	Breiman [43]; Agarwal et al. [45]; Zhang et al. [46]	Robust to noisy engineered features and useful for feature importance.	Less suitable for very sparse text without feature reduction.
Gradient Boosting / XGBoost	Lexical, stylistic, metadata, and structured features	Chen and Guestrin [44]; Zhang et al. [46]	Strong predictive performance with engineered features.	Requires tuning and may overfit noisy or biased labels.

1.3.3 Advantages and Limitations

The main advantages of classical models are interpretability, computational efficiency, and reproducibility. Linear models can show which words or n-grams contribute most strongly to a prediction, and tree-based models can provide feature-importance scores. This explainability is valuable in fake-news detection because users, journalists, and moderators often require a justification for why a claim is classified as suspicious. Classical models are also useful in low-resource settings where large Transformer models may be computationally expensive or prone to overfitting.

Despite these advantages, classical models have limited contextual understanding. They often rely on surface-level lexical cues and may fail when two texts express the same false claim using different vocabulary, or when the same words have different meanings in different contexts. They are also sensitive to domain shift: a model trained on COVID-19 misinformation may not generalize to political, financial, or geopolitical fake news if the vocabulary and narrative structure change. For this reason, classical machine learning is treated in this dissertation as a transparent baseline and explanatory reference point, while the main detection layer is based on Transformer architectures and calibrated probability outputs.

1.4 Transformer-Based Models for Fake-News Detection

Transformer-based models form the dominant family of modern neural architectures for fake-news detection because they learn contextual representations rather than relying only on isolated lexical features. In contrast to classical models, which typically represent text through sparse counts or engineered indicators, Transformers use attention mechanisms to model relationships among tokens, sentences and claims across an input sequence. This capacity is particularly important for fake-news detection, where misleading content may depend on implicit contradiction, selective framing, emotional language, missing context or semantic similarity between a claim and verified evidence [6], [7].

1.4.1 Transformer Architecture and Attention Mechanism

The Transformer architecture introduced by Vaswani et al. replaces recurrence with self-attention, allowing all tokens in a sequence to interact directly during representation learning [7]. Self-attention computes the relevance of one token to another through query, key and value projections. For fake-news detection, this mechanism enables the model to assign higher weight to entities, numerical claims, emotionally charged expressions, source cues and context-dependent phrases that may influence the credibility decision.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1.2}$$

In Equation (1.2), Q, K and V denote the query, key and value matrices, while d_k denotes the dimensionality of the key vectors. The equation is retained in the literature review because attention is the main theoretical mechanism that explains why Transformer models can capture contextual relations more effectively than Bag-of-Words or TF-IDF features.

The resulting contextual representation is valuable for Natural Language Processing because the meaning of a word or phrase is no longer treated as fixed. A claim about a vaccine, a political actor or a crisis event may receive different contextual interpretation depending on surrounding evidence, modality, sentiment and discourse structure.

1.4.2 BERT and RoBERTa-Based Detection

BERT is one of the most influential Transformer-based models for text classification. It uses a bidirectional encoder trained with masked language modelling and can be fine-tuned for downstream tasks by attaching a classification layer to the contextual representation of the input sequence [6]. In fake-news detection, BERT is useful because it can capture semantic relations between the headline, body text, named entities and implicit argumentative cues.

RoBERTa improves the BERT training procedure by using larger training data, dynamic masking and optimized pre-training choices [8]. In fake-news detection, RoBERTa is commonly used as a stronger contextual baseline because it often provides more stable representations across domains. Both BERT and RoBERTa are typically fine-tuned on labelled fake/real datasets, where the final classifier estimates a probability distribution over credibility labels.

$$p(y \mid x) = \mathrm{softmax}\big(Wh_{[\mathrm{CLS}]} + b\big) \tag{1.3}$$

In Equation (1.3), $h_{[\mathrm{CLS}]}$ denotes the contextual representation of the input sequence, W and b are trainable parameters of the classification layer, and $p(y \mid x)$ denotes the predicted probability distribution over labels. In this chapter, the equation is used only to describe the general classification principle; the detailed experimental fine-tuning settings are placed in Chapter 2 and Chapter 3.

1.4.3 Multilingual Transformer Models

Multilingual fake-news detection requires models that can represent several languages with shared subword vocabularies and cross-lingual semantic spaces. Multilingual BERT (mBERT) was trained across many languages and can support transfer to lower-resource settings, although its performance depends on language coverage, tokenization quality and domain similarity [9]. XLM-RoBERTa extends this idea by using large-scale multilingual CommonCrawl data and often provides stronger cross-lingual representations [11].

These models are relevant for the present thesis because fake-news detection is not limited to high-resource English corpora. Low-resource languages, including Albanian, may have fewer labelled datasets, smaller fact-checking infrastructures and greater variation in spelling or morphology. Multilingual Transformers therefore provide a practical mechanism for evaluating fake-news detection under language scarcity while preserving the same binary fake/real experimental logic.

1.4.4 Long-Sequence Models

Standard BERT-style models use full self-attention and are commonly constrained by a maximum input length. When fake-news articles are long, truncation can remove important evidence from the body of the article, including quotations, references, internal contradictions or misleading transitions between headline and text. Long-sequence models address this limitation by modifying the attention pattern so that longer inputs can be processed more efficiently.

BigBird is particularly relevant because it combines local, random and global attention, producing a sparse-attention pattern with substantially lower computational cost than full attention [12]. Local attention captures nearby context, random attention supports long-range information flow, and global attention allows selected tokens to aggregate document-level information. This makes BigBird suitable for long fake-news articles, where credibility cues may appear outside the first 512 tokens.

$$\text{Complexity}_{\text{full}} = O(n^2 d) \tag{1.4}$$

$$\text{Complexity}_{\text{sparse}} = O(nrd), \quad r \ll n \tag{1.5}$$

Equations (1.4) and (1.5) summarize the computational motivation for using sparse attention. Full self-attention scales quadratically with sequence length, whereas sparse attention reduces the number of attended token pairs and therefore makes long-document classification more feasible. This discussion is retained because it explains why BigBird is considered for full-article fake-news detection, while implementation-level details are reserved for Chapter 2.

For a token i, the BigBird attention set can be expressed in compact form as:

$$A_i = W_i \cup R_i \cup G \tag{1.6}$$

Here, W_i denotes a local window, R_i denotes a set of random tokens, and G denotes global tokens. This formal notation is included only to clarify the sparse-attention principle; the thesis does not treat the BigBird architecture itself as an original contribution.

$$h_i = \text{Attention}\left(q_i, K_{A_i}, V_{A_i}\right) \tag{1.7}$$

Equation (1.7) indicates that the contextual representation of a token is computed over the restricted sparse-attention set rather than over the complete sequence. The practical implication for fake-news detection is that more article-level evidence can be preserved without applying aggressive truncation.

1.4.5 Sentence-Level and Similarity Models

Sentence-level Transformer models provide another important direction for fake-news detection. SBERT modifies BERT through siamese or triplet network structures in order to produce sentence embeddings that can be compared efficiently through semantic similarity [10]. This is useful when the task involves comparing a claim against a verified statement, a fact-checking article, a headline, or a set of evidence sentences.

Similarity models are particularly useful for claim comparison and retrieval-based verification. A fake-news detector may first identify whether a new claim is semantically close to previously debunked claims and then use this similarity as evidence for credibility assessment. However, similarity alone is not sufficient: two claims may be semantically close but differ in important details such as time, location, quantity or causal framing.

1.4.6 Large Language Models and Emerging Directions

Large Language Models represent an emerging direction in fake-news research. They can support explanation generation, claim rewriting, weak supervision, zero-shot classification and assisted fact-checking workflows [47]-[49]. In low-resource settings, any LLM-assisted process must be separated from test data and manually checked for label consistency; it is not used as a core evidence source in this thesis.

At the same time, LLMs introduce specific risks. They may generate fluent but unsupported statements, reproduce social biases, overstate certainty or produce hallucinated explanations [49]. For this reason, LLM-generated outputs are not treated as ground truth in this dissertation. LLMs are used only as auxiliary or support tools for controlled augmentation, exploratory analysis or explanation drafting; final labels, evaluation sets and diffusion interpretations remain evidence-based, manually auditable and reproducible.

Recent instruction-following, retrieval-augmented and hallucination-checking research reinforces this boundary: large language models can support explanation, retrieval and weak-supervision workflows, but generated claims require external evidence, provenance and consistency checks before they can be used in a labelled benchmark or diffusion interpretation [48], [135]-[136], [143]-[145]. More recent fake-news and factuality studies reinforce this interpretation: LLMs can act as advisors or rationale generators for smaller detection models, fake-news detectors must be adapted to human-written, machine-paraphrased and machine-generated news, and factuality evaluation remains difficult for open-ended LLM outputs [181], [182], [183].

Table 1.3 supports the content-detection layer of the framework by identifying how textual evidence is prepared or interpreted before diffusion analysis.

Table 1.3. Main Transformer-based models relevant to fake-news detection.

Model	Core idea	Relevance to fake-news detection

BERT	Bidirectional Transformer encoder pre-trained with masked language modelling.	Strong baseline for contextual fake-news classification [6].
RoBERTa	Optimized BERT training with more data, dynamic masking, and no next-sentence prediction.	Often improves robustness and classification performance [8].
mBERT	Multilingual BERT trained across 104 languages.	Useful for cross-lingual or multilingual fake-news detection [9].
BigBird	Sparse-attention Transformer that combines local/window, random, and global attention to model long sequences with linear attention complexity.	Useful for full-article fake-news detection because it reduces information loss caused by truncating long articles to the first 512 tokens [12].
SBERT	Sentence-level embeddings based on siamese/triplet BERT networks.	Useful for semantic similarity, claim comparison, and sentence-level classification [10].
XLM-RoBERTa	Cross-lingual RoBERTa trained on large multilingual CommonCrawl data.	Useful for multilingual detection and transfer across languages [11].
Large Language Models	Generative Transformer models used for augmentation, explanation, and zero-shot tasks.	Support augmentation and zero-shot detection; require caution due to hallucination and bias.

1.4.7 Limitations of Transformer-Based Detection

Despite their strong performance, Transformer-based fake-news detectors have important limitations. First, they are computationally more expensive than classical models because they require tokenization, GPU-supported fine-tuning and careful validation. Second, they are less interpretable: attention weights and embedding representations do not automatically provide faithful explanations for why a model assigns a fake or real label.

Third, Transformers may be vulnerable to adversarial attacks, paraphrasing, topic shift and strategically manipulated language [50]. A model trained on COVID-19 claims may not generalize to political disinformation, financial rumours or Albanian-language news without adaptation. Domain adaptation is therefore a central challenge, especially when training and deployment data differ in topic, platform, language, source style or label policy [51].

Finally, Transformer-based detection remains dataset-dependent. If annotations are inconsistent, politically biased, delayed or incomplete, the model can learn these weaknesses as if they were reliable signals. For this reason, Transformer outputs should be interpreted as probabilistic credibility estimates rather than final truth judgments. The methodological chapters later connect these probabilities with temporal-network and epidemic-diffusion analysis, but the risk-based transmission equations are developed in Chapter 2 rather than in this literature-review section.

Model selection and implementation details therefore remain part of the methodological evidence. DeBERTa, library-level Transformer implementations and tokenizer-quality studies show that architecture, checkpoint choice, tokenization and sequence policy can change robustness, multilingual coverage and calibration, especially for long or low-resource fake-news texts [118]-[120].

1.5 Ensemble and Hybrid Approaches for Fake-News Detection

Ensemble and hybrid approaches form an important contextual part of the fake-news detection literature because they attempt to combine the strengths of several learners, feature spaces or modelling assumptions. In a narrow sense, an ensemble combines multiple classifiers and aggregates their outputs through voting, averaging or a meta-learner. In a broader sense, a hybrid system joins different types of evidence, such as lexical features, Transformer representations, user metadata, stance information or network signals. These approaches are relevant to this dissertation because they show one possible route for improving predictive performance, but they also clarify why the implemented thesis framework does not adopt an ensemble classifier as its core experimental detector.

The distinction between ensemble detection and the proposed dissertation framework is important. Ensemble methods usually operate inside the detection layer: several models are trained for the same classification task and their predictions are combined into a final fake/real decision. The framework developed in this dissertation instead links different analytical layers: a Transformer detector estimates content-level fake-news probability, temporal cascade analysis reconstructs how the fake-news claim spreads, SIR/SEIZ models represent diffusion mechanisms, and ABC/ABC-SMC estimates uncertainty. Therefore, the dissertation is hybrid at the framework level, but it does not use an ensemble classifier as the operational detection model.

1.5.1 Voting-Based Ensembles

Voting-based ensembles are among the simplest forms of model combination. In hard voting, several classifiers each produce a class label and the final decision is obtained by majority vote. In soft voting, the predicted probabilities are averaged or weighted, and the final label is selected from the combined probability distribution. For fake-news detection, voting may combine models trained on TF-IDF vectors, stylistic features, classical classifiers, neural models or different Transformer checkpoints. The motivation is that individual models may make different errors, so a collective decision can reduce variance and produce more stable predictions across heterogeneous examples.

Voting can be useful when the underlying data contain diverse linguistic patterns. A claim-based COVID-19 dataset, a long-form article dataset and a low-resource language dataset may emphasize different textual signals. A voting ensemble could combine a lexical classifier that captures explicit cue words, a Transformer model that captures contextual semantics, and a length-aware model that handles long documents. This logic is consistent with the general fake-news survey literature, where detection is commonly described as a multi-feature and multi-source task rather than a purely lexical classification problem [2]. Classical ensemble learners such as Random Forests and boosting-based methods also demonstrate how aggregating many weak or unstable predictors can improve robustness [43], [44].

However, voting-based ensembles also introduce limitations that are significant in a PhD methodology. First, they require careful calibration because the probability output of one model may not be directly comparable with the probability output of another. Second, they can hide the source of an error: if the ensemble misclassifies a fake-news claim, it may be difficult to determine whether the error comes from lexical features, Transformer semantics, dataset bias or class imbalance. Third, the ensemble decision may increase performance without improving theoretical understanding of how content risk becomes diffusion risk. For this dissertation, this limitation is central because the goal is not only to maximize a classification score, but also to connect detection with temporal spreading and epidemic inference.

1.5.2 Stacking and Meta-Learning

Stacking and meta-learning represent a more flexible ensemble family. Instead of directly voting among base classifiers, stacking trains a second-level model that learns how to combine the predictions of several first-level models. The base learners may include classical machine learning classifiers, deep neural networks, Transformer models or models trained on different feature groups. Their outputs are then used as inputs for a meta-classifier, which estimates the final fake/real decision. This approach can be attractive when different datasets or feature spaces contain complementary information, because the meta-learner can assign different importance to different model outputs.

In fake-news detection, stacking can theoretically combine content-level evidence with social and contextual evidence. For example, one base model may detect suspicious wording, another may evaluate stance toward a known claim, another may use source-level credibility and another may exploit network information. Such a design can improve classification performance when training data are sufficiently large and when the validation protocol is carefully separated from the training process. It can also support domain adaptation if the meta-learner learns which signals are more reliable for short claims, long articles or multilingual text. For this reason, stacking is a plausible research direction for large-scale operational fake-news systems.

Despite these advantages, stacking is not adopted as an experimental component in this dissertation. The main reason is methodological clarity. A stacked detector would introduce an additional layer of learned decision-making between the Transformer output and the diffusion model. This would make it harder to interpret how the content-level probability contributes to cascade analysis and to the proposed Baseline Posterior Fake-News Diffusion-Risk Functional. Stacking also increases the risk of validation leakage if meta-features are generated incorrectly, and it requires additional ablation studies to prove that the meta-learner improves generalization rather than simply memorizing dataset-specific artifacts. These requirements would shift the focus away from the dissertation's main contribution: connecting Transformer detection with temporal-network and epidemic-model inference.

1.5.3 Transformer-Based Hybrid Systems

Transformer-based hybrid systems extend ensemble thinking by combining contextual language models with additional architectures or evidence sources. A Transformer may be combined with convolutional or recurrent layers, with graph neural networks, with metadata features, or with fact-checking retrieval modules. In long-document settings, models such as Longformer and BigBird address the input-length limitation of standard Transformers and may be joined with domain-specific classification heads or auxiliary features [12]. In multilingual or low-resource settings, contextual models such as multilingual BERT can be combined with language-specific preprocessing, translation, data augmentation or cross-lingual transfer strategies [9].

These hybrid systems are relevant because fake-news detection is rarely a purely textual problem. A fake-news article or claim may depend on who posted it, how it was shared, whether it matches previously debunked information, and whether it appears during a coordinated diffusion event. Transformer-based hybrids therefore attempt to incorporate richer signals than text alone. At the same time, they often remain detection systems: the final output is still a class label or probability score. The dissertation's framework goes beyond this by using Transformer detection as the entry point into a temporal and epidemiological analysis, rather than as the final endpoint of the study.

Graph, multimodal and explainable fake-news models strengthen the same argument. GNN foundations, GCN/GAT/GraphSAGE variants and fake-news-specific graph models show how users, publishers, news items and propagation links can be represented as relational evidence, while multimodal and explainable systems such as EANN, SAFE, dEFEND, CSI and CredEye combine textual, visual, social-response and explanation signals [121]-[127], [137]-[139], [146]-[152], [171]-[172], [176].

The author's related work on BigBird-assisted SEIZ and ABC modelling is hybrid in a different sense: the Transformer component supports content-level identification, while the epidemic model and ABC/ABC-SMC layer explain and calibrate diffusion behaviour [86], [89]. This is not equivalent to building a Transformer ensemble. It is an integrated detection-diffusion-inference framework. The distinction is important for the thesis argument because it preserves interpretability at each layer. A Transformer score remains a content-risk signal; temporal-network measures remain cascade evidence; SEIZ parameters remain diffusion mechanisms; and ABC/ABC-SMC remains the uncertainty-estimation layer.

1.5.4 Advantages and Limitations

The main advantage of ensemble and hybrid approaches is predictive complementarity. Different models may capture different aspects of fake-news text: lexical repetition, sensational wording, source style, semantic contradiction, topic-specific misinformation, or long-range document structure. Combining these signals can improve robustness and reduce the chance that a single weak model dominates the final prediction. Ensembles may also stabilize results across random seeds, class imbalance and noisy examples. In practical systems where classification performance is the only target, this improvement can be valuable.

A second advantage is flexibility. Voting and stacking can combine classical machine learning, Transformer embeddings, topic features, user features and propagation features within one decision architecture. This allows the same general design to be adapted to multiple datasets and tasks. For example, a short-claim dataset may benefit from semantic similarity and stance cues, while a long-article dataset may benefit from long-sequence Transformers or document-level representations. Hybrid systems can therefore be attractive for applied fake-news monitoring tools.

The limitations are equally important. Ensembles increase architectural complexity, computational cost and reproducibility burden. They require additional hyperparameter tuning, multiple training runs, careful probability calibration and more complex error analysis. They may also reduce interpretability: if an ensemble improves F1 score but the contribution of each component is unclear, it becomes harder to explain why a particular fake-news claim was classified as fake. This matters in a dissertation that later uses detection outputs as inputs for temporal cascade analysis and epidemic modelling. In such a framework, interpretability and methodological traceability are not secondary concerns; they are necessary for connecting the classification result to the diffusion result.

Another limitation is that ensemble performance can be dataset-specific. A model combination that performs well on a COVID-19 claim dataset may not generalize to political articles, healthcare misinformation or Albanian-language news. If the ensemble is optimized only for classification metrics, it may obscure whether the detected fake-news claim actually produces a meaningful cascade. This dissertation therefore treats ensemble approaches as relevant background, but not as the main experimental strategy. The central research question is not whether an ensemble can marginally improve detection accuracy, but whether a coherent framework can connect detection, diffusion, posterior uncertainty and intervention-oriented reasoning.

1.5.5 Why Ensemble Methods Are Not Adopted in this Dissertation

The methodological decision not to use ensemble classifiers follows directly from the structure of the dissertation. The detection layer is required to produce a clear and interpretable probability that a text item or fake-news claim is fake. This probability is then linked to temporal cascade reconstruction, SIR/SEIZ diffusion modelling and ABC/ABC-SMC posterior inference. If the detection layer were itself a complex ensemble, the source of the probability score would be less transparent and the connection between content evidence and diffusion evidence would become harder to defend. Individual Transformer architectures therefore provide a more controlled operational layer for the experimental design.

This choice does not imply that ensemble methods are unimportant. They remain relevant in the literature and may be useful in future operational systems, especially when the goal is only to maximize classification accuracy across several benchmark datasets. In this thesis, however, the detection model must be evaluated together with temporal-network outputs such as $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, cascade reach and posterior predictive growth. The emphasis is on a transparent link between detection and diffusion rather than on a highly optimized classifier that is difficult to interpret in downstream epidemic modelling.

Although ensemble approaches often improve predictive performance, they introduce additional architectural complexity and reduce interpretability. Since the main objective of this dissertation is to investigate the interaction between Transformer-based detection, temporal diffusion modelling, and epidemic inference, individual Transformer architectures are selected as the primary operational detection layer. Ensemble methods are therefore treated as contextual literature rather than experimental components.

1.5.6 Why GNN and XAI Are Not Implemented as Core Experimental Layers

Graph Neural Networks and explainable AI are important modern directions for fake-news research because they can incorporate social context, propagation structure and explanation-oriented evidence [121]-[127], [115]-[117], [149], [152], [163], [164], [172]. They are included in the literature review to show how relational and interpretable approaches extend purely textual detection.

They are not implemented as core experimental layers in this dissertation for methodological reasons. GNN-based fake-news detection requires richer user-level graph data, stable follower or interaction networks, propagation labels across claims, and stronger privacy controls than are available consistently across the retained content-based datasets. XAI methods require a separate explanation-validation protocol to test whether explanations are faithful and useful rather than only visually plausible. For this thesis, the priority is therefore a traceable Transformer-ABC-Epidemic pipeline in which the Transformer score, temporal cascade, SIR/SEIZ parameters and ABC/ABC-SMC posterior uncertainty can be interpreted separately. GNN and XAI extensions are retained as future work rather than as unvalidated additions to the implemented methodology. Social-context graph models such as FANG show the value of graph representation for fake-news detection, but their full use depends on user-level interaction histories and stable social-context features that are not consistently available in the retained datasets [127].

1.6 Temporal-Network Approaches to Fake News Diffusion

Fake news diffusion cannot be understood only from textual content. A claim becomes socially influential when it moves through users, communities and platforms over time. Temporal-network approaches therefore represent fake news as a dynamic process in which nodes correspond to users, posts, claims or sources, and edges correspond to time-stamped interactions such as reposts, replies, mentions, comments, quotes or shares [52], [13], [14].

The key advantage of temporal-network modelling is that it preserves the order of interactions. In a static graph, all observed edges are aggregated into one structure, even when they occur at different times. In a temporal graph, an edge is active only at a given time or interval, making it possible to reconstruct chronologically valid diffusion paths and to distinguish potential exposure from actual time-respecting transmission.

1.6.1 Static vs Temporal Networks

Static networks are useful for describing the overall structure of a social platform, but they may misrepresent fake-news diffusion because they ignore temporal ordering. If user A interacts with user B after user C has already shared the same claim, then a static path A -> B -> C may appear possible even though it is causally invalid. Temporal networks avoid this problem by requiring interactions to follow chronological order.

$$G_T = (V, E_T), \quad E_T = \{(u_i, u_j, t_k, w_{ij})\} \tag{1.8}$$

In Equation (1.8), V denotes the set of users, posts or information sources, E_T denotes the set of time-stamped interactions, t_k denotes the interaction time, and w_{ij} denotes an optional weight representing interaction strength, repeated exposure or influence. This representation is the basis for reconstructing fake-news cascades from retweets, replies, comments, propagation trees or other time-dependent platform traces.

A time-respecting path is a sequence of edges whose timestamps increase along the path. This condition is essential for fake-news diffusion because information cannot flow backward in time. Time-respecting paths also make it possible to distinguish the implemented temporal cascade from optional path-memory extensions that would require richer ordered interaction paths [13]-[15].

1.6.2 Information Cascades and Diffusion Dynamics

An information cascade is the temporally ordered set of interactions generated by a fake-news item, related rumour, claim, URL or social-media post. Cascade analysis captures not only whether fake news is shared, but also how it spreads. The most common descriptive properties include final cascade size, depth, breadth, time to peak, maximum growth rate, inter-event time and cascade lifetime [53], [54], [55].

Cascade depth indicates the longest chain of propagation from the original source to later users. Cascade breadth indicates the number of users reached at comparable diffusion levels. Spreading speed measures how quickly the cascade grows, while viral diffusion describes rapid and large-scale propagation driven by repeated sharing, high centrality, emotional framing or cross-community transmission. In this review chapter, these measures are introduced as theoretical descriptors of diffusion; their operational extraction is specified in Chapter 2.

Figure 1.1 supports the temporal-network layer by showing how timestamped user interactions are converted into cascade evidence for Chapter 3.

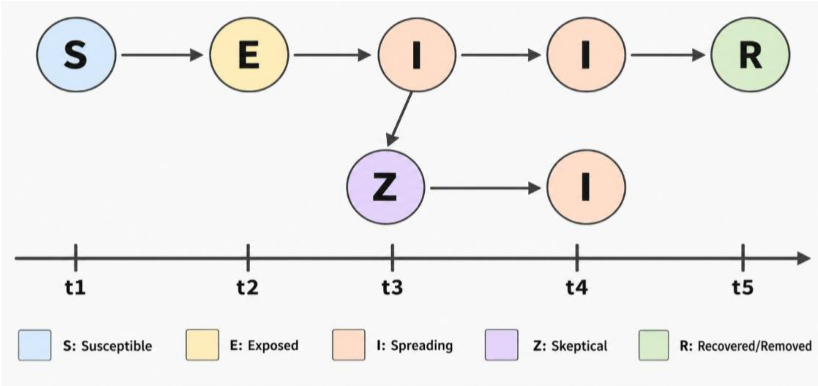

Figure 1.1. Temporal fake-news cascade with time-ordered user states.

1.6.3 Community Structure and Echo Chambers

Community structure strongly affects fake-news diffusion. Users are often organized into groups defined by political preference, language, geography, topic interest or platform behaviour. Within these groups, homophily can increase repeated exposure to similar claims, while polarization can reduce contact with corrective information. Echo chambers intensify this effect by reinforcing content that is consistent with group identity and filtering out opposing evidence [41], [16], [17].

The literature on online political communication shows that algorithmic ranking, selective exposure and social reinforcement can shape what information users encounter [56], [57], [18]. In fake-news diffusion, this means that the same claim may behave differently across communities: it may remain local inside a closed group, cross from one community to another through bridge users, or become amplified when influencers and recommendation systems increase visibility.

Influencer effects are especially relevant because highly central users can accelerate diffusion even when the original content is weakly credible. Recommendation systems may also increase exposure by repeatedly presenting related content, thereby creating indirect pathways that are not always visible in the observed dataset. For this reason, community and influence measures are not merely descriptive; they affect how temporal cascades should be interpreted and how epidemic-style parameters should be estimated.

1.6.4 Temporal-Network Metrics

Temporal-network metrics quantify the structural and chronological conditions under which fake news spreads. Centrality measures identify influential users, bridge nodes and repeated amplifiers. Reachability measures how many users can be reached through valid temporal paths. Temporal path measures evaluate whether a cascade can move across the network given the order of interactions, while diffusion efficiency measures how quickly information can travel from one part of the network to another [58]-[65].

Path-memory temporal-network metrics can be useful when diffusion depends on previous exposure, but they are treated in this dissertation as theoretical background rather than as an implemented Chapter 3 component. A user may react differently depending on whether a claim was received from a trusted peer, an influencer, a fact-checker or a highly polarized community. Higher-order network models capture such dependencies by representing transitions between edges rather than only transitions between nodes [13], [14], [66], [38].

In the present thesis, temporal-network approaches provide the theoretical bridge between content-level fake-news detection and epidemic modelling. Transformer models estimate the credibility of content, whereas directed temporal cascade networks describe how the detected content moves through users and communities. Section 1.6 therefore remains theoretical and literature-oriented; the operational construction of temporal cascades is developed in Section 2.5, while optional path-memory networks are retained only as future extensions.

Dynamic graph-learning work further supports this boundary. Temporal graph surveys, Temporal Graph Networks, TGAT, DyRep and EvolveGCN all treat graphs as evolving event sequences rather than fixed adjacency matrices, which is conceptually aligned with the thesis decision to keep timestamps, directed interactions and time-respecting cascades central to diffusion modelling [128]-[129], [153]-[155].

Although temporal graph-learning models such as Temporal Graph Networks, TGAT, DyRep and EvolveGCN are relevant to dynamic graph representation learning, they are not adopted as core experimental models in this dissertation. Their use typically requires richer user-level graph histories, stable node and edge features, repeated temporal events, supervised graph-learning objectives and substantially larger interaction datasets. The aim of this thesis is different: it requires an interpretable temporal-network representation that preserves timestamps, directed interactions and time-respecting cascades so that observable quantities such as $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, cascade depth, time to peak and spreading rate can be mapped directly into SIR/SEIZ modelling and ABC/ABC-SMC summary-statistic extraction. Therefore, temporal-network representation is selected as the methodological bridge between Transformer-based content detection and epidemic-parameter inference, while temporal graph neural models are retained as related literature and future work [128]-[129], [153]-[155]. Recent benchmark work on temporal graph learning also shows that robust evaluation requires standardized node- and edge-level tasks, large-scale event histories and reproducible temporal-graph protocols; this further supports treating temporal graph neural learning as a future extension rather than the core epidemic-inference layer of this dissertation [184].

1.7 Epidemic and Fake-News Spreading Models

Epidemic and rumor-spreading models provide a theoretical language for analysing fake news as a dynamic process of exposure, adoption, amplification, correction and decline. In this perspective, the unit of analysis is not only the textual credibility of a news item, but also the changing state of users and communities as the item circulates through a network. The models reviewed in this section therefore complement the content-based detection methods discussed earlier by explaining how fake-news cascades may grow, persist or disappear over time.

The epidemic analogy must be used critically. Biological infection and online information sharing are not identical processes: human interpretation, platform ranking, selective exposure, social identity and fact-checking all influence whether a user observes, accepts, rejects or retransmits a claim. Nevertheless, compartmental models remain useful because they offer interpretable states and parameters for describing diffusion, epidemic thresholds, delayed adoption and the effect of correction mechanisms [31]-[35].

For this thesis, epidemic and fake-news spreading models are reviewed through a cross-disciplinary literature base. Classical epidemic theory establishes the compartmental logic of SI, SIS, SIR and SEIR models [31], while rumor-spreading theory formalizes how information may stop spreading when users become stiflers or skeptics [32], [67]. Network science then shows that epidemic thresholds, reachability and cascade size depend strongly on topology and temporal ordering [70]-[72], [52]. Recent fake-news and misinformation studies extend this foundation to social-media cascades, correction processes, and platform-mediated amplification [34]-[35], [36], [37], [39].

Table 1.4 supports the epidemic-modelling layer by linking fake-news spreading concepts with the compartments and parameters later evaluated in Chapter 3.

Table 1.4. Mapping between epidemic terminology and fake-news propagation.

Epidemic concept	**Disease interpretation**	**Fake-news interpretation**	**Observable proxy in social data**
Susceptible	Individual who can be infected	User who has not yet engaged with the claim but may be exposed	Follower relation, community membership, timeline exposure
Exposed	Individual infected but not yet infectious	User who has seen a claim but has not yet shared it	View/impression when available; delayed reply; time lag before repost

Infected	Individual currently transmitting disease	User actively spreading the claim	Retweet, repost, quote, comment, share
Recovered/Removed	Individual no longer infectious	User stops spreading, loses interest, or becomes corrected	No further sharing after a correction or after cascade decay
Skeptic	Not standard in SIR; used in rumor models	User exposed to the claim but not spreading it, or rejecting/debunking it	Debunking reply, correction, non-engagement after exposure

1.7.1 Classical Epidemic Models

Classical epidemic models divide a population into compartments and describe how individuals move between them over time. The SI, SIS, SIR and SEIR families are important because they provide the basic mathematical vocabulary for later fake-news spreading models: susceptibility, active spreading, recovery, repeated exposure and delayed activation.

The foundational SIR logic is associated with Kermack and McKendrick [31], while the role of network topology in spreading thresholds and diffusion persistence is developed in the work of Pastor-Satorras and Vespignani [71] and in the broader review of epidemic processes in complex networks by Pastor-Satorras et al. [72]. These studies are essential for this thesis because social platforms cannot be treated as homogeneous populations; central users, hubs and community structure can change both the speed and reach of a fake-news cascade.

The adaptation of epidemic thinking to social behaviour and information diffusion is further supported by Centola's online social-network experiment [74] and by Wang et al.'s study of social contagions with memory of nonredundant information [75]. These works show that online diffusion may depend on reinforcement, repeated exposure and memory effects. In this thesis, such observations motivate directed temporal cascade representation and leave path-memory extensions as future work rather than as implemented Chapter 3 experiments.

In fake-news and rumor contexts, Jin et al. modelled news and rumors on Twitter using epidemiological concepts [33], while PLOS studies by Kauk et al. [36] and Murayama et al. [39] demonstrate that fake-news spreading processes can be analysed through epidemic-style dynamics. Therefore, SI, SIS and SEIR are retained as theoretical reference models, whereas SIR and SEIZ are used as the main operational diffusion models because they provide an interpretable balance between mathematical clarity and relevance to observable fake-news cascades.

Table 1.5 supports the epidemic-modelling layer by linking fake-news spreading concepts with the compartments and parameters.

Table 1.5. Classical epidemic models and their interpretation for fake-news diffusion.

Model	Core states	Representative literature	Fake-news interpretation	Main strength and limitation
SI	S -> I	Pastor-Satorras and Vespignani [71]; Pastor-Satorras et al. [72]	Users move from potential exposure to active spreading.	Useful for early uncontrolled diffusion, but has no correction, forgetting or recovery state.
SIS	S -> I -> S	Pastor-Satorras et al. [72]	Users may stop spreading and later become susceptible again after repeated exposure.	Useful for recurrent sharing, but does not represent permanent correction.
SIR	S -> I -> R	Kermack and McKendrick [31]; Jin et al. [33]; Murayama et al. [39]	Users share fake news and later stop, forget, become corrected or are removed from the active cascade.	Interpretable baseline with transmission and recovery parameters, but no explicit exposure delay or skepticism.
SEIR	S -> E -> I -> R	Pastor-Satorras et al. [72]; Centola [74]; Wang et al. [75]	Users may observe content before deciding whether to share it.	Captures exposure delay, but does not explicitly represent skeptical or debunking behaviour.

1.7.2 Rumor-Spreading Models

Rumor-spreading models are important for the theoretical foundation of this thesis because they describe the circulation of unverified information rather than biological infection. Unlike classical disease models, rumor models explicitly include social communication, repeated contact, loss of interest, contradiction and the possibility that a user stops spreading after meeting someone who already knows or rejects the rumor [32], [67].

The Daley-Kendall model divides the population into three conceptual states: ignorants, spreaders and stiflers. Ignorants have not yet heard the rumor; spreaders actively transmit it; and stiflers know the rumor but no longer spread it. The transition from spreader to stifler occurs when a spreader contacts another spreader or stifler, because the interaction reduces novelty and weakens the motivation to continue transmission. For fake-news diffusion, this model is useful because it represents the decay of sharing activity after saturation, correction or social challenge [32].

The Maki-Thompson model preserves the same broad logic but simplifies the contact mechanism by assuming that spreading may cease when a spreader contacts someone who already knows the rumor. It is often used as a mathematically convenient rumor model and provides a bridge between epidemic-style compartmental modelling and communication-driven diffusion. In the context of fake news, the Maki-Thompson formulation highlights that the disappearance of a cascade may result not only from platform intervention, but also from conversational saturation and user-level rejection [67].

These classical rumor models are valuable because they introduce social states that are closer to fake-news behaviour than purely biological infection. However, they remain limited for modern online platforms because they do not explicitly model algorithmic ranking, bots, recommendation loops, fact-checking latency, cross-community amplification or hidden exposure. These limitations motivate the use of temporal networks, SIR/SEIZ variants and ABC-based uncertainty estimation in the proposed thesis framework.

Additional rumour and information-diffusion studies show why this limitation matters in online settings: small-world rumour processes, complex-network rumour theory, influence maximization, viral marketing, blogspace diffusion and outbreak-detection formulations all demonstrate that cascade reach depends on network structure, seed placement, timing and repeated exposure rather than only on message content [68]-[69], [156]-[159].

1.7.3 SIR-Based Fake-News Diffusion

The SIR model is retained as an interpretable baseline because it maps fake-news diffusion into three modelling states: susceptible users, active spreaders and removed or corrected users. Its main parameters are the transmission coefficient beta and the recovery/removal coefficient gamma, while R0 = beta/gamma summarizes early cascade growth under the homogeneous approximation.

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{1.9}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \tag{1.10}$$

$$\frac{dR}{dt} = \gamma I \tag{1.11}$$

Equations (1.9)-(1.11) describe susceptible decline, active-spreader growth and removal, and transition into the recovered, inactive or corrected state.

$$R_0 = \frac{\beta}{\gamma} \tag{1.12}$$

Equation (1.12) gives the homogeneous SIR reproduction number, interpreted here as the early-stage spreading potential of one active fake-news spreader.

$$\Pr[X_i(t + \Delta t) = I \mid X_i(t) = S] = 1 - \exp\{-\beta \Delta t \sum_j A_{ij}(t) \mathbf{1}[X_j(t) = I]\} \tag{1.13}$$

Equation (1.13) links SIR diffusion with a temporal network by making activation depend on active neighbours and the time-varying adjacency structure A(t).

Figure 1.2 supports the epidemic-modelling layer by linking fake-news spreading concepts with the compartments and parameters later evaluated in Chapter 3.

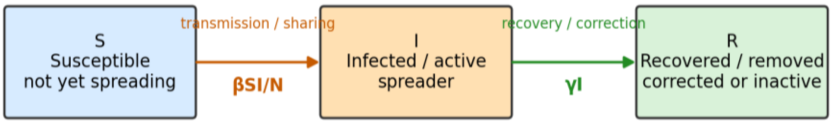

Figure 1.2. SIR state-transition graph for fake-news diffusion.

1.7.4 SEIZ and Advanced Fake-News Models

SEIZ extends SIR by adding exposed users and skeptical or non-spreading users. This makes it more suitable for fake-news cascades in which exposure, delayed adoption, questioning and correction can occur before or instead of reposting.

The SEIZ states are modelling proxies rather than direct psychological measurements: reposts indicate observable spreading, while silence or non-engagement cannot be equated directly with belief or skepticism [33], [73].

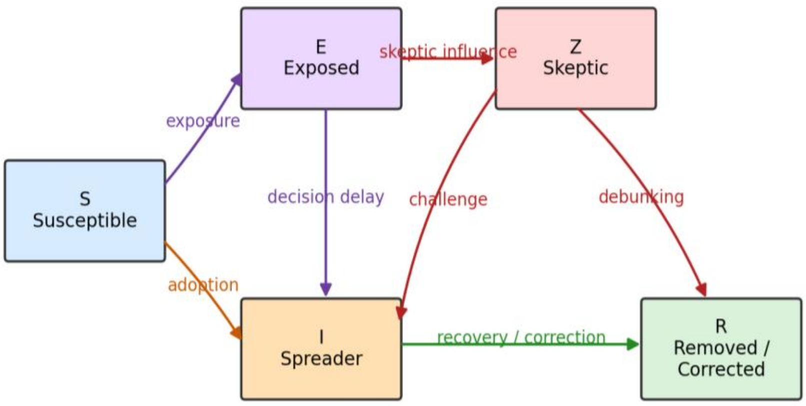

Figure 1.3. SEIZ state-transition graph for fake-news diffusion.

$$\frac{dS}{dt} = -\frac{\beta SI}{N} - \frac{bSZ}{N} \tag{1.14}$$

$$\frac{dE}{dt} = (1 - p)\frac{\beta SI}{N} + (1 - l)\frac{bSZ}{N} - \varepsilon E - \frac{\rho EI}{N} \tag{1.15}$$

$$\frac{dI}{dt} = p\frac{\beta SI}{N} + \varepsilon E + \frac{\rho EI}{N} - \gamma I - \delta I \quad (1.16)$$

$$\frac{dZ}{dt} = l\frac{bSZ}{N} + \delta I \quad (1.17)$$

Equations (1.14)-(1.17) summarize immediate adoption, delayed adoption, skeptical influence and correction. Thesis-specific content-aware SEIZ choices are defined in Chapter 2.

Table 1.6 supports the epidemic-modelling layer by linking fake-news spreading concepts with the compartments and parameters later evaluated in Chapter 3.

Table 1.6. Main coefficients in SIR and SEIZ fake-news diffusion models.

Coefficient	**Model**	**Meaning in fake-news diffusion**	**Possible empirical proxy**
β	SIR / SEIZ	Transmission rate from active spreaders to susceptible users	Repost rate, exposure rate, edge weight, influencer centrality
γ	SIR / SEIZ	Recovery, forgetting or stopping rate	Time until no further sharing; cascade decay
b	SEIZ	Contact rate between susceptible users and skeptical users	Replies by debunkers; corrective interactions
p	SEIZ	Probability of direct adoption after contact with a spreader	Immediate repost probability after first exposure
l	SEIZ	Probability of becoming skeptical after contact with skeptics	Correction replies, fact-checking exposure, non-engagement after debunking
ε	SEIZ	Transition rate from exposed to active spreading	Delay from exposure or reply to repost
ρ	SEIZ	Influence of active spreaders on exposed users	Repeated exposure; discussion threads that lead to adoption
δ	Advanced SEIZ	Correction or debunking rate	Warning labels, fact-checks, official corrections

1.7.5 Network-Based Epidemic Diffusion

Network-based epidemic diffusion extends compartmental modelling by replacing homogeneous mixing with structured contact. In online platforms, users are connected through follower links, retweets, replies, mentions, comments, quote posts and community membership. The probability of fake-news spread therefore depends not only on beta and gamma, but also on network position, temporal ordering and repeated exposure.

Scale-free and complex networks are especially important because highly connected users can create many exposure opportunities. A fake-news item shared by a central user, influencer or bridge node can cross communities more easily than the same item shared by a peripheral user. Temporal diffusion adds another layer: a static path is not sufficient for transmission unless the interactions occur in a valid chronological order.

For this reason, epidemic models used in fake-news research should be interpreted together with temporal-network metrics such as reachability, time-respecting paths, cascade depth, cascade breadth, spreading speed, centrality and community crossing. These metrics support the transition from descriptive cascade analysis to parameter inference in later chapters.

Empirical misinformation studies also show that exposure and sharing may be highly concentrated among specific users, platforms or polarized communities; this supports the thesis use of user influence, community structure, early propagation paths and cross-platform diffusion as contextual evidence rather than treating the population as homogeneous [60], [62]-[64], [168]-[170], [174]-[175], [179].

1.7.6 Limitations of Epidemic Analogies

Although epidemic models are useful, their analogy with fake-news diffusion has important limitations. Human behaviour is strategic, emotional and context-dependent. Users may share content ironically, reject it silently, read it without reacting, or engage with it only because the platform recommends it repeatedly. These processes are more complex than biological infection and cannot be captured fully by a small number of compartments.

A second limitation is hidden exposure. Most datasets record visible actions such as reposts, replies or comments, but they do not record all impressions, private messages, deleted posts or recommendation-system exposures. As a result, the observed cascade is only a partial trace of the underlying diffusion process. This makes fixed-parameter epidemic modelling vulnerable to overconfidence.

A third limitation concerns algorithmic amplification and domain dependence. Platform ranking systems, bots, coordinated behaviour and echo chambers may amplify fake news independently of its intrinsic textual credibility. Epidemic models are therefore most useful when they are combined with temporal-network analysis, Transformer-based content detection and uncertainty-aware inference. This motivates the use of Approximate Bayesian Computation in the following section.

1.8 Approximate Bayesian Computation for Diffusion Inference

Approximate Bayesian Computation (ABC) is a family of likelihood-free Bayesian methods designed for models that can be simulated but whose likelihood function is unavailable, analytically intractable, or computationally expensive to evaluate. In the context of fake-news diffusion, this distinction is essential: temporal cascades can be simulated under candidate spreading parameters, but the exact probability of observing a particular cascade is difficult to write because user exposure, private sharing, deletion, recommendation effects and belief change are only partially observed [19]-[25].

This section reviews ABC as part of the theoretical foundation of the dissertation. It intentionally avoids the detailed calibration workflow, prior design, distance functions, posterior prediction and intervention methodology, because these elements are part of the proposed methodological framework developed in Chapter 2. The purpose of the present section is to explain why ABC and ABC-SMC are appropriate for stochastic epidemic and social-network diffusion models, and what limitations remain in the literature.

1.8.1 Bayesian Inference and Intractable Likelihoods

Bayesian inference combines prior assumptions with observed data in order to estimate posterior uncertainty over model parameters. For a fake-news spreading model, the parameter vector may include transmission, recovery, exposure, skepticism and debunking rates. In principle, the posterior distribution can be written as:

$$p(\theta \mid D_{\text{obs}}) \propto p(D_{\text{obs}} \mid \theta)p(\theta) \quad (1.18)$$

In Equation (1.18), θ denotes the unknown parameter vector and D_{obs} denotes the observed cascade. The main difficulty is the likelihood term p(D_{obs} | θ). For temporal social-network cascades, the observed data rarely contain every exposure event, every private transmission, every deleted post, or every recommendation-driven view. Consequently, exact likelihood-based inference can be unrealistic even when the underlying spreading process is conceptually clear [24], [25], [76].

This problem places fake-news diffusion within the broader class of simulation-based inference problems. Reviews of ABC and simulation-based inference emphasize that such methods are useful when the forward simulator is available but the likelihood is not [23], [40], [26]. For the present thesis, this means that SIR- and SEIZ-style fake-news models can be simulated on temporal networks even when the complete probability model of the observed cascade cannot be evaluated directly.

1.8.2 ABC for Stochastic Epidemic Models

The basic ABC idea is to sample parameters from a prior distribution, simulate data from the model, compare simulated and observed data through summary statistics, and retain parameters that produce sufficiently similar simulations. In compact form, the ABC posterior approximation can be expressed as:

$$p_{\mathrm{ABC}}(\theta \mid D_{\mathrm{obs}}) \propto p(\theta)\Pr\{\rho\big(S(D_{\mathrm{sim}}), S(D_{\mathrm{obs}})\big) \leq \varepsilon \mid \theta\} \tag{1.19}$$

Equation (1.19) approximates the posterior by retaining parameter values whose simulated summary statistics are close to the observed cascade under a distance function and tolerance threshold. This is why ABC is relevant for stochastic epidemic models with difficult likelihoods [24], [25].

For fake-news diffusion, ABC is therefore used as a principle for uncertainty-aware comparison between observed and simulated cascades, not as a claim that hidden exposure or belief change is fully observed.

1.8.3 ABC-SMC Methods

ABC-SMC improves basic rejection ABC by moving through particle populations with decreasing tolerances, so later simulations concentrate on parameter regions that better reproduce the observed cascade [21], [22]. A simplified tolerance sequence can be written as:

$$\varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_T, \quad p_T(\theta) \approx p\big(\theta \mid \rho\big(S(D_{\mathrm{sim}}), S(D_{\mathrm{obs}})\big) \leq \varepsilon_T\big) \tag{1.20}$$

Equation (1.20) states the principle of progressive approximation. Pesonen et al. emphasize that future ABC applications require scalable simulation, careful summary-statistic design, diagnostic checks and transparent uncertainty reporting; this supports the use of ABC-SMC as an inference layer rather than a deterministic truth mechanism [77], [180].

1.8.4 ABC for Social-Network Diffusion

ABC is particularly relevant for social-network diffusion because information-spreading data are incomplete by design. Social platforms may record posts, reposts, comments, quote posts, replies and timestamps, but they rarely expose all impressions, algorithmic recommendations, private messages or silent exposures. This creates a gap between the visible cascade and the hidden diffusion mechanism. Network diffusion research has shown that reconstructing transmission pathways is difficult even when adoption times are observable, because the underlying influence network or exact sender-recipient pathway may be unknown [28].

Related work by Gomez-Rodriguez, Leskovec and Krause on inferring diffusion and influence networks, and by Farajtabar et al. on joint information diffusion and network co-evolution, demonstrates that information spread is not only a static graph problem but also a time-dependent stochastic process [28], [29]. These ideas are important for the present thesis because fake-news cascades require inference over both temporal structure and uncertain spreading parameters. ABC offers a way to connect the two: simulated cascades can be compared with observed cascades through summary statistics such as final size, time to peak, cascade depth, cascade breadth and spreading rate.

The literature also shows that false or misleading information can spread differently from reliable information and that online diffusion may be shaped by social reinforcement, platform structure and user behaviour [54], [53], [78]. This strengthens the motivation for combining content-based fake-news detection with uncertainty-aware diffusion inference. ABC does not solve annotation ambiguity or platform opacity by itself, but it provides a principled framework for representing uncertainty when the observed cascade is only a partial proxy for the underlying spreading process.

1.8.5 Challenges and Limitations of ABC

The first major limitation of ABC is summary-statistic selection. If the chosen statistics do not preserve the information needed to identify the parameters, the resulting posterior may be misleading. Work on semi-automatic ABC and dimension reduction emphasizes that summary statistics are not a neutral technical detail; they determine what aspects of the data are used for inference [79], [80]. In fake-news diffusion, cascade size alone is insufficient because two cascades may have the same final size but different speeds, depths, community transitions or correction dynamics.

The second limitation is identifiability. Different combinations of transmission, exposure, recovery, skepticism and debunking parameters may produce similar visible cascades, especially when exposure is hidden and timestamps are incomplete. The third limitation is computational complexity: ABC requires repeated simulation, and simulation cost increases when the model includes temporal networks, heterogeneous users, path-memory extensions or intervention scenarios [25], [26], [27].

The fourth limitation is model misspecification. An ABC posterior is conditional on the simulator used to generate synthetic cascades. If the SIR or SEIZ simulator omits important platform mechanisms, coordinated behaviour or recommendation effects, the posterior may fit observed summary statistics while still misrepresenting the causal process. Therefore, in this dissertation, ABC is treated as an uncertainty-aware inference tool, not as a guarantee of complete causal explanation. This review motivates the methodological choices developed in Chapter 2, where ABC is operationalized for epidemic parameter inference from fake-news temporal cascades.

The uncertainty interpretation is consistent with Bayesian and approximate Bayesian deep-learning literature, where posterior or ensemble uncertainty is used to represent predictive variability rather than a single deterministic truth. Dropout-as-Bayesian approximation, deep ensembles, Bayes-by-Backprop and SWAG therefore support the thesis distinction between point predictions, posterior parameter regions and uncertainty-aware diffusion forecasts [130]-[132], [165]-[167].

The literature review is therefore kept focused on concepts and methods that directly support the implemented Transformer, temporal-network, SIR/SEIZ and ABC/ABC-SMC framework. Detailed experimental material is reserved for the methodology and results chapters, where the implemented datasets and modelling choices are defined explicitly.

1.9 Research Gaps and Open Problems

The reviewed literature demonstrates substantial progress in fake-news detection, temporal-network analysis, epidemic spreading models and simulation-based Bayesian inference. Nevertheless, these research directions remain only partially connected. The resulting gap is not simply the absence of a single algorithm, but the absence of a coherent modelling structure that links semantic credibility, temporal diffusion, parameter uncertainty and intervention-oriented prediction. This section synthesizes the main open problems that motivate the proposed framework developed in Chapter 2.

Building on Section 1.5, the proposed methodology does not include ensemble learning as a central methodological component. Ensemble and hybrid systems remain part of the literature review because they motivate robustness and calibration, but the operational detection layer in this thesis is based on selected individual Transformer models. The only methodological connection to the ensemble literature is the use of calibrated fake-news probability scores, so that $p_f(x) = P(y = \text{fake} \mid x)$ can serve as a reliable content-risk signal for temporal-network construction, SIR/SEIZ diffusion modelling and ABC/ABC-SMC inference.

1.9.1 Separation Between Detection and Diffusion

A first limitation is the persistent separation between content-level fake-news detection and network-level diffusion analysis. Detection systems typically estimate whether a text, claim or article is fake, while diffusion studies often analyse how information spreads without using modern semantic credibility signals. This disconnect weakens early-warning capacity: a post may have a high fake-news probability but limited diffusion potential, whereas another post with moderate textual uncertainty may become socially harmful if it reaches influential users or polarized communities early in the cascade. A complete risk model therefore has to connect content risk with spread risk rather than treating them as independent outputs.

1.9.2 Temporal-Network Limitations

A second open problem concerns the simplification of social diffusion through static graph representations. Aggregated graphs are useful for describing general connectivity, but they remove the chronological order of interactions. Fake-news propagation depends on time-respecting paths, interaction delays, user activity bursts and repeated exposures across communities. Static representations may therefore suggest diffusion paths that are impossible in time, or they may hide early cross-community transitions that explain rapid cascade growth. Temporal and higher-order network representations are necessary for modelling fake-news diffusion as a dynamic process rather than as a static topology.

1.9.3 Uncertainty of Epidemic Parameters

A third gap is the uncertainty of epidemic-style parameters. SIR and SEIZ models require transmission, exposure, recovery, skepticism and debunking coefficients, but these quantities are rarely directly observable in social-network data. A repost is visible, but silent exposure, private sharing, recommendation-system exposure, reading without engagement and belief change are usually hidden. Consequently, fixed-parameter diffusion models may produce apparently precise results while underrepresenting uncertainty. ABC and ABC-SMC address this limitation by estimating posterior parameter distributions from simulated and observed cascade summaries, but their integration with Transformer-based fake-news detection remains underdeveloped.

1.9.4 Dataset and Annotation Limitations

A fourth limitation is dataset and annotation fragmentation. Content-oriented datasets provide labels and text but often lack user-level temporal interactions. Diffusion-oriented datasets provide propagation structures but may have weaker content labels, incomplete hydration, missing timestamps or rumour-oriented labels that do not correspond exactly to fake/real news classification. Annotation ambiguity further complicates evaluation because doubtful, partially false, satirical and evolving claims may be collapsed into binary categories. These issues show that dataset construction is not only a technical preprocessing problem; it is a methodological problem that affects model validity, reproducibility and generalization.

Existing benchmark datasets remain fragmented. Some datasets are strong for textual fake-news classification, while others are useful for temporal diffusion and propagation analysis. However, few datasets simultaneously provide reliable fake/real labels, full textual content, temporal interactions, cascade structure, fact-checking signals, and user-level diffusion traces. This limitation motivates the proposed integrated framework, which combines content-based Transformer detection with temporal-network, SIR/SEIZ, and ABC-based diffusion inference.

1.9.5 Lack of Intervention-Oriented Frameworks

A fifth gap is the limited connection between prediction and actionable intervention modelling. Many detection systems return a label or probability score, and many diffusion models estimate cascade growth, but practical mitigation requires asking how fact-checking, warning labels, visibility reduction, debunking responses or early containment may change future spread. Intervention analysis requires uncertainty-aware counterfactual simulation, because the effect of an intervention depends on temporal position, user influence, cascade phase and unknown epidemic parameters. Existing work often treats these components separately, which limits its usefulness for decision support.

1.9.6 Motivation for the Proposed Thesis

The synthesis of these gaps motivates the central direction of this dissertation: fake-news analysis should combine content-level detection, temporal-network diffusion, epidemic-state modelling and Bayesian uncertainty estimation. The integrated perspective can be summarized conceptually by combining a content-risk component with a cascade-risk component:

$$R(x, C) = \alpha P(\text{fake} \mid x) + (1 - \alpha)\Phi(C, G, \theta) \tag{1.21}$$

In Equation (1.21), R(x,C) denotes conceptual integrated risk, P(fake|x) denotes the fake-news probability estimated from textual content x, C denotes the observed temporal cascade, G denotes the temporal network, θ denotes uncertain epidemic parameters and Φ(.) denotes a normalized diffusion-risk function derived from cascade and network properties. The equation is retained as a compact conceptual expression of the research gap; the operational modelling choices, calibration procedures and intervention scenarios are developed in Chapter 2.

Synthesis Statement 1.8.1 (Need for integrated detection-diffusion inference). If two fake-news items have similar content-level fake-news probabilities but different temporal-network positions and epidemic parameters, then content classification alone cannot determine which item has higher diffusion risk. Conversely, if two cascades have similar early sizes but different content credibility probabilities, diffusion modelling alone cannot determine which cascade should be prioritized for fact-checking or mitigation.

The purpose of this synthesis statement is to clarify why the thesis cannot be reduced either to a Transformer classifier or to an epidemic simulator. The proposed framework must jointly consider what the content is, where it appears in the network, how it spreads over time, and how uncertain the diffusion parameters are.

Figure 1.4 supports the content-detection layer of the framework by identifying how textual evidence is prepared or interpreted before diffusion analysis.

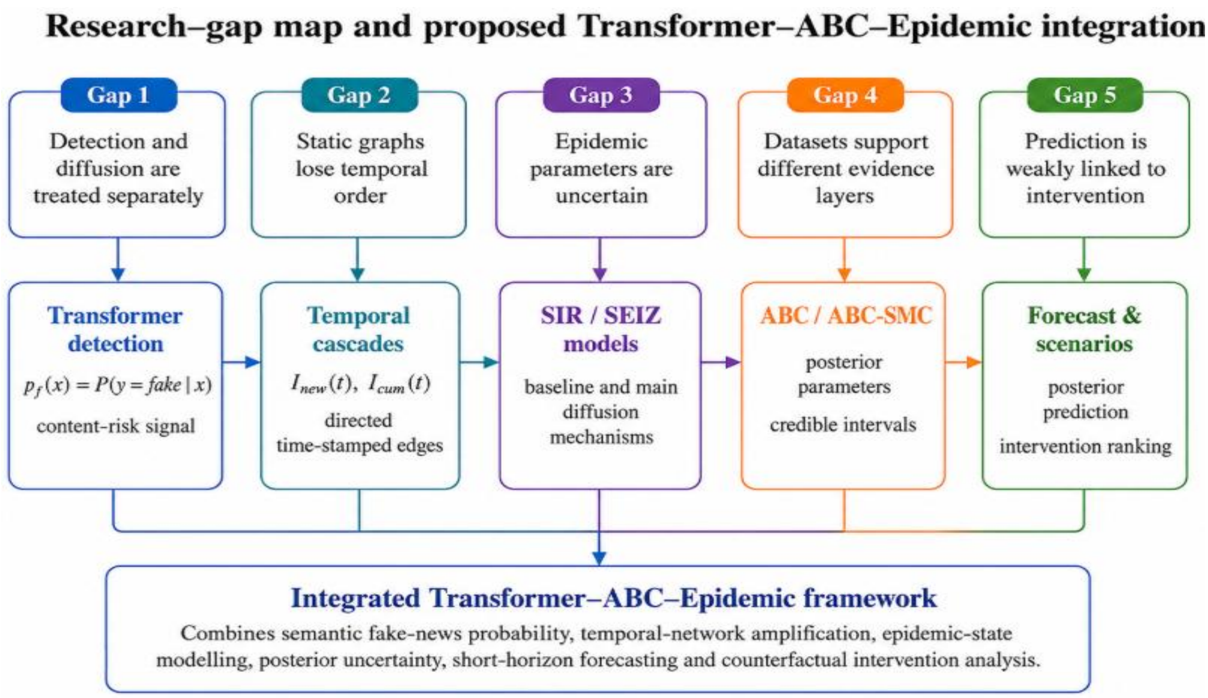

Figure 1.4. Research-gap map and proposed Transformer-ABC-Epidemic integration.

1.9.7 Research Questions Derived from the Literature Gaps

The literature gaps identified in Sections 1.9.1-1.9.6 lead to the following research questions, which guide the methodology in Chapter 2 and the experimental evidence in Chapter 3:

RQ1. How effectively can Transformer models classify fake and real news across short-text, long-text, multilingual and low-resource datasets?

RQ2. How can fake-news claims be represented as temporal cascades suitable for epidemic-style modelling?

RQ3. Does SEIZ provide a more suitable representation than SIR for fake-news diffusion when exposure delay and skepticism are relevant?

RQ4. Can ABC and ABC-SMC estimate uncertainty in fake-news diffusion parameters from partially observed cascades?

RQ5. Can a combined Transformer-temporal-network-ABC framework support posterior prediction and intervention-oriented fake-news risk analysis?

1.9.8 Research Gap, Methodological Response and Expected Evidence

Table 1.7 makes the traceability between the literature gaps, the Chapter 2 methodology and the Chapter 3 evidence explicit.

This traceability table clarifies that the dissertation does not add independent methods only for coverage. Each methodological component is retained because it answers a specific gap and produces evidence that can be evaluated in the results chapter.

Table 1.7. Research gaps, methodological responses and expected evidence.

Research gap	Methodological response in Chapter 2	Expected evidence in Chapter 3	Contribution to the dissertation
Detection and diffusion are often treated separately.	Use Transformer fake-news probability as a content-risk signal and link it to temporal cascade construction.	Detection metrics, claim-level probabilities and cascade eligibility outputs.	Connects content credibility with diffusion-risk analysis.
Static graphs can hide time-respecting diffusion paths.	Represent claims as directed temporal cascades with ordered interactions and time-indexed spreading curves.	$I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, cascade size, time to peak and temporal-network figures.	Preserves the chronology required for fake-news diffusion modelling.
SIR/SEIZ parameters are uncertain and only partially observable.	Estimate parameter uncertainty through ABC and ABC-SMC using observed cascade summaries.	Posterior estimates, credible intervals and posterior predictive calibration.	Provides uncertainty-aware epidemic-model inference for fake-news cascades.
Content datasets and diffusion datasets support different forms of evidence.	Separate content-oriented datasets from diffusion-oriented datasets and define explicit dataset roles.	Distinct detection experiments and diffusion experiments with non-interchangeable dataset use.	Improves methodological validity and prevents unsupported dataset claims.
Mitigation analysis requires more than classification labels.	Use posterior prediction and counterfactual intervention scenarios under ABC-calibrated diffusion models.	Forecasts, intervention scenarios and baseline diffusion-risk functional outputs.	Supports intervention-oriented reasoning from the integrated framework.

1.10 Chapter Conclusion

This chapter reviewed the theoretical foundations required for the dissertation. It showed that fake-news research has developed through several complementary directions: conceptual definitions and annotation theory, classical machine learning, Transformer-based text classification, contextual ensemble and hybrid approaches, temporal-network diffusion, epidemic-state modelling and likelihood-free Bayesian inference.

The synthesis developed in this chapter shows that the main limitation of existing research is not the absence of strong individual techniques. Rather, the limitation is the insufficient integration of techniques that operate at different analytical levels. This motivates the dissertation framework, which connects Transformer-based detection, temporal-network cascade construction, SIR/SEIZ diffusion modelling and ABC/ABC-SMC uncertainty-aware inference.

Table 1.8. Compact synthesis of reviewed literature and relevance to the proposed thesis.

Literature direction	**Main strength**	**Remaining limitation**	**Relevance to this thesis**
Fake-news definitions and annotation	Clarifies fake, doubtful, misleading and related concepts.	Labels may be subjective, incomplete or unstable over time.	Motivates careful binary mapping and discussion of annotation ambiguity.
Classical machine learning	Interpretable, efficient and useful as baseline methodology.	Limited contextual understanding and weak transfer under domain shift.	Provides historical background and explainable comparison points.
Transformer-based detection	Captures contextual semantics and supports probability scoring.	Requires computation, data quality and domain adaptation.	Forms the main content-detection layer of the proposed framework.
Probability calibration and robustness	Supports reliable interpretation of model scores and cross-dataset stability.	Does not replace the need for temporal diffusion modelling.	Motivates use of calibrated Transformer probability as an input to the diffusion-risk layer.
Temporal-network analysis	Preserves ordering, delays and time-respecting paths.	Often disconnected from content credibility labels.	Supports cascade construction and temporal diffusion and path-memory features.
Epidemic and rumor models	Represent spreading states, thresholds and cascade dynamics.	Human behaviour and hidden exposure are only approximated.	Provides SIR and SEIZ foundations for fake-news propagation analysis.

ABC and ABC-SMC inference	Estimates uncertain parameters when likelihoods are intractable.	Sensitive to summaries, identifiability and computational cost.	Enables uncertainty-aware calibration of SIR/SEIZ diffusion models.
Intervention-oriented modelling	Connects prediction with mitigation strategies.	Often absent from standard detection-only approaches.	Motivates posterior prediction and intervention analysis in Chapter 2.

ABC provides this calibration only when it is connected to observable content signals, temporal cascades and epidemic-state assumptions.

The review has established that fake-news analysis requires more than isolated textual classification. It has shown that content-based detection, fake-news temporal diffusion modelling, epidemic-state representation, and uncertainty-aware Bayesian inference address complementary aspects of the same problem. At the same time, the reviewed literature also demonstrates that existing approaches remain limited by the separation between detection and diffusion, the simplification of dynamic social interactions into static graphs, the uncertainty of epidemic parameters, the incomplete alignment between content datasets and cascade datasets, and the limited ability of many systems to support intervention-oriented analysis.

These limitations justify the need for a new integrated model. The proposed dissertation framework therefore aims to connect Transformer-based fake-news probability estimation with temporal-network representation, SIR/SEIZ spreading dynamics, and ABC/ABC-SMC parameter inference. This synthesis directly leads to the dissertation goal and research tasks: to develop, implement, and evaluate an integrated Transformer–ABC–Epidemic framework for fake-news detection, diffusion modelling, uncertainty-aware prediction, and intervention analysis.

❖ The goal of the dissertation is:

To develop and evaluate an integrated framework for fake-news detection and diffusion analysis on social networks by combining Transformer-based language models, temporal-network representation, SIR/SEIZ epidemic modelling, and Approximate Bayesian Computation for uncertainty-aware parameter inference. The dissertation aims to analyse fake news not only as a textual classification problem, but also as a dynamic diffusion process in which false or misleading content spreads through time-dependent social interactions.

The goal encompasses the analysis of existing research methodologies for fake-news detection, temporal-network diffusion, epidemic-style propagation modelling, and likelihood-free

Bayesian inference. It is interconnected with the development of a hybrid Transformer–ABC–Epidemic framework, where Transformer models provide content-level fake-news probability scores, temporal networks represent the structure and timing of social spread, SIR and SEIZ models describe diffusion behaviour, and ABC/ABC-SMC methods estimate uncertain epidemic parameters from partially observed cascades. The objective is to provide a reproducible and interpretable framework that supports fake-news classification, cascade modelling, posterior prediction, uncertainty-aware analysis, and intervention-oriented scenario evaluation.

In this regard, the goal of the dissertation is addressed through the following research tasks:

Task 1: To evaluate the state of the art in fake-news detection, temporal-network analysis, epidemic-style information diffusion, and Approximate Bayesian Computation, emphasizing their relevance to social-network misinformation analysis.

Task 2: To analyse existing fake-news datasets and distinguish between content-based datasets suitable for Transformer classification and diffusion-oriented datasets suitable for temporal cascade reconstruction, SIR/SEIZ modelling, and ABC/ABC-SMC inference.

Task 3: To implement and evaluate a Transformer-based binary fake-news detection layer using models such as BERT, RoBERTa, BigBird, mBERT, and XLM-RoBERTa across short-text, long-text, healthcare-related, English, and low-resource Albanian fake-news datasets.

Task 4: To construct directed temporal fake-news cascades from social-network interaction data and extract key diffusion quantities, including new active spreaders, cumulative active spreaders, cascade size, time to peak, and temporal-network structure.

Task 5: To compare SIR and SEIZ epidemic models for fake-news diffusion modelling and determine whether exposed and skeptic/non-spreading states provide a more suitable representation of fake-news propagation than a simpler SIR baseline.

Task 6: To apply ABC and ABC-SMC methods for uncertainty-aware inference of epidemic diffusion parameters from partially observed fake-news cascades, including posterior parameter estimation, credible intervals, posterior predictive validation, and short-horizon forecasting.

Task 7: To integrate Transformer fake-news probability, temporal-network amplification, SEIZ-based diffusion modelling, and ABC-SMC posterior predictive growth into a Baseline Posterior Fake-News Diffusion-Risk Functional, and to use the calibrated framework for cautious counterfactual intervention scenario analysis.

Through these tasks, the dissertation demonstrates that fake-news analysis should move beyond isolated classification accuracy and should instead be studied as an integrated content-to-cascade process involving textual credibility, temporal diffusion, epidemic-state modelling, posterior uncertainty, and intervention-oriented reasoning.

Chapter 2. Proposed Models and Methodology develops this framework in operational form. It defines the data layer, Transformer-based detection layer, temporal-network representation, SIR and SEIZ diffusion models, ABC/ABC-SMC inference procedure, posterior prediction strategy, and intervention-oriented methodology used in the remainder of the dissertation.

Chapter 2. Proposed Transformer–ABC–Epidemic Framework and Methodology.

2.1 Overall Proposed Framework

This chapter presents the proposed model and methodology of the dissertation. Its purpose is to transform the theoretical foundations reviewed in Chapter 1 into an operational research framework for detecting fake news, reconstructing temporal cascades, modelling diffusion and estimating uncertain parameters.

The proposed framework is organized into four connected layers. The first layer is the data layer, where content-based and diffusion-oriented datasets are selected, cleaned, harmonized and represented. The second layer is the Transformer-based binary detection layer, where models such as BERT, RoBERTa, mBERT, XLM-RoBERTa, SBERT and BigBird generate fake-news probability scores. The third layer is the temporal-network and epidemic modelling layer, where directed temporal cascades are mapped to SIR and SEIZ states. The fourth layer is the ABC inference layer, where uncertain diffusion parameters are estimated from observed cascade summaries.

The proposed methodology uses single Transformer and classical ML baselines rather than voting or stacking combinations. The operational pipeline therefore uses the Transformer output as a model-estimated content-risk probability, $p_f(x) = P(y = \text{fake} \mid x)$, for the later temporal diffusion layer. When validation behaviour supports calibration, this score can be interpreted as a calibrated probability; otherwise it is treated as a risk score that must not be confused with ground truth.

The methodological contribution is not a simple sequence of independent models. Instead, the framework links content risk with diffusion risk: Transformer outputs are used as content-level signals, temporal networks describe how information moves through users and communities, SIR and SEIZ represent spreading dynamics, and ABC/ABC-SMC quantifies uncertainty in the estimated epidemic parameters.

The chapter also defines dataset selection, preprocessing, binary label harmonization, temporal cascade construction, network representation, temporal cascade network modelling, ethical handling of social-network data and evaluation methodology.

Table 2.1. Methodological traceability matrix for the implemented thesis framework.

Research question / objective	**Methodological component**	**Dataset type**	**Chapter 3 evidence or output**

RQ1: binary fake-news detection across text types and languages	BERT, RoBERTa, BigBird, mBERT, XLM-RoBERTa and classical baselines	Constraint@AAAI2021, FRN, ISOT, CoAID and Albanian Fake News Corpus	Accuracy, precision, recall, F1-score, ROC-AUC and error analysis
RQ2: temporal cascade representation of fake-news claims	Timestamp normalization, directed edges, $I_{\mathrm{new}}(t)$, $I_{\mathrm{cum}}(t)$ and cascade features	COVID-19 Twitter/X cascade, Hoaxy / OSoMeNet, Hoaxy-imported cascade and synthetic Albanian cascade	Cascade size, depth, breadth, time to peak and temporal-network graphs
RQ3: SIR versus SEIZ diffusion modelling	SIR baseline and SEIZ state mapping for active spreading, exposure and skepticism	Principal empirical fake-news cascades	Fitted spreading curves, compartment trajectories and model-comparison tables
RQ4: uncertainty-aware epidemic parameter inference	ABC and ABC-SMC posterior inference using observed cascade summaries	Empirical diffusion-oriented cascades	Posterior estimates, credible intervals, convergence and calibration summaries
RQ5: posterior prediction and intervention analysis	Posterior predictive simulation, short-horizon forecasting and counterfactual parameter perturbation	Empirical cascades with extracted time series	Forecast outputs, intervention scenarios and baseline diffusion-risk functional evidence

Algorithm 2.1. Transformer-ABC-Epidemic pipeline.

Input: content-based fake-news datasets, diffusion-oriented cascade records and documented preprocessing rules.

1. Clean textual fields and harmonize labels into an auditable binary fake/real representation.

2. Tokenize content according to the selected Transformer family and evaluate the binary detection layer.

3. Extract or record the model-estimated fake-news probability $p_f(x)$ for eligible claims.

4. Preserve timestamps, user or source-target identifiers, cascade identifiers and action types for diffusion-oriented records.

5. Construct directed temporal cascades and compute $I_{\mathrm{new}}(t)$, $I_{\mathrm{cum}}(t)$ and network summary statistics.

```
6.    Fit SIR and SEIZ diffusion models to the empirical spreading curves.
7.    Run ABC and ABC-SMC inference using observed cascade summaries and
simulator-generated trajectories.
8.    Generate posterior predictive curves, short-horizon forecasts and
counterfactual intervention scenarios.
Output: classification evidence, cascade representation, epidemic parameter
estimates, posterior uncertainty and intervention-oriented diffusion-risk
analysis.
```

2.2 Dataset Selection Criteria

Dataset selection is a critical methodological stage because it determines the validity, reproducibility, and interpretability of the experiments carried out in this thesis. The revised thesis is binary-focused at the experimental level: the main detection task is the classification of news items as fake or real. Multiclass and doubtful-news classification are retained as background concepts in the literature review, but they are not treated as the main experimental contribution. Therefore, the selected datasets must primarily support reliable binary fake-news detection and, where possible, temporal-network modelling for SIR, SEIZ, ABC, and ABC-SMC analysis.

The dataset strategy follows the revised four-chapter organization of the thesis. Chapter 2 defines the empirical foundation and methodology, while Chapter 3 reports the real experimental results for binary Transformer-based detection, temporal cascade construction, epidemic modelling and Bayesian inference. The selection criteria therefore separate datasets appropriate for supervised text classification from datasets appropriate for reconstruction of social-network diffusion.

For content-based binary fake-news detection, a suitable dataset should contain labelled textual content, a transparent annotation or fact-checking procedure, sufficient representation of fake and real classes, stable train-validation-test partitions, and adequate coverage of topic domain, text length, and language. In the binary-focused version of this thesis, the principal content-oriented datasets are Constraint@AAAI2021 for COVID-19 misinformation, FRN for long-form fake and real news, ISOT for large-scale binary fake-news benchmarking, CoAID for COVID-19 health-related fake news, and the Albanian Fake News Corpus for low-resource multilingual evaluation [30], [81]-[84], [85].

For diffusion-oriented analysis, dataset requirements are stricter. SIR, SEIZ, ABC and ABC-SMC modelling require observable temporal signals, user interactions, cascade membership and network structure. The retained Chapter 3 diffusion layer therefore uses the COVID-19 real Twitter/X “The coronavirus is not only affecting the way we live, it’s also dramatically affecting the way we die.” fake-news claim cascade, the Hoaxy / OSoMeNet political fake-news claim cascade, the Hoaxy-imported social fake-news claim “Should One Tweet End a Career?”, and the synthetic Albanian temporal cascade. These datasets are retained because they support time-windowed spreading curves, directed user-to-user relationships and the $I_{\text{new}}(t)$ and $I_{\text{cum}}(t)$ quantities required for SEIZ and ABC/ABC-SMC modelling [90], [86]-[89].

Dataset documentation is also treated as a selection criterion. A dataset should make clear how the data were collected, how labels were assigned, what biases may exist, how the data may be accessed, and which ethical or platform restrictions apply. Documentation is essential in fake-news research because a poorly described dataset may lead to unreliable label interpretation, weak reproducibility, inappropriate generalization, or ethically problematic handling of user-level social-network records [91]-[94]. This selection logic follows datasheet and data-card principles, in which each dataset is documented by provenance, intended use, label construction, access constraints, ethical limits and known limitations [91], [185].

This documentation requirement is reinforced by recent dataset and misinformation reviews, including dataset-assessment surveys and social-media disinformation reviews, which show that text, labels, social context and temporal traces are unevenly available across fake-news corpora and COVID-19 multilingual/multimodal repositories [5], [113]-[114], [140]-[142].

Figure 2.1 clarifies how the selected data source is represented in the thesis and why it is suitable for the implemented detection or diffusion layer.

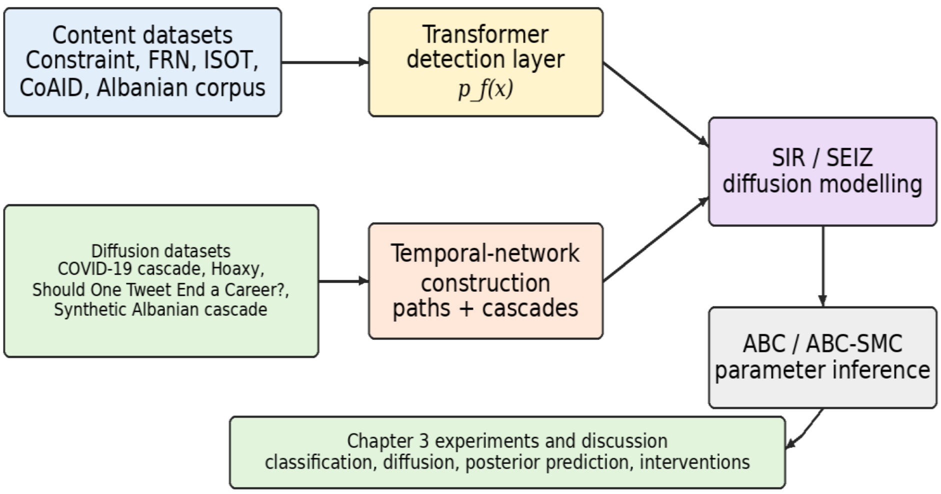

Figure 2.1. Dataset and methodology selection logic for the four-chapter Transformer-ABC-Epidemic framework.

In addition to the qualitative criteria in Table 2.2, a simple suitability score can be used as a guiding rubric. The score is not intended to replace expert judgement; rather, it helps compare datasets according to the requirements of binary detection and diffusion modelling.

Table 2.2. Dataset selection criteria and relevance to the proposed binary-focused thesis framework.

Criterion	Selection requirement	Role in binary detection	Role in diffusion / ABC
Label and text quality	Reliable fake/real labels and accessible text, title, claim, post, or article content.	Required for supervised training and model comparison.	Defines which cascades are treated as misinformation or reliable content.
Binary class balance	Sufficient representation of fake and real classes, or a documented imbalance strategy.	Improves F1-score, precision, recall, and error analysis.	Avoids modelling only the largest or most visible fake-news cascades.
Domain, language, and length	Coverage of health, politics, general news, long articles, short posts, and low-resource languages.	Supports cross-domain, long-text, and multilingual Transformer evaluation.	Enables comparison of diffusion behaviour across misinformation domains.
Temporal information	Publication time, tweet time, retweet delay, reply time, or event chronology.	Useful for early-detection experiments.	Essential for cascade growth, time-to-peak, and posterior prediction.
Interaction and cascade structure	Retweets, replies, reposts, comments, propagation trees, or user-engagement records.	Can be used as metadata when labels and text are available.	Required for temporal networks, state transitions, and SIR/SEIZ calibration.
Ethics and reproducibility	Licensing, privacy constraints, platform policy, anonymization, and dataset documentation.	Ensures responsible use of labels and text.	Ensures responsible use of temporal-network and user-level data.

$$Q(D) = \sum_k w_k \, C_k(D), \quad k \in \{\text{text}, \text{label}, \text{binary}, \text{time}, \text{network}, \text{ethics}\} \tag{2.1}$$

In Equation (2.1), Q(D) represents the overall suitability score of dataset D. The term $C_{\text{text}}(D)$ measures textual availability and quality; $C_{\text{label}}(D)$ measures label reliability; $C_{\text{binary}}(D)$ measures compatibility with the binary fake/real task; $C_{\text{time}}(D)$ measures temporal information; $C_{\text{network}}(D)$ measures whether user interactions or cascades can be reconstructed; and $C_{\text{ethics}}(D)$ measures documentation, access, privacy, and responsible-use constraints. The weights can be adjusted depending on whether the experiment is primarily a Transformer detection experiment or a diffusion-inference experiment.

Datasets were selected according to text quality, label reliability, temporal information, diffusion suitability and reproducibility constraints. The supplementary suitability classification is provided in Appendix F.

2.3 Principal Datasets

2.3.1 Content-Based Datasets for Transformer Detection

This section briefly introduces the datasets used in the dissertation and their methodological role.

The content-oriented datasets support the Transformer fake-news detection layer. Constraint@AAAI2021 provides short COVID-related misinformation examples suitable for evaluating Transformer models on short social-media texts and social-media-style classification. Its short and noisy social-media style also reflects realistic pandemic-information conditions. FRN and ISOT provide larger collections of fake and real news articles with different document lengths. CoAID focuses on health misinformation, while the Albanian Fake News Corpus supports multilingual evaluation in a low-resource setting.

Table 2.3. Comparative overview of principal binary content-based datasets used in the thesis.

Dataset	Key characteristics	Main use in this thesis
Constraint@AAAI2021	COVID misinformation, short text	Binary fake-news detection in a crisis-information setting.
FRN	general fake/real classification	Long-text fake-news detection and BigBird evaluation.
ISOT	long-form news articles	Large-scale binary benchmarking and cross-domain robustness testing.
CoAID	COVID-related content	Healthcare misinformation detection and partial social-context analysis.
Albanian Corpus	low-resource language	Multilingual Transformer evaluation in a low-resource setting.

2.3.2 Diffusion-Oriented Datasets for Temporal Cascade Modelling

Section 2.3.2 briefly introduces the diffusion-oriented datasets used in Chapter 3. The retained datasets are: the COVID-19 real Twitter/X “The coronavirus is not only affecting the way we live, it’s also dramatically affecting the way we die.” fake-news claim cascade; the Hoaxy / OSoMeNet “Three million votes in the Presidential Election were cast by illegal aliens.” fake-news claim cascade; the Hoaxy-imported "Should One Tweet End a Career?" social fake-news claim cascade; and the synthetic Albanian fake-news claim "Shqipëria do të presë 100 mijë palestinezë nga Gaza." These datasets provide timestamped user-to-user or claim-level records for directed temporal cascades, $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, SIR/SEIZ inputs and ABC/ABC-SMC summaries. Detailed descriptions, filtering rules and extended metadata are provided in Appendix D.

2.3.3 Summary of Dataset Roles

The selected datasets are intentionally divided into content and diffusion groups. Section 2.3.2 presents four different datasets that are related to diffusion-oriented analysis. This prevents the methodological error of assuming that all fake-news datasets are equally suitable for both classification and epidemic modelling. Data augmentation through back-translation and controlled paraphrasing was used only for training support. Detailed dataset descriptions, dataset cards, and extended metadata are provided in Appendix F.

2.4 Data Preprocessing and Experimental Preparation

Data preprocessing is a central methodological step in this dissertation because the proposed framework combines binary fake-news detection with temporal-network diffusion modelling and ABC-based inference. The purpose of preprocessing is not only to clean textual inputs for Transformer-based classifiers, but also to preserve the interaction and timestamp information required for later cascade construction. Therefore, preprocessing is treated as an operational bridge between the dataset descriptions in Section 2.3 and the Chapter 3 experiments on Transformer-based detection, temporal cascade construction, SIR/SEIZ diffusion modelling and ABC/ABC-SMC posterior inference.

The revised thesis direction adopts binary fake/real classification as the main experimental detection task. Multiclass and doubtful-news classification remain relevant in the literature review, but they are not treated as the principal experimental contribution. The preprocessing design therefore has two coordinated outputs: content-ready records for Transformer-based fake-news detection and diffusion-ready records in which user identifiers, timestamps, interactions and cascade identifiers are preserved for temporal modelling.

A single preprocessing policy is applied throughout this section. Ambiguous labels such as unverified are excluded from the main binary experiments and used only in sensitivity analysis. Generated or augmented examples are restricted to the training subset to prevent leakage into validation or test data. This policy is stated once here to avoid repetition across the subsection-level descriptions.

Table 2.4 shows that preprocessing is not a uniform text-cleaning operation. It is a methodological layer that prepares two different forms of evidence: content evidence for binary detection and temporal evidence for diffusion modelling.

Table 2.4. Preprocessing objectives for binary fake-news detection and temporal diffusion analysis.

Preprocessing objective	Relevant datasets	Methodological purpose
Text reliability	Constraint@AAAI2021, FRN, ISOT, CoAID, Albanian Fake News Corpus	Remove textual noise while preserving credibility signals for binary classification.
Label consistency	All content-based datasets	Map heterogeneous labels into a common fake/real representation for Chapter 3 experiments.
Long-text readiness	FRN and ISOT	Prepare long-form articles for Transformer input, including BigBird or window-based strategies.
Low-resource normalization	Albanian Fake News Corpus	Preserve Albanian-specific characters, named entities and local vocabulary for multilingual Transformers.
Temporal integrity	Diffusion-oriented datasets	Preserve timestamps, event order and user-interaction structure for cascade construction.
Reproducibility and leakage prevention	All datasets	Document cleaning, split and augmentation rules so that experimental results can be reproduced.

2.4.1 Text Cleaning and Normalization

Text cleaning and normalization are required because fake-news datasets are collected from heterogeneous sources, including social-media posts, fact-checking sites, online news articles, Twitter/X records and low-resource language corpora. Such data may contain duplicates, broken links, usernames, hashtags, emojis, irregular punctuation, inconsistent casing, incomplete fields and platform-specific metadata. These characteristics are common in online and social-media language, which is often informal and noisy [86], [87].

The goal of cleaning is not to remove all irregularity from the data. Some irregular elements may carry important information: a URL can indicate a linked article, a hashtag can signal a topic or campaign, and repeated punctuation can reflect sensational framing. Therefore, this thesis applies conservative cleaning: irrelevant noise is removed or standardized, while credibility-related and diffusion-related signals are preserved when they may support modelling [86], [88], [89].

Table 2.5 is retained in compact form because it states the main cleaning decisions that affect both Transformer input quality and diffusion-record integrity. More detailed dataset-specific cleaning rules are implementation details and should be stored in supplementary appendices rather than repeated in the main methodology.

Table 2.5. Core text cleaning and normalization operations retained in the main methodology.

Operation	Treatment	Reason
Duplicate and empty-text control	Remove exact duplicates and empty essential text fields.	Prevents inflated performance and invalid training records.
URL, mention and hashtag handling	Standardize rather than automatically delete when the token may carry credibility or diffusion meaning.	Preserves topic, source and platform signals.
Punctuation and casing normalization	Normalize excessive punctuation and casing while preserving sentence boundaries.	Improves tokenization without removing stylistic evidence.
Language-specific character preservation	Preserve Albanian characters such as ë and ç.	Maintains low-resource language information.

2.4.2 Label Harmonization

Binary label mapping is necessary because the selected content datasets use heterogeneous annotation schemes. Some datasets provide explicit fake and real classes, while others contain credibility or claim-type labels that must be mapped cautiously. The harmonized binary label is used only for binary content-based detection; the original label is retained as metadata so that the mapping remains auditable [81], [83].

$$y_{bin} = \begin{cases} 1, & \textit{if the item is fake-news-oriented,} \\ 0, & \textit{if the item is real or reliable.} \end{cases} \tag{2.2}$$

In Equation (2.2), y_{bin} denotes the harmonized binary target used in the Transformer detection experiments. The mapping does not erase the original dataset label; instead, the original label is preserved as raw-label metadata. This prevents a methodological error in which uncertain, auxiliary or diffusion-specific labels are treated as if they were identical to standard fake/real classification labels.

Table 2.6 keeps only the mapping logic required in the main text. Ambiguous labels such as unverified are excluded from the main binary experiments and retained only for sensitivity analysis. Generated examples inherit labels only inside the training subset, preventing train-test leakage.

Table 2.6. Compact binary mapping examples retained in the main methodology.

Dataset	Original labels	Binary mapping
Constraint@AAAI2021	fake / real	fake -> 1; real -> 0
ISOT	Fake.csv / True.csv	Fake.csv -> 1; True.csv -> 0
Hoaxy / OSoMeNet	claim and fact_checking retweet records claim can be treated as misinformation-oriented; fact_checking as corrective/reliable in diffusion analysis	claim-> 1 ; fact_checking->0

Figure 2.2 shows how raw dataset labels are transformed into a harmonized binary target while preserving the original label metadata. This workflow connects the dataset descriptions with the Transformer detection experiments without assuming that every dataset label has the same evidential meaning.

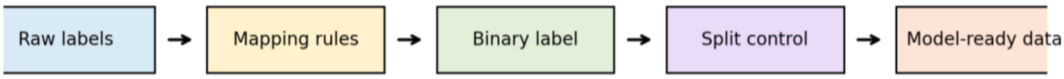

Figure 2.2. Binary label mapping and dataset harmonization workflow.

2.4.3 Tokenization and Transformer Input Preparation

After cleaning and label harmonization, textual records are converted into model-ready inputs using the tokenizer associated with each Transformer architecture. Tokenization is the point at which cleaned text becomes numerical model input. It must therefore preserve enough textual information for fake-news detection while respecting the sequence limits and input format of the selected model family.

$$\text{Tokenizer}_m(x_i) \rightarrow (\text{input\_ids}_i, \text{attention\_mask}_i, \dots) \quad (2.3)$$

In Equation (2.3), Tokenizer_m denotes the tokenizer associated with model family m, x_i is the cleaned text record, input_ids_i is the sequence of subword identifiers, and attention_mask_i marks valid input positions. Model-specific technical details such as token_type_ids, dynamic padding, special token handling, [CLS], [SEP] and truncation variants are moved to Appendix F because they are implementation specifications rather than core methodological arguments.

Figure 2.3 presents the transformation from cleaned text to Transformer-ready input. The figure is retained because it clarifies where text preprocessing ends and model-specific input preparation begins.

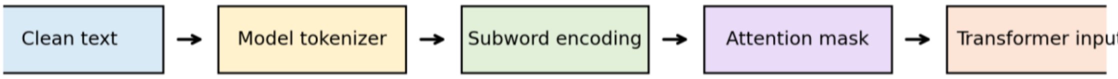

Figure 2.3. Transformer tokenization and model-input preparation workflow.

Long texts such as FRN and ISOT were processed using BigBird or window-based strategies, whereas short social-media texts were tokenized directly after cleaning and normalization. Padding, truncation and attention-mask policies followed the tokenizer associated with each Transformer model. The model-ready schema and tokenizer quality-control checklist are moved to Appendix F to keep the methodology focused and avoid unnecessary implementation detail in the main text.

Table 2.7. Tokenization strategy by model family.

Model family	Main tokenizer logic	Use in this dissertation
BERT / mBERT	WordPiece tokenization with fixed maximum sequence length.	Short and medium-length fake-news detection, including multilingual support.
RoBERTa / XLM-R	Byte-level BPE or SentencePiece-style subword modelling.	Robust subword handling for noisy or multilingual text.
BigBird	Sparse-attention long-sequence input.	Long-form article-style datasets such as FRN.

2.4.4 Handling Noisy and Imbalanced Data

Noise and class imbalance are treated as methodological risks rather than minor implementation issues. Noise sources include textual artefacts, duplicates, incomplete records and uncertain labels. Ambiguous samples are excluded from the main binary experiments and retained only for sensitivity analysis. This rule prevents noisy records from being forced into the fake/real target space [100].

$$R(D) = \frac{\max(n_{\text{fake}}, n_{\text{real}})}{\min(n_{\text{fake}}, n_{\text{real}})} \tag{2.4}$$

Equation (2.4) defines the class-imbalance ratio for a binary dataset D, where n_{fake} and n_{real} denote the number of fake and real examples. A value close to 1 indicates approximate balance, while higher values indicate stronger imbalance. This ratio guides whether class weighting, controlled augmentation or imbalance-aware evaluation should be used [101].

$$\mathcal{L}_{\text{WBCE}} = -[w_1 y \log(p) + w_0(1-y)\log(1-p)] \tag{2.5}$$

Equation (2.5) gives a weighted binary cross-entropy objective that can be used when imbalance is substantial. The weights w_1 and w_0 compensate for class-frequency differences, while p denotes the predicted fake-news probability. This objective is treated as a training option rather than as a mandatory choice for every dataset.

Figure 2.4 presents the workflow used to identify noise sources and class imbalance before binary Transformer training. The workflow supports stable training while preserving a clear separation between original evidence and training-only augmentation.

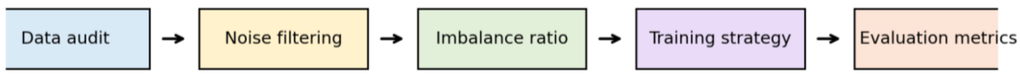

Figure 2.4. Workflow for handling noisy and imbalanced binary fake-news data.

Table 2.8 is compressed from a longer list of strategies into the three decisions that are most relevant for the dissertation. More detailed noise, imbalance and quality-control tables are better placed in Appendix F because they document implementation choices rather than define the core methodology.

Table 2.8. Compact strategy set for noisy and imbalanced binary fake-news data.

Strategy	Use
Class-weighted loss	Default option when class imbalance is visible.
Controlled augmentation	Training-support option for low-resource settings only.
Macro-F1 and PR-AUC evaluation	Evaluation options that reduce over-reliance on accuracy.

When imbalance becomes severe, focal loss and related hard-example objectives may be considered as robustness checks, but they are not used to replace transparent reporting of class distribution, fake-class recall, macro-F1 and probability calibration [101].

Low-resource Albanian text preparation

The Albanian Fake News Corpus is included to evaluate whether the Transformer detection layer can operate in a low-resource and multilingual setting. Albanian preprocessing requires particular care because excessive normalization may remove language-specific characters, named entities or local political vocabulary. The thesis therefore preserves Albanian orthography, including ë and ç, as the primary textual representation [84], [85], [98], [120].

Low-resource preparation follows three compact stages. The first stage verifies encoding, duplicate records and text completeness. The second stage applies conservative content cleaning while preserving Albanian-specific vocabulary and named entities. The third stage prepares the text for multilingual Transformer tokenization using mBERT or XLM-R, both of which are suitable for cross-lingual subword modelling [9], [11], [119], [120].

Figure 2.5 shows the reduced Albanian preprocessing workflow. The workflow emphasizes preservation of language-specific information, controlled cleaning and compatibility with multilingual Transformer tokenizers.

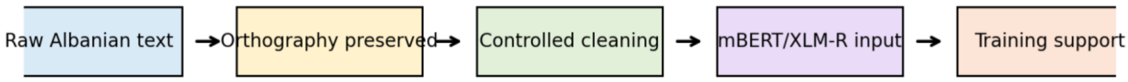

Figure 2.5. Low-resource Albanian fake-news preprocessing workflow for multilingual Transformer input.

Only one compact Albanian preprocessing table is retained in the main text. The previous separate multilingual Transformer preparation table is removed to avoid repetition, and the Albanian quality-control checklist is moved to Appendix F.

Table 2.9. Compact Albanian preprocessing challenges and responses.

Challenge	Recommended response	Reason
Language-specific characters	Preserve ë and ç; avoid forced ASCII conversion.	Prevents loss of Albanian orthographic information.
Local names and political vocabulary	Retain named entities and local terms.	Maintains domain-specific credibility cues.
Small dataset size	Use controlled augmentation only inside training data.	Supports training without altering test evidence.

General non-standard-word normalization and short-text normalization research also supports the conservative cleaning policy adopted here: normalization can reduce noise, but excessive normalization can remove credibility cues, named entities or negation markers that are important for fake-news classification [99], [112].

2.4.5 Preprocessing Workflow

Diffusion-oriented preprocessing is included only to define the bridge from raw interaction records to later temporal cascade construction. It does not duplicate the temporal-network methodology. The task at this stage is to retain claim identifiers, user identifiers, timestamps, interaction fields and quality flags so that Section 2.6 can construct temporal cascades, directed edges and infected-time series for SIR/SEIZ and ABC/ABC-SMC analysis.

For diffusion-oriented datasets, cleaning is deliberately less aggressive than in pure text classification. User identifiers, timestamps, retweet or interaction relations, and cascade membership must be preserved even when the associated text is incomplete. Missing timestamps, unavailable tweets and inconsistent user identifiers are assigned quality flags rather than silently removed. This preserves reproducibility and prevents artificial changes in cascade size.

Figure 2.6 presents the shortened diffusion-oriented preprocessing workflow. It shows that the output of this stage is not the final temporal network, but a quality-controlled event layer that becomes input for temporal cascade construction.

Figure 2.6. Preprocessing workflow for diffusion-oriented data before temporal cascade construction.

Table 2.10 is retained because it defines the minimum diffusion-ready output required by the later modelling stages. Dataset-specific diffusion schemas, quality-control decisions and state-proxy mappings are moved to the appendices to avoid repeating the same pipeline logic in both preprocessing and temporal cascade construction sections.

Table 2.10. Main objectives of preprocessing diffusion-oriented data.

Objective	Methodological role
Preserve cascade identity	Ensures that each event remains attached to its fake-news claim or thread.
Preserve users and interactions	Allows directed source-target relations to be constructed later.
Preserve timestamps	Supports temporal ordering, $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, SIR/SEIZ fitting and ABC distance calculation.
Assign quality flags	Documents missing, duplicated or inconsistent records without silently changing the evidence.

The final output of Section 2.4 is therefore a pair of coordinated model-ready representations. The first is a content-ready representation for Transformer-based binary fake-news detection. The second is a diffusion-ready representation in which claim labels, user identifiers, timestamps and directed interactions are preserved for temporal cascade construction. This separation supports the central logic of the dissertation without turning preprocessing into a stand-alone implementation textbook.

Methodological transition to Chapter 3 experiments

The preprocessing design also protects the empirical interpretation of the author-generated results reported in Chapter 3. For content-based datasets, the cleaned and harmonized records enter the Transformer and classical machine-learning experiments as binary fake/real examples. The preprocessing layer therefore determines whether reported classification metrics reflect model behaviour rather than artefacts of inconsistent labels, duplicated text, broken records or uncontrolled augmentation. For diffusion-oriented datasets, the same stage has a different objective: it must preserve the interaction fields that are later required to build temporal cascades. This distinction is important because the thesis does not treat a cleaned news article and a timestamped user-to-user interaction as equivalent methodological objects [86], [89].

A second reason for keeping preprocessing as a separate methodological layer is leakage control. Generated or augmented records are restricted to the training subset, and validation and test subsets remain grounded in original dataset records. This rule applies to English and Albanian content-based datasets and prevents artificial examples from inflating reported detection performance. In the diffusion-oriented layer, leakage control has a temporal meaning: later events cannot be used to construct earlier cascade features. Preserving this chronological boundary is necessary for the out-of-sample forecasting and ABC/ABC-SMC posterior prediction steps in Chapter 3.

The quality-control material placed in Appendix F should therefore be read as supporting documentation for the compact methodology in this section. The main text retains only the operations that explain the logic of the experimental pipeline, while the appendix preserves implementation schemas, checklists and tokenizer-specific details. This organization keeps Section 2.4 focused on the methodological argument while still allowing the experiments to be audited and reproduced. The result is a balanced preprocessing design: sufficiently detailed for PhD-level transparency, but not overloaded with implementation tables that would interrupt the progression from datasets to temporal cascade construction.

Finally, the two outputs of preprocessing are intentionally synchronized. The content-ready output produces a probability-based detection signal, while the diffusion-ready output preserves the time-indexed interaction structure used to compute cascade features. When these two outputs are interpreted together, the dissertation can move from a textual fake-news judgement to a model-ready representation of how a fake-news claim spreads. This is the reason why preprocessing is placed before temporal cascade construction: it creates the common evidential foundation for Transformer detection, temporal-network analysis, SIR/SEIZ modelling and ABC/ABC-SMC posterior inference.

2.5 Temporal Cascade Construction and Network Representation

Section 2.5 defines how the retained diffusion-oriented records are transformed into model-ready temporal cascades. The methodology specifies how timestamped claim-level events will be preprocessed, how directed temporal networks will be constructed, and how cascade features will be extracted for SIR/SEIZ diffusion modelling and ABC/ABC-SMC inference. The purpose is methodological: this section defines the operational representation used by the empirical work rather than repeating the results reported later in the thesis. The representation follows temporal-network principles and prior empirical studies of online misinformation diffusion, where the timing and direction of interactions are essential for distinguishing static connectivity from observable spreading dynamics [52]-[55].

Temporal-network representation is selected instead of temporal graph neural architectures because the dissertation requires interpretable cascade reconstruction and epidemic-parameter inference rather than learned dynamic node embeddings or link-prediction models. The available diffusion-oriented datasets provide timestamped interactions and cascade traces suitable for time-respecting paths, $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, SIR/SEIZ state mapping and ABC/ABC-SMC inference. Temporal graph neural models such as TGN, TGAT, DyRep and EvolveGCN are therefore treated as future extensions that would require richer user-level temporal graph data, stable node and edge attributes and additional supervised graph-learning objectives [128]-[129], [153]-[155].

A temporal cascade is treated as a chronologically ordered sequence of observable interactions associated with the same fake-news claim, rumour, source post, URL or external item. Each event preserves a source node, a target node, a timestamp, an action type, a cascade identifier and, where available, a label or claim type. Cascades support both descriptive analysis (size, depth and time-to-peak) and inferential tasks (SIR/SEIZ calibration and ABC summary extraction).

The basic event representation is defined as follows:

$$e_k = (u_s, u_t, t_k, a_k, c_k, y_k) \tag{2.6}$$

In Equation (2.6), u_s denotes the source or previous node, u_t denotes the target or responding node, t_k is the event time, a_k is the observed action type, c_k is the cascade or fake-news claim identifier, and y_k is the available technical label such as claim, covid_fake_claim or fact_checking. This representation keeps content labels and temporal evidence connected without treating all records as equivalent to epidemic states.

Figure 2.7 presents the operational sequence used to transform diffusion-oriented data into modelling inputs. Raw records are standardized, grouped by fake-news claim or source item, ordered by timestamp, converted into directed cascades and summarized for SIR/SEIZ and ABC/ABC-SMC analysis.

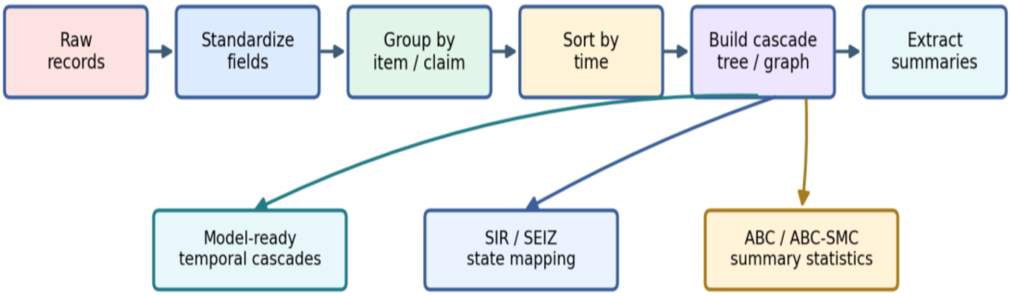

Figure 2.7. Temporal cascade construction workflow from diffusion-oriented records to model-ready cascades.

2.5.1 Cascade Preprocessing

Cascade preprocessing begins by filtering each diffusion-oriented dataset to the fake-news claim or external item under analysis and by preserving only the fields required for temporal reconstruction. In the COVID-19 real Twitter/X dataset, the operational fake-news cascade is obtained from rows with site_type = covid_fake_claim. In the Hoaxy / OSoMeNet datasets, the retained records are claim-scoped retweet interactions for the political fake-news claim "Three million votes in the Presidential Election were cast by illegal aliens." and the social fake-news claim "Should One Tweet End a Career?". The synthetic Albanian fake-news claim is retained only as a validation and simulation case for the low-resource temporal pipeline. Detailed dataset-specific field mappings remain in Appendix D.

For a fake-news claim c, the cascade is represented as:

$$C_c = \left(V_c, E_c^T, x_c, y_c, p_f(x_c)\right) \tag{2.7}$$

In Equation (2.7), V_c is the set of users or source nodes involved in the cascade, E_c^T is the set of directed time-stamped interactions, x_c is the associated textual fake-news claim or article, y_c is the harmonized claim label, and $p_f(x_c)$ is the Transformer-based fake-news probability used as a content-risk signal in the broader framework.

Table 2.11 defines the minimal temporal-network objects used by the methodology. It keeps the representation sufficiently general for the retained Twitter/X, Hoaxy / OSoMeNet and synthetic Albanian cascades, while also restricting the modelling layer to observable events and timestamps.

Table 2.11. Core elements of the temporal-network representation.

Element	**Definition in this thesis**	**Example**	**Use in later modelling**

Node	A user, post, source claim, article or reaction depending on the dataset.	Twitter/X user in Hoaxy; source tweet in a retweet cascade.	Defines the units that can become susceptible, exposed, spreading, recovered or skeptical in epidemic interpretation.
Directed edge	A time-ordered interaction from a source node to a target node.	from_user_id -> to_user_id in retweet data.	Represents possible information-transmission paths.
Timestamp	The time at which an interaction occurs or the delay from the source event.	tweet_created_at in Twitter/X or Hoaxy-style records.	Allows temporal ordering, time-to-peak analysis and time-respecting paths.
Weight	A value measuring interaction strength, frequency or relevance.	Number of retweets between two users or repeated interactions.	Supports weighted diffusion modelling and user-influence analysis.
Cascade identifier	A label linking all events belonging to the same fake-news claim or source item.	Claim ID, source tweet or external item label.	Groups events for cascade-level statistics and ABC summary features.

Table 2.12 brings the unified diffusion-oriented schema into the main methodology because it explains how raw interaction records become model-ready cascade records. The same schema also keeps detailed reproducibility information compatible with the supplementary preprocessing tables in Appendix D.

Table 2.12. Unified model-ready schema for diffusion-oriented records.

Field	Definition	Required role in the thesis
cascade_id	Identifier linking all records that belong to the same fake-news claim or source item.	Groups records before feature extraction and SIR/SEIZ preparation.
source	Originating user, source post, article or previous node.	Defines the starting side of a directed temporal edge.
target	Receiving, reposting, replying or interacting user/node.	Defines the activated or affected side of a directed temporal edge.

time	Timestamp or relative event time.	Allows chronological ordering and construction of time-respecting cascades.
action_type	Observed interaction type, such as retweet, reply, repost, claim or fact-checking record.	Supports interpretation of spreading, exposure or correction proxies.
claim_label	Technical dataset value such as claim, covid_fake_claim or fact_checking.	Maintains the distinction between fake-news claim rows and auxiliary/corrective rows.
text_or_title	Textual claim, article title or external-item description when available.	Links the temporal cascade to the Transformer-based content layer.

2.5.2 Temporal Network Construction

After cascade preprocessing, each claim-level event sequence is converted into a directed temporal network. Nodes represent users or source objects, while edges represent observed interactions ordered by time. Edge direction follows the observable flow of interaction; for example, a retweet-style edge is oriented from the source or retweeted user to the user who amplifies the item. Observable interactions are treated as behavioral proxies rather than direct latent belief states. This design is consistent with temporal-network methodology, where an interaction is meaningful for diffusion analysis only when its direction and event time are preserved [52], [55].

Figure 2.8 visualizes the basic network objects used in the temporal-cascade layer. The purpose of the representation is not to infer private belief directly, but to preserve the observable structure needed for diffusion modelling: users or posts as nodes, interactions as directed edges, timestamps as ordering constraints and weights as repeated or stronger interactions.

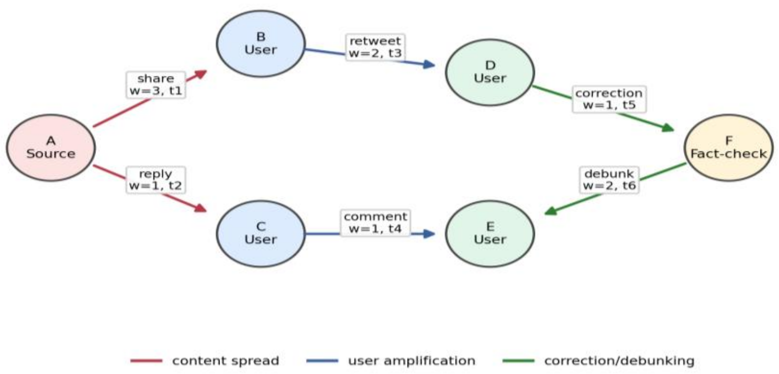

Figure 2.8. Illustration of nodes, directed edges, weights, and timestamps in a fake-news diffusion network.

The validity of a propagation path depends on temporal order. A path is accepted only if the timestamps increase along the sequence of edges:

$$P = [(u_0, u_1, t_1), \dots, (u_{m-1}, u_m, t_m)], \quad t_1 < t_2 < \dots < t_m \tag{2.8}$$

Equation (2.8) prevents impossible information-flow paths that can appear after static aggregation. It also ensures that cascade depth, breadth, time to peak, inter-event time and time-respecting reachability are computed from temporally valid events.

Figure 2.9 clarifies why the methodology uses temporal representation for diffusion analysis. Static aggregation can summarize that users interacted during the observation window, but it does not preserve the order in which a fake-news claim moves through the network. Temporal representation is therefore required for cascade reconstruction, SIR/SEIZ input construction and ABC/ABC-SMC summary statistics.

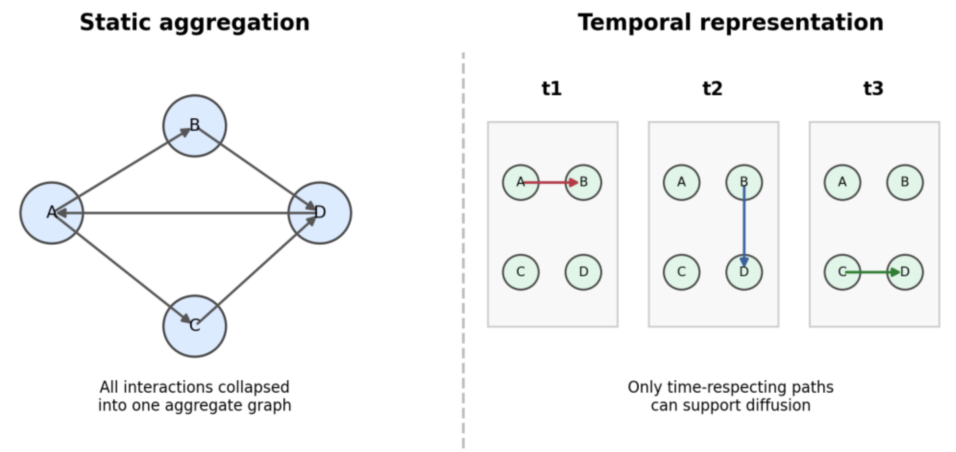

Figure 2.9. Difference between static network aggregation and temporal-network representation.

Table 2.13 formalizes the methodological reason for using temporal networks. The retained diffusion datasets are not used only to identify whether interactions occurred; they are used to reconstruct the order of activation that later generates $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, model fitting and posterior predictive analysis.

Table 2.13. Comparison between static and temporal network representations.

Aspect	**Static network**	**Temporal network**	**Implication for this thesis**
Edge interpretation	An edge exists if an interaction occurs at least once during the observation window.	An edge exists at a specific time or time interval.	Temporal networks are more suitable for cascade reconstruction and diffusion timing.
Path validity	All connected paths may appear valid after aggregation.	Only time-respecting paths are valid.	Prevents impossible information-flow paths.

Useful features	Degree, density, clustering and communities.	Inter-event time, time to peak and time-respecting reachability.	Temporal features support SIR/SEIZ simulation and ABC summary statistics.
Limitation	Loses the order of events.	Requires accurate timestamps and event ordering.	Temporal modelling is more demanding but more realistic for fake-news spread.

Figure 2.10 provides the only main-text visualization of user influence and community spread in this methodology section. Influential users, bridge nodes, local communities and fact-checking or corrective actors are interpreted as structural positions in the observed network, while detailed network-property tables are retained in Appendix D.

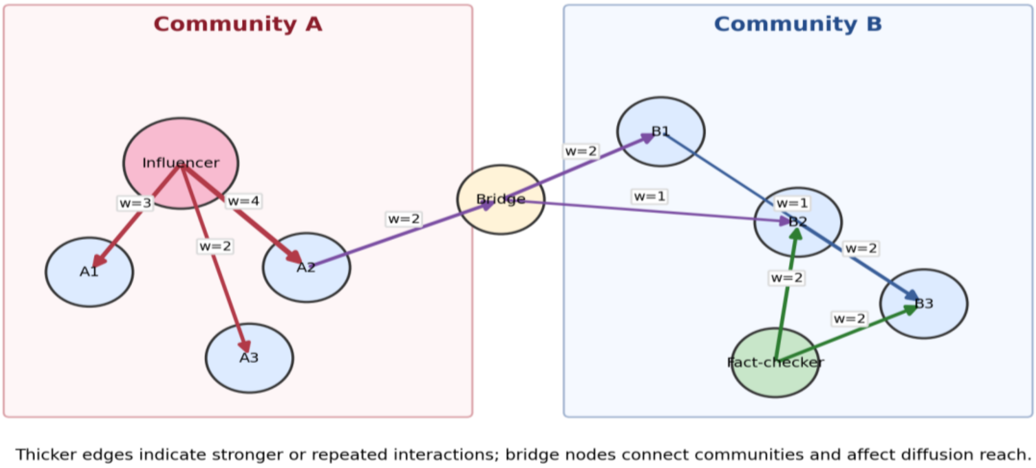

Figure 2.10. Illustration of user influence and community spread in a weighted temporal fake-news diffusion network.

A compact influence measure used in the temporal-network layer is the weighted outgoing activity of a user:

$$s_{\text{out}}(i) = \sum_j \sum_t w_{ij}\,(t) \tag{2.9}$$

where $s_{\text{out}}(i)$ represents the weighted outgoing activity of user i and $w_{ij}(t)$ is the weight of the interaction from user i to user j at time t. This measure supports the identification of users who repeatedly amplify a fake-news claim through observable interactions.

A second compact measure is the cross-community spread ratio:

$$R_{\text{cc}} = \frac{E_{\text{cross}}}{E_{\text{total}}} \tag{2.10}$$

where E_{cross} is the number of cascade edges connecting different communities and E_{total} is the total number of cascade edges. The ratio is used as a structural indicator of whether a fake-news claim remains localized or moves across communities. Centrality and community detection are treated as established network-analysis tools, while their detailed theoretical background is retained in the literature review and Appendix E [53], [58]-[61].

2.5.3 Cascade Features for SIR/SEIZ/ABC

The final output of temporal cascade construction is a set of cascade features and infected-time series that can be used by the epidemic and ABC layers. The methodology extracts both descriptive features and model-ready summaries. Descriptive features characterize the observed fake-news cascade, while inferential features provide the quantities compared with SIR/SEIZ simulations during ABC and ABC-SMC calibration.

The two central temporal activation measures are $I_{\text{new}}(t)$ and $I_{\text{cum}}(t)$. $I_{\text{new}}(t)$ is the number of new unique active spreaders observed in a given time window, while $I_{\text{cum}}(t)$ is the cumulative number of unique active spreaders observed up to that time. These two quantities form the direct bridge between temporal-network construction and epidemic-model fitting. This bridge is important because ABC and ABC-SMC later compare simulated and observed spreading curves through summary statistics rather than through a closed-form likelihood [76], [106].

Table 2.14 summarizes the cascade features retained for downstream modelling. The table separates the quantities that describe the observed cascade from the quantities that can be used as ABC summary statistics, making the connection between temporal-network construction and SIR/SEIZ calibration explicit.

Table 2.14. Cascade features for SIR/SEIZ/ABC summary extraction.

Feature	Definition	Use in SIR/SEIZ/ABC
Cascade size	Total number of participating nodes or unique active spreaders.	Used to compare final epidemic size and posterior predictive saturation.
$\boldsymbol{I_{\text{new}}(t)}$	Number of newly active unique spreaders in time window t.	Used as the observed infected-time increment for diffusion dynamics.
$\boldsymbol{I_{\text{cum}}(t)}$	Cumulative number of unique active spreaders up to time t.	Used as the main cumulative curve for model fitting and forecasting.
Time to peak	Time between cascade start and maximum $I_{\text{new}}(t)$.	Supports ABC summary statistics and timing comparison.
Inter-event time	Time gaps between consecutive events in the cascade.	Describes burstiness and delayed activation.
Reach and breadth	Number of users or branches reached by the cascade.	Supports descriptive comparison and posterior predictive checking.

Weighted outgoing activity	Temporal sum of outgoing edge weights for each user.	Identifies influential spreaders and possible relay hubs.
Cross-community ratio	Share of edges crossing community boundaries.	Summarizes structural spread beyond the initial community.

When the cascade is passed to the epidemic layer, the directed temporal cascade network for a fake-news claim c is defined as follows:

$$G_c^T = (V_c, E_c^T) \tag{2.11}$$

In Equation (2.11), G_c^T denotes the directed temporal cascade network associated with fake-news claim c. The set V_c contains the users, source accounts or source objects that appear as nodes in the cascade. The set E_c^T contains the directed time-stamped interactions observed for the same claim, where an edge may be written as (u_i, u_j, t, $w_{ij}(t)$) to indicate an interaction from node u_i to node u_j at time t with optional weight $w_{ij}(t)$. The superscript T indicates that the edge set is time-indexed and temporally ordered; it is not used as a matrix transpose. This formulation provides the formal object from which $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, cascade-level features and SIR/SEIZ inputs are derived [52]-[55], [76].

The ordered activations of target users generate $I_{\text{new}}(t)$ and $I_{\text{cum}}(t)$, while network properties such as weighted outgoing activity and community crossing provide optional structural summaries. At this stage, retweets, replies, reposts and fact-checking records are not interpreted as direct observations of latent belief. They are observable proxies that support modelling assumptions and posterior comparison.

The methodological boundary of Section 2.5 is intentionally narrow. The dissertation does not implement memory-dependent temporal-network models as empirical components. Instead, it uses directed temporal cascades because they are supported by the retained datasets and connect directly to the experimental pipeline: temporal-network visualization, $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, SIR/SEIZ fitting, ABC/ABC-SMC posterior inference and short-horizon predictive evaluation. More complex path-memory approaches remain possible future extensions rather than unvalidated components of the implemented methodology [52], [76].

2.6 SIR/SEIZ, ABC/ABC-SMC and Posterior Prediction Methodology

Section 2.6 defines the epidemic-modelling and posterior-inference layer of the proposed Transformer-ABC-Epidemic framework. The previous methodological stages define the cleaned textual inputs, the claim-level temporal cascades, the directed interaction structure and the temporal activation measures used in the empirical analysis. The present section explains how those outputs are converted into SIR and SEIZ diffusion models, how uncertain parameters are inferred with Approximate Bayesian Computation (ABC) and ABC-SMC, and how posterior samples are used for predictive simulation, short-horizon forecasting and counterfactual scenario reasoning.

The main methodological difficulty is that fake-news cascades are only partially observed. Platform records show visible posts, reposts, replies, timestamps, user-to-user edges and sometimes correction or fact-checking events. They do not reveal every exposure, every non-response, every private share or every user decision not to spread. For this reason, the likelihood of the full observed cascade under a stochastic epidemic model is not available in closed form. This is the same type of problem for which likelihood-free methods such as ABC have been developed in population genetics, stochastic epidemic modelling and simulation-based inference [19], [22]-[26].

Within this dissertation, ABC is used as an inferential layer rather than as a replacement for the epidemic simulator. Transformer models supply the content-risk signal, the temporal cascade supplies the observed diffusion summaries, SIR and SEIZ define the mechanistic simulators, and ABC/ABC-SMC estimates the posterior parameter regions that can reproduce the observed fake-news claim trajectories. This structure is consistent with the use of epidemic and rumour-spreading models for online information diffusion [31]-[36], temporal-network studies of misinformation [52], [54], and the author's related work on temporal fake-news networks, SIR/SEIZ comparison and BigBird-assisted SEIZ/ABC modelling [86]-[89].

Figure 2.11 summarizes the operational role of ABC in the dissertation. The observed temporal cascade is not used to compute a closed-form likelihood. Instead, prior parameter values are sampled, synthetic SIR/SEIZ cascades are generated, summary statistics are compared with the observed cascade, and accepted simulations approximate the posterior distribution of parameters capable of reproducing the fake-news spreading trajectory. The figure therefore links the observed cascade layer of Section 2.5 with the posterior-inference results reported in Chapter 3.

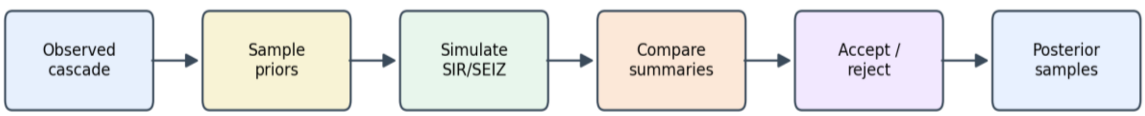

Figure 2.11. ABC workflow for calibrating fake-news epidemic parameters from temporal cascades.

2.6.1 SIR Baseline Model

The SIR model is retained as the baseline epidemic representation because it provides the simplest interpretable mapping from social exposure to active spreading and then to inactivity. In the fake-news setting, susceptible users are those who have not yet actively spread the fake-news claim, infected users are active spreaders during the observation window, and recovered or removed users are those who no longer contribute new spreading events. This interpretation follows the classical epidemic model introduced by Kermack and McKendrick [31], while adapting the compartments to information diffusion rather than biological infection.

The baseline parameter vector is written as:

$$\theta_{\mathrm{SIR}} = (\beta, \gamma, I_0) \tag{2.12}$$

In Equation (2.12), β is the transmission or spreading coefficient, γ is the recovery, forgetting or stopping coefficient, and I_0 is the initial active-spreader seed. SIR is useful for benchmarking because it tests whether the observed cumulative spreading curve can be explained with only transmission and recovery dynamics. However, fake-news diffusion often includes delayed exposure, hesitation, skepticism, fact-checking and non-spreading behaviour. These mechanisms motivate the use of SEIZ as the main model while preserving SIR as a transparent baseline in Chapter 3.

2.6.2 SEIZ Main Diffusion Model

SEIZ is selected as the main diffusion model because it is better aligned with fake-news dynamics than the simpler SIR baseline. The exposed state represents users who have encountered the fake-news claim but have not yet become active spreaders. The skeptic or non-spreading state represents users who reject, ignore, question or are corrected about the claim. This distinction is important because fake-news spreading is not an immediate homogeneous transmission process: users may observe the claim, hesitate, verify, encounter corrections, or stop before spreading. The model is therefore more compatible with misinformation and rumour-spreading settings than an immediate infection-only model [32]-[36].

The SEIZ parameter vector used in the methodology is defined as:

$$\theta_{\mathrm{SEIZ}} = (\beta, b, p, l, \rho, \varepsilon, \gamma, \delta, S_0, E_0, I_0, Z_0) \tag{2.13}$$

In Equation (2.13), β is the rate at which active spreaders influence susceptible users, b represents contact or influence involving skeptical users, p is the probability of immediate adoption after exposure, l is the probability of becoming skeptical or non-spreading, ρ captures interaction between exposed and active-spreading states, ε is the transition rate from exposed to active spreading, γ is the recovery or stopping rate, and δ is the correction, debunking or removal-related parameter used in the modified fake-news interpretation. The initial conditions S_0, E_0, I_0 and Z_0 define the starting state of the simulated cascade.

The SEIZ formulation is not introduced as a new epidemic model in this dissertation. Its methodological role is to provide a mechanistic structure that can represent exposed and skeptical/non-spreading states, while ABC and ABC-SMC estimate which parameter regions are compatible with the observed temporal fake-news cascades. This distinction is important for the logic of the thesis: the contribution is the integrated detection-diffusion-inference framework, not the invention of a separate epidemic compartment model.

2.6.3 ABC Inference

Let D_{obs} denote the observed fake-news cascade and let θ denote the epidemic parameter vector. In a standard Bayesian formulation, the posterior distribution is proportional to the product of the likelihood and the prior:

$$p(\theta \mid D_{\text{obs}}) \propto p(D_{\text{obs}} \mid \theta)p(\theta) \tag{2.14}$$

The difficulty in Equation (2.14) is the likelihood term $p(D_{\text{obs}} \mid \theta)$. For a fake-news cascade, this term would require the probability of the observed sequence of user activations, inter-event times, reposts, corrections and hidden non-responses. Because these elements are not fully observed and are not easily represented by a closed-form probability model, ABC replaces likelihood evaluation with simulation and summary-statistic comparison [19], [23]-[26].

$$p_{\text{ABC}}(\theta \mid D_{\text{obs}}) \propto p(\theta)\int \mathbf{1}\{\rho(S_{\text{sim}}, S_{\text{obs}}) \le \varepsilon\} p(D_{\text{sim}} \mid \theta)\, dD_{\text{sim}} \tag{2.15}$$

In Equation (2.15), D_{sim} is a simulated cascade generated by the SIR or SEIZ simulator, S(D) is a vector of summary statistics, ρ is a distance function and ε is the ABC tolerance threshold. A simulated parameter value is accepted when the distance between simulated and observed summaries is sufficiently small. The accepted parameter values approximate the posterior distribution over epidemic parameters that can reproduce the observed fake-news spreading pattern.

Table 2.15 is retained in the main methodology because prior ranges define the parameter space explored by ABC and ABC-SMC. The table also clarifies that the parameters are interpreted for fake-news propagation rather than for biological disease transmission. More detailed supplementary parameter and summary-statistic tables are placed in Appendix D.3 so that Section 2.6 remains focused on the operational inference logic.

Table 2.15. Candidate prior distributions for ABC-based inference of fake-news epidemic parameters.

Parameter	Model	Prior example	Interpretation in fake-news propagation

β	SIR/SEIZ	Uniform(0.01, 0.80)	Baseline rate at which active spreaders influence susceptible users.
γ	SIR/SEIZ	Uniform(0.01, 0.50)	Rate at which spreaders stop sharing, forget the content or lose interest.
ε	SEIZ	Uniform(0.01, 0.50)	Rate at which exposed users become active spreaders.
p	SEIZ	Beta(2, 2)	Probability of immediate sharing after contact with an active spreader.
l	SEIZ	Beta(2, 2)	Probability of becoming skeptical or non-spreading after exposure.
δ	SEIZ-modified	Uniform(0.00, 0.30)	Correction or debunking rate caused by warnings, fact-checks or counter-information.
ρ	SEIZ	Uniform(0.00, 0.60)	Interaction coefficient between exposed and infected users.

The summary-statistic vector is selected to preserve the cascade features used later in Chapter 3: final cascade size, time to peak, maximum spreading rate, cascade depth, cascade breadth, mean inter-event delay and the average Transformer-derived fake-news probability. This vector can be written as:

$$S(D) = \left[C_T, t_{\text{peak}}, r_{\max}, d_c, b_c, \tau_{\text{mean}}, \bar{p}_f\right] \tag{2.16}$$

In Equation (2.16), C_T is final cascade size, t_{peak} is time to peak, $r_{\max}$ is the maximum spreading rate, d_c is cascade depth, b_c is cascade breadth, τ_{mean} is the mean inter-event time and $\bar{p}_f$ is the mean Transformer-derived fake-news probability for the claim or cascade item. The distance between observed and simulated summaries is expressed as:

$$\rho(S_{\text{obs}}, S_{\text{sim}}) = \sum_k w_k \left|S_k^{\text{obs}} - S_k^{\text{sim}}\right| \tag{2.17}$$

The weights w_k in Equation (2.17) determine the importance of each summary statistic. If final cascade size is the primary target, C_T receives more weight. If early-warning behaviour is more important, t_{peak} and $r_{\max}$ can be emphasized. In Chapter 3, these summary statistics support posterior calibration, posterior predictive simulation and short-horizon out-of-sample forecasting. The full candidate summary-statistic table is provided in Appendix D.3.

Connection with Transformer-Based Detection and BigBird-Assisted SEIZ

Following BigBird-assisted SEIZ concepts [12], [86], [89], Transformer output probabilities may be incorporated into diffusion coefficients. For a text item x_i, the calibrated Transformer fake-news probability is defined as:

$$p_f(x_i) = P(y_i = \text{fake} \mid x_i) \tag{2.18}$$

In Equation (2.18), $p_f(x_i)$ denotes the probability assigned by the Transformer detector that textual item x_i belongs to the fake-news class. The score is not used as an ensemble component and does not replace temporal diffusion modelling. Instead, it provides a calibrated content-risk signal that can modulate the spreading pressure assigned to an observed temporal interaction.

$$\beta_{ij}(t) = \beta_0\, w_{ij}(t)\, [1 + \lambda p_f(x_i)]\, [1 + \eta c_j] \tag{2.19}$$

In Equation (2.19), $\beta_{ij}(t)$ is the effective transmission coefficient from user j to user i at time t, β_0 is the baseline transmission coefficient, $w_{ij}(t)$ is the temporal interaction weight, λ controls the contribution of the Transformer fake-news probability, c_j is a centrality or influence score of the spreading user j, and η controls the effect of user influence. The equation is used as a methodological bridge between the Transformer detection layer and the SEIZ diffusion layer: higher fake-news probability, stronger interaction weight and greater sender influence increase the effective spreading pressure along the observed temporal edge. The extended BigBird-assisted SEIZ discussion is retained in Appendix G.1.

The observed-data mapping used by ABC is therefore deliberately conservative. Visible posts or claim text contribute to the Transformer fake-news probability, shares and retweets contribute to cascade size and infected-curve reconstruction, timestamps determine the peak time and inter-event delays, and user-to-user edges provide the temporal interaction structure. Fact-checks, debunking replies or correction links are not interpreted as complete observations of belief change; they are treated only as partial evidence for correction, skepticism or non-spreading behaviour. This distinction prevents the methodology from over-interpreting platform traces while still allowing the ABC simulator to compare observable cascade summaries with simulated trajectories.

This design is also important for identifiability. Several different parameter combinations may reproduce a similar cumulative curve, especially when exposed and skeptical states are latent. ABC therefore does not claim to recover a single true parameter vector. It estimates a posterior region of plausible parameters, conditional on the chosen simulator, prior ranges and summary statistics. This interpretation is consistent with ABC applications to stochastic epidemics and spreading processes on networks [24], [25], [76], [77]. The detailed observed-data mapping and candidate summary-statistic tables are kept in Appendix D.3 so that the main section remains readable while preserving reproducibility for examination purposes.

2.6.4 ABC-SMC Inference

Standard rejection ABC is useful for explaining likelihood-free inference, but it can be inefficient when the prior is broad and the tolerance threshold must be small. ABC-SMC addresses this limitation by using a sequence of particle populations with progressively decreasing tolerance thresholds. Early populations explore broad regions of the parameter space, while later populations concentrate on parameter values that generate simulations closer to the observed cascade. This sequential structure is particularly useful for fake-news diffusion because SIR and SEIZ parameters interact nonlinearly and because the same cascade can often be reproduced by several plausible parameter combinations [22], [77], [79], [80].

Figure 2.12 illustrates the ABC-SMC procedure used to refine the posterior distribution and compare SIR with SEIZ under progressively decreasing tolerance thresholds. The method starts with a population of particles, simulates candidate cascades, compares summary statistics, weights accepted particles and resamples toward a refined posterior. In Chapter 3, this logic supports the comparison between SIR, SEIZ, ABC-SEIZ and ABC-SMC-SEIZ posterior predictive behaviour.

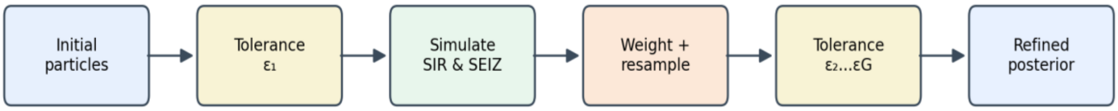

Figure 2.12. ABC-SMC workflow for comparing SIR and SEIZ on fake-news temporal cascades.

The candidate model set is written as:

$$\mathcal{M} = \{\mathrm{SIR}, \mathrm{SEIZ}\} \tag{2.20}$$

ABC-SMC can then compare the approximate posterior support of candidate models through:

$$P_{\mathrm{ABC}}(M \mid D_{\mathrm{obs}}) \propto P(M) P_{\mathrm{ABC}}(D_{\mathrm{obs}} \mid M) \tag{2.21}$$

Equation (2.21) is not interpreted as proof that one model is universally superior. Instead, the dissertation evaluates whether the SEIZ structure provides better posterior predictive calibration for the empirical fake-news cascades considered in Chapter 3. The accepted distances, posterior parameter distributions and posterior predictive curves are used to compare the practical calibration of SIR and SEIZ under the same observed $I_{\mathrm{new}}(t)$ and $I_{\mathrm{cum}}(t)$ trajectories.

The ABC-SMC implementation is also aligned with the computational tools used later in Chapter 3. In practice, the algorithm produces weighted particles rather than a single fitted parameter set. These particles represent candidate SIR or SEIZ parameter values that survived successive tolerance levels. The final population is then summarized through posterior medians, credible intervals and posterior predictive simulations. This workflow is compatible with ABCpy and related high-performance implementations of Approximate Bayesian Computation [107], while the thesis keeps the mathematical explanation independent of any single software package.

The model-comparison role of ABC-SMC should be interpreted carefully. A lower distance or better posterior predictive fit for SEIZ does not prove that SEIZ is universally the best model for all misinformation processes. It shows that, under the retained fake-news cascades, summary statistics and prior assumptions, the SEIZ state structure provides stronger posterior predictive calibration than the simpler SIR baseline. This wording is important because Chapter 3 evaluates empirical calibration for two principal fake-news claim cascades, not a universal theorem about all social-media diffusion datasets.

2.6.5 Posterior Predictive Simulation and Forecasting

After ABC or ABC-SMC has produced accepted parameter samples, posterior predictive simulation is used to assess how well the calibrated simulator reproduces the observed cascade and how plausible future trajectories behave under uncertainty. This step is essential because posterior parameter estimates alone do not show whether the model can reproduce the time profile of the fake-news claim. Posterior predictive simulation therefore transforms parameter uncertainty into distributions over future $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, peak time and final cascade size.

Posterior predictive simulation can be written as:

$$D_{\text{future}}^{(m)} \sim M\big(\theta^{(m)}\big), \quad \theta^{(m)} \sim p_{\text{ABC}}(\theta \mid D_{\text{obs}}) \tag{2.22}$$

In Equation (2.22), θ^(m) is a posterior parameter draw and M is the SIR or SEIZ simulator. Repeating the simulation over many posterior draws produces a distribution of fitted or future spreading curves. Chapter 3 uses this principle for posterior predictive fit, ABC-SEIZ versus ABC-SMC-SEIZ calibration comparison and out-of-sample short-horizon forecasting of $I_{\text{cum}}(t)$.

Figure 2.13 shows the full methodological integration used in the dissertation. Transformer output supplies $p_f(x)$, temporal cascade construction supplies summary statistics and interaction structure, SIR/SEIZ provides the simulator, ABC/ABC-SMC estimates posterior parameter uncertainty and posterior prediction produces fitted, forecasted or scenario-based trajectories. The figure explains why the ABC layer is not independent of the detection and temporal-network layers; it is the posterior-inference stage of the same framework.

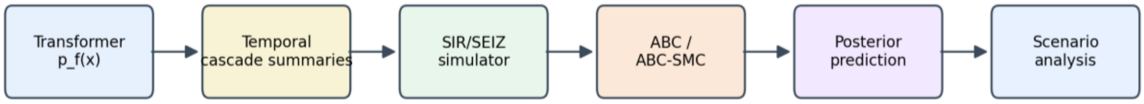

Figure 2.13. ABC calibration of a fake-news cascade using Transformer outputs and temporal-network summaries.

When scenario analysis is required, interventions are represented as parameter perturbations rather than as causal claims about real platform policy:

$$\beta_{\text{intervention}} = \alpha_\beta \beta, \quad 0 < \alpha_\beta < 1; \qquad \delta_{\text{intervention}} = \alpha_\delta \delta, \quad \alpha_\delta > 1 \qquad (2.23)$$

Equation (2.23) states the basic counterfactual interpretation used later in the thesis. Lower β represents reduced visibility or reduced exposure to the fake-news claim, while higher δ represents stronger correction or debunking pressure. In Chapter 3, such scenarios are interpreted as model-based counterfactual comparisons under a calibrated posterior baseline, not as causal estimates of real platform-policy effects. The detailed intervention-parameter interpretation table is retained in Appendix D.3.

Posterior predictive simulation has three methodological roles in the dissertation. First, it checks whether the posterior parameter values reproduce the observed in-sample spreading trajectory. Second, it generates uncertainty bands around the fitted or forecasted $I_{\text{cum}}(t)$ curve, which is necessary because a point estimate alone may hide parameter uncertainty. Third, it provides a bridge to counterfactual reasoning: once a posterior baseline is obtained, controlled parameter perturbations can be applied to explore how visibility reduction, correction or early mitigation would change the same calibrated cascade.

The distinction between fitting and forecasting is preserved throughout the methodology. In-sample posterior predictive fit evaluates whether the model can reconstruct the observed cascade under the same observation window. Out-of-sample short-horizon forecasting evaluates whether the model can predict a held-out segment of $I_{\text{cum}}(t)$ after calibration on an earlier part of the cascade. Chapter 3 uses this second form as a stronger validation check for the two main empirical cascades. The forecast is not presented as an unconditional prediction of future platform behaviour; it is a held-out posterior predictive test under the same dataset and modelling assumptions.

The same principle applies to counterfactual scenario analysis. A scenario such as reduced visibility is not interpreted as evidence that a real platform policy caused a measured historical reduction. It is a controlled model-based perturbation applied to the calibrated posterior baseline. This allows the thesis to rank intervention mechanisms by their expected effect on $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, peak spreading and final cascade size, while avoiding causal claims that cannot be supported by observational cascade data alone. The detailed separation between ABC-SMC inference and posterior predictive application is documented in Appendix D.4.

For the two principal cascades used in Chapter 3, this workflow creates a consistent methodological chain. The COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die." fake-news claim supplies a large health-misinformation cascade with repeated temporal activation. The Hoaxy / OSoMeNet "Three million votes in the Presidential Election were cast by illegal aliens." fake-news claim supplies a large political misinformation cascade with strong network amplification. Both are converted into observed $I_{\text{new}}(t)$ and $I_{\text{cum}}(t)$ trajectories, calibrated through SIR/SEIZ and ABC/ABC-SMC, and then evaluated through posterior predictive and forecast metrics.

2.6.6 Baseline Posterior Fake-News Diffusion-Risk Functional

To formalize the integration of the proposed framework, this thesis introduces a Baseline Posterior Fake-News Diffusion-Risk Functional. The functional is not a replacement for SIR or SEIZ and is not a new epidemic compartment model. Instead, it is a mathematical synthesis that connects Transformer-based fake-news probability, temporal-network amplification and ABC-SMC-SEIZ posterior predictive growth into one baseline diffusion-risk quantity.

For a fake-news claim c, textual input x_c, observed cascade S_c^{obs}, prediction horizon h and observed cascade population size N_c, the baseline risk is defined as:

$$\mathcal{R}_{\text{base}}(c,h) = \hat{p}_f(x_c)[1+\lambda A_c(t)]\mathbb{E}_{\theta\sim\pi_\varepsilon^{\text{SMC}}}\left[\frac{\Delta I_{\text{cum},c}(t,h;\theta)}{N_c}\right] \tag{2.24}$$

In Equation (2.24), $\hat{p}_f(x_c)$ is the Transformer-based probability that the textual item is a fake-news claim or fake-news article. $A_c(t)$ is the temporal-network amplification score computed from the observed cascade, λ controls the influence of this amplification term, $\pi_\epsilon^{\text{SMC}}(\theta \mid S_c^{\text{obs}})$ is the ABC-SMC posterior distribution of SEIZ parameters, $I_{\text{cum},c}^{\text{SEIZ}}(t+h;\theta)$ is the predicted cumulative number of active spreaders at horizon h, $I_{\text{cum},c}^{\text{obs}}(t)$ is the observed cumulative number of active spreaders up to time t, and N_c normalizes the score by the observed cascade population size. The functional therefore measures expected normalized future cumulative growth under the baseline posterior, scaled by content risk and network amplification.

A simple amplification term can be defined from the temporal cascade and network layer as:

$$A_c(t) = \omega_1 I_{\text{new},c}^{\text{norm}}(t) + \omega_2 H_c^{\text{norm}}(t) + \omega_3 B_c^{\text{norm}}(t), \quad \omega_1+\omega_2+\omega_3 = 1 \tag{2.25}$$

In Equation (2.25), I_{new},c^norm(t) is the normalized number of new unique active spreaders in the current time window, $H_c^{\text{norm}}(t)$ is a normalized hub-activity score, and $B_c^{\text{norm}}(t)$ is a normalized relay or betweenness-based activity score. When a minimal operational score is required, $A_c(t)$ may be approximated using only the normalized $I_{\text{new}}(t)$ signal. This keeps the functional compatible with the empirical cascades in Chapter 3, where $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, source hubs and relay users are the observable diffusion outputs.

If ABC-SMC returns M posterior particles $\{(\theta_m, w_m)\}$ for m = 1...,M, Equation (2.24) can be evaluated in particle form as:

$$\hat{\mathcal{R}}_{\text{base}}(c,h) = \hat{p}_f(x_c)[1+\lambda A_c(t)] \sum_m w_m \left[\frac{\Delta I_{\text{cum},c}(t,h;\theta_m)}{N_c}\right] \tag{2.26}$$

Equation (2.26) shows how the functional can be computed from the same ABC-SMC posterior particles that are used for posterior predictive simulation. In Chapter 3, the functional is linked to the two principal empirical cascades: the COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die." fake-news claim and the Hoaxy / OSoMeNet "Three million votes in the Presidential Election were cast by illegal aliens." fake-news claim. The equation is therefore defensible as a baseline risk synthesis because every term corresponds to an implemented output of the dissertation framework.

2.6.7 Framework Integration

The final part of Section 2.6 defines the common evaluation layer used in Chapter 3 and summarizes the logical connection between detection, diffusion and posterior inference. The experimental chapter reports the numerical values of the error metrics, but the mathematical definitions belong to the methodology. For an observed cumulative spreading curve $I_{\text{obs}}(t)$ and a simulated or posterior predictive curve $I_{\text{sim}}(t)$, the principal curve-distance measure is the relative 2-norm distance:

$$d_{L2}(I_{\text{obs}}, I_{\text{sim}}) = \frac{\|I_{\text{sim}}(t) - I_{\text{obs}}(t)\|_2}{\|I_{\text{obs}}(t)\|_2} \tag{2.27}$$

Equation (2.27) is scale-normalized, allowing the COVID-19 real Twitter/X cascade and the Hoaxy / OSoMeNet "Three million votes in the Presidential Election were cast by illegal aliens." cascade to be compared despite different final cascade sizes. Lower values indicate closer agreement between simulated and observed cumulative active-spreader trajectories.

The root mean squared error and mean absolute error are used as complementary absolute-error measures:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(I_{\text{sim}}(t) - I_{\text{obs}}(t)\right)^2} \tag{2.28}$$

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|I_{\text{sim}}(t) - I_{\text{obs}}(t)| \tag{2.29}$$

For out-of-sample posterior predictive forecasting, the normalized RMSE is reported as:

$$\text{nRMSE} = \frac{\text{RMSE}}{I_{\text{obs}}(T)} \tag{2.30}$$

Equations (2.27)-(2.30) provide the common evaluation layer for fitted curves, posterior predictive simulations and short-horizon forecasts. They are not introduced as theoretical contributions; they are methodological definitions used to make the Chapter 3 comparisons reproducible. The new methodological contribution remains the integrated baseline risk functional in Equations (2.24)-(2.26), which combines Transformer content risk, temporal-network amplification and ABC-SMC-SEIZ posterior predictive growth.

The role of the ABC layer in the dissertation can be summarized in four connected steps. First, Transformer models provide calibrated content-risk probabilities. Second, temporal cascade construction produces the $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$ and network summaries needed for simulation. Third, SIR and SEIZ translate those summaries into interpretable diffusion mechanisms, with SIR serving as the baseline and SEIZ serving as the main model. Fourth, ABC and ABC-SMC estimate parameter uncertainty and generate posterior predictive trajectories. This integrated methodology is the basis for the empirical results, forecasts and counterfactual scenario analysis in Chapter 3.

The placement of detailed materials in the appendices is part of this methodological design. Appendix D preserves the supplementary cascade summaries, prior ranges, summary-statistic details, observed-data mappings and intervention-parameter interpretations used to reproduce the ABC and ABC-SMC results. Appendix G retains the extended theoretical discussion of likelihood-free inference, ABC rejection targets, network-spreading interpretation and SIR-SEIZ model-comparison equations. This separation allows Section 2.6 to remain a readable methodology section while still providing sufficient technical depth for audit, examination and reproducibility.

The final methodological boundary is therefore explicit. Section 2.6 defines the models, inference logic, posterior predictive use and baseline risk functional that are actually evaluated in Chapter 3. It does not claim that all hidden states of user belief are observed, nor does it claim that counterfactual scenarios are causal platform experiments. Instead, it defines a coherent simulation-based inference pipeline for partially observed fake-news cascades: observed temporal interactions are summarized, SIR and SEIZ simulators generate comparable trajectories, ABC and ABC-SMC estimate uncertainty, and posterior predictive checks evaluate whether the fitted model reproduces or forecasts the observed cumulative spreading process.

This boundary is important for a high-level PhD methodology because it keeps the contribution both ambitious and defensible. The ambition lies in connecting Transformer-based detection, temporal-network cascade construction, epidemic modelling and likelihood-free posterior inference within one framework. The defensibility lies in the careful separation between observed data, latent states, model assumptions, posterior uncertainty and counterfactual interpretation. Chapter 3 operationalizes this methodology without restating all formulas, using the equations and definitions established in Section 2.6 as the reference layer for the empirical results.

2.7 Dataset Suitability, Ethics and Reproducibility Boundaries

Section 2.7 synthesizes the dataset-suitability, ethical and reproducibility boundaries of the proposed methodology. Its purpose is not to repeat the dataset descriptions already provided in Section 2.3, nor to reproduce the implementation details preserved in Appendices D and G. Instead, it explains why the selected datasets are methodologically appropriate for the two evidence layers of the dissertation: content-based fake-news detection and diffusion-oriented fake-news claim modelling. This synthesis is necessary because a dataset that is useful for binary textual classification is not automatically suitable for temporal-network construction, SIR/SEIZ modelling or ABC/ABC-SMC inference.

The final thesis design therefore uses a two-layer suitability logic. The first layer assesses whether a dataset provides labelled text that can be cleaned, tokenized and used for Transformer-based fake-news detection. The second layer assesses whether a dataset provides timestamped interactions, directed source-target relations, cascade identifiers or claim-level temporal traces that can be transformed into temporal fake-news cascades. This distinction protects the methodology from treating all fake-news datasets as interchangeable and keeps the Chapter 3 experiments aligned with the available evidence.

This section also defines the ethical and reproducibility boundary of the data layer. Social-network diffusion data may contain user identifiers, timestamps and interaction traces; therefore, the thesis uses them only for aggregate cascade modelling and does not attempt to identify, profile or evaluate individual users. Reproducibility is supported by documenting label mapping, preprocessing, cascade filtering, temporal aggregation and ABC/ABC-SMC configuration in the main methodology and appendices [91]-[94], [133], [134].

The suitability synthesis is therefore summarized through two complementary questions:

1) Can the dataset support reliable binary fake-news detection from text?
2) Can the dataset support temporal cascade reconstruction and epidemic-model inference from ordered interactions?

The detailed audit dimensions behind these questions are preserved in Appendix F.4, while the main text retains only the final role decisions required for Chapter 3.

2.7.1 Suitability for Transformer-Based Fake-News Detection

Transformer-based fake-news detection requires datasets with reliable textual fields, interpretable labels, sufficient class coverage and limited risk of train-test leakage. In the operational design of this dissertation, the content-based layer consists of Constraint@AAAI2021, FRN, ISOT, CoAID and the Albanian Fake News Corpus. These datasets support author-generated Transformer and classical machine-learning results in Chapter 3. They are not used as empirical temporal-diffusion evidence unless user-level temporal interactions are available and can be reconstructed responsibly.

For systematic comparison, the suitability of a content-based dataset D_i for Transformer detection can be expressed qualitatively as:

$$S_T(D_i) = w_L L_i + w_T T_i + w_B B_i + w_Q Q_i + w_N N_i \tag{2.31}$$

In Equation (2.31), $S_T(D_i)$ denotes the Transformer-detection suitability of dataset D_i. The term L_i represents label reliability, T_i represents text availability and document length suitability, B_i represents binary class balance, Q_i represents text quality after preprocessing, and N_i represents language or domain suitability. The weights w_L, w_T, w_B, w_Q and w_N are researcher-defined qualitative weights. The expression is not used as a strict numerical score; it formalizes the criteria used when deciding which datasets should remain in the operational content-based layer.

Table 2.16 retains only the content-based datasets that are actually used in the Chapter 3 detection layer. The table also states the methodological boundary of each dataset so that detection evidence is not confused with temporal-diffusion evidence.

Table 2.16. Final content-based datasets retained for Transformer and classical machine-learning detection.

Dataset	**Primary detection role**	**Reason for retention**	**Methodological boundary**
Constraint@AAAI2021	Short COVID-related fake/real text detection.	Supports short social-media-style misinformation classification under pandemic-information conditions [30].	Used for content detection, not for temporal diffusion modelling.

FRN	Long-form fake/real article detection.	Provides long news articles and supports long-sequence Transformer evaluation, including BigBird-style modelling [12], [81], [95].	Used for article-level detection; no cascade reconstruction is inferred from the corpus alone.
ISOT	Large English fake/real article benchmark.	Provides a large binary benchmark for classical and Transformer comparison.	Used for content-level robustness analysis only.
CoAID	Healthcare misinformation detection.	Connects fake-news detection with health-related misinformation and partial social-context evidence [83], [97].	Used mainly as a content dataset; social signals are not treated as complete temporal cascades.
Albanian Fake News Corpus	Low-resource Albanian fake-news detection.	Supports multilingual and low-resource Transformer evaluation [84], [85], [98].	Used for detection and synthetic validation support, not as empirical Twitter/X or Hoaxy evidence.

The practical output of this suitability decision is a controlled content-risk layer. Transformer models trained or evaluated on the selected content datasets can produce a fake-news probability score $p_f(x)$, but this score is only one component of the full thesis framework. It becomes operationally meaningful for diffusion analysis only when it is linked to a fake-news claim that also has temporal-interaction evidence.

2.7.2 Suitability for Temporal Diffusion Modelling, SIR/SEIZ and ABC/ABC-SMC

Temporal diffusion datasets require a different suitability logic. They must provide timestamped interactions, directed user-to-user or source-target relations, cascade membership and extractable summary statistics. These properties are necessary because SIR, SEIZ and ABC/ABC-SMC do not operate directly on raw text; they operate on time-indexed spreading curves and summary statistics derived from the temporal cascade.

For diffusion-oriented data, a qualitative suitability expression can be written as:

$$S_D(D_i) = w_T T_i + w_C C_i + w_E E_i + w_M M_i + w_A A_i + w_Q Q_i \quad (2.32)$$

In Equation (2.32), $S_D(D_i)$ denotes the temporal-diffusion suitability of dataset D_i. The term T_i represents timestamp availability, C_i represents cascade or claim-level structure, E_i represents directed edge availability, M_i represents compatibility with SIR/SEIZ state mapping, A_i represents availability of ABC/ABC-SMC summary statistics, and Q_i represents data quality after temporal preprocessing. As with Equation (2.31), the expression is a methodological guide rather than a fixed numerical rule.

Table 2.17 shows that only two diffusion-oriented datasets are used as the principal empirical cases for SIR/SEIZ, ABC/ABC-SMC, posterior prediction and counterfactual analysis. The additional Hoaxy social cascade and the synthetic Albanian cascade support comparison and validation without being treated as equivalent empirical bases.

Table 2.17. Final diffusion-oriented datasets retained for temporal, epidemic and ABC/ABC-SMC modelling.

Dataset / cascade	**Fake-news claim or item**	**Main diffusion support**	**Role**
COVID-19 real Twitter/X	“The coronavirus is not only affecting the way we live, it’s also dramatically affecting the way we die.”	Timestamped user-to-user interactions filtered to site_type = covid_fake_claim.	Principal empirical health misinformation cascade for temporal-network, SEIZ/ABC-SMC and $I_{\text{cum}}(t)$ forecasting.
Hoaxy / OSoMeNet political	“Three million votes in the Presidential Election were cast by illegal aliens.”	Large-scale retweet cascade with directed user-to-user interactions.	Principal empirical political fake-news claim cascade for SIR/SEIZ, ABC/ABC-SMC and forecast comparison.
Hoaxy-imported social cascade	“Should One Tweet End a Career?”	Claim-labelled Hoaxy social cascade with timestamped spreading records.	Additional comparative social fake-news claim cascade for temporal-network interpretation.
Synthetic Albanian temporal cascade	“Shqipëria do të presë 100 mijë palestinezë nga Gaza.”	Generated timestamped cascade with claim and fact-checking records.	Validation and simulation support for the low-resource Albanian pipeline; not treated as

			empirical Twitter/X evidence.

For ABC/ABC-SMC inference, the observed cascade is reduced to a vector of summary statistics. A generic representation of this vector is:

$$S(D_i) = \left(C_T, t_{\text{peak}}, r_{\max}, d_c, b_c, q_f, n_Z\right) \tag{2.33}$$

In Equation (2.33), C_T is the final cascade size, t_{peak} is the time to peak diffusion, $r_{\max}$ is the maximum spreading rate, d_c is cascade depth, b_c is cascade breadth, q_f is the Transformer fake-news probability associated with the item, and n_Z is the number of skeptical, corrective or fact-checking-related events when such information is available. The exact vector is adapted to each dataset, but the role of the vector remains the same: it translates an observed fake-news claim cascade into summary evidence that can be compared with simulated SIR/SEIZ trajectories in the ABC and ABC-SMC procedures [19]-[26], [35], [86]-[89].

The suitability boundary is therefore explicit. A content-only dataset may provide strong Transformer evidence but cannot by itself support epidemic diffusion modelling. A diffusion-oriented dataset may provide rich interaction evidence but should not be treated as a standard text-classification corpus unless the textual item and its label are sufficiently defined. This is the reason why Chapter 3 separates Transformer detection experiments from temporal-network, SIR/SEIZ, ABC/ABC-SMC and forecasting experiments.

2.7.3 Ethical, Privacy and Reproducibility Considerations

Ethical handling includes anonymization of user identifiers, avoidance of harmful content amplification, and adherence to dataset licensing constraints. Reproducibility is supported through documentation of preprocessing, label mapping, train-validation-test splits, temporal-cascade construction, and ABC configuration parameters.

The same reproducibility boundary is supported by data statements, datasheets, data cards, model cards, dataset-development audits, FAIR data principles and ACM artifact-review guidance, which emphasize transparent provenance, documented preprocessing and clear separation between original data, derived features and evaluation artifacts [91]-[94], [133]-[134], [185].

2.7.4 Methodological Assumptions and Validity Boundaries

The proposed framework rests on explicit validity boundaries. The Transformer score $p_f(x)$ is not treated as absolute truth; it is a model-estimated content-risk signal. A repost, retweet, reply or share is treated as observable spreading activity, not as proof that the user believes the claim.

Similarly, non-engagement is not direct evidence of skepticism. SEIZ states are therefore modelling proxies for observable diffusion behaviour rather than psychological states.

The ABC and ABC-SMC posterior distributions are conditional on the selected simulator, priors, tolerance schedule, distance function and summary statistics. They quantify uncertainty under the stated modelling assumptions; they do not prove a complete causal mechanism for misinformation adoption. The synthetic Albanian cascade is used as validation and pipeline-support evidence, not as an empirical cascade with the same evidential weight as the principal real-world diffusion datasets.

GNNs and XAI are not excluded because they lack scientific value. GNNs are retained as future work because they require more complete user-level graph data, stable social-context links and stronger privacy safeguards. XAI is also retained as future work because faithful explanations require a separate validation design. LLMs are used only as auxiliary or support tools and not as ground-truth sources for labels, test sets or diffusion interpretations.

This boundary prevents the methodology from overclaiming. The dissertation evaluates the implemented Transformer, temporal-cascade, SIR/SEIZ and ABC/ABC-SMC pipeline, while positioning GNN, XAI and broader LLM-based extensions as future research directions that require additional data and validation.

Table 2.18. Implemented components and literature-only or future-work boundaries.

Component	Status in this thesis	Reason for status
BERT, RoBERTa, BigBird, mBERT and XLM-RoBERTa	Implemented and evaluated	They provide the operational Transformer detection layer and generate the model-estimated fake-news probability used by the framework.
SIR and SEIZ diffusion models	Implemented and evaluated	They provide interpretable epidemic-state representations for active spreading, stopping, exposure and skepticism.
ABC and ABC-SMC	Implemented and evaluated	They quantify uncertainty in diffusion parameters when exact likelihood evaluation is not feasible.
GNN and temporal graph learning	Literature review and future work	They require richer user-level graph data, reliable follower or interaction networks and consistent propagation labels than are available across all retained datasets [121]-[129], [153]-[155].

Explainable AI	Future explanatory extension	LIME, SHAP and explainable fake-news systems are relevant, but a full XAI layer would require a separate explanation-validation protocol [115]-[117], [152], [172].
Large Language Models	Auxiliary or support role only	LLMs can assist augmentation or analysis, but final labels, evaluation sets and diffusion interpretations are not generated from LLM output [47]-[49], [135]-[136], [143]-[145].
Voting, stacking and classifier ensembles	Literature review only	They may improve accuracy but would reduce traceability between the Transformer score and the downstream diffusion-risk interpretation.

2.8 Chapter Summary

This chapter presented the methodological framework of the dissertation. It defined the selected datasets, preprocessing procedures, temporal-network representation, SIR/SEIZ diffusion modelling and ABC-based parameter inference. It also added explicit methodological traceability from research questions to Chapter 3 evidence, clarified the boundary between model-estimated fake-news probability and ground truth, and separated implemented components from literature-only or future-work extensions such as GNN, XAI and auxiliary LLM support. These methodological elements provide the basis for the experiments presented in Chapter 3.

The methodological elements defined in this chapter are evaluated in Chapter 3 in the same sequence: first through Transformer-based binary detection, then through temporal cascade reconstruction, followed by SIR/SEIZ diffusion modelling, ABC/ABC-SMC posterior inference, posterior predictive forecasting and counterfactual intervention analysis.

Chapter 3. Experimental Results and Discussion

3.1 Introduction

This chapter reports the empirical results obtained after executing the Transformer-ABC-Epidemic framework defined in Chapter 2. The organizing principle is intentionally strict: Chapter 2 explains how the framework works, whereas Chapter 3 reports what happened when the framework was run on the selected content-based and diffusion-oriented evidence. Consequently, this chapter does not repeat full dataset descriptions, preprocessing recipes, tokenization rules, label-mapping procedures or mathematical derivations already established in the methodology chapter and the appendices. It concentrates on numerical outputs, visual evidence, model comparisons and interpretation boundaries.

The empirical design follows the detection-diffusion-inference sequence of the thesis. First, Transformer-based and baseline text-classification experiments evaluate whether fake and real news can be separated reliably across short-text, long-document, healthcare and multilingual settings. Second, selected fake-news claims are converted into temporal cascades so that user activation, cascade growth and time-respecting diffusion patterns can be observed. Third, SIR and SEIZ epidemic models are fitted to the principal empirical cascades. Finally, ABC and ABC-SMC are used to estimate posterior parameter uncertainty, posterior predictive behaviour and intervention-oriented outcomes.

This structure is necessary because fake-news risk is not a single empirical quantity. A detector may achieve high F1-score or ROC-AUC, but this does not prove that a false claim spreads widely, peaks quickly or remains active after corrective information. Conversely, a cascade may show strong temporal amplification even when the originating content is only one item within a larger misinformation environment. Chapter 3 therefore evaluates content-level and diffusion-level evidence separately before integrating them in the final discussion.

The dataset boundary is preserved from Chapter 2. Content-based datasets are used for binary fake-news detection and model comparison, whereas diffusion-oriented datasets are used for temporal-cascade reconstruction, SIR/SEIZ modelling and ABC/ABC-SMC inference. This prevents classification accuracy, cascade size, posterior uncertainty and synthetic validation from being treated as interchangeable evidence. In practical terms, the five content datasets provide item-level fake/real labels, while the principal diffusion cases provide timestamped spreading traces for claim-level temporal modelling.

For content detection, the chapter reports author-generated experimental outputs rather than importing performance values from the original dataset papers. The emphasis is placed on accuracy, precision, recall, F1-score, ROC-AUC, confusion matrices and ROC curves because these outputs show what the implemented detection layer produced under the thesis pipeline. The goal is not to claim that one Transformer architecture is universally superior across all misinformation domains, but to identify how model behaviour changes across short posts, long news articles, healthcare misinformation and low-resource Albanian content.

For diffusion analysis, the chapter focuses on observable temporal behaviour. The COVID-19 Twitter/X fake-news cascade and the Hoaxy / OSoMeNet political fake-news cascade provide the two principal empirical cases for modelling. They are used to examine new-spreader activity, cumulative spreading curves, cascade shape, temporal reachability and the fit of SIR, SEIZ and ABC-calibrated models. Additional temporal cases are retained only as supplementary or validation-oriented evidence and are not given the same inferential status as the two principal empirical cascades.

The diffusion and Bayesian results are interpreted as modelling approximations rather than literal biological claims about social behaviour. SIR and SEIZ test how well epidemic-style state transitions approximate observed spreading curves, while ABC and ABC-SMC quantify posterior uncertainty and posterior predictive behaviour. The baseline posterior fake-news diffusion-risk functional is used only as an integrative demonstration linking Transformer-estimated fake-news probability, temporal-network amplification and forecast growth.

Detailed cleaning, tokenization, harmonization, split documentation, run logs and output inventories are referenced rather than repeated; they remain in Chapter 2 and Appendix C, with supplementary cascade and intervention material in Appendix D and extended dataset cards in Appendix F.

The remainder of the chapter follows this execution sequence: Section 3.2 summarizes the setup, Section 3.3 reports detection results, Sections 3.4-3.6 report cascade, epidemic and posterior outputs, Section 3.7 reports interventions, and Sections 3.8-3.9 synthesize the empirical conclusions.

3.3 Transformer-Based Binary Detection Results and Comparative Analysis

This section reports the content-detection results produced by the Transformer-ABC-Epidemic framework. The purpose is not to repeat the preprocessing, tokenization, split construction or implementation settings defined in Chapter 2 and Appendix C. Instead, Section 3.3 presents what happened when the binary detection layer was executed on the selected content-based datasets: Constraint@AAAI2021, FRN, ISOT, CoAID and the Albanian Fake News Corpus.

The section is organized as a compact results block. Section 3.3.1 reports the short-text COVID-19 detection result. Section 3.3.2 synthesizes the long-text and domain-specific evidence from FRN, ISOT and CoAID. Section 3.3.3 reports the multilingual Albanian result. Section 3.3.4 integrates the main comparative patterns and explains why the detection outputs are suitable as content-risk signals for the temporal cascade analysis that follows.

3.2 Experimental Setup

This section summarizes only the experimental settings required for interpreting the results reported in Chapter 3. Detailed preprocessing, tokenization, dataset harmonization, temporal-cascade construction and ABC/ABC-SMC implementation procedures are presented in Chapter 2 and Appendix C. Therefore, this section does not repeat full dataset descriptions or preprocessing blocks. Its purpose is to state the evidence used in the experiments, the role assigned to each dataset, the train-validation-test and evaluation logic, and the reproducibility conditions under which the reported results were generated.

The experimental setup follows the same four-block structure as the thesis methodology: detection, temporal cascade analysis, epidemic modelling and Bayesian inference. Each block produces a different type of evidence. Detection experiments produce classification metrics and probability outputs. Temporal experiments produce cascades and time-ordered spreading curves. Epidemic experiments produce SIR and SEIZ fitted trajectories. Bayesian experiments produce posterior parameter estimates and posterior predictive summaries. The blocks are connected, but they are not interchangeable. This experimental bridge is also aligned with the author's CSIT 2025 work on fake-news classification and Twitter temporal-network spreading dynamics [108].

Constraint@AAAI2021 was used as the short-text benchmark; FRN and ISOT represented long-document settings; CoAID provided healthcare misinformation; and the Albanian corpus supported multilingual evaluation. These datasets were evaluated as content-based evidence only. Their role in Chapter 3 is to test the detection layer across different text lengths, domains and languages, not to provide direct temporal diffusion evidence. The dataset sources and corpus descriptions are documented in the bibliography and Chapter 2 [30], [81]-[85], [95]-[98], [109], [110].

Diffusion experiments relied on two principal empirical cases: a COVID-19 Twitter/X fake-news cascade and the Hoaxy / OSoMeNet “Three million votes” cascade. These cases provide the principal temporal evidence for cascade reconstruction, SIR/SEIZ comparison, ABC/ABC-

SMC posterior inference, posterior prediction and intervention-oriented interpretation. Additional temporal examples, including the Hoaxy-imported "Should One Tweet End a Career?" case and the synthetic Albanian cascade, are retained only for supplementary comparison or pipeline validation; they are not interpreted with the same inferential weight as the two principal empirical cascades. The Hoaxy/OSoMeNet and temporal-diffusion data sources are documented in the bibliography and Chapter 2 [52]-[55], [90].

Table 3.1. Experimental evidence and datasets.

Block	Datasets	Output
Detection	Constraint, FRN, ISOT, CoAID, Albanian	F1, ROC
Temporal	Twitter COVID, Hoaxy	Cascades
Epidemic	Twitter COVID, Hoaxy	SIR/SEIZ
Bayesian	Twitter COVID, Hoaxy	Posterior

All datasets used the preprocessing pipeline described in Chapter 2 and Appendix C. For the content datasets, this means that text cleaning, binary label harmonization, tokenizer preparation, split handling and output-file storage followed the documented procedure. For the diffusion-oriented datasets, timestamp parsing, interaction filtering, cascade construction and temporal aggregation followed the temporal-network preparation rules defined in the methodology chapter. The full preprocessing blocks are not repeated here because the results chapter is reserved for outputs and interpretation.

The content-detection experiments were evaluated at item level. Each news article, post, claim or text record was treated as one classification instance with a binary fake/real label. When an official train-validation-test split was available, the original split was preserved. Where a fixed split had to be created, one documented split was used consistently for the reported experiment. This rule ensures that the performance values in Section 3.3 can be traced back to stored prediction files and metric outputs rather than to an undocumented ad hoc split.

The central leakage-prevention rule was that the same article, source record, hydrated tweet group, duplicated claim or augmented variant must not appear simultaneously in training and test evidence. This rule is especially important in fake-news detection because duplicated or near-duplicated claims can inflate model performance if they are distributed across partitions. Controlled augmentation, where used for low-resource support, was restricted to the training side only and was not counted as independent validation or test evidence.

For the content datasets, the main evaluation measures were accuracy, precision, recall, F1-score and ROC-AUC. F1-score is emphasized because several fake-news datasets may contain class imbalance or domain-specific label distributions. ROC-AUC is retained because it evaluates ranking behaviour across probability thresholds and is useful for interpreting fake-class probability scores. Confusion matrices are used to support false-positive and false-negative analysis, while ROC curves show whether the selected model separates fake and real examples consistently across thresholds.

The Transformer experiments were compared with classical machine-learning baselines where this added interpretive value. TF-IDF, Logistic Regression, SVM, Naive Bayes and related baselines establish the lexical reference level, while BERT, RoBERTa, BigBird, mBERT and XLM-RoBERTa test contextual, long-sequence and multilingual settings. The comparison is reported as empirical evidence from the thesis implementation rather than as a general ranking of all possible fake-news detectors.

The diffusion-oriented experiments were evaluated at cascade or time-series level rather than by random row-level train-test partitions. A retweet, repost, reply or directed interaction belongs to a time-ordered spreading process; randomly splitting individual interaction rows would break the chronological structure required for temporal networks, SIR/SEIZ simulation and ABC/ABC-SMC inference. For this reason, the temporal experiments use claim-level cascades, time-aggregated spreading curves, calibration windows and short-horizon prediction windows instead of ordinary row-level classification splits.

For the COVID-19 Twitter/X cascade and the Hoaxy / OSoMeNet cascade, the observable temporal signal is represented through quantities such as new active spreaders, cumulative spreaders, cascade size, time to peak and temporal reachability. These values are reported in Section 3.4 before epidemic modelling is applied. This order is deliberate: the empirical cascade is first described as an observed temporal process, and only then is it interpreted through SIR, SEIZ and Bayesian posterior models.

SIR and SEIZ are evaluated using the reconstructed empirical spreading curves as observed diffusion evidence. SIR is treated as the baseline epidemic model because it provides a simple susceptible-infected-recovered structure. SEIZ is treated as the richer diffusion model because it includes exposed and skeptic/non-spreading states. The comparison therefore evaluates whether the additional state structure improves the approximation of observed fake-news diffusion under the selected empirical cascades.

For SIR, SEIZ and ABC/ABC-SMC evaluation, the principal fitting measure follows the relative L2 distance defined in Chapter 2. RMSE, MAE, normalized RMSE, posterior predictive error and visual-fit interpretation are reported only where they add explanatory value. These measures are not introduced as new methodology in Chapter 3; they are used to interpret the quality of the model runs produced by the pipeline.

The ABC and ABC-SMC experiments evaluate uncertainty in diffusion parameters rather than producing only a single deterministic fit. ABC samples candidate parameter values that generate simulated cascades sufficiently close to the observed cascade summaries. ABC-SMC refines this process by sequentially reducing the tolerance level and updating the posterior sample population. The relevant outputs are posterior parameter estimates, credible intervals, accepted particles, posterior predictive simulations, convergence behaviour and comparative calibration between ABC-SEIZ and ABC-SMC-SEIZ.

The Bayesian diffusion analysis is reported only for the two principal empirical cascades. This boundary is necessary because posterior inference requires a sufficiently structured observed cascade for summary-statistic extraction, distance evaluation and repeated simulation. The synthetic Albanian cascade is useful for checking whether the temporal-network pipeline can operate under a low-resource scenario, but it is not interpreted as empirical evidence of real Twitter/X or Hoaxy diffusion. Similarly, supplementary Hoaxy-imported visualisation data support comparative temporal interpretation but do not carry the same inferential weight as the principal empirical cases.

The implementation uses a reproducible Python-based pipeline. Transformer experiments rely on saved model outputs, metric files, prediction files, confusion-matrix counts and ROC-curve data. The appendix materials preserve split summaries, run logs and output files so that the values reported in the chapter can be checked without relying only on manually copied tables. The Bayesian diffusion component is stored separately from the Transformer detection outputs because the experimental unit differs: detection operates on news items or posts, while diffusion inference operates on claim-level temporal cascades.

For the ABC and ABC-SMC runs, the recorded outputs include the observed cascade, selected summary statistics, prior ranges, distance function, tolerance rule, accepted particles and posterior predictive trajectories. The implementation uses ABCpy for likelihood-free sampling and posterior output management, following the computational perspective on Approximate Bayesian Computation used in ABCpy [107]. This software boundary is included only to clarify reproducibility; the methodological details are not repeated here because they belong to Chapter 2 and the supplementary materials.

The fixed reproducibility requirements for Chapter 3 are as follows: each dataset uses one documented split or experimental boundary; augmentation is excluded from validation and test evidence; timestamp parsing is recorded for diffusion-oriented data; user identifiers are anonymized or handled according to dataset constraints; the output files used to produce tables and figures are archived; and ABC/ABC-SMC settings are saved together with posterior outputs. These requirements allow the reported metrics, cascade curves and posterior estimates to be traced back to documented experimental artefacts.

The evaluation protocol also defines an interpretation boundary. Accuracy, F1-score and ROC-AUC describe how well a model classifies content as fake or real. Cascade size, time to peak and temporal reachability describe how a claim spread through a temporal interaction structure. SIR/SEIZ fitting error describes how well an epidemic-style model approximated that spreading curve. ABC/ABC-SMC posterior summaries describe uncertainty in diffusion parameters and posterior predictive behaviour. The outputs are complementary, but they must not be interpreted as the same type of evidence.

Model-specific execution is reported only insofar as it affects interpretation. Short-sequence Transformers are evaluated where the input is naturally claim-level or post-level, while long-sequence modelling is considered where article-level evidence would otherwise be truncated. Multilingual modelling is evaluated only for the Albanian corpus, where the purpose is to test whether the same detection layer can operate under low-resource language conditions. These distinctions are experimental boundaries, not separate methodological frameworks.

The classification tables that follow therefore report outputs generated by the implemented pipeline under the documented split and evaluation rules. They should be read as evidence of behaviour under the thesis configuration rather than as universal benchmark rankings. This is particularly important for long-document and healthcare misinformation datasets, where class balance, text length, source artefacts and domain-specific vocabulary may influence performance. The interpretation of each table therefore focuses on the observed result, the error pattern and its relevance to the detection-diffusion bridge.

The temporal and Bayesian result tables follow the same evidence principle. A cascade table describes the observed spreading trace; a SIR or SEIZ table describes the quality of an epidemic-style approximation; an ABC or ABC-SMC table describes posterior uncertainty under the selected summary statistics and simulator assumptions. These outputs are connected through the framework, but each answers a different empirical question. The reporting sequence is designed to prevent posterior inference from being interpreted as a substitute for observed cascade evidence.

This setup also defines how supplementary material should be read. Appendix C supports reproducibility of detection runs, while Appendix D supports detailed cascade and intervention interpretation. The main chapter retains only the outputs required to judge the empirical execution of the framework. This separation is deliberate: it allows Chapter 3 to remain a results chapter while preserving the audit trail needed for a high-level PhD evaluation.

Accordingly, Section 3.3 reports what happened in the Transformer-based detection experiments, Section 3.4 reports what happened when fake-news claims were represented as temporal cascades, Section 3.5 reports what happened when SIR and SEIZ were fitted to the empirical spreading curves, Section 3.6 reports what happened when ABC and ABC-SMC were used for posterior inference, and Section 3.7 reports what happened under counterfactual intervention scenarios. The remainder of Chapter 3 therefore follows the same order as the implemented framework while avoiding the methodological repetition that belongs to Chapter 2.

3.3.1 Results on Content-Based Datasets

Table 3.2 summarizes the author-generated detection results on Constraint@AAAI2021 Task-A. BERT achieved the best overall performance, reaching 98.11% accuracy and 98.00% F1-score. The strongest non-Transformer baseline was SVM + TF-IDF with 94.39% F1-score. Compared with this baseline, BERT improved performance by approximately 3.6 percentage points, indicating the benefit of contextual representations for short COVID-19 misinformation messages.

Table 3.2. Author-generated detection performance on Constraint@AAAI2021 Task-A.

Model / method	**Accuracy**	**Precision**	**Recall**	**F1-score**
CNN + BiLSTM	92.01%	92.01%	92.01%	92.01%
SVM + TF-IDF	94.35%	94.42%	94.39%	94.39%
SVM + Word2Vec (ED = 200)	92.66%	92.67%	92.66%	92.66%
SVM + Word2Vec (ED = 150)	92.94%	92.94%	92.94%	92.94%
BERT	98.11%	98.11%	98.00%	98.00%
Logistic Regression (LR)	92.07%	92.11%	92.00%	92.00%
Support Vector Machine (SVM)	91.09%	91.13%	91.00%	91.00%
Naive Bayes (NB)	93.01%	93.06%	93.00%	93.00%
NB + LR	93.01%	93.05%	93.00%	93.00%

The confusion matrix and ROC curve in Figure 3.1 further confirm strong separation between fake and real classes. The BERT classifier correctly identified 1,002 real and 1,098 fake messages, with only 18 false positives and 22 false negatives. The ROC-AUC value of 0.9943 suggests highly reliable ranking performance across thresholds. These results support the use of BERT as the short-text content-detection layer within the proposed framework.

Figure 3.1 combines the two complementary diagnostics of the best-performing BERT detector on Constraint@AAAI2021 Task-A. Panel (a) shows a low-error confusion matrix, with 1,002 real messages and 1,098 fake messages correctly classified, and only 18 false positives and 22 false negatives. Panel (b) shows strong threshold-independent separation, with ROC-AUC = 0.9943. Together, the two panels confirm that the short-text detection layer provides a reliable content-risk signal before temporal cascade analysis.

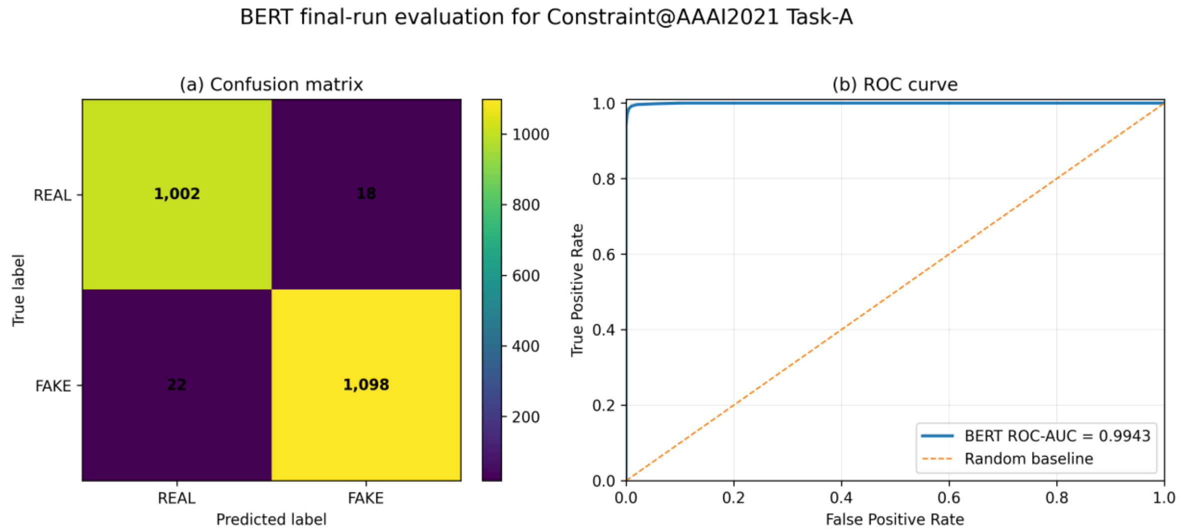

Figure 3.1. BERT final-run evaluation for Constraint@AAAI2021 Task-A: (a) confusion matrix; (b) ROC curve.

3.3.2 Long-Text and Domain-Specific Detection Results

FRN, ISOT and CoAID were used to evaluate model behaviour under long-document and domain-specific conditions. FRN and ISOT represent article-level fake-news detection, while CoAID evaluates COVID-19 healthcare misinformation across short posts/claims and longer news items. Detailed preprocessing and run-log material is retained in Appendix C and Appendix F; the main text keeps only the comparative results needed for interpreting the detection layer.

Table 3.3. Compact comparison of strongest classical baselines across FRN, ISOT and CoAID.

Dataset / subset	**Best classical baseline**	**Fake F1**	**Macro-F1**	**ROC-AUC**
FRN	Linear SVM + TF-IDF	0.9531	0.9526	0.9903
ISOT	Linear SVM + TF-IDF	0.9924	0.9921	0.9996
CoAID social posts / claims	Logistic Regression + TF-IDF	0.9358	0.9488	0.9830
CoAID longer news	Logistic Regression + TF-IDF	0.4615	0.7243	0.9328

Table 3.4. Comparative Transformer detection performance across FRN, ISOT and CoAID.

Dataset / subset	**Text type**	**Best Transformer setting**	**Fake F1**	**Macro-F1**	**ROC-AUC**

FRN	Long political/general articles	BigBird-RoBERTa-base	0.9691	0.9692	0.9946
ISOT	Large English article benchmark	RoBERTa title-body	0.9948	0.9945	0.99997
CoAID social posts / claims	Short healthcare claims/posts	RoBERTa-base	0.9615	0.9701	0.9915
CoAID longer news	Longer healthcare news items	BigBird-RoBERTa-base	0.8205	0.9081	0.9786

For FRN, BigBird-RoBERTa achieved the strongest performance, with fake-class F1 = 0.9691 and ROC-AUC = 0.9946. This confirms the advantage of long-sequence modelling when relevant evidence may appear beyond the first 512 tokens. Figure 3.2 confirms that the FRN BigBird configuration produces a balanced error profile for real and fake long-form articles.

Figure 3.2 confirms strong class separation for the FRN BigBird model. Only 18 real articles are incorrectly flagged as fake and 21 fake articles are missed. This error pattern indicates strong class separation and a balanced trade-off between fake-class recall and false-positive control, which is important when the detection output is later interpreted as a content-risk signal.

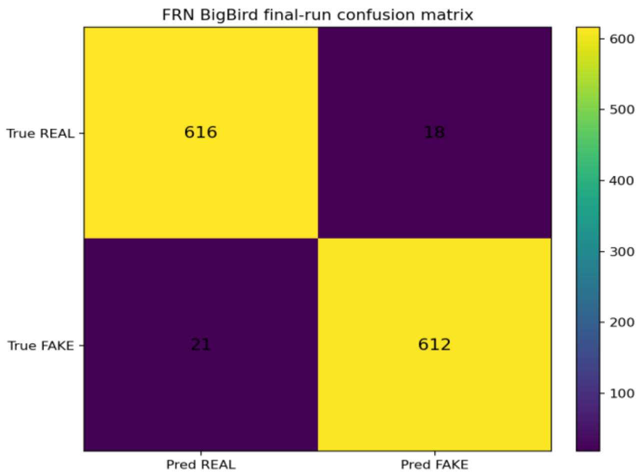

Figure 3.2. BigBird final-run confusion matrix for FRN.

ISOT produced the highest overall scores. RoBERTa achieved fake-class F1 = 0.9948 and ROC-AUC = 0.99997. This very strong result is retained as benchmark evidence, but it should be interpreted cautiously because ISOT contains source and style regularities that may simplify article-level fake/real separation. Figure 3.3 therefore confirms ranking strength on the benchmark rather than universal cross-domain generalization.

Figure 3.3 shows that the best-performing ISOT configuration (RoBERTa title-body) achieves both near-perfect ranking separation and a very low error profile. The ROC curve in panel (a) indicates excellent discriminative ability (ROC-AUC = 0.99997), while the confusion matrix in panel (b) shows only 22 real articles misclassified as fake and 27 fake articles misclassified as real. Taken together, these plots confirm that the model performs extremely strongly on the ISOT benchmark, although the result should still be interpreted with caution because of dataset-specific source and style regularities.

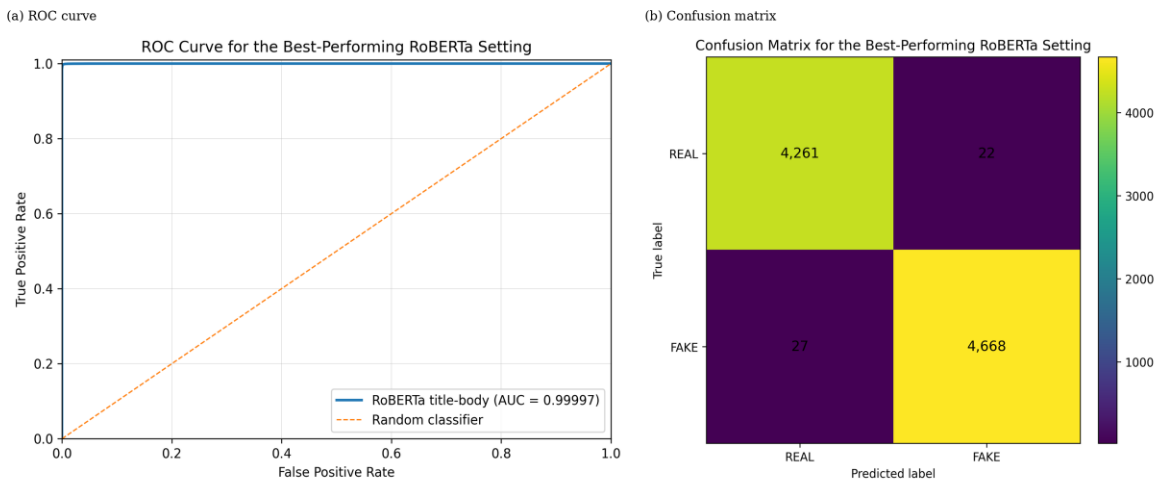

Figure 3.3. Complementary evaluation plots for the best-performing ISOT setting: (a) ROC curve; (b) confusion matrix.

For CoAID, different behaviour was observed across subsets. RoBERTa performed best on short healthcare posts and claims, while BigBird produced the strongest result for longer news items. This supports the hypothesis that sequence length influences fake-news detection performance. Overall, Transformer models consistently exceeded classical TF-IDF baselines, while BigBird showed advantages for long-document settings.

Figure 3.4 summarizes the strongest CoAID Transformer ranking results for short healthcare posts/claims and longer news items.

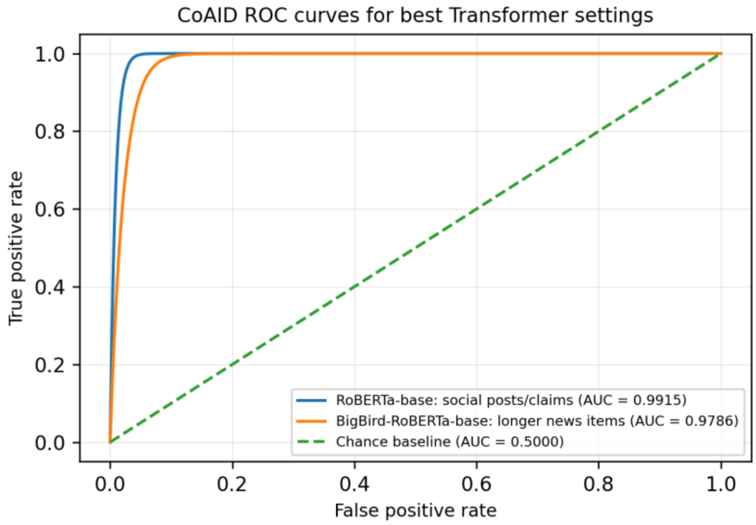

Figure 3.4. ROC curves for the best CoAID Transformer settings.

3.3.3 Multilingual and Low-Resource Results

The Albanian Fake News Corpus was included to evaluate detection performance in a low-resource multilingual setting. Table 3.5 reports the compact comparison between the strongest classical baseline and multilingual Transformer models. XLM-RoBERTa achieved the strongest result, with fake-class F1 = 0.9285 and ROC-AUC = 0.9857, outperforming both the classical baseline and mBERT. These results indicate that multilingual contextual representations capture Albanian credibility patterns more effectively than sparse lexical features.

Table 3.5. Albanian multilingual detection performance.

Model	Fake F1	ROC-AUC
Linear SVM + word/char TF-IDF	0.8802	0.9516
mBERT-base multilingual	0.9057	0.9741
XLM-RoBERTa-base	0.9285	0.9857

Figure 3.5 presents the confusion matrix and ROC curve of the best-performing XLM-RoBERTa configuration.

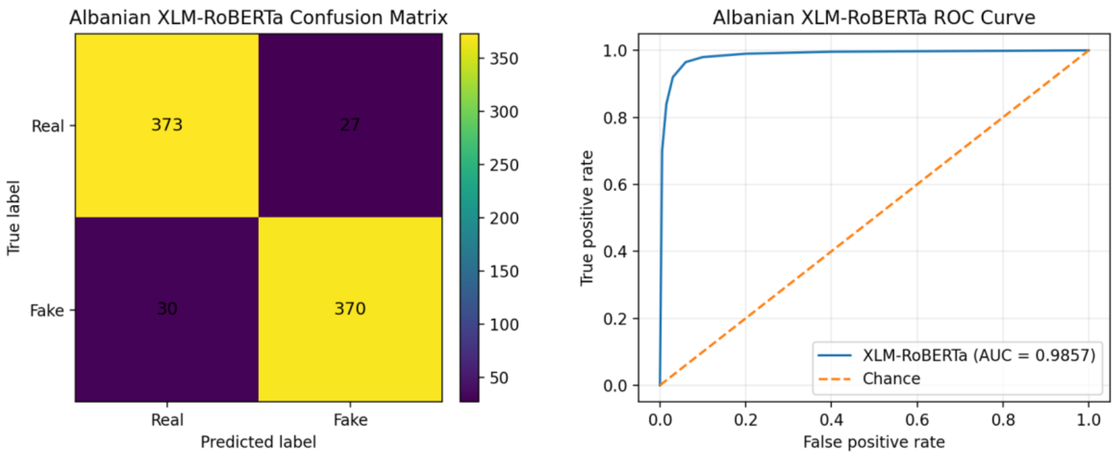

Figure 3.5. Albanian XLM-RoBERTa evaluation results: confusion matrix and ROC curve.

3.3.4 Comparative Detection Analysis

The cross-dataset results indicate three consistent patterns. First, performance depends on dataset type and domain: short COVID-19 posts, long political articles, healthcare misinformation and Albanian news contain different credibility cues. Second, long-text datasets benefit from sequence-preserving strategies such as sliding windows and BigBird. Third, multilingual datasets require language-aware Transformer representations. Across all experiments, false negatives remain more critical than false positives because missed fake items weaken downstream diffusion-risk estimation.

Table 3.6 confirms that no single input policy is optimal across all settings. BERT is sufficient for short COVID-19 posts, RoBERTa is highly effective for article-style benchmarks and short healthcare claims, BigBird is preferable when longer article evidence must be preserved, and XLM-RoBERTa is the strongest setting for low-resource Albanian detection. The result therefore supports the thesis decision to treat the Transformer layer as a controlled content-risk module rather than as a one-size-fits-all classifier.

Table 3.6. Author-generated cross-dataset comparison of the strongest Transformer detection settings.

Dataset / subset	**Text type**	**Best Transformer setting**	**Fake F1**	**Macro-F1**	**ROC-AUC**
Constraint@AAAI2021	Short COVID-19 posts	BERT-base	0.9800	0.9800	0.9943
FRN	Long political/general articles	BigBird-RoBERTa-base	0.9691	0.9692	0.9946

ISOT	Large English article benchmark	RoBERTa title-body	0.9948	0.9945	0.99997
CoAID social posts/claims	Short healthcare claims/posts	RoBERTa-base	0.9615	0.9701	0.9915
CoAID longer news	Longer healthcare news items	BigBird-RoBERTa-base	0.8205	0.9081	0.9786
Albanian Fake News Corpus	Low-resource Albanian news	XLM-RoBERTa-base	0.9285	0.9287	0.9857

Table 3.7. Main false-positive and false-negative patterns across the content-based detection datasets.

Dataset / group	**Main false-positive pattern**	**Main false-negative pattern**	**Framework implication**
Short COVID-19 posts and claims	Real health posts with alarmistic vocabulary can resemble misinformation.	Fake claims written in cautious or official-sounding language may be missed.	High fake-class recall is required before temporal monitoring.
Long English articles	Opinionated real articles may resemble unreliable sources.	Long fake articles may hide misleading evidence outside truncated input.	Sliding windows and BigBird reduce length-related evidence loss.
Healthcare longer news	Legitimate COVID-19 reports may contain risk vocabulary similar to fake claims.	Neutral biomedical wording can mask misleading content.	Longer health news requires both ranking quality and fake-class recall.
Low-resource Albanian news	Emotionally loaded real Albanian texts may be flagged as fake.	Formal journalistic style can make fake Albanian texts appear credible.	Language-aware tokenization is necessary for robust multilingual detection.

This comparative analysis completes the content-detection part of the empirical evaluation. The models provide high-quality fake-class probability outputs, but classification metrics alone do not measure social harm or diffusion risk. The next section therefore moves from item-level detection to temporal cascade and network evidence, where selected fake-news claims are evaluated according to their observed spreading behaviour.

3.4 Temporal Cascade Results

This section reports the temporal-cascade evidence that connects the Transformer detection results in Section 3.3 with the SIR/SEIZ diffusion models in Section 3.5. The section is intentionally results-focused. It does not repeat the full preprocessing, timestamp parsing or graph-construction procedure from Chapter 2; instead, it reports only the outputs needed for the transition from content-level detection to epidemic-style modelling: the detection gate, the selected cascades, the observed infected-user trajectories and two representative temporal-network structures.

The central distinction is that a high fake-news probability is not, by itself, diffusion evidence. Transformer output is used as a screening signal, whereas temporal modelling requires timestamped interaction records, user or node identifiers and a claim-scoped cascade boundary. This rule prevents the chapter from treating text-classification accuracy, cascade reach and synthetic validation as equivalent forms of evidence. It also makes clear why only the two principal empirical cascades are carried forward to the main SIR/SEIZ and ABC/ABC-SMC analyses.

Accordingly, Section 3.4 is reorganized into three compact result blocks. Section 3.4.1 states the detection gate and cascade-selection decision. Section 3.4.2 summarizes the observed infected-user trajectories in one table and one combined figure. Section 3.4.3 retains only the two principal empirical network examples and moves auxiliary network visualizations to Appendix D. This structure keeps the section aligned with the purpose of a results chapter: to show what was observed and how those observations justify the next modelling step.

3.4.1 Detection gate and cascade selection

Transformer outputs were used as a screening mechanism before temporal-network modelling. A claim was transferred to cascade analysis only when the fake-class probability exceeded the decision threshold, $p(\text{fake}|x) >= 0.50$, and when the item also had sufficient temporal interaction evidence. This two-part gate separates content-level credibility detection from diffusion evidence and prevents content-only datasets from being used as epidemic-model inputs.

Table 3.8 summarizes the selected cases. The COVID-19 Twitter/X and Hoaxy / OSoMeNet cases are the two principal empirical cascades used later for SIR, SEIZ and ABC/ABC-SMC fitting. The Hoaxy-imported “Should One Tweet End a Career?” case is retained only as an auxiliary comparative trajectory, while the Albanian case is retained as a synthetic validation example for checking that the pipeline can support low-resource claim inputs when an appropriate temporal schema is available.

Table 3.8. Detection gate and cascade-selection decision before temporal-network modelling.

Case	p(fake|x)	Interaction evidence	Decision
COVID-19 Twitter/X fake-news claim	0.914	Timestamped directed Twitter/X interactions	Cascade analysis; transferred to SIR/SEIZ
Hoaxy / OSoMeNet "Three million votes" claim	0.963	Claim-scoped retweet cascade	Cascade analysis; transferred to SIR/SEIZ
"Should One Tweet End a Career?" Hoaxy-imported claim	0.874	Small Hoaxy-style temporal cascade	Comparative trajectory only
Synthetic Albanian fake-news claim	0.918	Generated Hoaxy-style temporal schema	Validation only

All selected items exceeded the decision threshold and were therefore eligible for temporal inspection. However, eligibility does not mean that all cases carry the same evidential weight. The two large empirical cascades support model fitting because they contain sufficient observed interaction volume and chronological structure. The auxiliary Hoaxy-imported case and the Albanian synthetic case are useful for comparison and validation, but they are not treated as equal empirical bases for the posterior-inference and intervention results later in the chapter.

The gate therefore performs a methodological narrowing function. It moves the chapter from the general detection problem, where the unit of analysis is a labelled text item, to the cascade problem, where the unit of analysis is a time-ordered set of user interactions associated with one fake-news claim. This narrowed set supplies the observed I(t) curves required for SIR and SEIZ modelling.

3.4.2 Observed infected-user trajectories

After cascade selection, the temporal evidence is reduced to two directly observed series. $I_{\text{new}}(t)$ denotes the number of users who first appear as active spreaders within the selected aggregation window, while $I_{\text{cum}}(t)$ denotes the cumulative number of unique active spreaders up to that time. In the empirical Twitter/X and Hoaxy cases, active spreading is derived from retweet, quote or directed sharing evidence; in the Albanian case, it is derived from the synthetic claim-labelled records used only for validation.

The reporting emphasis is the comparative cascade shape rather than a repeated case-by-case reconstruction. For this reason, the principal numerical descriptors are consolidated in Table 3.9 and the four observed curves are merged in Figure 3.6. Detailed node counts, edge counts, component sizes and auxiliary graph summaries can be retained in Appendix D, where they support reproducibility without weakening the compact results narrative in the main chapter.

Table 3.9 summarizes the principal cascade characteristics used later for SIR and SEIZ fitting. The key point is not that every selected item becomes a full epidemic-inference case. Rather, the table separates the empirical cascades with sufficient scale and temporal structure from the auxiliary and validation examples that support interpretation but should not dominate the main model-comparison claims.

Table 3.9. Cascade summary statistics used for the SIR/SEIZ transition.

Case	**Peak I_{new}**	**Final I_{cum}**	**Records**	**Role in later analysis**
COVID-19 Twitter/X fake-news claim	5,892	95,599	133,374	Principal empirical cascade
Hoaxy / OSoMeNet “Three million votes” claim	4,660	149,079	1,048,575	Principal empirical cascade
“Should One Tweet End a Career?”	638	797	817	Auxiliary comparison
Synthetic Albanian fake-news claim	39	520	520	Pipeline validation only

This consolidation is important for assessment because it makes the evidence hierarchy auditable. The record volume indicates whether the curve is estimation-ready or only diagnostic; the final cumulative size supports normalization; and the peak incremental activity summarizes burst intensity without requiring four separate textual reconstructions.

Figure 3.6 summarizes the observed incremental and cumulative infected-user trajectories extracted from the selected cascades. The Hoaxy case exhibits the largest diffusion reach, while the COVID-19 case shows repeated activity bursts over a longer observation period. The “Should One Tweet End a Career?” cascade demonstrates delayed burst behaviour, whereas the Albanian example is retained only for synthetic pipeline validation.

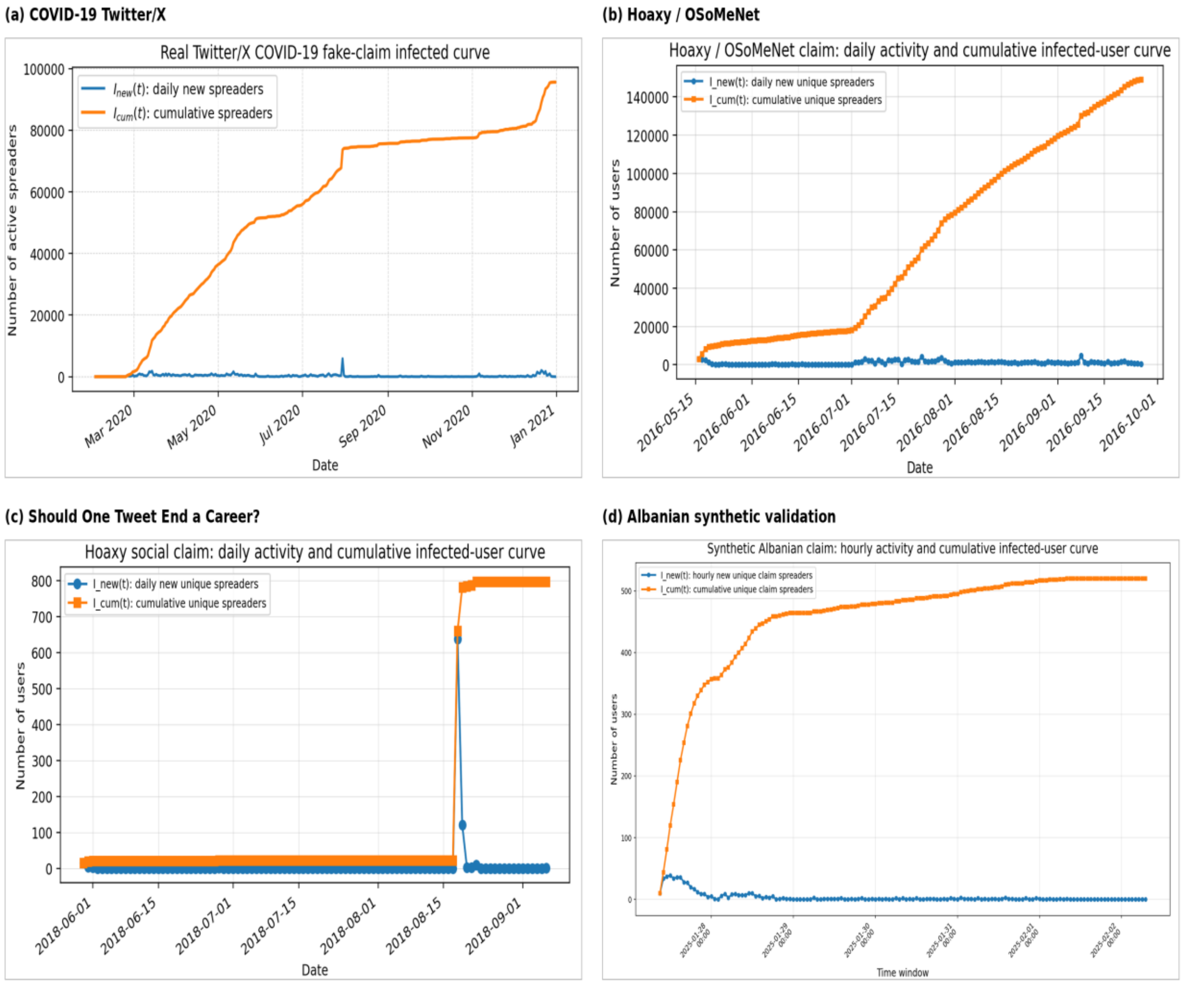

Figure 3.6. Observed infected-user trajectories for selected fake-news cascades: (a) COVID-19 Twitter/X, (b) Hoaxy / OSoMeNet, (c) “Should One Tweet End a Career?”, and (d) Albanian synthetic validation.

This combined presentation removes the repeated figure-by-figure interpretation that previously made the section appear methodological rather than empirical. The four panels show that fake-news diffusion cannot be evaluated from final size alone. Some cascades grow through sustained large-scale exposure, others through delayed reactivation, and the validation case shows only that the pipeline can generate a model-ready trajectory under a controlled schema.

For the transition to Section 3.5, the operational consequence is straightforward. The COVID-19 Twitter/X and Hoaxy / OSoMeNet curves become the observed infected trajectories for the SIR baseline and the SEIZ model. The auxiliary Hoaxy-imported and Albanian trajectories remain outside the primary SIR/SEIZ fitting set because their role is comparative or validation-oriented rather than posterior-inference evidence.

3.4.3 Representative temporal-network structures

Directed temporal networks were constructed from timestamped user interactions following the methodology defined in Chapter 2. The resulting structures preserve chronological interaction order and provide the infected-user trajectories used later for epidemic modelling. In this results section, the network discussion is kept deliberately compact: the figures verify the existence of directed diffusion structures, while the quantitative modelling uses the full time-ordered cascade series rather than the reduced visual cores.

Only the two principal empirical networks are retained in the main text. The network visualizations for the “Should One Tweet End a Career?” and Albanian synthetic cases are moved to Appendix D because they repeat the same visual logic but do not provide the main empirical basis for SIR, SEIZ, ABC or ABC-SMC inference. This truncation keeps the chapter focused on the two cascades that carry the strongest modelling evidence.

The retained visual evidence has a limited but necessary role. It shows that the observed curves are not isolated count sequences; they are generated by directed interaction structures in which source hubs and relay users can concentrate exposure. The figure is therefore used to motivate model transfer to SIR/SEIZ, not to introduce a second round of network-method explanation.

Figure 3.7 (a)(b) illustrates representative temporal-network structures extracted from the two principal empirical cascades. Both networks exhibit hub-centred diffusion patterns, with a small number (400 nodes) of dominant source nodes and a large number of downstream spreading users. The figure is included as a structural check that the observed $I(t)$ trajectories are generated by directed user-to-user diffusion rather than by unstructured aggregate counts.

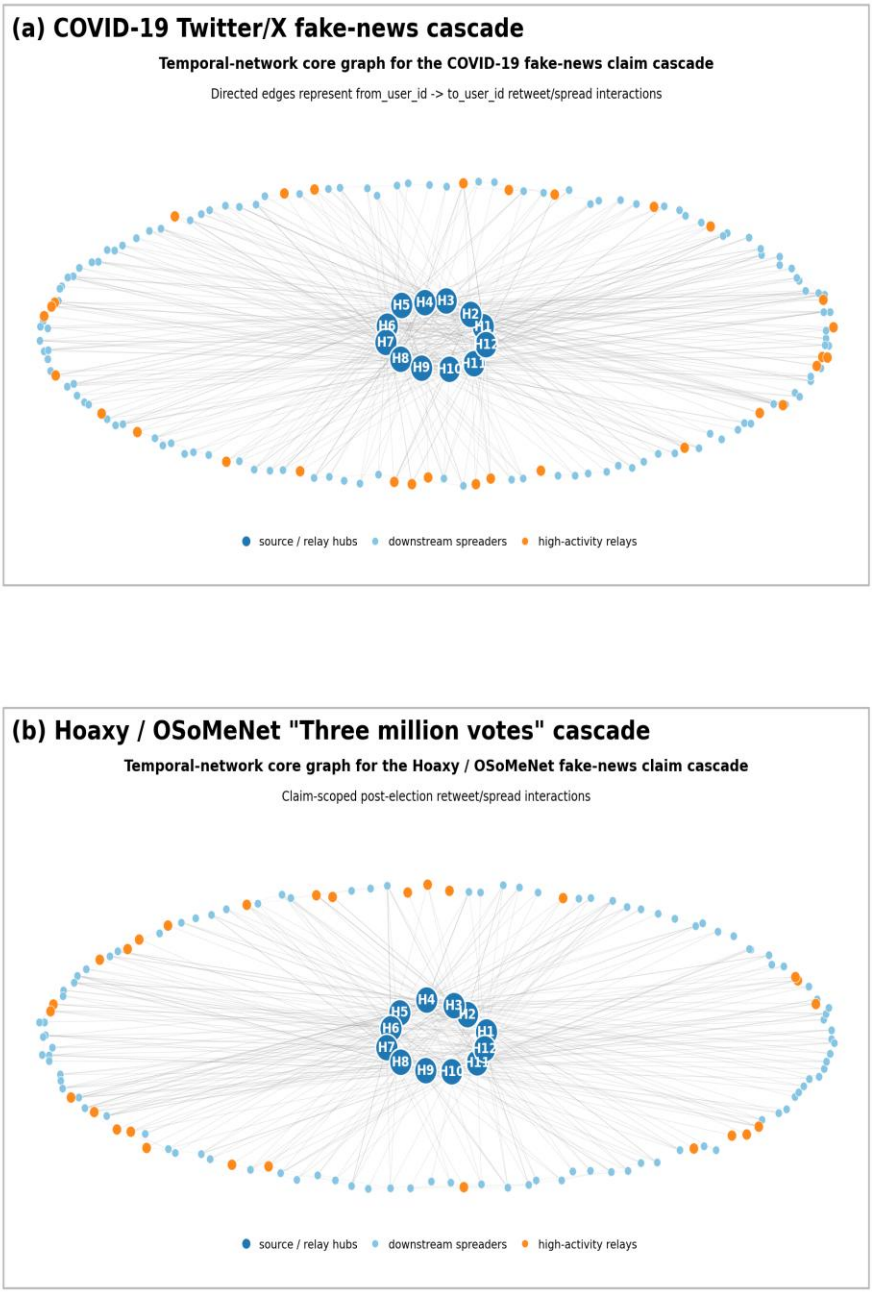

Figure 3.7. Core temporal-network examples for the two principal empirical cascades: (a) COVID-19 Twitter/X and (b) Hoaxy / OSoMeNet.

The visual cores are not used as substitutes for the full cascade data. They are reduced views designed for readability. The SIR and SEIZ fitting in the next section is based on the observed infected-user trajectories derived from the complete selected cascades, while the network visualizations help interpret why hub and relay activity can accelerate the observed curve before the epidemic model is fitted.

Section 3.4 therefore completes the bridge from detection to temporal modelling. Transformer probabilities establish which claims pass the fake-news screening gate; timestamped interactions convert those claims into cascades; $I_{\text{new}}(t)$ and $I_{\text{cum}}(t)$ provide the empirical spreading curves; and the representative networks confirm that the principal cases have directed diffusion structures suitable for SIR/SEIZ analysis. The next section uses the COVID-19 and Hoaxy / OSoMeNet trajectories as observed $I_{\text{obs}}(t)$ curves for epidemic-model comparison.

This reporting choice preserves the role of Section 3.4 as a transition section rather than a second methods chapter. Detailed cascade-summary tables, node/edge counts and auxiliary network visualizations should remain in Appendix D, while the main chapter should retain only the detection gate, cascade-selection table, consolidated trajectory summary and representative empirical network evidence needed for the SIR/SEIZ transition.

3.5 Epidemic Diffusion Modelling Results

Section 3.5 reports the epidemic-modelling results obtained after the temporal cascades in Section 3.4 were transformed into cumulative active-spreader trajectories. The purpose of this section is empirical: it reports how the SIR baseline and the SEIZ model behaved when they were fitted to the observed $I_{\text{cum}}(t)$ curves of the two principal fake-news cascades. The methodological definitions of the compartments, transition rules, equations, fitting distance and simulator assumptions are not repeated here because they were established in Chapter 2. The SIR-SEIZ comparison reported here extends the author's SoftCOM 2025 study on fake-news identification and epidemic-model comparison on Twitter temporal networks [88].

Both epidemic models were fitted to the cumulative infected-user trajectories extracted from the COVID-19 Twitter/X cascade and the Hoaxy / OSoMeNet “Three million votes in the Presidential Election were cast by illegal aliens” cascade. The two cases provide different diffusion conditions. The COVID-19 cascade is longer and more irregular, with repeated bursts and late reactivation. The Hoaxy cascade is larger and more politically concentrated, with a smoother cumulative growth pattern after the main activation phase. This contrast is useful because it tests whether the models behave consistently across different misinformation domains and temporal shapes.

The SIR model is retained as a transparent baseline because it provides a simple reference curve for direct spreading and subsequent inactivity. SEIZ is evaluated as the main fake-news diffusion model because its fitted trajectory can represent delayed activation and non-spreading behaviour through latent states. In Chapter 3 these states are interpreted only through model fit and trajectory shape. They are not treated as direct observations of private belief, because the empirical data contain visible retweets, timestamps and user-to-user interactions rather than complete exposure or belief histories.

The results are therefore reported through curve-level performance rather than through a new theoretical derivation. The key quantities are relative L2 fitting error, RMSE, MAE and the final fitted cumulative active-spreader count. Lower errors indicate that the model more closely reconstructs the observed $I_{\text{cum}}(t)$ trajectory. The central question is whether the additional SEIZ structure produces an empirical improvement large enough to justify its use in the subsequent ABC and ABC-SMC calibration stage.

3.5.1 Model Fitting Results

Table 3.10 summarizes the principal epidemic-model fitting results for SIR and SEIZ across the two empirical cascades. The table replaces the separate setup and result tables used in the earlier draft, because the experimental unit is the same in all cases: the observed daily cumulative trajectory of unique active spreaders. This compact form also keeps Chapter 3 aligned with its results-oriented role, while the modelling details remain in Chapter 2 and the supplementary materials.

Table 3.10. Summary of epidemic-model fitting results for the two empirical fake-news cascades.

Dataset / empirical cascade	Model	Relative L2	RMSE	MAE	Final observed / fitted I_{cum}
COVID-19 Twitter/X fake-news claim	SIR	0.1626	9,958.4	7,793.3	95,599 / 73,510
COVID-19 Twitter/X fake-news claim	SEIZ	0.0463	2,838.9	2,092.6	95,599 / 84,921
Hoaxy / OSoMeNet “Three million votes” fake-news claim	SIR	0.0511	4,150.2	3,560.2	149,079 / 143,396
Hoaxy / OSoMeNet “Three million votes” fake-news claim	SEIZ	0.0469	3,809.7	3,200.6	149,079 / 140,408

The COVID-19 cascade shows the clearest difference between the two models. The SIR baseline produces a relative L2 error of 0.1626 and predicts a final cumulative value of 73,510 active spreaders, compared with the observed 95,599. This underestimation indicates that the simple baseline captures the early acceleration only partially and reaches saturation too early.

The result is consistent with the visual structure of the observed curve in Section 3.4, where the cascade includes delayed bursts rather than one smooth wave.

SEIZ substantially improves the reconstruction of the same COVID-19 cascade. The relative L2 error decreases to 0.0463, while RMSE and MAE fall to 2,838.9 and 2,092.6 respectively. The fitted final cumulative value of 84,921 remains below the observed endpoint, but the trajectory follows the middle and late portions of the empirical curve much more closely than SIR. This improvement is important because the COVID-19 case is the more difficult of the two empirical cascades: it is longer, less regular and more affected by reactivation across the observation window.

The Hoaxy / OSoMeNet political fake-news cascade presents a different pattern. The SIR baseline already provides a relatively strong approximation, with a relative L2 error of 0.0511 and a fitted endpoint of 143,396 compared with the observed 149,079. This suggests that the cumulative growth of the Hoaxy cascade is closer to a single large-scale diffusion process than the COVID-19 cascade. Nevertheless, the SIR trajectory still smooths over the main acceleration phase and compresses the behavioural interpretation into a single active-spreading process.

The SEIZ fit for the Hoaxy cascade produces a modest but still meaningful improvement. The relative L2 error decreases to 0.0469, with lower RMSE and MAE than the SIR baseline. The endpoint is slightly less close than the SIR endpoint, but the overall curve distance is better, which is the main criterion used for model comparison in this section. This distinction matters because the modelling objective is not only to reproduce the final count, but also to match the shape of the observed cumulative trajectory across time.

Across both cascades, the table supports a consistent result: SEIZ improves the trajectory-level fit even when the magnitude of improvement differs by dataset. The improvement is large for the multi-burst COVID-19 cascade and smaller for the smoother Hoaxy cascade. This pattern indicates that delayed exposure and non-spreading behaviour are most important when the misinformation process is temporally irregular, but they remain useful even when the cumulative curve is more regular.

Figure 3.8 compares the observed cumulative spreading trajectories with the SIR and SEIZ predictions for both empirical cascades. In the COVID-19 case, the SIR curve reaches saturation too early, whereas the SEIZ curve follows the observed trajectory more closely across the main growth interval. In the Hoaxy case, both models are closer to the empirical curve, but the SEIZ trajectory still gives a slightly lower overall fitting error. The figure therefore visually confirms the numerical pattern reported in Table 3.10.

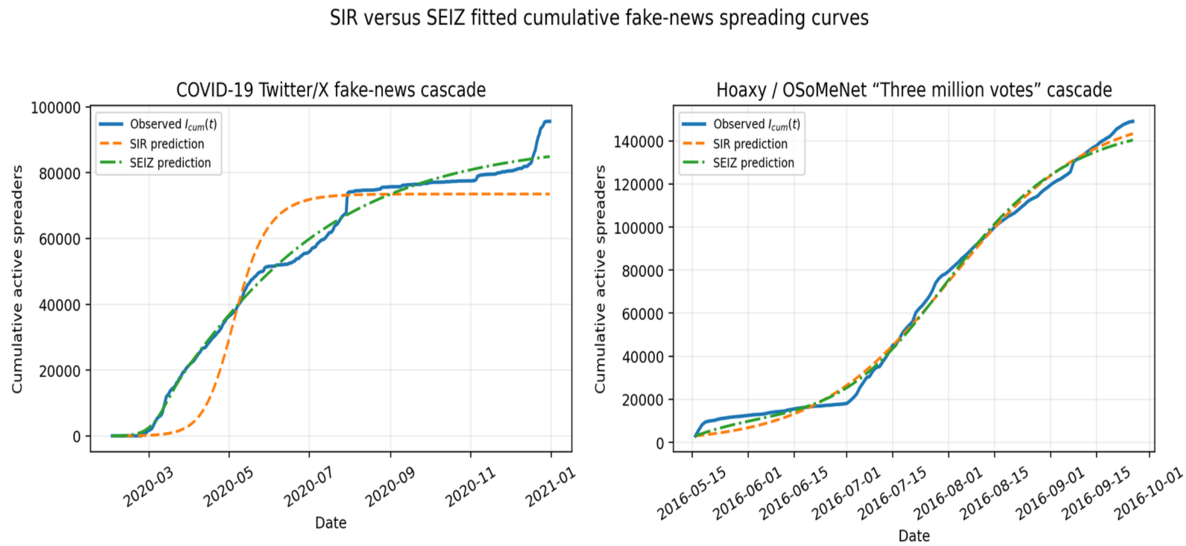

Figure 3.8. SIR versus SEIZ fitted cumulative fake-news spreading curves for the two empirical cascades.

The fitted curves also clarify why endpoint accuracy alone is not sufficient. A model may produce a final cumulative value close to the observed endpoint while still misrepresenting the timing and shape of growth. For fake-news diffusion, timing is central because early acceleration and late reactivation affect forecasting and intervention analysis. For this reason, the subsequent ABC and ABC-SMC stages use the curve-level fit and summary statistics rather than relying only on final cascade size.

3.5.2 SIR-SEIZ Comparative Analysis

SEIZ consistently outperformed the simpler SIR baseline in trajectory-level error. For the COVID-19 cascade, the relative L2 error decreased from 0.1626 to 0.0463, corresponding to an approximately 71.5% reduction. This is the strongest result of Section 3.5, because it shows that the richer diffusion structure is not only theoretically plausible but empirically necessary for this public-health misinformation cascade. The observed COVID-19 trajectory contains delayed growth and repeated activation patterns that cannot be represented adequately by a single direct spreading-and-removal baseline.

For the Hoaxy / OSoMeNet cascade, the improvement is smaller but still directionally consistent. The relative L2 error decreases from 0.0511 under SIR to 0.0469 under SEIZ, an approximately 8.2% reduction. This smaller difference suggests a more regular diffusion process in which the SIR baseline is already competitive. However, the SEIZ result remains preferable because it improves the full-trajectory distance and provides a more realistic representation of delayed activation and non-spreading behaviour, which are central to misinformation diffusion.

The different size of improvement across the two cascades is scientifically useful. It shows that the value of SEIZ is not uniform across all fake-news processes. When the cascade is irregular, delayed and multi-burst, the benefit is substantial. When the cascade is smoother and dominated by a large political activation process, the improvement is more moderate. This distinction strengthens rather than weakens the thesis argument, because it shows that the model comparison is sensitive to empirical cascade shape rather than imposed as a purely theoretical preference.

The SIR baseline remains important because it provides an interpretable lower-complexity reference. Its role in the thesis is not to be discarded, but to establish how much explanatory and predictive benefit is obtained by moving to SEIZ. In both empirical cases, SIR identifies the broad cumulative growth pattern, while SEIZ reduces the mismatch that arises when exposure, delayed adoption or non-spreading behaviour influence the observed curve. The comparison therefore supports a layered modelling strategy: use SIR as a baseline, but use SEIZ as the main model for posterior inference and intervention analysis.

The results also support the methodological boundary between Chapter 2 and Chapter 3. Chapter 2 defined the SIR and SEIZ mechanisms and explained why SEIZ is theoretically appropriate for fake-news diffusion. Chapter 3 now shows what happened when those models were fitted to empirical cascades. The observed improvement of SEIZ provides the result-level justification for using the SEIZ structure in the ABC and ABC-SMC calibration stage reported in Section 3.6.

From a practical perspective, the comparative result means that a fake-news cascade should not be interpreted only as immediate transmission from one active spreader to another. The fitted SEIZ behaviour suggests that a visible retweet curve may contain latent delay, hesitation, correction, non-sharing and reactivation effects. These effects are especially important for public-health misinformation, where users may encounter a claim multiple times, delay sharing, or react after later events renew attention to the topic.

The comparison therefore has a direct implication for downstream risk analysis. If SIR underestimates the late cumulative growth of a cascade, then intervention timing and forecasted risk may also be underestimated. SEIZ reduces this risk by allowing the fitted trajectory to remain responsive to delayed activation. This makes SEIZ the preferred epidemic model for the posterior predictive and counterfactual stages, while SIR remains a compact baseline for auditability and model comparison.

3.5.3 Compartment-Transition Interpretation

The fitted compartment trajectories provide an additional result-level interpretation of the SIR-SEIZ comparison. They do not introduce new methodology; they show how the fitted models allocate the observed diffusion process across internal states. The purpose is to make the difference between the baseline and the main model visible without repeating the equations or theoretical definitions already given in Chapter 2.

The most informative example is the COVID-19 cascade, because this is the case where the numerical improvement from SEIZ is strongest. The SIR trajectory compresses the cascade into a direct movement between active spreading and inactivity, which creates an early saturation pattern. The SEIZ trajectory distributes part of the process into exposed and skeptic or non-spreading states, allowing the model to represent delayed activation and incomplete adoption more explicitly. This allocation explains why SEIZ follows the empirical cumulative curve more closely.

Figure 3.9(a)(b) illustrates representative compartment dynamics for the COVID-19 fake-news cascade. Compared with SIR, SEIZ allocates users into exposed and skeptic/non-spreading states, allowing delayed activation and non-spreading behaviour to be represented explicitly. The figure is included as a compact representative example because the Hoaxy compartment plots produce the same interpretive conclusion with smaller marginal benefit; the detailed Hoaxy compartment trajectories can therefore remain in the supplementary material rather than consuming additional space in the main results chapter. Supplementary Hoaxy / OSoMeNet compartment-transition plots are provided in Figures H.8 and I.9 of Appendix H.

The figure should be interpreted as a model-based decomposition of the fitted cascade, not as direct evidence of individual belief states. The empirical data show visible spreading events and timestamps. They do not show all impressions, private exposures, silent rejections or individual psychological decisions. Therefore, the exposed and skeptic components are useful latent modelling categories, but they should not be overinterpreted as directly observed user attitudes.

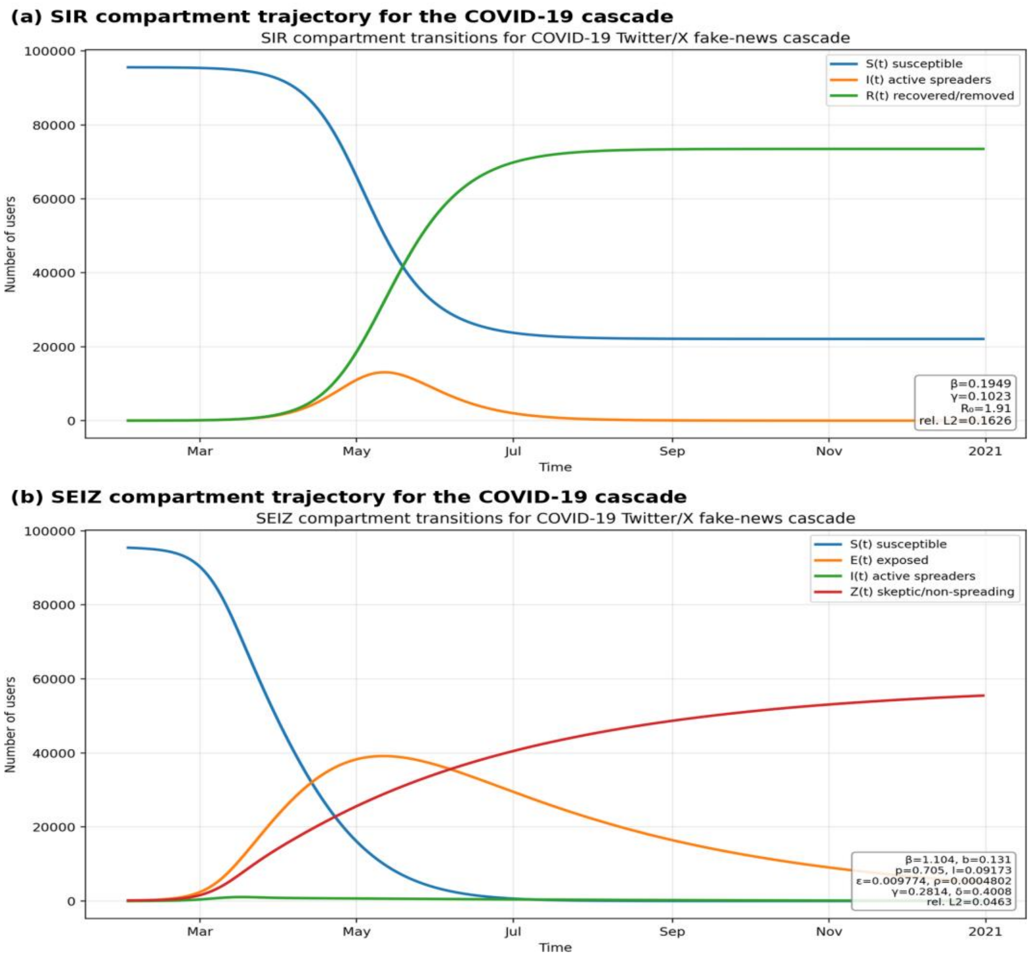

Figure 3.9. Representative compartment-transition interpretation for the COVID-19 cascade: SIR and SEIZ fitted trajectories.

The compartment interpretation reinforces the error results in Table 3.10. The COVID-19 cascade is not well represented by a simple direct-spreading curve because part of its growth occurs after delays and secondary activation periods. SEIZ can reproduce this behaviour more effectively because it does not require every interaction to become immediate active spreading. In this sense, the compartment curves provide an explanatory bridge between the observed temporal cascade in Section 3.4 and the posterior inference in Section 3.6.

For the Hoaxy / OSoMeNet cascade, the fitted model comparison is more balanced. SIR already reconstructs much of the cumulative trajectory, which is consistent with the smoother large-scale growth pattern. However, the SEIZ fit remains preferable because it provides lower curve-level error and a more detailed interpretation of delayed exposure before active spreading. The smaller improvement for Hoaxy therefore does not contradict the COVID-19 result; rather, it indicates that the empirical value of the additional SEIZ states depends on the cascade shape.

Taken together, the fitting results and compartment interpretation support the use of SEIZ as the primary epidemic model in the remainder of Chapter 3. SIR remains valuable for transparency, but SEIZ provides the stronger empirical basis for ABC and ABC-SMC calibration. The next section therefore estimates uncertainty around the epidemic parameters using likelihood-free posterior inference, rather than relying only on deterministic point fits.

This conclusion is deliberately limited to the two principal empirical cascades used in the thesis. It does not claim that SEIZ is universally superior for every misinformation dataset or platform. It shows that, under the retained COVID-19 and Hoaxy / OSoMeNet cascades, the SEIZ structure gives better or more informative reconstruction of fake-news diffusion than the SIR baseline. That result is sufficient for the thesis pipeline because the ABC and intervention analyses require a calibrated diffusion model that can represent exposure delay and non-spreading behaviour under uncertainty.

The model-comparison result also clarifies how Section 3.5 supports the wider dissertation argument. Transformer detection identifies candidate fake-news content, while the temporal-network layer shows whether that content develops into an observable cascade. The epidemic layer then tests whether the resulting $I_{\text{cum}}(t)$ curve can be reconstructed by a simple baseline or whether additional latent diffusion structure is required. In both empirical cases, SEIZ provides the stronger basis for this bridge because it preserves the observed cumulative trajectory more effectively than SIR and remains interpretable for later posterior analysis.

Accordingly, Section 3.5 establishes the empirical transition to Section 3.6. The next stage does not re-fit the models only as deterministic curves; it estimates uncertainty around the diffusion parameters and evaluates posterior predictive behaviour. This is necessary because the observed cascades are incomplete traces of platform activity. The deterministic fits reported here show which epidemic structure should be carried forward, while ABC and ABC-SMC test how stable that structure remains when parameter uncertainty and hidden exposure are considered.

3.6 ABC and ABC-SMC Inference Results

The deterministic SIR and SEIZ results in Section 3.5 provide useful point estimates, but they do not report uncertainty around the diffusion parameters. Section 3.6 therefore applies Approximate Bayesian Computation (ABC) and sequential ABC-SMC to the two principal empirical cascades: the COVID-19 Twitter/X fake-news claim and the Hoaxy / OSoMeNet “Three million votes” claim. The section is intentionally compact and results-oriented: it reports the inference setup, the posterior summaries, the posterior predictive validation, and the convergence evidence needed before the intervention scenarios in Section 3.7. The ABC-based uncertainty layer is directly connected to the author's TechDefense 2025 work on BigBird-assisted SEIZ modelling with Approximate Bayesian Computation [89].

3.6.1 Inference setup and posterior estimation

ABC was applied to account for partially observed diffusion processes. The observed object is the infected-user trajectory extracted from the temporal-network layer, while the simulator generates SIR or SEIZ trajectories under candidate parameter vectors. Candidate particles are accepted when the simulated curve and the observed curve are sufficiently close under a combined distance criterion.

The inference target remains the SEIZ model because Section 3.5 showed that SEIZ better represents delayed exposure and non-spreading behaviour in fake-news cascades. SIR is retained as a transparent baseline for posterior comparison. Table 3.11 reports only the observed cascade statistics needed to define the ABC target; longer summary vectors and run logs are reserved for the appendices.

Table 3.11 keeps the ABC target evidence at cascade level. The COVID-19 case has a smaller final reach but a longer and more irregular activity profile, while the Hoaxy case reaches a larger final size over a shorter window. This difference motivates dataset-specific posterior inference rather than transferring one parameter set across both cascades.

Table 3.11. Observed cascade summary statistics used for ABC inference.

Dataset	Final size	Peak day	Peak I_{new}	Time-to-peak
COVID-19 Twitter/X fake-news claim	95,599	179	5,892	179
Hoaxy / OSoMeNet “Three million votes” claim	149,079	115	4,660	115

The ABC distance combines trajectory agreement with standardized summary statistics. The curve component received greater weight because the main objective is to reproduce the full infected-user trajectory, not only the endpoint or peak.

$$d_{\text{curve}}(\theta) = \frac{\lVert I_{\text{sim}}(t|\theta) - I_{\text{obs}}(t) \rVert_2}{\lVert I_{\text{obs}}(t) \rVert_2} \tag{3.1}$$

$$d_{\text{ABC}}(\theta) = w_{\text{curve}} d_{\text{curve}}(\theta) + w_{\text{sum}} d_{\text{summary}}(\theta), \quad w_{\text{curve}} > w_{\text{sum}} \tag{3.2}$$

The posterior summaries in the main chapter are restricted to the quantities required for interpretation: accepted particles, mean accepted distance, final tolerance, and key parameters. Detailed posterior intervals are provided in Appendix G; the main chapter reports only the principal parameter summaries.

Table 3.12 shows the same pattern observed in the deterministic modelling stage. ABC-SIR is adequate as a low-dimensional baseline, especially for the smoother Hoaxy cascade, but its accepted distance is higher for the COVID-19 cascade. ABC-SEIZ produces lower accepted distances and retains a parameter structure that can represent delayed adoption, partial rejection and stopping from active spreading.

Table 3.12. Posterior estimate summary for ABC-SIR and ABC-SEIZ.

Dataset	Model	Particles	Mean distance	Tolerance	Key parameters
COVID	ABC-SIR	300	0.175	0.186	β=.195; γ=.102; R_0=1.92
COVID	ABC-SEIZ	400	0.071	0.092	β=1.076; b=.129; p=.698; l=.0915; γ=.271; δ=.399
Hoaxy	ABC-SIR	300	0.060	0.075	β=.055; γ=.005; R_0=10.6
Hoaxy	ABC-SEIZ	400	0.055	0.064	$\beta \approx 0$; b=2.000; p=.0118; l=.0359; ε=.0428; γ=.0086

The posterior values are not interpreted as direct psychological measurements of individual users. They are uncertainty-aware diffusion parameters that reproduce the observed cascade under the SEIZ state assumptions and are therefore suitable for posterior predictive checking.

3.6.2 Posterior predictive validation

Posterior predictive validation checks whether the accepted particles reproduce the empirical infected-user curves. The comparison is made against the same observed cumulative trajectories used in Section 3.5, but the presentation is compressed into one combined figure instead of repeating separate COVID and Hoaxy analyses.

Figure 3.10(a)(b) compares deterministic and uncertainty-aware epidemic trajectories. ABC-SEIZ preserves the SEIZ fit while adding posterior uncertainty intervals.

The combined figure summarizes the posterior predictive validation without repeating separate trajectory discussions for each cascade. In the COVID-19 cascade, SIR saturates too early and underestimates late cumulative growth, while SEIZ and ABC-SEIZ remain closer to the observed curve. In the Hoaxy cascade, SIR is already competitive because the growth pattern is smoother, but ABC-SEIZ still provides the preferred uncertainty-aware representation. The result-bearing point is therefore the same for both datasets: posterior uncertainty strengthens the SEIZ interpretation without reopening the full methodological explanation of ABC.

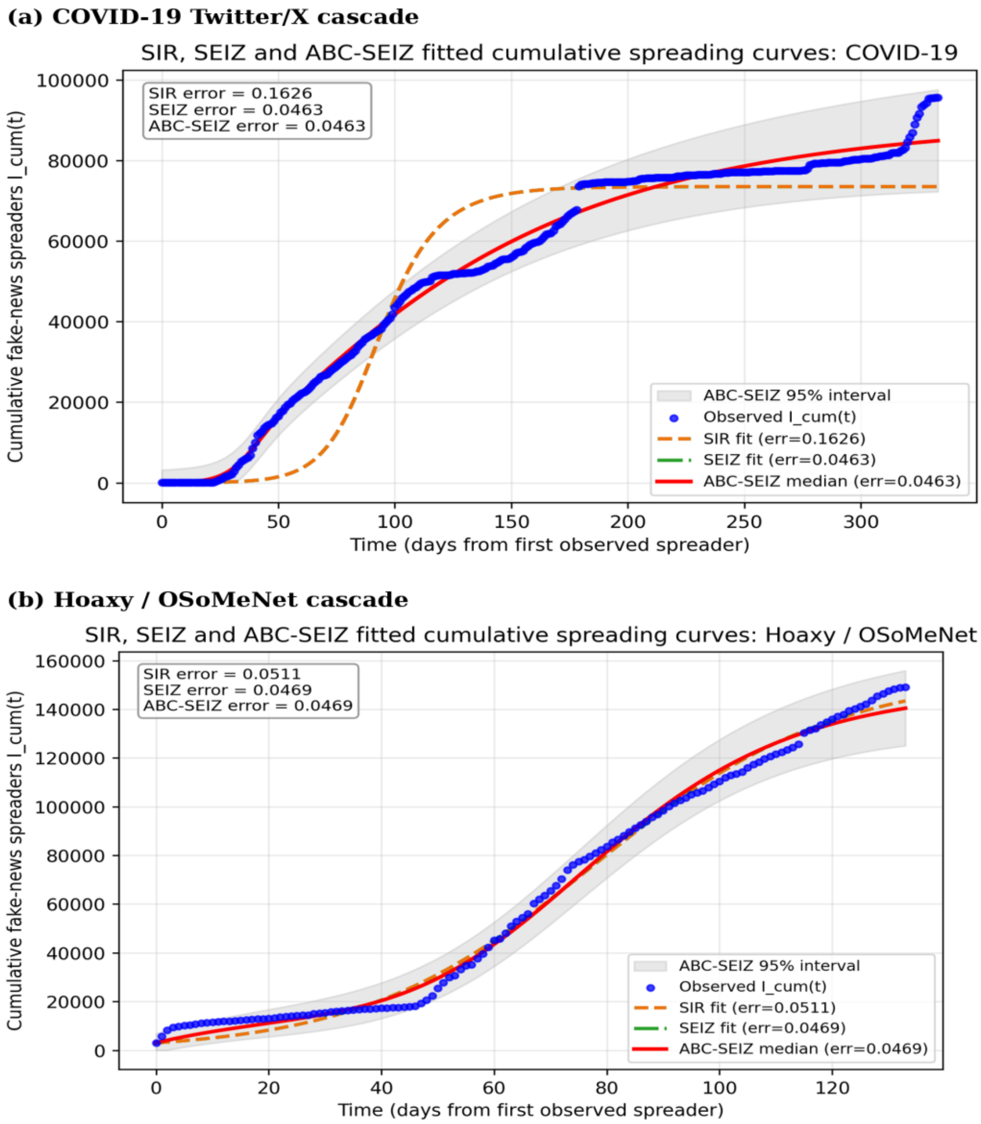

Figure 3.10. SIR, SEIZ and ABC-SEIZ fitted cumulative fake-news spreading curves for the two empirical cascades.

Figure 3.11 compares fitting-error distributions across models. ABC-SEIZ yields the narrowest distribution and the lowest overall error, while SIR is shifted toward larger reconstruction errors. The figure therefore adds new evidence beyond a single trajectory plot because it summarizes the stability of the error reduction across posterior predictive simulations.

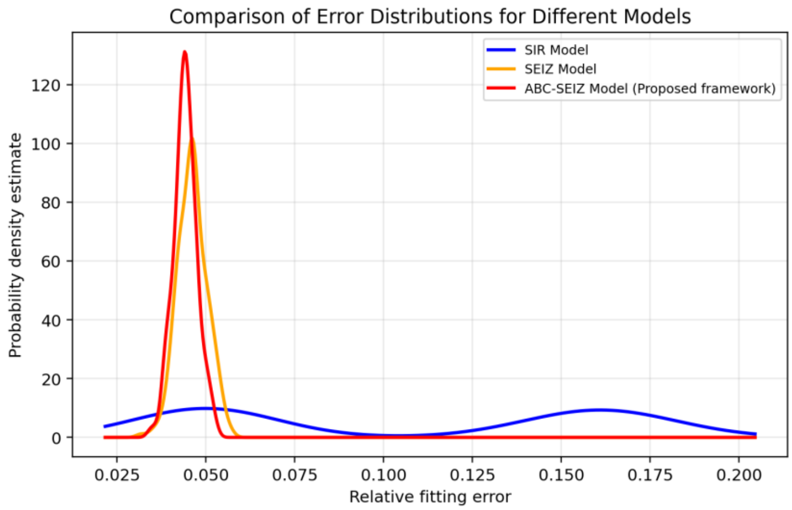

Figure 3.11. Fitting-error distributions for SIR, SEIZ and ABC-SEIZ models.

The posterior predictive evidence supports a concise conclusion. SIR remains useful as a reference model, but the ABC-SEIZ layer is the main inferential output of Section 3.6 because it combines the stronger SEIZ trajectory fit with uncertainty-aware calibration. Detailed posterior predictive error values are provided in Appendix D rather than repeated as another main-chapter table.

3.6.3 ABC-SMC convergence and forecast calibration

ABC-SMC refines the ABC posterior population by reducing the tolerance across generations. The purpose of the convergence check is not to introduce another epidemic model, because the state structure remains SEIZ; it is to verify that sequential calibration moves the accepted particles toward closer posterior predictive reconstruction.

Table 3.13 reports the convergence quantities needed in the main chapter. Both datasets use four generations, and both end with final accepted distances below the final tolerance. The COVID-19 cascade retains a higher final distance because its multi-burst trajectory is harder to reproduce than the smoother Hoaxy cascade.

Table 3.13. ABC-SMC convergence summary for the SEIZ fake-news diffusion model.

Dataset	Generations	Initial ε	Final ε	Final distance
COVID	4	0.30	0.092	0.070
Hoaxy	4	0.18	0.063	0.055

The convergence evidence is intentionally limited to accepted distance and tolerance reduction. No additional network construction details are repeated here, because the temporal reconstruction has already been reported in Section 3.4 and the deterministic SIR/SEIZ comparison in Section 3.5.

Substantively, the convergence pattern shows that posterior uncertainty is narrowed by the cascade evidence rather than left at the prior level. The COVID-19 schedule still closes at a wider distance, which is consistent with repeated bursts, while the Hoaxy schedule closes more tightly, which is consistent with smoother saturation.

These convergence results are used only as a calibration check. They do not duplicate the SIR/SEIZ model comparison and do not introduce new cascade statistics. The following figures therefore move directly to aggregate predictive error and held-out forecast performance.

Figure 3.12.(a)(b) shows that ABC-SMC progressively reduced accepted distances across generations, indicating stable posterior convergence. The Hoaxy profile is tighter by the final generation, while the COVID-19 profile remains wider because repeated activity bursts make the trajectory less regular. This supports the use of ABC-SMC as a calibration refinement before the forecast and intervention analyses.

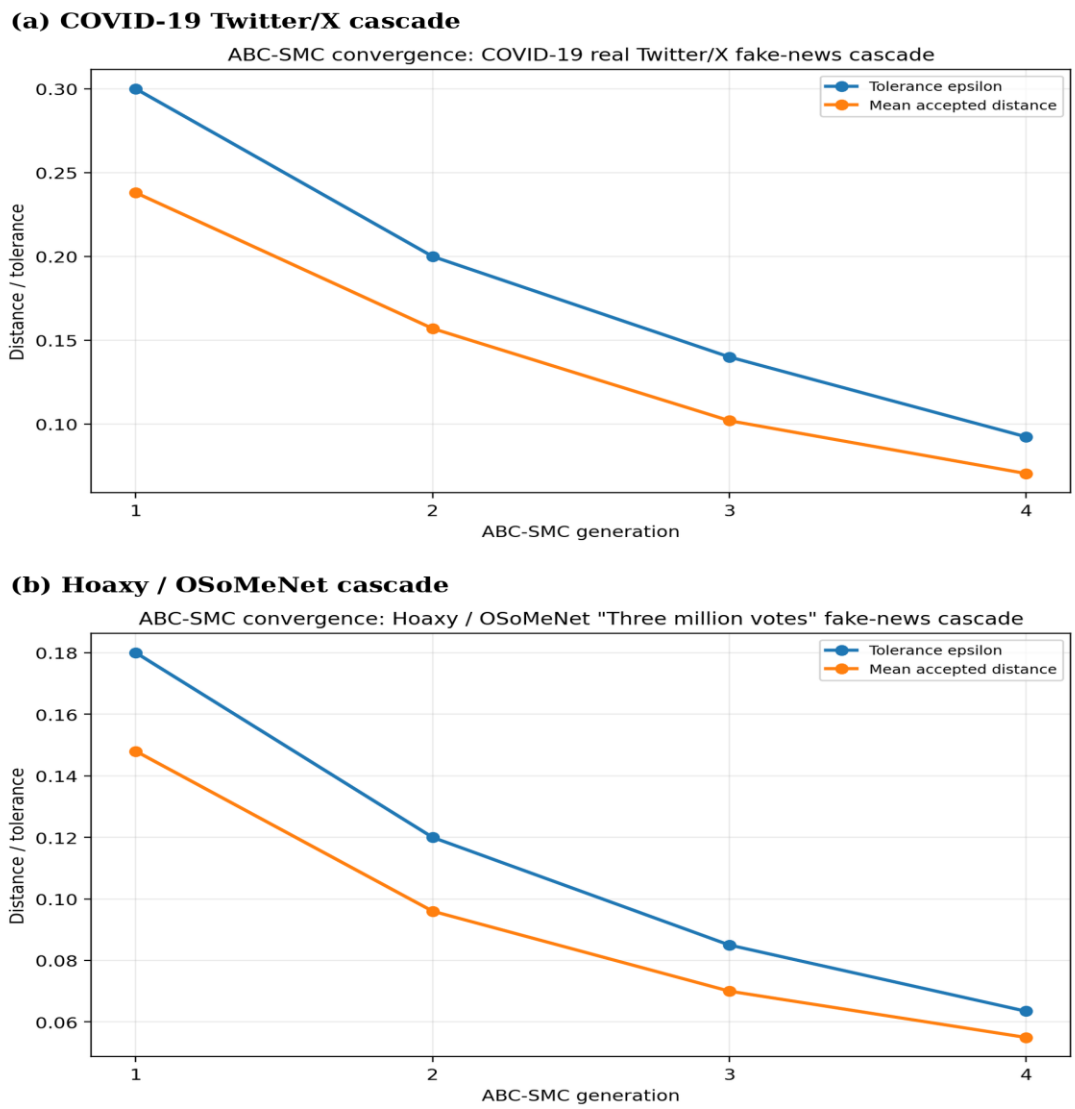

Figure 3.12. ABC-SMC convergence profiles for the COVID-19 and Hoaxy / OSoMeNet cascades.

The convergence result should be read as evidence of sequential calibration, not as a claim that ABC-SMC changes the underlying diffusion mechanism. The mechanism remains SEIZ; the improvement comes from a more selective posterior population under a decreasing tolerance schedule.

The final calibration comparison is most clearly reported through the aggregated posterior predictive error figure. The two separate ABC-SEIZ versus ABC-SMC-SEIZ trajectory figures are omitted because they repeat information already visible in the main fit plots. The aggregated comparison in Figure 3.13 summarizes the calibration advantage more concisely.

Figure 3.13 shows that sequential calibration reduces the normalized predictive errors for both cascades. For the COVID-19 case, ABC-SMC-SEIZ reduces the relative L2 error by about 16.2% compared with ABC-SEIZ. For the Hoaxy case, the corresponding reduction is about 16.4%. Detailed RMSE and MAE values are provided in Appendix D, while the main text retains only the result-bearing comparison.

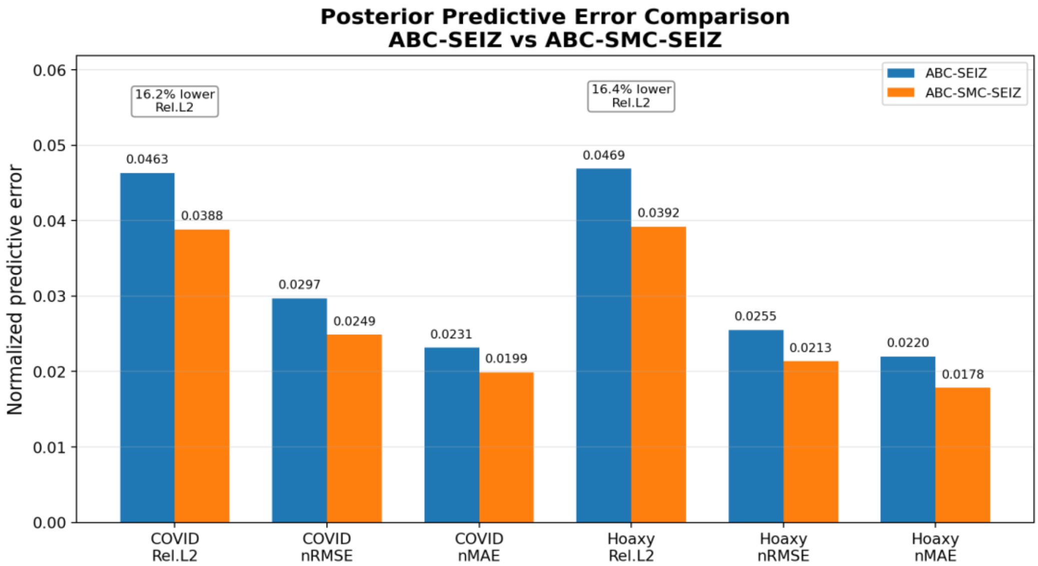

Figure 3.13. Comparative posterior predictive error of ABC-SEIZ and ABC-SMC-SEIZ across the two principal empirical fake-news cascades.

The last validation step evaluates whether the calibrated ABC-SMC-SEIZ layer supports short-horizon forecasting. The model is calibrated on an earlier segment of each cascade and then used to forecast the held-out cumulative trajectory. This is a limited near-term forecast under uncertainty, not an unconditional prediction of future platform behaviour.

Figure 3.14 evaluates out-of-sample short-horizon forecasting. The calibrated ABC-SMC-SEIZ model follows the observed saturation trend, although the COVID-19 forecast underestimates late growth. The result demonstrates that the calibrated framework supports limited near-term forecasting under uncertainty.

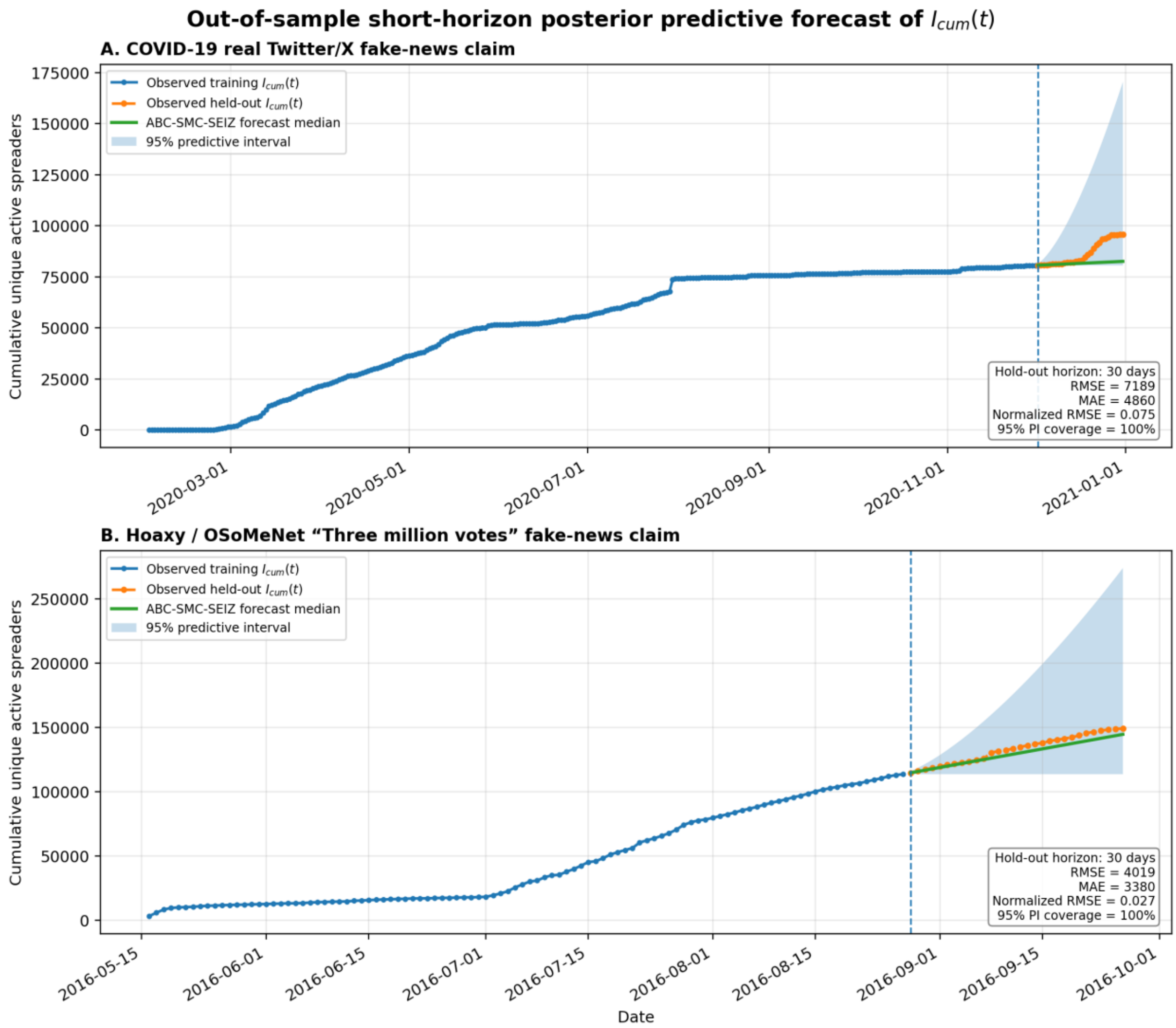

Figure 3.14. Out-of-sample short-horizon ABC-SMC-SEIZ posterior predictive forecast for the COVID-19 and Hoaxy / OSoMeNet cascades.

Overall, Section 3.6 demonstrates a compact result chain: ABC defines the likelihood-free posterior setup, Table 3.12 reports the main posterior estimates, Figure 3.10 and Figure 3.11 validate the predictive fit, and Table 3.13 with Figure 3.12 verifies ABC-SMC convergence. The aggregated error and forecast figures then justify using the calibrated SEIZ posterior as the basis for the counterfactual scenarios in Section 3.7. Supplementary posterior-predictive and ABC-SMC calibration plots are provided in Figures H.10-I.13 of Appendix H.

3.7 Intervention Scenario Analysis

The intervention component is retained as a concise counterfactual scenario analysis rather than as a separate causal policy experiment. The purpose is to test how the same calibrated fake-news claim cascades would change if selected SEIZ parameters were perturbed after an intervention trigger time. The results should therefore be read as controlled model-based scenario rankings, not as estimates of real platform policy effects.

The numerical values in this section remain relative to the ABC-SEIZ posterior-median baseline because the intervention simulations were generated from that baseline in the Chapter 3 run. The ABC-SMC-SEIZ calibration and the out-of-sample short-horizon forecast in Section 3.6.4 provide robustness evidence that the calibrated SEIZ structure can reproduce and forecast cumulative spreading behaviour, but they do not convert the intervention scenarios into direct ABC-SMC policy estimates. A future rerun with raw ABC-SMC posterior samples could update the same scenario table with posterior intervention intervals.

The analysis is restricted to the two principal empirical cascades used for SIR, SEIZ and ABC/ABC-SMC inference: the COVID-19 real Twitter/X fake-news claim and the Hoaxy / OSoMeNet "Three million votes" fake-news claim. Corrective debunking changes adoption and correction-related parameters, visibility reduction changes direct spreading and exposure pressure, and early combined mitigation applies both mechanisms before the main cascade acceleration. Table 3.14 condenses the intervention evidence into one main-chapter summary table, while the full parameter perturbation tables are retained in Appendix D.

Table 3.14 keeps the intervention evidence in the main chapter while avoiding an overly long policy-simulation section. The strongest result is the early combined mitigation scenario, which reduces final cumulative active spreaders by 34.2% in the COVID-19 cascade and by 16.6% in the Hoaxy / OSoMeNet cascade. The delayed scenarios show much smaller reductions and leave the observed peak unchanged, confirming that the timing of intervention is more important than the intervention label alone.

Table 3.14. Concise counterfactual intervention scenario summary under the ABC-SEIZ posterior-median baseline.

Dataset	**Strongest debunking scenario**	**Strongest visibility / mitigation scenario**	**Additional gain**	**Main interpretation**
COVID-19 Twitter/X fake-news claim	Early debunking - strong: final $I_{\text{cum}}(t)$ = 64,695; peak $I_{\text{new}}(t)$ = 343; 23.9% final-size reduction.	Early combined mitigation: final $I_{\text{cum}}(t)$ = 55,862; peak $I_{\text{new}}(t)$ = 327; 34.2% final-size reduction.	8,833 fewer final active spreaders than the strongest debunking scenario; +10.3 percentage points in final-size reduction.	The multi-burst COVID-19 cascade benefits most when corrective effects are combined with early visibility reduction.

Hoaxy / OSoMeNet "Three million votes" fake-news claim	Early debunking - strong: final $I_{\text{cum}}(t)$ = 135,071; peak $I_{\text{new}}(t)$ = 1,926; 6.1% final-size reduction.	Early combined mitigation: final $I_{\text{cum}}(t)$ = 119,984; peak $I_{\text{new}}(t)$ = 1,584; 16.6% final-size reduction.	15,087 fewer final active spreaders than the strongest debunking scenario; +10.5 percentage points in final-size reduction.	The political fake-news cascade is less responsive to correction-only changes and more responsive to early control of visibility and relay exposure.

This concise intervention analysis completes the empirical pipeline without overclaiming causal platform effects. Transformer detection supplies the warning signal, temporal-network monitoring identifies early acceleration, SEIZ translates the cascade into diffusion mechanisms, ABC/ABC-SMC quantifies uncertainty and short-horizon predictability, and the counterfactual layer tests which parameter changes would most strongly reduce the same fake-news claim cascade.

The comparison between the two cascades is also important for interpretation. The COVID-19 Twitter/X cascade remains sensitive to early combined action because its multi-burst pattern leaves several windows in which visibility reduction and corrective pressure can still reduce final reach. The Hoaxy / OSoMeNet political cascade is more resistant to debunking alone, indicating that highly amplified political claims may require early reduction of relay exposure in addition to corrective messaging. These results support the thesis argument that intervention reasoning must be linked to observed temporal dynamics rather than applied as a generic after-the-fact policy label.

The scenario ranking also shows that timing and mechanism interact. A correction-oriented intervention can reduce the final cumulative curve only if enough users have not yet entered the active-spreading process. Once the cascade has already reached its peak or passed through its main relay users, correction alone has limited effect on the observed spreading trajectory. Early combined mitigation performs better because it acts on both interpretation and exposure: it reduces the probability of continued adoption while also limiting the visibility conditions under which additional users become exposed to the claim.

For thesis interpretation, the intervention analysis should therefore be treated as the final result-bearing extension of the posterior diffusion framework. It does not claim to evaluate an actual moderation policy, and it does not infer individual user belief. It shows what happens inside the calibrated SEIZ mechanism when intervention-relevant parameters are changed in a controlled way. This is sufficient for Chapter 3 because the aim is to demonstrate how detection, temporal monitoring, epidemic modelling and Bayesian calibration can support cautious, reproducible and auditable scenario reasoning.

3.8 Integrated Discussion

The results indicate that fake-news risk cannot be explained through content detection alone. High classification performance provides an initial warning signal, but social risk emerges when rapid activation, network amplification and sustained diffusion occur simultaneously.
SEIZ consistently outperformed SIR by incorporating exposed and skeptic states, while ABC-SMC improved calibration and predictive reliability. Together, the framework integrates content detection, diffusion analysis and uncertainty-aware inference.
The first empirical lesson of Chapter 3 is that fake-news detection and fake-news diffusion measure different aspects of the same phenomenon. The Transformer experiments show whether a textual item, article or claim can be classified as fake or real with strong accuracy, F1-score and ROC-AUC. These metrics are necessary because an unreliable content detector would weaken every later stage of the framework. However, they do not show whether a false claim will remain local, reach a small audience, or become a persistent cascade. A high fake-class probability therefore becomes operationally meaningful only when it is connected to temporal monitoring.
The temporal-network results provide this second layer of evidence. By converting selected fake-news claims into $I_{\text{new}}(t)$, $I_{\text{cum}}(t)$, directed user-to-user edges, hubs, relay users and temporal reachability patterns, the chapter shows how diffusion risk becomes observable. The COVID-19 real Twitter/X fake-news claim and the Hoaxy / OSoMeNet "Three million votes" fake-news claim are the strongest empirical cases because they provide sufficient temporal structure for epidemic modelling and posterior inference. The additional Hoaxy-imported and synthetic Albanian cases remain useful for comparative and validation-oriented interpretation, but they are not treated as equivalent inferential evidence for ABC/ABC-SMC intervention conclusions.
The third empirical lesson is that the epidemic analogy is useful only when it is interpreted as a modelling approximation. SIR provides a simple baseline for cumulative spreading, but it assumes a direct susceptible-infected-recovered structure that is too restrictive for fake-news diffusion. Fake-news cascades contain delayed exposure, hesitation, non-sharing after exposure, correction, skepticism and repeated activation. SEIZ is better aligned with these mechanisms because it allows exposed users and skeptic or non-spreading users to be

represented explicitly. This explains why the SEIZ results are more suitable for the selected empirical cascades than the simpler SIR baseline.

ABC and ABC-SMC strengthen the analysis by moving the thesis away from single deterministic parameter estimates. The posterior results show that fake-news diffusion parameters should be interpreted with uncertainty, especially when cascades are noisy, incomplete or multi-burst. ABC provides a likelihood-free route to plausible parameter regions, while ABC-SMC refines this process through sequential tolerance reduction and improves posterior predictive calibration. The short-horizon $I_{\mathrm{cum}}(t)$ forecast further shows that the calibrated SEIZ structure has predictive value beyond retrospective curve fitting, although the forecast remains horizon-specific and should not be interpreted as unconditional platform prediction.

This figure is retained instead of the former duplicated pipeline table because the visual structure is sufficient to explain the methodological flow. It shows that the chapter is not a sequence of unrelated experiments: the Transformer layer supplies the content-risk signal, the temporal-network layer converts selected claims into observable cascades, SIR and SEIZ translate the spreading curves into diffusion mechanisms, and ABC/ABC-SMC estimates uncertainty and posterior predictive behaviour. The pipeline therefore gives the examiner a clear path from the classification results to the diffusion and intervention results.

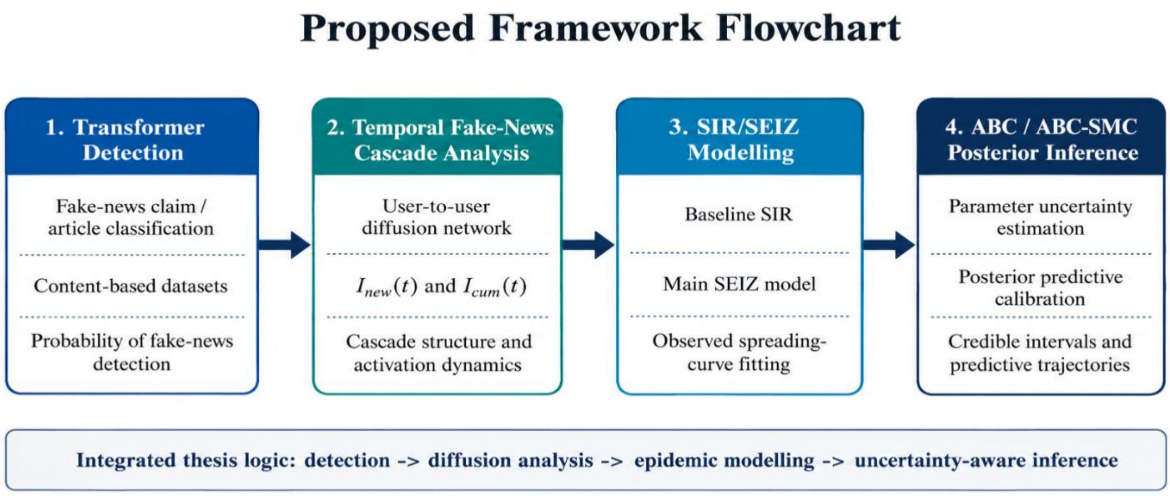

Figure 3.15. Hybrid framework flowchart of the Chapter 3 pipeline from Transformer-based fake-news detection to temporal cascade analysis, SIR/SEIZ diffusion modelling, and ABC/ABC-SMC posterior inference. The integrated framework links detection, temporal analysis, epidemic modelling and posterior inference into a unified pipeline.

The intervention findings must be read within this integrated logic. The counterfactual scenarios do not prove that a real social-media platform would produce the same effects after a moderation action. Instead, they rank how the calibrated cascades respond when key diffusion parameters are changed under controlled assumptions. The strongest reductions occur when visibility reduction and corrective pressure are applied early, before the cascade has reached its main acceleration phase. This supports the argument that fake-news mitigation should be evaluated in relation to timing, diffusion state and network amplification rather than through content classification alone.

Baseline Posterior Fake-News Diffusion-Risk Functional. The functional is retained only as a compact synthesis of Chapter 3 outputs, because its methodological definition has already been introduced in Chapter 2. Its role here is not to introduce a new epidemic model or to replace SEIZ. Instead, it connects three already reported quantities: the Transformer-estimated fake-news probability, the temporal-network amplification observed in the cascade, and the ABC-SMC-SEIZ posterior predictive growth over the selected forecast horizon. The value of the functional is therefore interpretive and traceable: each component can be inspected in the preceding results rather than treated as an unsupported composite score.

$$\mathcal{R}_{\text{base}}(c,h)=\hat{p}_f(x_c)[1+\lambda A_c(t)]\mathbb{E}_{\theta\sim\pi_\varepsilon^{\text{SMC}}}\left[\frac{\Delta I_{\text{cum},c}(t,h;\theta)}{N_c}\right] \tag{3.3}$$

Figure 3.16 shows both the conceptual construction of the score and the two empirical examples. The COVID-19 and Hoaxy / OSoMeNet cases should not be compared as universal platform-risk rankings. They are horizon-specific demonstrations showing how the thesis framework can convert detection, cascade amplification and posterior growth into a normalized baseline risk quantity for a selected claim and time window.

The Transformer block supplies the fake-news probability, the temporal-network block supplies amplification, and the ABC-SMC-SEIZ block supplies expected forecast growth. Multiplying these components gives a baseline posterior diffusion-risk score for a selected forecast horizon. The figure clarifies that the score is not an additional unvalidated model; it is a compact bridge between the empirical outputs already generated by the chapter.

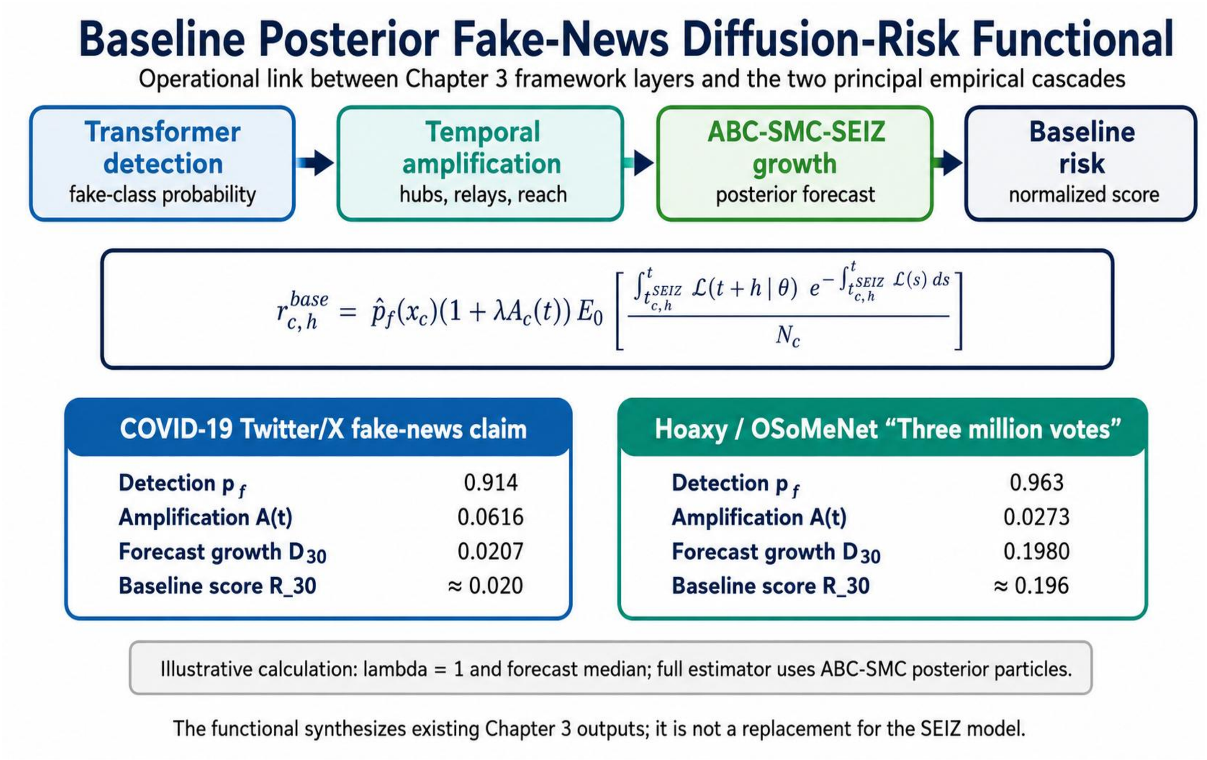

Figure 3.16. Operational demonstration of the baseline posterior fake-news diffusion-risk functional linking Transformer detection, temporal-network amplification and ABC-SMC-SEIZ forecast growth for the two principal empirical cascades.

The integrated discussion also clarifies the main limitation of the framework. The results support a content-to-cascade interpretation of fake-news risk, but the framework does not infer individual belief, intention or persuasion. A retweet, reply or repost is treated as an observable diffusion event, not as direct proof that a user accepted the claim as true. Similarly, the skeptic state in SEIZ is a modelling proxy for non-spreading or resistant behaviour, not a measured psychological state. These boundaries make the conclusions scientifically cautious while preserving the value of the temporal and Bayesian analysis.

Overall, the empirical evidence supports the central thesis claim: fake-news analysis becomes stronger when detection, temporal cascades, epidemic modelling and posterior inference are combined. The Transformer layer provides reliable content warning, the temporal-network layer identifies whether the warning becomes socially active, SEIZ improves the mechanistic representation of fake-news diffusion, ABC/ABC-SMC provides uncertainty-aware calibration, and the intervention analysis translates the calibrated model into cautious scenario reasoning. This integration is the main scientific value of Chapter 3.

For examination purposes, the most important conclusion is that each empirical layer answers a different question. The detection layer asks whether the content is likely to be fake. The temporal-network layer asks whether the claim is spreading and through which activation pattern. The SEIZ layer asks which diffusion mechanism best approximates the observed trajectory. The ABC/ABC-SMC layer asks how uncertain the parameters and predictive trajectories are. The intervention layer then asks which controlled parameter changes would reduce the same calibrated cascade. This sequence is what distinguishes the proposed framework from a stand-alone classifier or a purely descriptive network analysis.

The integrated framework also preserves methodological caution. It does not claim that every highly classified fake-news item will become a large cascade, nor that every cascade can be fully explained by epidemic compartments. It claims that combining content evidence with temporal and posterior evidence provides a stronger basis for analysing social risk than using any single layer alone. This is why Chapter 3 retains Figure 3.15 and Figure 3.16 as visual synthesis tools: they make the logic of the thesis transparent without repeating the duplicate tables that previously restated the same information in words.

3.9 Chapter Summary

This chapter evaluated the proposed framework across content-based and diffusion-oriented datasets. Transformer models produced strong detection performance, while temporal-network analysis identified spreading dynamics. SEIZ provided a stronger representation of fake-news diffusion than SIR, and ABC/ABC-SMC added uncertainty-aware calibration and forecasting capability. The results support studying fake-news risk as an integrated content-to-cascade process.

The chapter also showed why the evidence layers should remain separated during interpretation. Content-based datasets support binary fake-news classification and probability scoring, whereas diffusion-oriented datasets support temporal cascade reconstruction, epidemic modelling and posterior inference. This separation prevents classification metrics, cascade statistics and intervention scenarios from being treated as interchangeable evidence.

The final empirical contribution is the connection between posterior calibration and practical scenario reasoning. The intervention analysis suggests that early combined mitigation is more effective than delayed action, while the baseline posterior diffusion-risk functional gives a compact way to link Transformer detection, temporal amplification and ABC-SMC-SEIZ forecast growth. The concluding section builds on these findings by consolidating the conclusions, limitations and future research directions; the separate Thesis Contributions section then summarizes the dissertation's scientific and methodological contributions.

Table 3.15. Research-question evidence summary for Chapter 3.

RQ	Evidence in Chapter 3	Conclusion
RQ1	Section 3.3; Tables 3.2-3.7 and Figures 3.1-3.5	Transformer performance across datasets
RQ2	Section 3.4; Tables 3.8-3.9 and Figures 3.6-3.7	Temporal cascades reconstructed
RQ3	Section 3.5; Table 3.10 and Figures 3.8-3.9	SEIZ compared with SIR
RQ4	Section 3.6; Tables 3.11-3.13 and Figures 3.10-3.14	ABC/ABC-SMC uncertainty estimated
RQ5	Sections 3.7-3.8; Table 3.14 and Figures 3.15-3.16	Posterior prediction and intervention logic shown

Conclusion - a summary of the results obtained

This concluding section summarizes the results obtained in the dissertation and relates them to the dissertation goal, the research tasks and the research questions formulated in Chapter 1. The dissertation examined fake news as a combined content and diffusion phenomenon rather than as a single text-classification task. For this reason, the conclusion brings together the evidence from Transformer-based detection, temporal cascade reconstruction, SIR and SEIZ epidemic modelling, ABC and ABC-SMC posterior inference, forecasting and intervention-oriented scenario analysis.

The central dissertation goal was to develop and evaluate an integrated framework for fake-news detection and diffusion analysis on social networks. The obtained results support this objective by showing that textual credibility, temporal amplification and posterior uncertainty should be analysed together. A detection model can identify suspicious content, but it cannot determine social impact without temporal evidence. Similarly, a diffusion model can describe spread, but it remains incomplete without content-risk probability and uncertainty-aware parameter inference.

The final synthesis therefore emphasizes both the scientific value and the methodological boundary of the thesis. The proposed framework is not presented as a universal solution to all fake-news problems. It is presented as a traceable and reproducible content-to-cascade methodology that integrates detection, temporal diffusion, epidemic modelling and Bayesian uncertainty in a single analytical structure.

Conclusions of the Thesis

Firstly, Transformer-based models provide a strong and operationally useful layer for binary fake-news detection. BERT, RoBERTa, BigBird, mBERT and XLM-RoBERTa produced reliable fake-class probability outputs across short COVID-19 posts, long news articles, healthcare misinformation and low-resource Albanian text. BERT was effective for short posts, RoBERTa performed strongly in article-style and healthcare settings, BigBird was useful for longer documents, and XLM-RoBERTa produced the strongest low-resource Albanian result. This answers RQ1 and confirms that model choice must depend on text length, language and domain.

The second conclusion is that detection results must be connected with temporal cascade evidence before social risk can be evaluated. A fake-news probability score identifies suspicious content, but it does not show whether the claim becomes influential. The temporal-network layer addresses this limitation by reconstructing directed, timestamped cascades and producing $I_{\mathrm{new}}(t)$ and $I_{\mathrm{cum}}(t)$ trajectories for diffusion modelling. This answers RQ2 and supports the thesis argument that fake-news risk is a combined content and propagation problem.

The two principal empirical cascades demonstrate this point. The COVID-19 Twitter/X fake-news claim showed irregular and multi-burst behaviour, while the Hoaxy / OSoMeNet political claim produced a larger and smoother cumulative spread. These differences show that fake-news diffusion cannot be inferred from classification probability alone. The auxiliary Hoaxy-imported and synthetic Albanian cases further validate the pipeline structure, but they are correctly treated as supplementary or validation-oriented evidence.

The third conclusion concerns epidemic modelling. SIR is useful as a transparent baseline, but SEIZ gives a more suitable representation of fake-news dynamics because it includes exposed and skeptic or non-spreading states. In the two principal empirical cascades, SEIZ produced lower trajectory-level error than SIR, especially for the irregular COVID-19 cascade. This answers RQ3 and justifies using SEIZ as the main diffusion model for posterior inference and intervention-oriented analysis.

The fourth conclusion is that uncertainty-aware inference is necessary because social-network cascades are only partially observed. Platform data contain visible reposts, replies, timestamps and directed interactions, but they do not contain all exposures, private shares, deleted posts, recommendation-system impressions or silent decisions not to spread. ABC and ABC-SMC address this limitation by estimating plausible posterior regions of epidemic parameters instead of relying only on deterministic fitted values. The posterior predictive simulations and short-horizon forecasts show that ABC-SMC-SEIZ improves calibration and supports uncertainty-aware diffusion analysis. This answers RQ4.

The fifth conclusion is that the proposed framework can support cautious intervention-oriented reasoning. The counterfactual scenarios show that early combined mitigation is more effective than delayed action or correction-only scenarios. The results are especially important because they connect mitigation timing with the actual phase of the observed cascade. In both principal cascades, interventions applied before the main acceleration phase produced stronger reductions than interventions applied after the peak. This answers RQ5.

The baseline posterior fake-news diffusion-risk functional provides the final synthesis of the empirical pipeline. It links Transformer fake-news probability, temporal-network amplification and ABC-SMC-SEIZ posterior predictive growth into one auditable risk quantity. Its value is not that it replaces the individual models, but that it makes the connection between content risk, cascade activity and expected posterior growth explicit and inspectable.

Limitations of the study:

- The Transformer probability $p_f(x)$ is a model-estimated content-risk signal and should not be interpreted as an absolute fact-checking judgement.
- Retweets, reposts, replies and shares are observable diffusion events; they do not prove individual belief, intention or persuasion.
- The exposed and skeptic states in SEIZ are modelling proxies because hidden exposure, silent rejection and private sharing are not fully observed in platform data.
- The main posterior inference and intervention analysis are based on two principal empirical cascades; the additional Hoaxy-imported and synthetic Albanian cases are supplementary or validation-oriented.
- The intervention scenarios are controlled counterfactual simulations and should not be read as causal estimates of real platform moderation policies.
- The synthetic Albanian temporal cascade supports low-resource pipeline validation, but future work requires real Albanian temporal diffusion data for stronger empirical conclusions.

These limitations do not weaken the dissertation contribution. They define the conditions under which the results should be interpreted and make the framework scientifically defensible. The thesis demonstrates a reproducible way to connect fake-news detection, temporal diffusion and uncertainty-aware modelling while avoiding unsupported claims about private belief, platform causality or universal generalization.

Overall, the dissertation shows that fake-news analysis becomes stronger when detection, temporal diffusion, epidemic modelling and Bayesian uncertainty are combined within one traceable framework. The proposed approach moves fake-news research beyond isolated classification accuracy toward integrated, uncertainty-aware and intervention-oriented modelling.

Future Work

- ❑ Future research should validate the framework on a larger number of real-world fake-news cascades from different domains, platforms and languages. Broader evaluation would test whether the SEIZ and ABC-SMC behaviour observed in the COVID-19 and Hoaxy cascades also appears in elections, health crises, financial rumours, geopolitical misinformation and local-language contexts.

- A second direction is the extension of the temporal-network layer toward higher-order temporal networks, when ethically and legally richer social-network evidence is available. Follower graphs, reply trees, quote-post chains, repeated exposure histories, bot indicators, community membership and cross-platform links could be used to model path-dependent diffusion beyond simple pairwise timestamped edges. This would allow future versions of the framework to capture repeated exposure, memory effects and cross-community transitions, while enabling comparison with temporal graph-learning methods and graph neural networks.
- A third direction is the expansion of the uncertainty layer. ABC-SMC could be compared with neural posterior estimation, sequential likelihood-free inference or hierarchical Bayesian models in which parameters vary across communities, platforms and claim types. This would improve posterior calibration and make early-warning forecasts more robust.
- A fourth direction is the strengthening of the intervention component. Real warning-label timestamps, fact-checking events, visibility changes or moderation records could support stronger causal evaluation. This would move the framework from controlled counterfactual ranking toward more realistic policy-oriented analysis.
- A fifth direction is the development of an intervention-aware posterior diffusion-risk functional. Future work could extend the baseline risk functional by replacing the baseline posterior parameter vector θ with an intervention-modified parameter vector $\theta^{(a)}$, where a represents actions such as debunking, warning labels, visibility reduction or combined mitigation. This would allow the framework to estimate the posterior risk reduction produced by each intervention and compare mitigation strategies through expected reduction in future cumulative spread.
- Finally, future work should improve multilingual and low-resource coverage. Larger Albanian corpora, better fact-checking alignment, dialect-aware preprocessing and real Albanian temporal cascades would make the framework more useful for countries and languages where misinformation circulates actively but benchmark resources remain limited.

Thesis Contributions

This dissertation makes the following scientific and methodological contributions to fake-news detection and diffusion analysis.

1. Proposed and evaluated an integrated framework that connects Transformer-based fake-news detection, temporal cascade reconstruction, SIR/SEIZ epidemic modelling and ABC/ABC-SMC posterior inference. The contribution lies in treating fake news as a content-to-cascade process rather than as an isolated text-classification problem.

2. Established a clear methodological boundary between datasets used for binary fake-news detection and datasets used for temporal diffusion modelling. This distinction prevents content-only classification results from being incorrectly interpreted as epidemic or cascade evidence.
3. The dissertation implemented and evaluated BERT, RoBERTa, BigBird, mBERT and XLM-RoBERTa across short-text, long-text, healthcare and low-resource Albanian fake-news datasets. The resulting fake-news probability score is used as a content-risk signal that supports the transition from detection to temporal cascade analysis.
4. Fake-news claims was defined as directed, timestamped temporal cascades and extracts key diffusion quantities such as new active spreaders, cumulative active spreaders, cascade size, time to peak and network reach. This contribution provides the empirical bridge between content detection and epidemic-style modelling.
5. Comparison between SIR and SEIZ models on the principal empirical fake-news cascades and showed that SEIZ provides a more suitable representation when delayed exposure, hesitation, skepticism or non-spreading behaviour are relevant. SIR is retained as a transparent baseline, while SEIZ is used as the main diffusion model.
6. Applied ABC and ABC-SMC to estimate uncertain epidemic parameters from partially observed social-network cascades. This contribution moves the analysis beyond deterministic point estimates by reporting posterior parameter distributions, credible intervals, posterior predictive checks and calibration diagnostics.
7. The dissertation introduces a baseline risk functional that links Transformer fake-news probability, temporal-network amplification and ABC-SMC-SEIZ posterior predictive growth. The functional does not replace the individual models; instead, it provides an auditable mathematical synthesis of the detection, diffusion and posterior-inference layers.
8. To evaluate counterfactual intervention scenarios, the dissertation uses the calibrated diffusion framework, including debunking, visibility reduction and early combined mitigation. The contribution is not a causal claim about real platform policy, but a reproducible model-based ranking of intervention mechanisms under explicit assumptions.

Together, these contributions move fake-news analysis beyond isolated classification accuracy toward an integrated, uncertainty-aware and intervention-oriented framework for studying how false information is detected, amplified and potentially mitigated in social networks.

List of publications related to the thesis

Author Publications Connected to the Dissertation

1. Bozhiqi, K., & Guliashki, V. G. (2023). Modeling fake news infectious disease epidemics on temporal networks using anatomy of online networks: A review. In Proceedings of the 24th International Conference on Control Systems and Computer Science (CSCS 2023), Bucharest, Romania, May 24–26, 2023, pp. 206–212. IEEE. https://doi.org/10.1109/CSCS59211.2023.00040. https://ieeexplore.ieee.org/document/10214729 [87]
2. Bozhiqi, K. V., & Shtino, V. B. (2025). Classification of fake news using machine learning techniques and spreading dynamics with temporal network on Twitter. In Advances in Computer Science and Information Technology: Book of Proceedings of the 1st International Scientific Conference on Computer Science and Information Technology – CSIT 2025, Durrës, Albania, pp. 195–209. Luis Print. [108]
3. Bozhiqi Vula, K., & Guliashki, V. G. (2025). BigSEIZ: Combining BigBird with epidemic models to analyse spreading fake news on Twitter temporal network. In L. Barolli, T. Enokido, & I. Woungang (Eds.), Complex, Intelligent, and Software Intensive Systems: Proceedings of CISIS 2025 (pp. 47–60). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-96096-3_5. Lecture Notes on Data Engineering and Communications Technologies (LNDECT, volume 261), ISSN: 2367-4512 (print), ISSN: 2367-4520 (electronic), **SJR (0,123), Q4**, https://link.springer.com/book/10.1007/978-3-031-96096-3 [86]
4. Bozhiqi, K., & Guliashki, V. G. (2025). BigSEIZ Test Outcomes: Applying ML Techniques for Fake News Identification and Comparing SIR and SEIZ Epidemic Models on Twitter's Temporal Network. In Proceedings of the 33rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2025), ConTEL Symposium, Split, Croatia, pp. 1–7. https://doi.org/10.23919/SoftCOM66362.2025.11197375. [88]
5. Bozhiqi, K. (2025). A Hybrid NLP–Epidemiological Model for Fake News Network: BigBird-Assisted SEIZ Modeling with Approximate Bayesian Computation. In Proceedings of IEEE International Workshop on Technologies for Defense and Security (TechDefense 2025), Rome, Italy, pp. 95–100. IEEE. https://www.researchgate.net/publication/399822332_A_Hybrid_NLP_Epidemiological_Model_for_Fake_News_Network_BigBird_Assisted_SEIZ_Modeling_with_Approximate_Bayesian_Computation [89]

Citations and Award

- Bozhiqi, K., & Guliashki, V. G. (2023). Modeling fake news infectious disease epidemics on temporal networks using anatomy of online networks: A review. In Proceedings of the 24th International Conference on Control Systems and Computer Science (CSCS 2023), Bucharest, Romania, May 24–26, 2023, pp. 206–212. IEEE. https://doi.org/10.1109/CSCS59211.2023.00040. [87]

1) Cited by: Castiglione, A., Cimmino, L., Loia, V., Nappi, M., & Sorrentino, C. (2025). Dynamic Lip Motion Analysis for Deepfake Detection. In 2025 25th International Conference on Control Systems and Computer Science, pp. 452–459. DOI: 10.1109/CSCS66924.2025.00073 [186]. This citation is also listed in the IICT-BAS citation report [187].

2) Cited by: Girka, K. G., (2025), Master Thesis: Analysis of mathematical Markov models for predicting the spread of fake news in social networks, Kharkiv National University of Radio Electronics, Faculty of Information and Analytical Technologies and Management, Department of Applied Mathematics, (Гирка, К. Г. (2025). Аналіз математичних марковських моделей прогнозування поширення фейкових новин у соціальних мережах) [191].
https://openarchive.nure.ua/entities/publication/73c7dc37-3a51-429a-be15-30ef64228ed1, https://openarchive.nure.ua/server/api/core/bitstreams/8d4a6679-abbb-4c7e-aafc-28a2b527e6cb/content

- Bozhiqi Vula, K., & Guliashki, V. G. (2025). BigSEIZ: Combining BigBird with epidemic models to analyse spreading fake news on Twitter temporal network. In L. Barolli, T. Enokido, & I. Woungang (Eds.), Complex, Intelligent, and Software Intensive Systems: Proceedings of CISIS 2025 (pp. 47–60). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-96096-3_5. [86]

3) Cited by: Ding, W., & Bhuiyan, H. (2026). PLiTS-Net: A PLM-Conditioned Liquid Thread Network for Social Media Sentiment Analysis. *IEEE International Conference on Big Data and Smart Computing (BigComp* 2026*)*, (pp. 237-244). DOI: 10.1109/BigComp68355.2026.00044 [189].

- Bozhiqi, K., & Guliashki, V. G. (2025). BigSEIZ Test Outcomes: Applying ML Techniques for Fake News Identification and Comparing SIR and SEIZ Epidemic Models on Twitter's Temporal Network. In Proceedings of the 33rd International Conference on Software,

Telecommunications and Computer Networks (SoftCOM 2025), ConTEL Symposium, Split, Croatia, pp. 1–7. https://doi.org/10.23919/SoftCOM66362.2025.11197375. [88]
4) Cited by: Chung, J., Chaddha, R., Kaur, A., Das, D., Badhe, S., & Huang, N. (2026). Generative AI and digital ecosystem resilience: A proactive lifecycle-based survey. TechRxiv. https://doi.org/10.36227/techrxiv.177092219.91727099/v1 [190].

- Bozhiqi, K. (2025). A Hybrid NLP–Epidemiological Model for Fake News Network: BigBird-Assisted SEIZ Modeling with Approximate Bayesian Computation. In Proceedings of IEEE International Workshop on Technologies for Defense and Security (TechDefense 2025), Rome, Italy, pp. 95–100. IEEE. [89]

Award record: The IEEE Women in Engineering / TechDefense 2025 special-session record documents the **Best Poster** award for the BigBird-assisted SEIZ and ABC work [188]. The award is reported as presentation-level recognition and does not replace the empirical evaluation of the dissertation framework in Chapter 3.

Declaration of Originality

I declare that

1. This dissertation contains original research results obtained by me with the support and assistance of my supervisor. The presented dissertation work has been prepared by me.
2. Results that have been obtained, described and/or published by other scientists are duly and extensively cited in the bibliography. In the dissertation, no foreign texts are used directly or indirectly or, if parts of them are used, they are referred to / quoted and no part of my dissertation infringes the copyright of institution or person.
3. No part of my dissertation work has been presented in this form in the same or in another university, educational and/or scientific institution for the award of educational or scientific degree.

In the event of a discrepancy with the circumstances declared by me according to points 1, 2 and 3 of this declaration, I bear responsibility in accordance with the law and normative documents of IICT.

Signature:

Bibliography

[1] Shushkevich, E. (2024). Fake News Detection in Online Platforms. Doctoral thesis, Technological University Dublin. https://arrow.tudublin.ie/ittthedoc/16

[2] Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Computing Surveys, 53(5), Article 109. https://doi.org/10.1145/3395046

[3] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600

[4] Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

[5] Li, Y., Jiang, B., Shu, K., & Liu, H. (2020). MM-COVID: A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation. arXiv:2011.04088.

[6] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems.

[8] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692.

[9] Pires, T., Schlinger, E., & Garrette, D. (2019). How Multilingual is Multilingual BERT? Proceedings of ACL, 4996–5001. https://doi.org/10.18653/v1/P19-1493

[10] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Proceedings of EMNLP-IJCNLP, 3982–3992. https://doi.org/10.18653/v1/D19-1410

[11] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, É., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. Proceedings of ACL, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

[12] Zaheer, M., Guruganesh, G., Dubey, A., Ainslie, J., Alberti, C., Ontañón, S., Pham, P., Ravula, A., Wang, Q., Yang, L., & Ahmed, A. (2020). Big Bird: Transformers for Longer Sequences. Advances in Neural Information Processing Systems, 33, 17283–17297. https://proceedings.neurips.cc/paper/2020/hash/c8512d142a2d849725f31a9a7a361ab9-Abstract.html

[13] Scholtes, I. (2017). When is a network a network? Multi-order graphical model selection in pathways and temporal networks. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1037-1046. https://doi.org/10.1145/3097983.3098145

[14] Scholtes, I., Wider, N., Pfitzner, R., Garas, A., Tessone, C. J., & Schweitzer, F. (2014). Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. Nature Communications, 5, 5024. https://doi.org/10.1038/ncomms6024.

[15] Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D., & Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. Nature Communications, 5, 4630. https://doi.org/10.1038/ncomms5630.

[16] Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. Proceedings of the National Academy of Sciences, 113(3), 554-559. https://doi.org/10.1073/pnas.1517441113

[17] Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. Proceedings of the National Academy of Sciences, 118(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

[18] Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130-1132. https://doi.org/10.1126/science.aaa1160

[19] Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. Genetics, 162(4), 2025-2035. https://doi.org/10.1093/genetics/162.4.2025

[20] Marjoram, P., Molitor, J., Plagnol, V., & Tavare, S. (2003). Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences, 100(26), 15324-15328. https://doi.org/10.1073/pnas.0306899100

[21] Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences, 104(6), 1760-1765. https://doi.org/10.1073/pnas.0607208104

[22] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface, 6(31), 187-202. https://doi.org/10.1098/rsif.2008.0172

[23] Sunnaker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian Computation. PLOS Computational Biology, 9(1), e1002803. https://doi.org/10.1371/journal.pcbi.1002803

[24] Kypraios, T., Neal, P., & Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. Mathematical Biosciences, 287, 42-53. https://doi.org/10.1016/j.mbs.2016.07.001

[25] Minter, A., & Retkute, R. (2019). Approximate Bayesian Computation for infectious disease modelling. Epidemics, 29, 100368. https://doi.org/10.1016/j.epidem.2019.100368

[26] Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. Proceedings of the National Academy of Sciences, 117(48), 30055-30062. https://doi.org/10.1073/pnas.1912789117

[27] Papamakarios, G., & Murray, I. (2016). Fast epsilon-free inference of simulation models with Bayesian conditional density estimation. Advances in Neural Information Processing Systems, 29.
[28] Gomez-Rodriguez, M., Leskovec, J., & Krause, A. (2010). Inferring networks of diffusion and influence. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1019-1028. https://doi.org/10.1145/1835804.1835933
[29] Farajtabar, M., Wang, Y., Gomez-Rodriguez, M., Li, S., Zha, H., & Song, L. (2015). COEVOLVE: A joint point process model for information diffusion and network co-evolution. Advances in Neural Information Processing Systems, 28.
[30] Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., & Chakraborty, T. (2021). Fighting an Infodemic: COVID-19 Fake News Dataset. In Combating Online Hostile Posts in Regional Languages during Emergency Situation. Springer. https://doi.org/10.1007/978-3-030-73696-5_3
[31] Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, 115(772), 700-721. https://doi.org/10.1098/rspa.1927.0118
[32] Daley, D. J., & Kendall, D. G. (1964). Epidemics and rumours. Nature, 204, 1118. https://doi.org/10.1038/2041118a0
[33] Jin, F., Dougherty, E., Saraf, P., Cao, Y., & Ramakrishnan, N. (2013). Epidemiological modelling of news and rumors on Twitter. Proceedings of the 7th Workshop on Social Network Mining and Analysis, 1-9. https://doi.org/10.1145/2501025.2501027
[34] Kucharski, A. (2016). Study epidemiology of fake news. Nature, 540(7634), 525. https://doi.org/10.1038/540525a
[35] Raponi, S., Khalifa, Z., Oligeri, G., & Di Pietro, R. (2022). Fake news propagation: A review of epidemic models, datasets, and insights. ACM Transactions on the Web, 16(3), Article 12. https://doi.org/10.1145/3522756
[36] Kauk, J., Kreysa, H., & Schweinberger, S. R. (2021). Understanding and countering the spread of conspiracy theories in social networks: Evidence from epidemiological models of Twitter data. PLOS ONE, 16(8), e0256179. https://doi.org/10.1371/journal.pone.0256179
[37] Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. Nature Communications, 10, 7. https://doi.org/10.1038/s41467-018-07761-2
[38] Lambiotte, R., Rosvall, M., & Scholtes, I. (2019). From networks to optimal higher-order models of complex systems. Nature Physics, 15, 313–320. https://doi.org/10.1038/s41567-019-0459-y.
[39] Murayama, T., Wakamiya, S., Aramaki, E., & Kobayashi, R. (2021). Modeling the spread of fake news on Twitter. PLOS ONE, 16(4), e0250419. https://doi.org/10.1371/journal.pone.0250419
[40] Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. Statistics and Computing, 22, 1167-1180. https://doi.org/10.1007/s11222-011-9288-2

[41] Ruffo, G., Semeraro, A., Giachanou, A., & Rosso, P. (2021). Studying Fake News Spreading, Polarisation Dynamics, and Manipulation by Bots: A Tale of Networks and Language. arXiv:2109.07909.

[42] Joachims, T. (2002). Learning to Classify Text Using Support Vector Machines. Springer.

[43] Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

[44] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785

[45] Agarwal, P., Reddivari, S., & Reddivari, K. (2022). Fake News Detection: An Investigation based on Machine Learning. Proceedings of the 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI), 61-62. https://doi.org/10.1109/IRI54793.2022.00025

[46] Zhang, C., Gupta, S., Kauten, C., Deokar, A. V., & Qin, X. (2019). Detecting fake news for reducing misinformation risks using analytics approaches. European Journal of Operational Research, 279(3), 1036-1052. https://doi.org/10.1016/j.ejor.2019.06.022

[47] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

[48] OpenAI. (2023). GPT-4 Technical Report. arXiv:2303.08774.

[49] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), Article 248. https://doi.org/10.1145/3571730

[50] Jin, D., Jin, Z., Zhou, J. T., & Szolovits, P. (2020). Is BERT Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment. Proceedings of the AAAI Conference on Artificial Intelligence, 34(5), 8018-8025. https://doi.org/10.1609/aaai.v34i05.6311

[51] Gururangan, S., Marasovic, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. Proceedings of ACL, 8342-8360. https://doi.org/10.18653/v1/2020.acl-main.740

[52] Holme, P., & Saramäki, J. (2012). Temporal networks. Physics Reports, 519(3), 97-125. https://doi.org/10.1016/j.physrep.2012.03.001

[53] Vosoughi, S., Roy, D., & Aral, S. (2018). The Spread of True and False News Online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

[54] Shao, C., Hui, P.-M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. PLOS ONE, 13(4), e0196087. https://doi.org/10.1371/journal.pone.0196087

[55] Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. Proceedings of

the 20th International Conference on World Wide Web, 695–704. https://doi.org/10.1145/1963405.1963503.

[56] Sunstein, C. R. (2001). Republic.com. Princeton University Press.

[57] Pariser, E. (2011). The Filter Bubble: What the Internet Is Hiding from You. Penguin.

[58] Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. Social Networks, 1(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7.

[59] Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the Web. Stanford InfoLab Technical Report.

[60] Newman, M. E. J. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103.

[61] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

[62] Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on Twitter. Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 65–74. https://doi.org/10.1145/1935826.1935845.

[63] Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, 591–600. https://doi.org/10.1145/1772690.1772751.

[64] Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). TwitterRank: Finding topic-sensitive influential Twitterers. Proceedings of the Third ACM International Conference on Web Search and Data Mining, 261–270. https://doi.org/10.1145/1718487.1718520.

[65] Pfitzner, R., Scholtes, I., Garas, A., Tessone, C. J., & Schweitzer, F. (2013). Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. Physical Review Letters, 110(19), 198701. https://doi.org/10.1103/PhysRevLett.110.198701.

[66] Benson, A. R., Gleich, D. F., & Leskovec, J. (2016). Higher-order organization of complex networks. Science, 353(6295), 163–166. https://doi.org/10.1126/science.aad9029.

[67] Maki, D. P., & Thompson, M. (1973). Mathematical Models and Applications: With Emphasis on the Social, Life, and Management Sciences. Prentice-Hall.

[68] Zanette, D. H. (2002). Dynamics of rumor propagation on small-world networks. Physical Review E, 65(4), 041908. https://doi.org/10.1103/PhysRevE.65.041908

[69] Nekovee, M., Moreno, Y., Bianconi, G., & Marsili, M. (2007). Theory of rumour spreading in complex social networks. Physica A: Statistical Mechanics and its Applications, 374(1), 457-470. https://doi.org/10.1016/j.physa.2006.07.017

[70] Newman, M. E. J. (2002). Spread of epidemic disease on networks. Physical Review E, 66(1), 016128. https://doi.org/10.1103/PhysRevE.66.016128

[71] Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic spreading in scale-free networks. Physical Review Letters, 86(14), 3200-3203. https://doi.org/10.1103/PhysRevLett.86.3200

[72] Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. Reviews of Modern Physics, 87(3), 925-979. https://doi.org/10.1103/RevModPhys.87.925

[73] Maleki, M., Mead, E., Arani, M., & Agarwal, N. (2021). Using an epidemiological model to study the spread of misinformation during the Black Lives Matter movement. arXiv:2103.12191.

[74] Centola, D. (2010). The spread of behavior in an online social network experiment. Science, 329(5996), 1194-1197. https://doi.org/10.1126/science.1185231

[75] Wang, Y., Xiao, G., Liu, Y., & Jin, C. (2019). Dynamics of social contagions with memory of nonredundant information. Physical Review E, 100(6), 062308.

[76] Dutta, R., Mira, A., & Onnela, J.-P. (2018). Bayesian inference of spreading processes on networks. Proceedings of the Royal Society A, 474(2215), 20180129. https://doi.org/10.1098/rspa.2018.0129

[77] Lintusaari, J., Gutmann, M. U., Dutta, R., Kaski, S., & Corander, J. (2017). Fundamentals and recent developments in Approximate Bayesian Computation. Systematic Biology, 66(1), e66-e82. https://doi.org/10.1093/sysbio/syw077

[78] Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094-1096. https://doi.org/10.1126/science.aao2998

[79] Blum, M. G. B., Nunes, M. A., Prangle, D., & Sisson, S. A. (2013). A comparative review of dimension reduction methods in approximate Bayesian computation. Statistical Science, 28(2), 189-208. https://doi.org/10.1214/12-STS406

[80] Fearnhead, P., & Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. Journal of the Royal Statistical Society: Series B, 74(3), 419-474. https://doi.org/10.1111/j.1467-9868.2011.01010.x

[81] McIntire, G. (2017). Fake Real News Dataset. GitHub repository. https://github.com/GeorgeMcIntire/fake_real_news_dataset

[82] Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments (pp. 127-138). Springer.

[83] Cui, L., & Lee, D. (2020). CoAID: COVID-19 healthcare misinformation dataset. arXiv:2006.00885.

[84] Canhasi, E., Shijaku, R., & Berisha, E. (2022). Albanian Fake News Detection. ACM Transactions on Asian and Low-Resource Language Information Processing, 21(5). https://doi.org/10.1145/3487288

[85] Rexha, G. (2021). Albanian Fake News Corpus. Kaggle dataset derived from Albanian Fake News Detection.

[86] Bozhiqi Vula, K., & Guliashki, V. G. (2025). BigSEIZ: Combining BigBird with epidemic models to analyse spreading fake news on Twitter temporal network. In L. Barolli, T. Enokido, & I. Woungang (Eds.), Complex, Intelligent, and Software Intensive Systems: Proceedings of the 19th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2025), Fukuoka, Japan, July 2–4, 2025 (pp. 47–60). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-96096-3_5

[87] Bozhiqi, K., & Guliashki, V. (2023). Modelling fake news infectious disease epidemics on temporal networks using anatomy of online networks: A review. In Proceedings of the 24th International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, May 24–26, 2023 (pp. 206–212). IEEE. https://doi.org/10.1109/CSCS59211.2023.00040

[88] Bozhiqi, K., & Guliashki, V. (2025). BigSEIZ Test Outcomes: Applying ML Techniques for Fake News Identification and Comparing SIR and SEIZ Epidemic Models on Twitter's Temporal Network. Proceedings of the ConTEL Symposium within the 33rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2025), Split, Croatia, September 18-20, 2025. https://doi.org/10.23919/SoftCOM66362.2025.11197375

[89] Bozhiqi, K. (2025). A Hybrid NLP–Epidemiological Model for Fake News Network: BigBird-Assisted SEIZ Modelling with Approximate Bayesian Computation. IEEE International Workshop on Technologies for Defense and Security (TechDefense 2025), Rome, Italy, November 5–7, 2025.

[90] Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. Proceedings of the 25th International Conference Companion on World Wide Web, 745-750. https://doi.org/10.1145/2872518.2890098

[91] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

[92] Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. Transactions of the Association for Computational Linguistics, 6, 587-604. https://doi.org/10.1162/tacl_a_00041

[93] Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. Patterns, 2(11), 100336. https://doi.org/10.1016/j.patter.2021.100336

[94] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229. https://doi.org/10.1145/3287560.3287596

[95] Joolsa. (2022). FRN / Fake and Real News Dataset. Dataset resource based on George McIntire's fake and real news corpus.

[96] ISOT Research Group. (2022). ISOT Fake News Dataset. University of Victoria. https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/

[97] Cui, L., & Lee, D. (2020). CoAID GitHub Repository. https://github.com/cuilimeng/CoAID
[98] Shijaku, R. (2021). Albanian Fake News Corpus GitHub Repository. https://github.com/rexshijaku/alb-fake-news-corpus
[99] Han, B., & Baldwin, T. (2011). Lexical normalisation of short text messages: Makn sens a #twitter. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 368-378. https://aclanthology.org/P11-1038/
[100] Frenay, B., & Verleysen, M. (2014). Classification in the presence of label noise: A survey. IEEE Transactions on Neural Networks and Learning Systems, 25(5), 845-869. https://doi.org/10.1109/TNNLS.2013.2292894
[101] Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision, 2980-2988. https://doi.org/10.1109/ICCV.2017.324
[102] Carrignon, S., Bentley, R. A., & Ruck, D. J. (2019). Modelling rapid online cultural transmission: Evaluating neutral models on Twitter data with approximate Bayesian computation. Palgrave Communications, 5, Article 83. https://doi.org/10.1057/s41599-019-0295-9
[103] Ma, J., Gao, W., & Wong, K.-F. (2018). Rumor Detection on Twitter with Tree-Structured Recursive Neural Networks. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1980-1989. https://doi.org/10.18653/v1/P18-1184
[104] Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics, 41, 379-406. https://doi.org/10.1146/annurev-ecolsys-102209-144621
[105] Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLOS ONE, 11(3), e0150989. https://doi.org/10.1371/journal.pone.0150989
[106] Csillery, K., Blum, M. G. B., Gaggiotti, O. E., & Francois, O. (2010). Approximate Bayesian Computation (ABC) in practice. Trends in Ecology & Evolution, 25(7), 410-418. https://doi.org/10.1016/j.tree.2010.04.001
[107] Dutta, R., Schoengens, M., Pacchiardi, L., Ummadisingu, A., Widmer, N., Kunzli, P., Onnela, J.-P., & Mira, A. (2021). ABCpy: A High-Performance Computing Perspective to Approximate Bayesian Computation. Journal of Statistical Software, 100(7), 1-38. https://doi.org/10.18637/jss.v100.i07
[108] Bozhiqi, K. V., & Shtino, V. B. (2025). Classification of fake news using machine learning techniques and spreading dynamics with Temporal Network on Twitter. In E. Xhaferra, U. Cerma, V. Shtino, E. Tabaku, R. Kapciu, & F. Bushi (Eds.), Advances in Computer Science and Information Technology: Book of Proceedings of the 1st International Scientific Conference on Computer Science and Information Technology - CSIT 2025 (pp. 195-209). Durres, Albania: Luis Print. Available at https://uamd.edu.al/wp-content/uploads/2025/07/Book-of-Proceedings-1.pdf

[109] Sharif, O., Hossain, E., & Hoque, M. M. (2021). Combating Hostility: Covid-19 Fake News and Hostile Post Detection in Social Media. arXiv:2101.03291. https://doi.org/10.48550/arXiv.2101.03291

[110] Sharif, O. (2021). CONSTRAINT-AAAI2021: Code and dataset for the CONSTRAINT shared task in AAAI-2021. GitHub repository. https://github.com/Omar-Sharif/CONSTRAINT-AAAI2021

[111] Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception Detection for News: Three Types of Fakes. Proceedings of the Association for Information Science and Technology, 52(1), 1–4. https://doi.org/10.1002/pra2.2015.145052010083

[112] Sproat, R., Black, A. W., Chen, S., Kumar, S., Ostendorf, M., & Richards, C. (2001). Normalization of non-standard words. Computer Speech & Language, 15(3), 287-333. https://doi.org/10.1006/csla.2001.0169

[113] Harris, S., Hadi, H. J., Ahmad, N., & Alshara, M. A. (2024). Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas. Technologies, 12(11), 222. https://doi.org/10.3390/technologies12110222

[114] Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T. H., Ding, K., Karami, M., & Liu, H. (2020). Combating disinformation in a social media age. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(6), e1385. https://doi.org/10.1002/widm.1385

[115] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144. https://doi.org/10.1145/2939672.2939778

[116] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

[117] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv:1702.08608.

[118] He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. International Conference on Learning Representations.

[119] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38-45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[120] Rust, P., Pfeiffer, J., Vulic, I., Ruder, S., & Gurevych, I. (2021). How good is your tokenizer? On the monolingual performance of multilingual language models. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, 3118-3135. https://doi.org/10.18653/v1/2021.acl-long.243

[121] Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. International Conference on Learning Representations (ICLR).

[122] Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph Attention Networks. International Conference on Learning Representations (ICLR).

[123] Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. Advances in Neural Information Processing Systems.

[124] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4-24.

[125] Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake News Detection on Social Media using Geometric Deep Learning. arXiv:1902.06673.

[126] Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., & Huang, J. (2020). Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks. Proceedings of the AAAI Conference on Artificial Intelligence, 34(01), 549-556. https://doi.org/10.1609/aaai.v34i01.5393

[127] Nguyen, V.-H., Sugiyama, K., Nakov, P., & Kan, M.-Y. (2022). FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. Communications of the ACM, 65(4), 124-132. https://doi.org/10.1145/3517214

[128] Kazemi, S. M., Goel, R., Jain, K., Kobyzev, I., Sood, A., Wu, J., & Poupart, P. (2020). Representation learning for dynamic graphs: A survey. Journal of Machine Learning Research, 21(70), 1-73.

[129] Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., & Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs. arXiv:2006.10637.

[130] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Proceedings of the International Conference on Machine Learning, 1050-1059.

[131] Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? Advances in Neural Information Processing Systems, 30.

[132] Murphy, K. P. (2022). Probabilistic Machine Learning: An Introduction. MIT Press.

[133] M. D. Wilkinson et al., 'The FAIR Guiding Principles for scientific data management and stewardship,' Scientific Data, 2016.

[134] Association for Computing Machinery, Artifact Review and Badging policy, official ACM publication policy, accessed 2026.

[135] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Roziere, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv:2302.13971.

[136] Zhang, X., et al. (2024). Large language models for fake news detection: A survey. arXiv.

[137] Singhal, S., Kabra, A., Sharma, M., Shah, R. R., Chakraborty, T., & Kumaraguru, P. (2022). SpotFake+: A multimodal framework for fake news detection. IEEE International Conference on Multimedia Information Processing and Retrieval.

[138] Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). MVAE: Multimodal variational autoencoder for fake news detection. The World Wide Web Conference Companion, 291-298.
[139] Qi, P., Cao, J., Yang, T., Guo, J., & Li, J. (2021). User preference-aware fake news detection. Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval.
[140] Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. Information Sciences, 497, 38-55. https://doi.org/10.1016/j.ins.2019.05.035
[141] Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). Fake news detection in social media: A systematic review. IEEE Access, 9, 171-184.
[142] Pierri, F., Luceri, L., Jindal, N., & Ferrara, E. (2022). Fake news, misinformation and disinformation in social media: A review. Springer.
[143] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.
[144] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W.-t., Rocktaschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.
[145] Manakul, P., Liusie, A., & Gales, M. J. F. (2023). SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 9004-9017. https://doi.org/10.18653/v1/2023.emnlp-main.557
[146] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. IEEE Transactions on Neural Networks, 20(1), 61-80. https://doi.org/10.1109/TNN.2008.2005605
[147] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. Advances in Neural Information Processing Systems, 29.
[148] Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: Going beyond Euclidean data. IEEE Signal Processing Magazine, 34(4), 18-42. https://doi.org/10.1109/MSP.2017.2693418
[149] Lu, Y.-J., & Li, C.-T. (2020). GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 505-514. https://doi.org/10.18653/v1/2020.acl-main.48
[150] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., & Gao, J. (2018). EANN: Event adversarial neural networks for multi-modal fake news detection. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 849-857. https://doi.org/10.1145/3219819.3219903

[151] Zhou, X., Wu, J., & Zafarani, R. (2020). SAFE: Similarity-aware multi-modal fake news detection. In Advances in Knowledge Discovery and Data Mining (PAKDD 2020), Lecture Notes in Computer Science, 12085, 354-367. Springer. https://doi.org/10.1007/978-3-030-47436-2_27

[152] Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). dEFEND: Explainable fake news detection. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 395-405. https://doi.org/10.1145/3292500.3330935

[153] Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., & Achan, K. (2020). Inductive representation learning on temporal graphs. International Conference on Learning Representations.

[154] Trivedi, R., Farajtabar, M., Biswal, P., & Zha, H. (2019). DyRep: Learning representations over dynamic graphs. International Conference on Learning Representations.

[155] Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T. B., & Leiserson, C. E. (2020). EvolveGCN: Evolving graph convolutional networks for dynamic graphs. Proceedings of the AAAI Conference on Artificial Intelligence, 34(04), 5363-5370. https://doi.org/10.1609/aaai.v34i04.5984

[156] Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 137-146. https://doi.org/10.1145/956750.956769

[157] Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. ACM Transactions on the Web, 1(1), Article 5. https://doi.org/10.1145/1232722.1232727

[158] Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through blogspace. Proceedings of the 13th International Conference on World Wide Web, 491-501. https://doi.org/10.1145/988672.988739

[159] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., & Glance, N. (2007). Cost-effective outbreak detection in networks. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 420-429. https://doi.org/10.1145/1281192.1281239

[160] Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2931-2937. https://doi.org/10.18653/v1/D17-1317

[161] Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. Proceedings of NAACL-HLT, 809-819. https://doi.org/10.18653/v1/N18-1074

[162] Wadden, D., Lin, S., Lo, K., Wang, L. L., van Zuylen, M., Cohan, A., & Hajishirzi, H. (2020). Fact or fiction: Verifying scientific claims. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 7534-7550. https://doi.org/10.18653/v1/2020.emnlp-main.609

[163] Barredo Arrieta, A., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial

Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[164] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), Article 93. https://doi.org/10.1145/3236009

[165] Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural networks. Proceedings of the 32nd International Conference on Machine Learning, 37, 1613-1622.

[166] Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems, 30.

[167] Maddox, W. J., Garipov, T., Izmailov, P., Vetrov, D., & Wilson, A. G. (2019). A simple baseline for Bayesian uncertainty in deep learning. Advances in Neural Information Processing Systems, 32.

[168] Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. Science, 363(6425), 374-378. https://doi.org/10.1126/science.aau2706

[169] Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances, 5(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

[170] Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39-50. https://doi.org/10.1016/j.cognition.2018.06.011

[171] Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 797-806. https://doi.org/10.1145/3132847.3132877

[172] Popat, K., Mukherjee, S., Strötgen, J., & Weikum, G. (2018). CredEye: A credibility lens for analyzing and explaining misinformation. Companion Proceedings of the Web Conference 2018, 155-158. https://doi.org/10.1145/3184558.3186967

[173] Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. ACM Computing Surveys, 51(2), Article 32. https://doi.org/10.1145/3161603

[174] Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., & Quattrociocchi, W. (2016). Users polarization on Facebook and YouTube. PLOS ONE, 11(8), e0159641. https://doi.org/10.1371/journal.pone.0159641

[175] Liu, Y., & Wu, Y.-F. B. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1), 354-361. https://doi.org/10.1609/aaai.v32i1.11268

[176] Shu, K., Wang, S., & Liu, H. (2019). Beyond news contents: The role of social context for fake news detection. Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 312-320. https://doi.org/10.1145/3289600.3290994

[177] Kochkina, E., Liakata, M., & Zubiaga, A. (2018). All-in-one: Multi-task learning for rumour verification. Proceedings of the 27th International Conference on Computational Linguistics, 3402-3413.

[178] Gorrell, G., Kochkina, E., Liakata, M., Aker, A., Zubiaga, A., Bontcheva, K., & Derczynski, L. (2019). SemEval-2019 Task 7: RumourEval, determining rumour veracity and support for rumours. Proceedings of the 13th International Workshop on Semantic Evaluation, 845-854. https://doi.org/10.18653/v1/S19-2147

[179] Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Sirivianos, M., Stringhini, G., & Blackburn, J. (2017). The web centipede: Understanding how web communities influence each other through the lens of mainstream and alternative news sources. Proceedings of the 2017 Internet Measurement Conference, 405-417. https://doi.org/10.1145/3131365.3131390

[180] Pesonen, H., Simola, U., Köhn-Luque, A., Vuollekoski, H., Lai, X., Frigessi, A., Kaski, S., Frazier, D. T., Maneesoonthorn, W., Martin, G. M., & Corander, J. (2023). ABC of the future. International Statistical Review, 91, 243-268. https://doi.org/10.1111/insr.12522

[181] Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., & Qi, P. (2024). Bad Actor, Good Advisor: Exploring the Role of Large Language Models in Fake News Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 38(20), 22105-22113. https://doi.org/10.1609/aaai.v38i20.30214

[182] Su, J., Cardie, C., & Nakov, P. (2024). Adapting Fake News Detection to the Era of Large Language Models. Findings of the Association for Computational Linguistics: NAACL 2024, 1473-1490. https://doi.org/10.18653/v1/2024.findings-naacl.95

[183] Wang, Y., Wang, M., Manzoor, M. A., Liu, F., Georgiev, G. N., Das, R. J., & Nakov, P. (2024). Factuality of Large Language Models: A Survey. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 19519-19529. https://doi.org/10.18653/v1/2024.emnlp-main.1088

[184] Huang, S., Poursafaei, F., Danovitch, J., Fey, M., Hu, W., Rossi, E., Leskovec, J., Bronstein, M. M., Rabusseau, G., & Rabbany, R. (2023). Temporal Graph Benchmark for Machine Learning on Temporal Graphs. Advances in Neural Information Processing Systems, 36. https://proceedings.neurips.cc/paper_files/paper/2023/hash/066b98e63313162f6562b35962671288-Abstract-Datasets_and_Benchmarks.html

[185] Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 1776-1826. https://doi.org/10.1145/3531146.3533231

[186] Castiglione, A., Cimmino, L., Loia, V., Nappi, M., & Sorrentino, C. (2025). Dynamic Lip Motion Analysis for Deepfake Detection. In Proceedings of the 2025 25th International Conference on Control

Systems and Computer Science (CSCS), Bucharest, Romania, May 27-30, 2025 (pp. 452-459). IEEE. https://doi.org/10.1109/CSCS66924.2025.00073

[187] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences. (2025). Citations of IICT-BAS publications for 2025. https://www.iict.bas.bg/docs/2025-citations-IICT.pdf

[188] IEEE Women in Engineering Special Session #11. (2025). IEEE Women in Engineering with Focus in Cybersecurity and AI - PART II: TechDefense 2025 special-session record. IEEE vTools Events. https://events.vtools.ieee.org/m/528408

[189] Ding, W., & Bhuiyan, H. (2026). PLiTS-Net: A PLM-Conditioned Liquid Thread Network for Social Media Sentiment Analysis. In Proceedings of the 2026 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 237-244). IEEE. https://doi.org/10.1109/BigComp68355.2026.00044

[190] Chung, J., Chaddha, R., Kaur, A., Das, D., Badhe, S., & Huang, N. (2026). Generative AI and digital ecosystem resilience: A proactive lifecycle-based survey. TechRxiv. https://doi.org/10.36227/techrxiv.177092219.91727099

[191] Girka, K. G. (2025). Analysis of mathematical Markov models for predicting the spread of fake news in social networks. Master thesis, Kharkiv National University of Radio Electronics, Faculty of Information and Analytical Technologies and Management, Department of Applied Mathematics, Kharkiv. https://openarchive.nure.ua/entities/publication/73c7dc37-3a51-429a-be15-30ef64228ed1

Appendices

Appendix A - List of tables

Appendix B - List of figures

Appendix C. Reproducibility Materials

Appendix C consolidates reproducibility materials, run logs, saved artefacts and generated outputs used in Chapter 3. Detailed scripts, checkpoints, seeds and output files are archived separately.

Appendix C.1. FRN BigBird run logs and split summary

Table C.1. Verified 60/20/20 FRN split summary for the BigBird final run.

Split	Real	Fake	Total	Role in the BigBird run
Training	1,903	1,898	3,801	Model fitting
Validation	634	633	1,267	Checkpoint selection and hyperparameter checking
Test	634	633	1,267	Held-out final evaluation
Total	3,171	3,164	6,335	Complete FRN corpus used by the BigBird script; seed 42

Table C.2. Compact BigBird GPU final-run log for FRN.

Run-log field	Recorded value
Dataset	FRN / fake_or_real_news.csv
Model checkpoint	google/bigbird-roberta-base
Framework	Hugging Face Transformers with PyTorch
Input policy	Concatenated title and article body; max_length = 1,024 tokens
Split policy	Stratified 60/20/20 train-validation-test split after duplicate control
Seed / positive class	Seed 42; FAKE = positive class
Test set	1,267 articles: 633 fake and 634 real
Fake-class precision	0.9714
Fake-class recall	0.9668
Fake-class F1	0.9691
Macro-F1	0.9692
Balanced accuracy	0.9692
ROC-AUC	0.9946
Confusion matrix labels	Rows: true REAL, true FAKE; columns: predicted REAL, predicted FAKE
Confusion matrix	[[616, 18], [21, 612]]

Appendix C.2. FRN BERT/RoBERTa run logs

Table C.3. Compact BERT and RoBERTa GPU final-run performance log for FRN.

Model	Policy	Checkpoint	Fake P	Fake R	Fake F1	Macro-F1	Bal. acc.	ROC-AUC
BERT-base	Fixed truncation 512	bert-base-uncased	0.9395	0.9321	0.9358	0.9361	0.9361	0.9800
RoBERTa-base	Fixed truncation 512	roberta-base	0.9492	0.9447	0.9470	0.9471	0.9471	0.9857
BERT-base	Sliding window 512/256	bert-base-uncased	0.9558	0.9558	0.9558	0.9558	0.9558	0.9894
RoBERTa-base	Sliding window 512/256	roberta-base	0.9667	0.9621	0.9644	0.9645	0.9645	0.9938

Table C.4. BERT and RoBERTa confusion-matrix counts and ROC-AUC values for FRN.

Model	Policy	True REAL -> REAL	True REAL -> FAKE	True FAKE -> REAL	True FAKE -> FAKE	ROC-AUC
BERT-base	Fixed truncation 512	596	38	43	590	0.9800
RoBERTa-base	Fixed truncation 512	602	32	35	598	0.9857
BERT-base	Sliding window 512/256	606	28	28	605	0.9894
RoBERTa-base	Sliding window 512/256	613	21	24	609	0.9938

Appendix C.3. ISOT run logs

Table C.5. Compact ISOT BERT/RoBERTa title-body final-run log.

Run-log field	Recorded value
Dataset	ISOT Fake News Dataset
Input setting	Title + article body, tokenized with the selected Transformer tokenizer
Split and positive class	80/20 stratified split; FAKE = positive class
BERT final metrics	fake-F1 = 0.9935; macro-F1 = 0.9932; balanced accuracy = 0.9932; ROC-AUC = 0.99995
RoBERTa final metrics	fake-F1 = 0.9948; macro-F1 = 0.9945; balanced accuracy = 0.9946; ROC-AUC = 0.99997
Confusion matrices	BERT: [[4255, 28], [33, 4662]]; RoBERTa: [[4261, 22], [27, 4668]], ordered as [[true REAL, pred REAL/FAKE], [true FAKE, pred REAL/FAKE]].
Saved output files	isot_roberta_bert_title_body_metrics_final.json; isot_roberta_bert_title_body_predictions_final.csv; isot_roberta_bert_title_body_confusion_matrices_final.csv; isot_roberta_bert_title_body_roc_curve_points_final.csv.

Table C.6. Author-generated ISOT comparison classical TF-IDF baselines.

Model	Input / implementation setting	Fake precision	Fake recall	Fake F1	Macro-F1	Balanced accuracy	ROC-AUC
Logistic Regression + TF-IDF	Classical baseline; title + body, TF-IDF	0.9910	0.9894	0.9902	0.9897	0.9898	0.9994
Linear SVM + TF-IDF	Classical baseline; title + body, TF-IDF	0.9925	0.9923	0.9924	0.9921	0.9921	0.9996
Multinomial NB + TF-IDF	Classical baseline; title + body, TF-IDF	0.9388	0.9482	0.9435	0.9405	0.9403	0.9833

Appendix C.4. CoAID run logs

Table C.7. Compact CoAID detection run-log summary.

Run-log item	Recorded value
Dataset version	CoAID repository version 0.4; text-only detection experiment derived from news, post and claim files [97]
Positive class	FAKE = 1; REAL = 0
Social posts / claims split	584 training records and 146 test records; test set = 52 fake and 94 real

Longer-news split	3,346 training records and 837 test records; test set = 19 fake and 818 real
Best social-post/claim Transformer result	RoBERTa-base: fake F1 = 0.9615, macro-F1 = 0.9701, balanced accuracy = 0.9701, ROC-AUC = 0.9915
Best longer-news Transformer result	BigBird-RoBERTa-base: fake F1 = 0.8205, macro-F1 = 0.9081, balanced accuracy = 0.9186, ROC-AUC = 0.9786
Saved output files	coaid_detection_metrics.json; coaid_detection_comparison_thesis_ready.csv; coaid_detection_confusion_matrices.csv; CoAID prediction files for classical and Transformer settings

Appendix C.5. Albanian output files

Table C.8. Output files associated with the Albanian detection outputs.

Output file	Content	Use in Chapter 3
albanian_split_summary_60_20_20.csv	Train, validation and test counts for fake and real labels	Supports Table 3.12
albanian_detection_generated_metrics.csv	Metric rows for Logistic Regression, Linear SVM, Multinomial NB, mBERT and XLM-RoBERTa	Supports Table 3.12
albanian_detection_generated_metrics.json	Structured metric JSON and split settings	Supports reproducibility of the reported metrics
albanian_xlmr_generated_predictions.csv	Item-level true labels, predicted labels and fake probabilities for the best model	Supports ROC and error analysis
albanian_xlmr_confusion_matrix.png	Best-model confusion matrix figure	Supports Figure 3.9
albanian_xlmr_roc_curve.png	Best-model ROC curve figure	Supports Figure 3.10

Appendix C.6. Constraint@AAAI2021 output files

Table C.9. Output files associated with the Constraint@AAAI2021 detection outputs.

Saved artefact	Contents	Use in Chapter 3
constraint_aaai2021_split_summary.csv	Official train-validation-test counts and fake/real label totals.	Supports Table 3.6 and verifies the benchmark split.
constraint_aaai2021_classical_metrics.json	LR, SVM, NB and NB + LR metric dictionary.	Supports the classical rows reported in Table 3.3.
constraint_aaai2021_bert_metrics.json	BERT fake-class precision, fake-class recall, fake-class F1, macro-F1, balanced accuracy and ROC-AUC.	Supports the BERT row reported in Table 3.3.
constraint_aaai2021_predictions.csv	Record-level true label, predicted label and fake-class probability.	Allows recalculation of all metrics and error analysis.
constraint_aaai2021_confusion_matrix.csv	Confusion-matrix counts with REAL=0 and FAKE=1.	Supports false-positive and false-negative interpretation.
constraint_aaai2021_roc_curve_points.csv	FPR, TPR and threshold values.	Supports ROC-AUC verification and optional ROC plotting.
constraint_aaai2021_run_log.txt	Seed, library versions, model checkpoint, tokenizer and execution notes.	Documents reproducibility conditions for the final run.
constraint_aaai2021_confusion_matrix_figure.png	BERT confusion-matrix figure generated from the held-out prediction log.	Supports Figure 3.1.
constraint_aaai2021_roc_curve_figure.png	BERT ROC-curve figure generated from fake-class probability scores.	Supports Figure 3.1.

Appendix D. Supplementary Cascade, Posterior and Intervention Details

Appendix D retains complete cascade statistics, full intervention perturbation tables, posterior details and complete calibration metrics used to support Chapter 3. Detailed scenario values remain here rather than in the main results chapter. These tables are intentionally labelled as supplementary audit material. They extend the Chapter 2 methodology with dataset-specific schemas, quality-control checks and parameter-audit details, rather than restating the core methodological tables.

Tables D.1-D.4 confirm the suitability of the retained cascades for temporal diffusion modelling and separate principal empirical cases from supplementary validation cases.

Table D.1. Filtered real Twitter/X COVID-19 temporal-cascade summary for site_type = covid_fake_claim.

Feature	Observed value	How it was extracted	Interpretation for Chapter 3
Claim title	"The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die."	title column / fake-news claim text	Defines the filtered COVID-19 fake-news claim cascade.
Dataset status	Real Twitter/X temporal-network dataset filtered to site_type = covid_fake_claim	covid_real_mapped_from_user_to_user.csv	Moves the item from content screening to empirical temporal-network analysis.
Site type	covid_fake_claim only; covid_true_claim rows excluded	site_type filter	Focuses the temporal analysis on fake-news claim spreading.
Observed events	133,374 records; 7,731 unique tweet IDs	row count; tweet_id	Each retained record is one observed claim-spreading interaction.
User nodes	95,599 nodes: 6,088 source users and 95,599 active spreading users	from_user_id and to_user_id	Defines the directed temporal-network user set.
Directed edges	125,057 unique directed edges	from_user_id -> to_user_id	Edges represent information flow from source to spreader.
Observation window	2020-02-02 10:52:45 - 2020-12-31 08:48:41 UTC (332.91 days)	min/max tweet_created_at	Observation period for daily I(t).
Daily activity pattern	310 active days; peak daily $I_{\text{new}}(t)$ = 5,892 on 30 July 2020	daily first-time to_user_id aggregation	Identifies the strongest daily burst of fake-news claim spreading.
Final infected curve	$I_{\text{cum}}(t)$ = 95,599 unique active spreaders	cumulative first-time to_user_id	Empirical infected-user curve for SIR/SEIZ/ABC.
Largest component	1 weak component; largest = 95,599 nodes and 125,057 edges	NetworkX weak components	The filtered fake-news claim network is connected enough for temporal diffusion analysis.

Table D.2. Real Hoaxy / OSoMeNet temporal-cascade summary for the claim "Three million votes in the Presidential Election were cast by illegal aliens."

Feature	Observed value	How it was extracted	Interpretation for Chapter 3

Claim title	"Three million votes in the Presidential Election were cast by illegal aliens."	title field and claim-filtered Hoaxy / OSoMeNet file; blank title entries standardized to the selected fake-news claim	Defines the empirical fake-news claim cascade used for temporal-network and later SIR/SEIZ/ABC modelling.
Dataset status	Real Hoaxy / OSoMeNet Twitter/X temporal-network dataset supplied for this thesis case	uploaded CSV file aliens_retweet.postelection.claim.csv	Moves the claim from Transformer detection into empirical temporal-network analysis.
Site type	claim-only; blank site_type rows standardized to claim after claim-filter validation	site_type column and file scope	Supports claim-cascade analysis; it does not provide a direct claim-versus-fact-checking comparison.
Observed events	1,048,575 timestamped retweet records; 39,864 unique tweet IDs	row count and tweet_id uniqueness	Each row is treated as one observed retweet/spreading interaction.
User nodes	156,556 unique nodes: 22,761 source users and 149,079 retweeting users	union of from_user_id and to_user_id	Defines the user population observed in the directed temporal network.
Directed edges	523,231 unique directed edges	unique from_user_id -> to_user_id pairs	Edges represent information flow from the retweeted/source user to the retweeting user.
Observation window	2016-05-16 05:54 to 2016-09-26 14:58; duration = 133.38 days	minimum and maximum tweet_created_at	Supports daily temporal ordering and time-to-peak analysis.
Peak activity	2016-09-08 with 27,375 events; maximum daily new spreaders = 4,660; peak hour 2016-09-08 15:00 with 761 new spreaders	daily event aggregation and first appearance of to_user_id	Identifies the strongest fake-news diffusion burst in the observed period.
Final infected curve	$I_{\mathrm{cum}}(t)$ = 149,079 unique active spreaders	cumulative first-time to_user_id count	Provides the empirical infected-user curve for later SIR/SEIZ/ABC fitting.
Largest component	150,787 nodes and 519,392 edges; 96.3% of nodes and 99.3% of edges	largest weakly connected component in NetworkX	Indicates that most observed interactions belong to one dominant connected cascade.

Table D.3. Hoaxy-imported social fake-news claim cascade summary for "Should One Tweet End a Career?".

Feature	**Observed value**	**How it was extracted**	**Interpretation for Chapter 3**
Claim title	Should One Tweet End a Career?	title field and canonical URL grouping	Defines the cascade identifier for this external Hoaxy claim.
Source domain / site type	redstate.com; site_type = claim	domain, site_domain and site_type fields	The file contains claim records only; it is not a claim/fact-checking comparison.
Observation window	30 May 2018 18:34 UTC to 06 Sep 2018 22:29 UTC (99.2 days)	parsed tweet_created_at	Supports duration, time-to-peak and temporal ordering.
Observed events	817 records: 728 retweets, 87 quotes, 2 origin tweets	tweet_type counts	Separates active spreading and quote/origin events.
Unique users / edges	805 user nodes; 809 unique directed edges	union of from_user_id and to_user_id; unique source-target pairs	Defines the directed temporal network.
Largest weak component	774 nodes (96.1% of the graph)	NetworkX weak-component analysis	Shows that the cascade is mostly connected rather than fragmented.

Peak activity	18 Aug 2018: 653 events; peak hour 18 Aug 2018 14:00 UTC with 89 events	daily and hourly aggregation	Indicates a delayed burst after the initial article publication.
Time to peak	About 79.8 days after the first observed event	peak-hour timestamp minus first event timestamp	Provides a temporal summary for later SIR/SEIZ fitting.
Final $I_{\text{cum}}(t)$	797 unique active spreaders	first observed share by to_user_id	Used as the cumulative infected-user curve.
Dominant source pattern	One anonymized source node has 755 unique outgoing neighbours and 763 weighted outgoing events	out-degree and weighted out-degree	Suggests a broadcast-like cascade driven by a highly central source.

Table D.4. Synthetic Albanian augmented temporal-cascade summary for the claim “Shqipëria do të presë 100 mijë palestinezë nga Gaza.”

Feature	Observed / generated value	How it was extracted	Interpretation for Chapter 3
Dataset type	Synthetic / augmented Hoaxy-style temporal cascade	Generated from the Albanian claim and stored with the fields from_user_id, to_user_id, tweet_id, tweet_created_at, site_type and title.	Pipeline validation only; not empirical Twitter/X or Hoaxy evidence.
Claim title	“Shqipëria do të presë 100 mijë palestinezë nga Gaza.”	title column for records with site_type = claim.	External Albanian fake-news-risk claim after XLM-RoBERTa / mBERT detection.
Fact-checking title	“Pretendimi për pranimin e 100 mijë palestinezëve në Shqipëri u mohua.”	title column for records with site_type = fact_checking.	Provides a corrective/counter-claim curve for SEIZ-style interpretation.
Total events	720 records: 520 claim and 200 fact_checking.	Row count and site_type grouping.	Enough generated events for testing cascade extraction and I(t).
Nodes and edges	725 unique nodes and 720 directed edges.	Unique users from from_user_id/to_user_id and unique directed pairs.	Medium-size synthetic temporal network for pipeline testing.
Time range	2025-01-27 09:13:26 to 2025-02-02 07:14:30; duration = 142.02 hours.	Minimum and maximum tweet_created_at.	Supports hourly temporal aggregation.
Peak activity	Claim peak: 2025-01-27 12:00, $I_{\text{new}} = 39$.	Hourly aggregation by unique newly active users with site_type = claim.	Identifies the main fake-news claim burst used in Figure 3.15.
Final curves	I_cum_claim = 520.	Cumulative unique active users for site_type = claim.	Claim-only model-ready curve for SIR/SEIZ/ABC pipeline validation.

Tables D.5-D.6 provide the full intervention perturbation values supporting the concise scenario analysis in Section 3.7.

Table D.5. Detailed fact-checking and debunking parameter perturbations for the two empirical fake-news cascades.

Dataset	Scenario	Trigger	Parameter modification after trigger	Final $I_{\text{cum}}(t)$ / peak $I_{\text{new}}(t)$	Reduction vs baseline
COVID-19 real Twitter/X “The coronavirus is not only affecting the way we live, it’s also dramatically affecting the way we die.” fake-news claim	ABC-SEIZ baseline	No counterfactual change	Posterior-median SEIZ parameters from Section 3.6	84,963 / 680	Reference

COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die." fake-news claim	Early debunking - moderate	Day 7	p x 0.85; epsilon x 0.85; l x 1.10; delta x 1.25	77,475 / 435	8.8% lower
COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die." fake-news claim	Early debunking - strong	Day 7	p x 0.65; epsilon x 0.65; l x 1.25; delta x 1.50	64,695 / 343	23.9% lower
COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die." fake-news claim	Delayed strong debunking	Observed peak	p x 0.65; epsilon x 0.65; l x 1.25; delta x 1.50	81,429 / 680	4.2% lower
Hoaxy / OSoMeNet "Three million votes in the Presidential Election were cast by illegal aliens." fake-news claim	ABC-SEIZ baseline	No counterfactual change	Posterior-median SEIZ parameters from Section 3.6	143,837 / 2,124	Reference
Hoaxy / OSoMeNet "Three million votes in the Presidential Election were cast by illegal aliens." fake-news claim	Early debunking - moderate	Day 7	p x 0.85; epsilon x 0.85; l x 1.10; delta x 1.25	141,559 / 2,093	1.6% lower
Hoaxy / OSoMeNet "Three million votes in the Presidential Election were cast by illegal aliens." fake-news claim	Early debunking - strong	Day 7	p x 0.65; epsilon x 0.65; l x 1.25; delta x 1.50	135,071 / 1,926	6.1% lower
Hoaxy / OSoMeNet "Three million votes in the Presidential Election were cast by illegal aliens." fake-news claim	Delayed strong debunking	Observed peak	p x 0.65; epsilon x 0.65; l x 1.25; delta x 1.50	141,499 / 2,124	1.6% lower

Table D.6. Detailed visibility-reduction and early-mitigation parameter perturbations for the two empirical fake-news cascades.

Dataset	**Scenario**	**Trigger**	**Parameter modification after trigger**	**Final $I_{\text{cum}}(t)$ / peak $I_{\text{new}}(t)$**	**Reduction vs baseline**
COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die." fake-news claim	ABC-SEIZ baseline	No counterfactual change	No modification; all parameters fixed at the ABC-SEIZ posterior medians from Section 3.6.	84,963 / 680	Reference
COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically affecting the way we die." fake-news claim	Visibility reduction - moderate	Day 7	beta x 0.80; b x 0.80	72,903 / 489	14.2% lower
COVID-19 real Twitter/X "The coronavirus is not only affecting the way we live, it's also dramatically	Visibility reduction - strong	Day 7	beta x 0.60; b x 0.60	59,634 / 377	29.8% lower

affecting the way we die.” fake-news claim					
COVID-19 real Twitter/X “The coronavirus is not only affecting the way we live, it’s also dramatically affecting the way we die.” fake-news claim	Early combined mitigation	Day 3	beta x 0.60; b x 0.60; p x 0.90; epsilon x 0.90; l x 1.10	55,862 / 327	34.2% lower
COVID-19 real Twitter/X “The coronavirus is not only affecting the way we live, it’s also dramatically affecting the way we die.” fake-news claim	Delayed visibility reduction	Observed peak	beta x 0.60; b x 0.60	80,844 / 680	4.8% lower
Hoaxy / OSoMeNet “Three million votes in the Presidential Election were cast by illegal aliens.” fake-news claim	ABC-SEIZ baseline	No counterfactual change	No modification; all parameters fixed at the ABC-SEIZ posterior medians from Section 3.6.	143,837 / 2,124	Reference
Hoaxy / OSoMeNet “Three million votes in the Presidential Election were cast by illegal aliens.” fake-news claim	Visibility reduction - moderate	Day 7	beta x 0.80; b x 0.80	134,284 / 1,948	6.6% lower
Hoaxy / OSoMeNet “Three million votes in the Presidential Election were cast by illegal aliens.” fake-news claim	Visibility reduction - strong	Day 7	beta x 0.60; b x 0.60	124,913 / 1,720	13.2% lower
Hoaxy / OSoMeNet “Three million votes in the Presidential Election were cast by illegal aliens.” fake-news claim	Early combined mitigation	Day 3	beta x 0.60; b x 0.60; p x 0.90; epsilon x 0.90; l x 1.10	119,984 / 1,584	16.6% lower
Hoaxy / OSoMeNet “Three million votes in the Presidential Election were cast by illegal aliens.” fake-news claim	Delayed visibility reduction	Observed peak	beta x 0.60; b x 0.60	140,368 / 2,124	2.4% lower

Appendix D.1. Supplementary Diffusion-Oriented Preprocessing Tables

Table D.7. Dataset-specific preprocessing requirements for temporal cascade construction.

Dataset	Raw diffusion information	Required preprocessing	Model-ready output
Hoaxy / OSoMeNet	Directed retweet records with from_user_id, to_user_id, tweet_id, tweet_created_at and site_type.	Convert retweet records into directed temporal edges; preserve claim/fact_checking labels; sort events by timestamp.	Label-aware temporal retweet network for comparing misinformation and fact-checking diffusion.
Other candidates	Engagement records, hashtags, comments, reposts, replies or temporal social traces.	Check whether user interactions and timestamps are sufficient; define cascade membership and reliability of labels.	Auxiliary temporal records for robustness testing or external validation.

Table D.8. Supplementary diffusion-record schema retained for reproducibility audit.

Field	Description	Example
cascade_id	Identifier of the news item, rumour thread, source tweet, URL or claim cascade.	covid_fake_claim_001; aliens_claim_001; career_claim_001; alb_claim_001

source_node	User, tweet or parent node from which the observed interaction originates.	retweeted user; parent tweet; source tweet
target_node	User, tweet or child node that receives, retweets, replies to or reacts to the information.	retweeting user; child tweet; reaction tweet
event_time	Absolute timestamp or normalized relative time from the beginning of the cascade.	2016-05-16 05:54; 12 minutes after source
event_type	Observed interaction type or diffusion event.	retweet; reply; quote; reaction; fact_checking
content_label	Credibility or rumour label of the associated item, preserved from the source dataset.	fake; real; claim; fact_checking; false rumour; unverified
diffusion_role	Derived modelling role used cautiously for epidemic-state mapping.	spreading; corrective; questioning; skeptical-like; exposed-like
quality_flag	Indicator of missing values, duplicate status, hydration status or temporal consistency.	valid; duplicated; missing text; unavailable tweet

Table D.8. Quality-control decisions for diffusion-oriented preprocessing.

Issue	**Recommended decision**	**Reason**
Duplicate interaction records	Remove exact duplicates after preserving one verified copy.	Prevents artificial inflation of cascade size and retweet volume.
Missing timestamp	Exclude from temporal modelling or retain only for non-temporal descriptive statistics.	Temporal ordering is essential for cascade construction and time-respecting paths.
Missing user identifier	Exclude from edge-level modelling if source or target cannot be defined.	Network edges cannot be constructed without source and destination information.
Unavailable tweet after hydration	Mark as unavailable and retain tweet_id only if allowed by dataset policy.	Maintains reproducibility while respecting platform access limitations.
Conflicting label or event type	Resolve using documented dataset rules or exclude from model-fitting subsets.	Prevents ambiguous records from contaminating SIR, SEIZ or ABC summaries.
Out-of-order event time	Sort by timestamp and check whether event time precedes the cascade seed.	Detects impossible or inconsistent temporal paths.
Artificial or augmented text linked to cascades	Use only in training experiments, not as observed diffusion evidence.	Prevents synthetic text from being mistaken for real temporal social behaviour.

Appendix D.2. Supplementary Temporal Cascade Construction and Network-Property Tables

Table D.9. Dataset-specific sources of time-stamped interactions and claim-level cascade records.

Dataset	**Relevant fields or files**	**Cascade construction interpretation**
Hoaxy / OSoMeNet	from_user_id, to_user_id, tweet_id, tweet_created_at, site_type	Directed retweet edge with time and claim/fact-checking label.

Table D.11. Supplementary temporal-network element details retained for reproducibility audit.

Element	**Definition in this thesis**	**Example**	**Use in later modelling**
Weight	A value measuring interaction strength, frequency, or relevance.	Number of retweets between two users; interaction-type score.	Supports weighted diffusion modelling and user influence analysis.

Table D.10. Quality-control checklist for nodes, edges, weights, and temporal structure.

Quality issue	**Recommended action**	**Reason for evaluation quality**

Inconsistent timestamps	Parse to a common datetime format; compute delay from cascade root when needed.	Prevents invalid temporal ordering and incorrect time-to-peak values.
Duplicate records	Remove exact duplicates or convert repeated interactions into frequency weights.	Avoids artificial inflation of cascade size and edge strength.
Missing user identifiers	Drop unusable events or assign anonymous placeholders only when methodologically justified.	Protects network validity and reproducibility.
Ambiguous edge direction	Document direction as information flow or response flow before analysis.	Ensures correct interpretation of centrality, depth, and diffusion paths.
Mixed interaction types	Separate layers or encode action type as a feature.	Preserves differences between retweets, replies, quotes, and corrections.
Unavailable hydrated content	Keep identifier-level analysis and document hydration status.	Maintains reproducibility under platform limitations.
Train-test leakage	Split by cascade or article rather than by individual interaction when evaluating prediction.	Prevents the same cascade appearing in both training and testing.

Table D.11. Network properties used for fake-news diffusion analysis.

Property	**Definition**	**Interpretation for fake-news diffusion**	**Example use**
In-degree	Number of incoming interactions received by a user.	Indicates how often a user receives reposts, replies, mentions, or exposure signals.	A fact-checker receiving many replies may function as a correction or discussion hub.
Out-degree	Number of outgoing interactions produced by a user.	Indicates how actively a user spreads or forwards content to others.	A user with many retweets or reposts may represent an active spreader.
Weighted degree	Degree adjusted by interaction frequency or interaction importance.	Captures repeated amplification more accurately than unweighted degree.	Repeated retweets from the same source can be represented by a higher edge weight.
Betweenness centrality	Extent to which a node lies on paths between other nodes.	Identifies bridge users who connect communities and may enable cross-community diffusion.	A user linking political communities can increase the reach of misinformation.
PageRank / eigenvector influence	Node importance based on the importance of connected neighbours.	Captures influence beyond raw activity counts.	A user followed or retweeted by other influential users may receive a high influence score.
Clustering coefficient	Degree to which neighbours of a node are connected with each other.	Indicates local reinforcement and repeated exposure inside dense groups.	High clustering may support echo-chamber-style circulation.
Community membership	Assignment of nodes to groups or modules.	Helps identify whether fake news remains local or spreads between communities.	A claim moving from one political group to another represents cross-community spread.
Cascade breadth and depth	Breadth is maximum activity at one diffusion level; depth is longest propagation path.	Describes whether diffusion is wide, deep, or both.	A shallow but wide cascade may indicate rapid broadcasting; a deep cascade may indicate multi-step diffusion.

Table D.12. Dataset-specific network properties for diffusion analysis.

Dataset	**Useful network properties**	**Influence interpretation**	**Use in later modelling**
Hoaxy / OSoMeNet	Retweet in-degree, out-degree, edge weights, claim/fact-checking community spread.	Separates misinformation spreaders from fact-checking/correction actors.	Useful for SEIZ skepticism/debunking proxies and intervention analysis.

Table D.13. Quality-control considerations for network-property extraction.

Issue	Risk	Recommended handling
Bot-like amplification	May inflate degree and cascade size.	Report both raw and cleaned metrics where bot indicators are available.
Repeated interactions	May overstate influence if all repetitions are treated equally.	Use edge weights but also report unweighted degree.
Missing timestamps	May distort temporal centrality and time-to-peak analysis.	Exclude invalid events or impute only when justified.
Community detection instability	Different algorithms may produce different partitions.	Use one primary method and report the method consistently.
Source-user uncertainty	The first observed user may not be the true origin.	Interpret source nodes as observed seeds rather than absolute origins.

Appendix D.3. Supplementary ABC/ABC-SMC Parameter and Summary-Statistic Tables

Table D.16. Supplementary ABC prior ranges retained for posterior-sensitivity audit.

Parameter	Model	Prior example	Interpretation in fake-news propagation
beta	SIR/SEIZ	Uniform(0.01, 0.80)	Baseline rate at which active spreaders influence susceptible users.
gamma	SIR/SEIZ	Uniform(0.01, 0.50)	Rate at which spreaders stop sharing, forget the content, or lose interest.
epsilon	SEIZ	Uniform(0.01, 0.50)	Rate at which exposed users become active spreaders.
p	SEIZ	Beta(2, 2)	Probability of immediate sharing after contact with an infected user.
l	SEIZ	Beta(2, 2)	Probability of becoming skeptical after contact with a skeptic.
delta	SEIZ-modified	Uniform(0.00, 0.30)	Correction or debunking rate caused by fact-checking, warnings, or counter-information.
rho	SEIZ	Uniform(0.00, 0.60)	Contact coefficient between exposed users and infected users.

Table D.14. Observed cascade data and their role in ABC-based epidemic inference.

Observed cascade information	Hidden diffusion meaning	Use in ABC inference
Post or claim text	Credibility risk of the information item.	Used by a Transformer model to estimate P(fake|x), which can influence the initial infection seed or transmission risk.
Reposts, retweets, shares	Visible evidence of active spreading behaviour.	Used to reconstruct the infected curve I(t) and final cascade size.
Replies and comments	Reaction, disagreement, correction, or amplification.	Can support estimation of exposed, skeptical, or recovered states.
Timestamps of interactions	Temporal order and speed of propagation.	Used to calculate time to peak, growth rate, inter-event delays, and cascade lifetime.
Fact-checking links or debunking replies	Correction or skepticism process.	Used to estimate debunking or correction parameters, such as δ.
User centrality and network edges	Exposure opportunity and influence structure.	Used to simulate spread on real, synthetic, or hybrid temporal networks.

Table D.15. Candidate summary statistics for ABC inference on fake-news cascades.

Category	Summary statistic	Interpretation for fake-news cascade
Size	C_T	Final number of users, posts, or interactions involved in the cascade.
Speed	$r_{\max}$	Maximum spreading rate within a time window.

Timing	t_{peak}	Time required to reach the maximum spreading rate.
Structure	d_c	Cascade depth, or the longest time-respecting diffusion path.
Structure	b_c	Cascade breadth, or the largest number of active users at the same diffusion level.
Correction	τ_f	Delay between first misinformation spread and first visible fact-checking signal.
Correction	δ estimate	Rate at which infected users become corrected, skeptical, or inactive.
Transformer score	mean P(fake|x)	Average fake-news probability assigned by the detection model to cascade content.
Network influence	centrality of early spreaders	Influence of initial users or communities in the early cascade.

Table D.16. Intervention interpretation through ABC-calibrated epidemic parameters.

Intervention scenario	**Parameter effect**	**Expected cascade effect**
Warning label	Reduces β or p.	Fewer susceptible users become spreaders.
Early fact-checking	Increases δ and reduces τ_f.	Faster movement from infected to skeptical or corrected states.
Visibility reduction	Reduces β through lower exposure.	Slower growth rate and smaller final cascade size.
Influencer-based correction	Increases b and l through skeptical contact.	More users move toward Z rather than I.
Early detection by Transformer	Triggers ABC calibration earlier.	Improves forecast lead time and intervention timing.

Appendix D.4. Supplementary ABC-SMC Posterior Prediction Boundary Table

Table D.17. Boundary between ABC-SMC inference and posterior predictive application.

Aspect	**Section 2.7.3**	**Section 2.7.4**
Main purpose	Explains ABC-SMC as a methodological tool for posterior estimation and comparison of SIR and SEIZ.	Explains how ABC is applied to observed fake-news cascades and what information is extracted from them.
Main question	Which model and parameter values best explain the data?	Which cascade features should be matched, and how do posterior samples support prediction?
Avoided duplication	Does not repeat dataset-specific cascade construction.	Does not re-explain the full ABC-SMC algorithm or model-selection details.
Thesis role	Provides the inference method.	Connects the inference method to fake-news diffusion data and intervention analysis.

Appendix E. Supplementary Literature Mapping and Research Context

Appendix E provides a supplementary literature mapping and research-context bridge for the dissertation. It links fake-news detection, temporal-network diffusion, epidemic inference, ABC/ABC-SMC uncertainty estimation and the proposed Transformer-ABC-Epidemic framework. The appendix is retained as contextual support for Chapter 1 and Chapter 2, while mathematical derivations and implementation-oriented details are consolidated in Appendix G.

Table E.1. Supplementary literature mapping and research context for the proposed Transformer-ABC-Epidemic framework.

Literature direction / research area	Representative authors and reference index	Main contribution in the literature	Relevance to the proposed thesis
Fake-news definitions, theories and annotation	Zhou and Zafarani [2]; Allcott and Gentzkow [4]; Shu et al. [3]; Rubin et al. [111]; Harris et al. [113]	Defines fake news, disinformation, misinformation, satire, doubtful news, annotation ambiguity and benchmark constraints.	Establishes the conceptual basis for binary fake/real modelling while preserving discussion of doubtful and ambiguous cases.
Social-media fake-news impact and societal relevance	Allcott and Gentzkow [4]; Vosoughi, Roy and Aral [53]; Lazer et al. [78]	Shows that false information has political, social and epistemic consequences and may diffuse differently from true information.	Justifies studying fake news as both a content-classification problem and a social diffusion-risk problem.
Data-mining and social-context views of fake news	Shu et al. [3]; Shu, Bhattacharjee et al. [114]	Frames fake-news detection through text, source, user engagement, social context and platform traces.	Supports the thesis decision to connect Transformer probability scores with temporal-network and cascade-level information.
Early fake-news detection from propagation paths and social context	Liu and Wu [175]; Shu, Wang and Liu [176]; Ruchansky, Seo and Liu [171]; Popat et al. [172]	Shows that fake-news detection can be strengthened by propagation paths, temporal/social context, user behaviour and credibility-oriented explanation, rather than relying only on textual content.	Strengthens the thesis argument that fake-news risk should be studied as a content-to-cascade process. These works support the transition from Transformer-based content detection to temporal cascade modelling, while justifying why social context and propagation are treated as diffusion evidence rather than as ordinary text-classification features.
Classical machine-learning baselines	Breiman [43]; Joachims [42]; Chen and Guestrin [44]; Agarwal et al. [45]; Zhang et al. [46]	Introduces interpretable classifiers and strong sparse-feature baselines such as Random Forest, SVM and XGBoost.	Provides historical and comparative background for explaining why Transformer-based contextual representations are required.
Explainable and interpretable machine learning	Ribeiro et al. [115]; Lundberg and Lee [116]; Doshi-Velez and Kim [117]	Develops explanation frameworks such as LIME, SHAP and principles for interpretable machine learning.	Supports the discussion of transparency, feature importance and accountability in fake-news detection systems.
Transformer architecture and attention	Vaswani et al. [7]	Introduces self-attention and the Transformer architecture as a foundation for modern NLP.	Provides the theoretical basis for the Transformer-based content-detection layer of the dissertation.
BERT, RoBERTa and domain-adapted Transformers	Devlin et al. [6]; Liu et al. [8]; Gururangan et al. [51]; He et al. [118]	Introduces contextual encoders, optimized pretraining, domain adaptation and disentangled attention mechanisms.	Supports fine-tuning individual Transformer models for fake-news classification and domain-sensitive probability scoring.
Multilingual and sentence-level Transformer models	Pires et al. [9]; Reimers and Gurevych [10]; Conneau et	Develops multilingual representations, semantic	Supports Albanian and multilingual fake-news detection, semantic

	al. [11]; Wolf et al. [119]; Rust et al. [120]	sentence embeddings and practical Transformer tooling.	similarity, claim comparison and low-resource evaluation.
Long-sequence Transformer modelling	Zaheer et al. [12]	Introduces BigBird sparse attention for processing longer sequences more efficiently than full self-attention.	Supports long-form fake-news article analysis and motivates BigBird as an option for long textual inputs.
Graph neural networks and social-context detection	Kipf and Welling [121]; Velickovic et al. [122]; Hamilton et al. [123]; Wu et al. [124]; Monti et al. [125]; Bian et al. [126]; Nguyen et al. [127]	Introduces graph convolution, graph attention, inductive graph representation and graph-based fake-news/rumor detection.	Positions graph learning as an important related direction and future extension, while the thesis focuses on temporal-network and epidemic modelling.
Temporal networks and dynamic graphs	Holme and Saramäki [52]; Kazemi et al. [128]; Rossi et al. [129]	Reviews time-stamped interactions, dynamic graph representation and temporal graph neural architectures.	Justifies modelling fake-news diffusion through time-respecting paths rather than only aggregated static graphs.
Higher-order and memory-aware temporal diffusion	Scholtes [13]; Scholtes et al. [14]; Rosvall et al. [15]; Benson et al. [66]; Lambiotte et al. [38]	Shows that chronological order and path memory can alter diffusion speed, community detection and network flow.	Supports methodological background for possible path-memory extensions in fake-news diffusion analysis.
Community structure, echo chambers and polarization	Sunstein [56]; Pariser [57]; Del Vicario et al. [16]; Cinelli et al. [17]; Bakshy et al. [18]	Explains how selective exposure, echo chambers and polarized communities shape information consumption and spread.	Motivates the inclusion of community structure, homophily and cross-community spread in fake-news diffusion analysis.
Influence, centrality and platform-mediated spread	Freeman [58]; Page et al. [59]; Newman [60]; Blondel et al. [61]; Romero et al. [55]; Bakshy et al. [62]; Kwak et al. [63]; Weng et al. [64]	Provides centrality, PageRank, modularity, community detection and influence measures for social-network analysis.	Supports the operational extraction of network features such as influential users, cascade breadth, community spread and diffusion reach.
Classical epidemic models	Kermack and McKendrick [31]; Newman [70]; Pastor-Satorras and Vespignani [71]; Pastor-Satorras et al. [72]	Establishes compartmental and network-based epidemic theory, including SIR-style dynamics and epidemic thresholds.	Provides the mathematical foundation for modelling fake-news spread using susceptible, infected/spreading and recovered states.
Rumor-spreading models	Daley and Kendall [32]; Maki and Thompson [67]; Zanette [68]; Nekovee et al. [69]	Adapts epidemic logic to rumor dynamics, including ignorant, spreader and stifler-type states.	Justifies treating fake-news propagation as a social contagion process with hesitation, stopping and correction mechanisms.
SEIZ and misinformation-specific epidemic models	Jin et al. [33]; Maleki et al. [73]; Raponi et al. [35]	Applies exposed and skeptic states to news, rumor and misinformation diffusion and reviews epidemic fake-news propagation.	Supports the use of SEIZ as a richer model than SIR for fake-news cascades involving exposure, skepticism and debunking.
Advanced social contagion and memory effects	Centola [74]; Wang et al. [75]; Kauk et al. [36]; Murayama et al. [39]	Shows that social reinforcement, memory and platform behaviour affect contagion dynamics in online networks.	Motivates the use of temporal summaries, hidden exposure assumptions and memory-aware cascade statistics.
Empirical false-news and misinformation diffusion	Vosoughi, Roy and Aral [53]; Kucharski [34]; Shao et al. [90]; Shao et al. [54]; Bovet and Makse [37]	Provides empirical evidence on false-news spread, misinformation tracking and election-related diffusion networks.	Links content-level falsehood with observable cascades, fact-checking signals and temporal-network diffusion data.
Approximate Bayesian Computation foundations	Beaumont et al. [19]; Marjoram et al. [20]; Sisson	Develops likelihood-free Bayesian inference through	Supports parameter inference when fake-news

	et al. [21]; Sunnaker et al. [23]; Marin et al. [40]; Lintusaari et al. [77]	simulation, rejection sampling, sequential Monte Carlo and ABC foundations.	diffusion likelihoods are analytically intractable or partially observed.
ABC-SMC and stochastic epidemic inference	Toni et al. [22]; Kypraios et al. [24]; Minter and Retkute [25]	Introduces ABC-SMC for dynamical systems and reviews ABC for stochastic epidemic models.	Justifies sequential posterior estimation for SIR/SEIZ parameters such as transmission, recovery, exposure, skepticism and debunking rates.
Summary statistics and identifiability in ABC	Blum et al. [79]; Csillery et al. [106]; Beaumont [104]; Fearnhead and Prangle [80]	Studies summary-statistic selection, dimension reduction, practice-oriented ABC and identifiability issues.	Supports the use of cascade size, time to peak, depth, breadth, growth rate and correction delay as ABC summaries.
Bayesian uncertainty and probabilistic learning	Gal and Ghahramani [130]; Kendall and Gal [131]; Murphy [132]	Develops Bayesian views of uncertainty estimation, probabilistic learning and model uncertainty.	Strengthens the uncertainty-aware interpretation of posterior diffusion risk and calibrated probability outputs.
Simulation-based inference and network diffusion inference	Dutta, Mira and Onnela [76]; Cranmer et al. [26]; Papamakarios and Murray [27]; Gomez-Rodriguez et al. [28]; Farajtabar et al. [29]	Connects network spreading inference, simulator-based inference, diffusion networks and co-evolving temporal processes.	Provides methodological precedent for estimating hidden spreading parameters from partially observed online cascades.
Online cultural transmission with ABC	Carrignon et al. [102]	Uses ABC to compare simulated online transmission with observed Twitter data.	Supports the use of cascade-level summary statistics for calibration of online fake-news diffusion models.
Fake-news datasets and benchmark assessment	Patwa et al. [30]; Ahmed et al. [82]; Cui and Lee [83]; Canhasi et al. [84]; Rexha [85]; Harris et al. [113]	Documents the content-based datasets and benchmark constraints retained in the final thesis dataset layer.	Motivates the thesis distinction between content-based datasets for detection and diffusion-oriented fake-news claim cascades for temporal modelling.
Dataset documentation, ethics and reproducibility	Gebru et al. [91]; Bender and Friedman [92]; Paullada et al. [93]; Mitchell et al. [94]; Wilkinson et al. [133]; ACM [134]	Introduces datasheets, data statements, model cards, FAIR principles and artifact-review practices.	Supports responsible dataset reporting, reproducible preprocessing, and ethical treatment of user-level social-network traces.
Large language models and hallucination risk	Brown et al. [47]; OpenAI [48]; Ji et al. [49]; Touvron et al. [135]; Zhang et al. [136]	Reviews few-shot language models, generative LLMs, hallucination risks and LLM-assisted fake-news detection.	Positions LLMs as an emerging auxiliary direction for explanation and zero-shot analysis, not as a source of core Chapter 3 evidence.
Multimodal and user-context fake-news detection	Singhal et al. [137]; Khattar et al. [138]; Qi et al. [139]; Zhou et al.	Combines text, image, user preference, multimodal signals and COVID-19 credibility resources.	Identifies future extensions beyond the text-and-temporal-network focus of the current thesis.
Survey and systematic-review literature	Bondielli and Marcelloni [140]; Khan et al. [141]; Pierri et al. [142]; Harris et al. [113]	Synthesizes fake-news detection, rumor detection, misinformation/disinformation and benchmark limitations.	Provides high-level validation of the research gap and supports the integrated Transformer-ABC-Epidemic dissertation direction.
Thesis-specific integrated direction	Bozhiqi Vula and Guliashki [86]; Bozhiqi and Guliashki [87]; Bozhiqi and Guliashki [88]; Bozhiqi [89]; Bozhiqi Vula and Shtino [108]	Connects temporal-network fake-news epidemic modelling, BigSEIZ, SIR/SEIZ comparison and BigBird-assisted SEIZ-ABC analysis.	Directly aligns the author's research trajectory with the proposed integration of Transformer detection, epidemic modelling and ABC inference.

Appendix F. Extended Dataset Cards

Appendix F provides extended dataset cards that support dataset transparency and reproducibility. It includes raw examples, field interpretation, dataset schemas and sample records for the content-based and diffusion-oriented data used or documented in the thesis.

Appendix note. The table distinguishes datasets used for content-based Transformer detection from datasets suitable for temporal cascade reconstruction, SIR/SEIZ diffusion modelling and ABC/ABC-SMC posterior inference. This separation prevents content-only datasets from being interpreted as diffusion evidence.

Table F.1. Practical classification of dataset suitability under the binary-focused thesis design.

Suitability category	**Dataset characteristics**	**Typical examples in the thesis**	**Recommended use**
Primary binary detection datasets	Contain labelled text and a clear fake/real or reliable/unreliable target.	Constraint@AAAI2021, FRN, ISOT, CoAID, Albanian Fake News Corpus.	Main datasets for Chapter 3 binary Transformer detection and cross-dataset comparison.

The following supplementary tables provide the detailed content-dataset descriptions and representative examples that were moved from Section 2.3 to keep the main methodology concise.

Table F.2. Summary of Constraint@AAAI2021 dataset characteristics.

Characteristic	**Fake**	**Real**	**Combined / Comment**
Number of messages	5,600	5,100	10,700 total
Unique words	19,728	22,916	37,503 combined
Average words per post	22	32	27 combined
Average characters per post	143	218	183 combined
Primary label type	Fake	Real	Binary classification

Table F.3. Representative examples of fake and real news in the Constraint@AAAI2021 dataset.

Label	**Example content (shortened/paraphrased)**	**Main characteristics**
Fake	A social-media post claims that an Italian billionaire died by suicide after his family died from COVID-19.	Sensational and emotional wording; unverifiable personal tragedy; strong viral framing.
Fake	A post presents government lockdown advice as politically biased and deliberately confusing.	Political framing, sarcasm, and weak factual basis.
Fake	A message claims that millions of COVID-19 vaccine doses are already available after a sudden company announcement.	False medical claim, exaggerated certainty, and misleading public-health information.
Real	A verified public-health message reports that many COVID-19 vaccine candidates are in clinical and pre-clinical testing.	Evidence-oriented wording and institutional scientific context.
Real	An official update reports newly confirmed COVID-19 cases, discharged patients, and deaths.	Numerical reporting, public-health surveillance language, and verifiable update structure.

Table F.4. Representative examples of news from the FRN dataset.

Label	Example content (shortened/paraphrased)	Dataset relevance
Fake	An article claims that Brexit will push the United Kingdom toward stronger trade relations with Russia and other non-EU states.	Long political article with geopolitical framing and source credibility concerns.
Fake	A post presents the Clinton family as a “crime family” and frames the Clinton Foundation through highly accusatory language.	Highly polarised political framing and emotionally charged claim.
Fake	A sensational article claims that a secret alien base was found in the Moon’s Tycho crater.	Conspiracy-style narrative and unsupported extraordinary claim.
Real	An article discusses the continuation of a U.S. ground-force presence in Afghanistan and interprets the foreign-policy implications.	Long-form political analysis with conventional news structure.
Real	A news analysis discusses increasing political polarization in the United States and party divisions in Congress.	Formal political commentary and analytical framing.

Table F.5. ISOT Fake News Dataset characteristics.

Characteristic	Description	Relevance to thesis
Dataset type	Binary content-based fake-news dataset.	Supports fake/real Transformer classification.
Main domain	Politics and world news.	Provides a domain complementary to COVID-19 and Albanian datasets.
Approximate size	23,481 fake articles and 21,417 true articles in the commonly distributed version.	Large-scale benchmark for stable model comparison.
Typical fields	title, text, subject, date, label after preprocessing.	Suitable for BERT, RoBERTa, BigBird, and Transformer models.
Diffusion information	No direct social-network cascade information.	Requires external enrichment for SIR, SEIZ, or ABC analysis.

Table F.6. Representative ISOT data-format examples for binary fake-news detection.

Record type	Label	Example field content (shortened / schema-oriented)	Use in thesis
Fake.csv record	Fake	title: political headline from an unreliable source; text: article body; subject: politics/world news; date: publication date.	Training or testing a fake-news classifier on article-level false content.
True.csv record	Real	title: Reuters-style political or world-news headline; text: article body; subject: world/politics; date: publication date.	Reference class for binary comparison against fake articles.
Preprocessed record	Fake or real	title + text concatenated into one input sequence; label converted to 1 for fake and 0 for real.	Input format for Transformer and BigBird fine-tuning.
Long-text record	Fake or real	Full article text exceeding short-post length.	Useful for testing whether BigBird preserves article-level evidence better than truncated BERT input.

Table F.7. CoAID dataset characteristics.

Characteristic	Description	Relevance to thesis
Dataset type	COVID-19 healthcare misinformation dataset with news, social posts, and user engagements.	Supports binary detection and partial social-context analysis.

Reported paper version	4,251 news items, approximately 296,000 engagements, and 926 social-platform posts.	Provides a documented research baseline [83].
Repository version	5,216 news items, 296,752 engagements, and 958 social-platform posts.	Should be cited and documented if this version is used [97].
Typical fields	News title/content, source, date, label, social post information, and engagement records.	Useful for Transformer classification and partial engagement modelling.
Diffusion suitability	Moderate; depends on available engagement timestamps and relationship reconstruction.	Can support partial cascade features but may require additional processing.

Table F.8. Representative CoAID data-format examples for detection and partial diffusion analysis.

Dataset component	**Label / type**	**Example field content (shortened / schema-oriented)**	**Use in thesis**
News item	Fake / misinformation	title: COVID-19 health-related claim; content: article or post text; source: website or social platform; label: fake.	Input for Transformer-based healthcare misinformation detection.
News item	Real / credible	title: verified COVID-19 healthcare report; content: article text; source: credible public-health or news source; label: real.	Reference class for binary COVID-19 detection.
Social-platform post	Post related to news item	post text; platform identifier; associated news or claim; optional date or metadata.	Short-text classification and link between news content and social discussion.
User engagement record	Engagement with related content	user identifier or anonymized account reference; engagement type; related news/post; time if available.	Partial evidence for engagement-based diffusion analysis and cascade summarization.

Table F.9. Representative examples from the Albanian Fake News Corpus.

Label	**Example content (shortened from Albanian text)**	**Main signal for analysis**	**Thesis relevance**
Fake	“Vetëm duke përdorur sodën e bukës...” — a clickbait-style claim that baking soda can make skin appear much younger.	Sensational health/beauty claim, exaggerated promise, and promotional framing.	Useful for detecting Albanian clickbait and unsupported health-related claims.
Fake	A sensational article claims to explain effects during pregnancy and directs the reader through advertising before opening the full article.	Clickbait wording, weak evidence, and advertising-driven structure.	Useful for testing robustness against noisy low-resource web text.
True	“Studimi gjerman...” — a report on COVID-19 patients in German hospitals and mortality among ventilated cases.	Numerical reporting, institutional study reference, and public-health context.	Useful for multilingual health-news classification in Albanian.
True	“Qorraj: Projekti i Xhamisë së Madhe...” — a report quoting an engineer about the Great Mosque project in Prishtina.	Named source, detailed explanation, and local news reporting.	Useful for modelling authentic Albanian news discourse and named-entity patterns.

Appendix F.1. Raw Examples and Sample Records for Content-Based Datasets

The following tables preserve the detailed preprocessing, schema, sequence-policy and quality-control information moved from Section 2.4. They are retained here to keep the methodology chapter readable while maintaining full reproducibility of the preprocessing decisions.

Table F.10. Text cleaning and normalization operations used in the thesis.

Operation	**Recommended treatment**	**Reason for inclusion**

Duplicate removal	Remove exact duplicates; document near-duplicates.	Prevents inflated performance and repeated cascade counts.
Missing-text handling	Remove empty essential text; retain partial metadata only for diffusion use.	Protects classifier quality while preserving temporal evidence.
URL handling	Replace URLs with <URL> rather than deleting link evidence.	Preserves source and article-sharing signals.
User mentions	Replace usernames with <USER>; preserve IDs only for networks.	Reduces sparsity and protects privacy.
Hashtags	Preserve or segment meaningful hashtags; remove only redundant symbols.	Keeps topic, event, and campaign information.
Emojis and emoticons	Map common emojis to descriptive tokens where useful.	Preserves affective or sarcastic signals.
Casing	Lowercase only when a lexical preprocessing variant is used; preserve case for cased Transformers.	Balances sparsity reduction with model requirements.
Punctuation and whitespace	Normalize whitespace and excessive punctuation; preserve boundaries.	Improves tokenization and retains structure.
Non-standard spellings	Normalize obvious elongated or informal forms.	Reduces sparsity in noisy short texts [99], [112].
Language-specific characters	Preserve ë and ç; avoid forced ASCII conversion.	Maintains Albanian semantic and morphological information.

Table F.11. Dataset-specific considerations for text cleaning and normalization.

Dataset group	**Main text characteristics**	**Cleaning priority**
Constraint@AAAI2021	Short COVID-19 social-media posts with URLs, hashtags, and crisis-related vocabulary.	Normalize noise while preserving health-related named entities and claims.
FRN	Long political and general news articles with article-level structure.	Preserve paragraphs and context; avoid excessive truncation before BigBird input preparation.
ISOT	Long fake and real news articles with formal news style and political topics.	Remove metadata artifacts, duplicated article fragments, and source-format inconsistencies.
CoAID	COVID-19 news items, claims, social posts, and health-related content.	Preserve medical terms, institution names, and public-health references.
Albanian Fake News Corpus	Albanian news text with low-resource language characteristics.	Preserve Albanian characters, local names, and language-specific morphology.
Hoaxy / OSoMeNet	Retweet records with user IDs, tweet IDs, timestamps, and claim/fact-checking labels.	Do not remove identifiers or timestamps required for temporal-network analysis.

Table F.12. Quality-control checks after text cleaning.

Quality-control check	**Expected verification**	**Reason for high-quality evaluation**
Record count comparison	Compare number of records before and after cleaning.	Documents whether cleaning removed substantial data.
Class balance check	Recalculate fake/real distribution after cleaning.	Ensures that cleaning did not distort binary labels.
Text-length statistics	Report minimum, maximum, mean, and median text length.	Supports model selection between BERT-style models and BigBird.
Duplicate audit	Identify exact and near-duplicate records.	Reduces risk of train-test leakage.
Timestamp validation	Check format, missing values, and chronological order.	Preserves temporal evidence for cascade construction.
Identifier preservation	Verify user IDs, tweet IDs, and cascade IDs where applicable.	Maintains network structure for SIR/SEIZ and ABC modelling.
Augmentation separation	Keep generated or back-translated data out of validation/test partitions.	Prevents evaluation contamination and inflated performance.

Table F.13. Dataset-specific label harmonization strategy.

Dataset	Original label structure	Binary mapping used in thesis	Important precaution
Constraint@AAAI2021	fake / real COVID-19 posts	fake -> 1; real -> 0	Preserve noisy social-media markers when they may carry semantic value, but remove non-informative duplication.
FRN	fake / real long-form news articles	fake -> 1; real -> 0	Long articles should not be truncated before the chosen Transformer strategy is defined.
ISOT Fake News Dataset	Fake.csv and True.csv file-level separation	records from Fake.csv -> 1; records from True.csv -> 0	Remove duplicate or near-duplicate articles before splitting to avoid train-test leakage.
CoAID	COVID-19 fake/reliable news, social posts, and engagements	fake or misinformation-oriented records -> 1; real/reliable records -> 0	Separate content classification from engagement-based diffusion analysis.
Albanian Fake News Corpus	fake / true Albanian news folders or labels	fake -> 1; true -> 0	Maintain Albanian-specific text normalization and avoid aggressive preprocessing that removes linguistic information.
Hoaxy / OSoMeNet	claim and fact_checking retweet records	claim can be treated as misinformation-oriented; fact_checking as corrective/reliable in diffusion analysis	This mapping is used mainly for diffusion analysis, not as a direct supervised text label unless tweet/article content is retrieved.

Table F.14. Quality-control rules for binary label mapping.

Quality-control rule	Purpose	Application in this thesis
Preserve raw labels	Maintains transparency and allows later audit of mapping decisions.	Store both raw_label and binary_label.
Separate uncertain labels	Reduces label noise from ambiguous cases.	Unverified, doubtful, and partially false labels are excluded from main binary experiments or used only in sensitivity analysis.
Prevent train-test leakage	Protects validity of evaluation metrics.	Remove duplicates and split before augmentation where appropriate.
Keep original test data	Ensures realistic evaluation.	Use generated or back-translated data only for training augmentation.
Track language and domain	Supports cross-dataset and cross-language interpretation.	Record whether the item is English, Albanian, COVID-19, political, entertainment, or general news.
Avoid direct belief inference	Prevents overinterpretation of social-network labels.	Treat retweets and claims as observable diffusion signals, not confirmed psychological belief.

Table F.15. Recommended harmonized record schema after label mapping.

Field	Description	Use in later chapters
dataset_id	Name or code of the source dataset.	Supports dataset-wise model comparison.
item_id	Unique identifier of the article, claim, post, tweet, or cascade item.	Prevents duplication and supports traceability.
text	Cleaned textual content used for classification.	Input to Transformer and baseline classifiers.
raw_label	Original dataset label before mapping.	Audit trail for relabelling decisions.

binary_label	Canonical label: Fake = 1, Real = 0.	Target variable for binary classification experiments.
language	Language of the text, e.g., English or Albanian.	Supports multilingual and low-resource analysis.
domain	Topic domain such as COVID-19, politics, health, entertainment, or general news.	Supports cross-domain evaluation.
split	Train, validation, or test partition.	Ensures reproducible evaluation.
cascade_id / timestamp	Optional diffusion information for network datasets.	Used for temporal cascade construction and SIR/SEIZ/ABC modelling.

Table F.16. Recommended sequence policies for binary detection experiments.

Input type	**Preferred strategy**	**Example**	**Quality-control requirement**
Short post or tweet	Use complete text with padding to maximum sequence length.	Constraint@AAAI2021 social-media message.	Check that URLs, usernames, and hashtags are handled consistently.
Title plus body article	Concatenate title and body with separator token.	ISOT or FRN article.	Document maximum length and truncation/chunking policy.
Very long article	Use BigBird or split into overlapping windows; aggregate predictions where necessary.	Long FRN or ISOT article.	Ensure that all windows from one item remain in the same train/test split.
Multilingual / Albanian text	Use multilingual tokenizer and preserve original orthography.	Albanian Fake News Corpus article.	Avoid removing diacritics or named entities unless justified.

Table F.17. Main sources of noise in binary fake-news datasets and recommended handling procedures.

Noise source	**Typical manifestation**	**Risk for binary detection**	**Recommended handling**	**Example datasets**
Label noise	Incorrect fake/real labels, outdated labels, weakly verified labels, ambiguous claims.	The classifier learns contradictory supervision and may become poorly calibrated.	Check label source; flag uncertain records; exclude unresolved cases from core binary experiments when necessary.	All binary datasets

Table F.18. Recommended strategies for handling binary class imbalance.

Strategy type	**Method**	**When useful**	**Main risk**	**Recommended thesis use**
Data-level	Random oversampling of the minority class	Small or moderately imbalanced datasets.	May duplicate noisy examples and increase overfitting.	Use only on the training split; combine with validation monitoring.
Data-level	Random undersampling of the majority class	Very large majority class with redundant examples.	May discard useful examples and reduce domain coverage.	Use cautiously for large datasets such as ISOT if computation is limited.
Data-level	Controlled augmentation	Low-resource or minority-class shortage.	Generated examples may be unrealistic or label-inconsistent.	Use for training only; manually review generated examples.
Algorithm-level	Class-weighted loss	Moderate to high imbalance.	Can overemphasize noisy minority labels.	Preferred first-line strategy for Transformer models.
Algorithm-level	Focal loss	Many easy majority examples and hard minority cases.	Requires additional tuning and may be unstable on small datasets.	Use as a robustness experiment, not as the only training objective [101].
Decision-level	Threshold tuning	When probability calibration matters.	May optimize validation performance	Tune on validation data only and report the threshold.

			but fail under domain shift.	
Evaluation-level	Macro-F1, balanced accuracy, PR-AUC	All imbalanced binary settings.	Metrics may emphasize different trade-offs.	Report multiple metrics rather than accuracy alone.

Table F.19. Dataset-specific noise and imbalance considerations for the binary-focused thesis.

Dataset	**Likely noise issue**	**Likely imbalance issue**	**Recommended action**	**Evaluation note**
Constraint@AAAI2021	Short posts, platform-style language, hashtags, URLs, emojis.	Moderate fake/real difference may still appear in splits.	Use text normalization, stratified splitting, and macro-F1.	Report fake-class recall because missing fake health claims is costly.
FRN	Long article bodies, duplicate or source-specific phrasing.	Generally close to balanced but should be checked after filtering.	Deduplicate before splitting; use long-sequence policy.	Compare standard BERT truncation with BigBird or chunking.
ISOT	Publisher/source artifacts and possible topic-source correlations.	Large dataset but class balance must be checked after preprocessing.	Avoid source leakage; split carefully; report cross-domain results if possible.	Do not rely only on accuracy because source artifacts may inflate scores.
CoAID	Mixed formats: news, social posts, engagements, missing URLs.	May vary by news/social-post subset.	Separate news-level and social-post-level experiments.	Report subset-level results to avoid hiding imbalance.
Albanian Fake News Corpus	Low-resource language, local vocabulary, possible small sample size.	Potential class-size limitations.	Preserve Albanian orthography; use multilingual Transformers; consider training-only augmentation.	Report macro-F1 and confidence intervals if possible.
Hoaxy / OSoMeNet	Retweet-only visibility and claim/fact-checking source labels.	Claim and fact-checking records may be unequally distributed.	Treat primarily as diffusion data; analyze claim and fact-checking curves separately.	Do not interpret retweet absence as non-exposure.

Table F.20. Quality-control checklist for noisy and imbalanced binary fake-news data.

Check	**Required action**	**Reason**
Class distribution	Report fake and real counts for training, validation, and test splits.	Ensures imbalance is visible and not hidden by aggregate counts.
Duplicate control	Remove exact duplicates and control near-duplicates before splitting.	Prevents train-test leakage and inflated performance.
Label consistency	Audit ambiguous or contradictory labels and document mapping rules.	Reduces the impact of label noise on binary supervision.
Augmentation boundary	Apply augmentation only to training data.	Avoids contaminating validation or test evaluation.
Threshold selection	Tune decision threshold only on validation data.	Prevents test-set overfitting.
Metric reporting	Report macro-F1, fake-class recall, precision, and balanced accuracy.	Supports fair evaluation under imbalance.
Probability calibration	Check whether predicted probabilities are usable as risk signals.	Important for linking detection to diffusion modelling.
Documentation	Record all preprocessing and balancing decisions.	Improves reproducibility and thesis transparency.

Table F.21. Low-resource Albanian preprocessing challenges and recommended responses.

Challenge	**Risk for Transformer-based detection**	**Recommended preprocessing response**

Diacritics and character variants	Replacing or losing ë and ç may change tokenization and may reduce language-specific information.	Use Unicode normalization and preserve Albanian characters whenever possible. A no-diacritic auxiliary version may be stored only for analysis, not as a replacement for the original text.
Small labelled corpus	The model may overfit lexical patterns or source-specific wording.	Use stratified cross-validation, regularization, early stopping, and, if needed, training-only augmentation.
Local names and political entities	Removing named entities can remove important credibility cues.	Preserve person, party, institution, location, and media-source names.
Noisy web-style text	Clickbait punctuation, advertising fragments, URLs, and repeated symbols may dominate the input.	Remove URLs, duplicated punctuation, HTML fragments, and unrelated advertising text while retaining the main claim.
Negation and short function words	Removing words such as nuk, jo, or s' may reverse meaning.	Do not apply aggressive stop-word removal; preserve negation markers and modal expressions.
Long article bodies	Standard Transformer models may truncate important evidence.	Use title+lead+body strategies, sliding windows, or hierarchical pooling when articles exceed the model limit.
Generated or back-translated data	Synthetic variants may introduce label drift or leakage.	Use generated/paraphrased examples only in training and verify that the label meaning is preserved.

Table F.22. Multilingual Transformer preparation choices for Albanian fake-news detection.

Model / tokenizer	**Preparation strategy**	**Expected benefit**	**Precaution**
mBERT / WordPiece	Use conservative cleaning and preserve Albanian orthography before WordPiece tokenization.	Provides multilingual transfer and can be fine-tuned with limited Albanian labels.	May split rare Albanian words into many subwords; monitor sequence length and unknown-like fragmentation.
XLM-RoBERTa / SentencePiece	Use raw-text compatible normalization and avoid manual word segmentation.	Usually strong for cross-lingual transfer and low-resource languages.	Can still be sensitive to noisy spelling, malformed punctuation, and inconsistent casing.
BigBird-style long-input model	Use only if a suitable multilingual or translated long-document setup is available; otherwise use sliding windows with mBERT/XLM-R.	Useful for long Albanian articles if the model supports the language setting.	Do not assume English-pretrained long models automatically solve Albanian low-resource classification.
Back-translated augmentation	Generate paraphrases only from training examples and keep the original test set untouched.	Increases lexical diversity in a low-resource setting.	Manual or rule-based checks are needed to avoid label drift.

Table F.23. Quality-control checklist for Albanian preprocessing and augmentation.

Quality-control item	**Required action before model training**	**Purpose**
Original-text preservation	Store the raw Albanian text together with the cleaned version.	Allows auditability and prevents irreversible loss of meaning.
Unicode normalization	Use a consistent Unicode form and verify the rendering of ë and ç.	Reduces encoding inconsistencies without removing language-specific information.
Negation preservation	Do not remove Albanian negation words such as nuk, jo, or s'.	Prevents semantic reversal in fake/real classification.
Entity preservation	Keep names of people, institutions, locations, parties, and media sources.	Maintains source and context signals relevant to credibility.
Label consistency check	Review generated or paraphrased examples for semantic and label preservation.	Prevents artificial data from changing the truthfulness class.

Tokenizer diagnostics	Record average token length, truncation rate, and subword fragmentation.	Ensures that multilingual tokenization does not distort the dataset.
Reproducibility log	Document preprocessing rules, model tokenizer, maximum length, and split seed.	Supports reproducibility and high-quality thesis evaluation.

Appendix F.2. Field Interpretation, Evaluation Metrics and Model-Preparation Tables

Table F.24. Evaluation metrics recommended under noisy and imbalanced binary conditions.

Metric	What it measures	Why it is useful	Limitation
Accuracy	Overall proportion of correct predictions.	Simple and interpretable for balanced datasets.	Can be misleading under class imbalance.
Precision for fake class	How many predicted fake items are truly fake.	Useful when false accusations of fake news are costly.	Does not measure missed fake items.
Recall for fake class	How many true fake items are detected.	Important for health, election, or safety misinformation.	High recall may increase false positives.
F1-score	Harmonic mean of precision and recall.	Balances detection quality for the fake class.	May hide real-class performance if reported alone.
Macro-F1	Average F1 across fake and real classes.	More robust for imbalanced binary data.	Can be sensitive when one class has very few examples.
Balanced accuracy	Average recall across both classes.	Useful when class sizes differ strongly.	Does not show precision trade-offs.
PR-AUC	Precision-recall trade-off across thresholds.	Useful when fake items are rare or minority class matters.	Less intuitive than threshold-based metrics.
Calibration error	Agreement between predicted probabilities and observed labels.	Important because $p_f(x)$ is later used for diffusion-risk modelling.	Requires enough validation data and careful binning.

Appendix F.3. Dataset Schemas and Tokenization Implementation Details

Appendix F.3 retains compact supplementary schemas and quality-control checklists associated with Section 2.4. They are placed in the appendix to keep the main preprocessing methodology focused on the logic of the dissertation while preserving reproducibility material for implementation review.

Table F.17 documents the compact model-ready schema produced after tokenization. It supports reproducibility without overloading the main text of Section 2.4.

Table F.25. Model-ready record schema after tokenization.

Field	Description
record_id	Unique identifier of the cleaned textual record.
clean_text	Cleaned and normalized text used as tokenizer input.
raw_label	Original dataset label preserved for auditability.
$\boldsymbol{y}_{\mathbf{bin}}$	Harmonized binary fake/real target.
input_ids	Subword token identifiers produced by the selected tokenizer.
attention_mask	Mask marking non-padding tokens.
split	Training, validation or test split indicator.

Table F.18 summarizes the Transformer input checks that are applied after tokenization and before training. The checks are supplementary because they concern implementation control rather than the methodological argument.

Table F.26. Quality-control checklist for Transformer input preparation.

Check	Purpose

No empty tokenized records	Prevents invalid model inputs.
No split leakage	Ensures that duplicates or augmented texts do not cross train/test boundaries.
Label mapping verified	Checks that y_{bin} matches the documented mapping rule.
Tokenizer/model alignment	Confirms that the tokenizer belongs to the selected Transformer architecture.

Table F.17 records the quality-control steps for Albanian preprocessing and controlled augmentation. These steps support the low-resource detection experiments while preserving the original validation and test evidence.

Table F.27. Quality-control checklist for Albanian preprocessing and augmentation.

Check	Purpose
Albanian characters preserved	Protects language-specific information.
Named entities retained	Maintains local domain cues.
Augmentation restricted to training data	Prevents leakage and artificial test-set evidence.

Table F.18 keeps tokenizer-specific implementation details in the appendix, while Section 2.4 states only the model-level tokenization logic needed to understand the experimental pipeline.

Table F.28. Tokenization technical details retained as supplementary implementation material.

Detail	Role in implementation
token_type_ids	Used by selected BERT-style models when segment distinction is required.
Dynamic padding	Pads batches to the longest sequence in the batch to reduce unnecessary padding.
Special token handling	Adds architecture-specific markers such as classification and separator tokens.
[CLS] and [SEP]	BERT-style tokens used for classification input and sequence boundaries.
Truncation variants	Defines whether text is truncated from the end, split into windows or processed with long-sequence models.

Appendix F.4. Dataset Suitability, Ethics and Reproducibility Tables

Appendix F.4 preserves extended suitability and ethics tables that support the concise synthesis in Section 2.7. These tables are placed in the appendix so that the main methodology remains focused while the dataset-inclusion logic remains auditable.

Table F.17 supports the concise Section 2.7 discussion by documenting the final role assignment of each retained dataset or cascade family.

Table F.29. Supplementary benchmark dataset-role classification used to support Section 2.7.

Dataset / family	Dataset type	Main evidence available	Thesis role
Constraint@AAAI2021	Content-based benchmark	Short COVID-related fake/real text.	Primary short-text Transformer detection dataset.
FRN	Content-based benchmark	Long fake/real news articles.	Long-text detection and BigBird evaluation.
ISOT	Content-based benchmark	Large English fake/real news articles.	Binary classification and cross-domain robustness benchmark.
CoAID	Health misinformation dataset	COVID-19 healthcare misinformation and partial engagement evidence.	Health-related detection dataset and partial bridge to social context.
Albanian Fake News Corpus	Low-resource content dataset	Albanian fake/true news.	Multilingual and low-resource Transformer evaluation.

COVID-19 real Twitter/X fake-news claim	Diffusion-oriented cascade	Timestamped user-to-user fake-news claim interactions.	Principal empirical health misinformation cascade.
Hoaxy / OSoMeNet illegal-votes fake-news claim	Diffusion-oriented cascade	Large directed retweet cascade.	Principal empirical political fake-news cascade.
Hoaxy "Should One Tweet End a Career?"	Additional diffusion cascade	Claim-labelled social cascade.	Comparative temporal-network case.
Synthetic Albanian temporal cascade	Synthetic / validation cascade	Generated timestamped claim and fact-checking records.	Low-resource validation and simulation support.

Table F.18 expands Equation (2.31) and provides the qualitative checklist behind the content-based dataset selection process.

Table F.30. Detailed criteria for assessing Transformer-based fake-news detection suitability.

Criterion	High-suitability condition	Reason
Label reliability	Labels can be consistently mapped to fake/real.	Required for supervised binary detection.
Text availability	Each record contains usable text after preprocessing.	Required for tokenization and Transformer input.
Class coverage	Both fake and real classes are sufficiently represented.	Reduces unstable training and misleading accuracy.
Domain relevance	Dataset matches a thesis domain such as COVID-19, health, politics, general news or Albanian language.	Supports the empirical scope of Chapter 3.
Reproducibility	Splits, labels and preprocessing choices can be documented.	Supports author-generated experimental results.

Table F.17 expands Equation (2.32) and documents why diffusion datasets require more than labelled text.

Table F.31. Detailed criteria for assessing temporal diffusion, SIR/SEIZ and ABC/ABC-SMC suitability.

Criterion	High-suitability condition	Reason
Timestamp availability	Records contain event times or can be ordered reliably.	Required for $I_{\mathrm{new}}(t)$, $I_{\mathrm{cum}}(t)$ and time-to-peak analysis.
Directed interactions	Source-target or user-to-user edges are available.	Required for temporal-network reconstruction.
Claim-level scope	Events can be assigned to a specific fake-news claim.	Prevents mixing unrelated cascades.
State-proxy compatibility	Observable events can support active spreading, exposure or corrective proxies.	Supports SIR/SEIZ modelling without claiming to observe private beliefs.
ABC summaries	Cascade size, peak, breadth, depth or forecast quantities can be extracted.	Required for ABC/ABC-SMC distance and posterior predictive evaluation.

Table F.18 preserves the supplementary audit trail for the ethics and reproducibility decisions summarized in Section 2.7.3.

Table F.32. Supplementary ethics and reproducibility audit checklist for the retained data layer.

Audit item	Required documentation	Purpose
Dataset role	State whether the dataset supports detection, diffusion, validation or literature-only discussion.	Avoids methodological overreach.
User identifier handling	Use identifiers only as technical nodes and avoid user profiling.	Protects privacy and keeps the analysis aggregate.
Claim filtering	Document site_type and claim-level filtering rules.	Ensures reproducible cascade construction.

Temporal aggregation	Report the time window used for $I_{\text{new}}(t)$ and $I_{\text{cum}}(t)$.	Supports reproducible SIR/SEIZ and ABC inputs.
Synthetic data boundary	Identify synthetic or augmented data as support rather than empirical evidence.	Prevents overclaiming in validation experiments.

Appendix G. Supplementary Mathematical Derivations, Posterior Distributions and Credible Intervals

Appendix G consolidates supplementary mathematical derivations, full posterior-distribution records, credible-interval details and implementation-oriented notes for ABC/ABC-SMC inference.

Appendix G.1. Supplementary Theoretical Notes and Derivations for ABC-Based Fake-News Diffusion Methodology

Supplementary community-spread context. Political fake-news claims may circulate strongly inside ideologically aligned communities, while health-related fake-news claims may cross community boundaries through urgency, fear or public-health relevance. This literature-level context supports the use of community-spread indicators in Section 2.5, but it is not treated as an additional operational methodology in the main text [53], [55].

Appendix G.1 retains detailed theoretical notes moved from Section 2.6 so that the main methodology remains concise while preserving derivations and literature connections for audit and examination purposes.

Theorem 2.6.2 (ABC rejection target distribution). Assume that theta is sampled from the prior p(theta), simulated data D_{sim} are generated from the simulator p(D_{sim} | theta), and theta is accepted when the distance between simulated and observed summary statistics is below a tolerance epsilon. Then accepted samples are drawn from an approximate posterior distribution proportional to Equation (G.1).

$$p_{\text{ABC}}(\theta \mid D_{\text{obs}}) \propto p(\theta)\int \mathbf{1}\{\rho(S_{\text{sim}}, S_{\text{obs}}) \le \varepsilon\} p(D_{\text{sim}} \mid \theta)\, dD_{\text{sim}} \qquad \text{(G.1)}$$

Proof sketch. The ABC algorithm samples theta from p(theta), then samples D_{sim} from p(D_{sim} | theta). The joint density of accepted pairs is therefore the product p(theta)p(D_{sim} | theta) multiplied by an acceptance indicator. Marginalizing over all possible simulated datasets gives Equation (G.1). This identity is the basis of rejection ABC and its extensions [79], [76].

For fake-news cascades, the summary statistic vector can include temporal, structural, network, and content-derived variables. A suitable vector is shown in Equation (G.2).

$$S(D) = \left[C_T, t_{\text{peak}}, r_{\max}, d_c, b_c, \tau_{\text{mean}}, \bar{p}_f\right] \qquad \text{(G.2)}$$

Here C_T is final cascade size, t_{peak} is time to peak, $r_{\max}$ is maximum spreading rate, d_c is cascade depth, b_c is cascade breadth, τ_{mean} is mean inter-event time, and $\bar{p}_f$ is the mean Transformer fake-news probability across cascade items.

$$\rho(S_{\text{obs}}, S_{\text{sim}}) = \textstyle\sum_k w_k \left|S_k^{\text{obs}} - S_k^{\text{sim}}\right| \qquad \text{(G.3)}$$

Equation (G.3) defines a weighted distance function. The weights w_1..., w_5 control the importance of each feature. For example, if early warning is the main objective, t_{peak} and r_{max} may receive higher weights; if final cascade size is the main objective, C_T may receive higher weight.

Theorem 2.6.3 (Convergence under sufficient summaries). If S(D) is sufficient for theta and the ABC tolerance epsilon tends to zero, then the ABC posterior based on S(D) converges to the exact Bayesian posterior conditioned on the observed summaries. If S(D) is also sufficient for the full data D, this limiting distribution is equivalent to the exact posterior p(theta | D_{obs}).

Interpretation for this thesis. This theorem explains why the choice of summary statistics is central. If the selected cascade summaries do not capture the important features of diffusion, ABC will estimate parameters that match only superficial properties of the observed data. Therefore, the summaries used in this thesis must represent cascade size, timing, depth, growth rate, network structure, and Transformer-based content risk. The importance of summary-statistic selection and dimension reduction is emphasized in ABC methodological reviews [76], [86].

- *BigBird-Assisted SEIZ Modelling*

The BigSEIZ framework proposed by Vula and Guliashki is directly relevant because it explicitly connects a long-sequence Transformer model, BigBird, with SEIZ epidemic modelling for fake-news spread on Twitter temporal networks [83]. The methodological value of this work for the present thesis is that it treats fake-news analysis as a hybrid problem: semantic detection is performed at the content layer, while temporal network diffusion is analysed at the propagation layer.

In a thesis implementation, BigBird can process long news articles, claims, or threads and produce a fake-news probability score. This score can be denoted as:

$$p_f(x_i) = P(y_i = \text{fake} \mid x_i) \tag{G.4}$$

where x_i is the text of post or article i. This probability can then be incorporated into the transmission component of a SEIZ model. A content-aware transmission coefficient may be written as:

$$\beta_{ij}(t) = \beta_0\, w_{ij}(t) \left[1 + \lambda p_f(x_i)\right] \left[1 + \eta c_j\right] \tag{G.5}$$

where $\beta_{ij}(t)$ is the time-dependent transmission coefficient from user j to user i, β_0 is the baseline transmission rate, $w_{ij}(t)$ is the temporal interaction weight, $p_f(x_i)$ is the BigBird fake-news probability, c_j is the centrality or influence of the spreading user, and λ and η regulate the contribution of content risk and user influence. This formulation is thesis-specific, but it follows the methodological logic of connecting BigBird-based semantic representation with SEIZ-based temporal diffusion [83].

The IEEE TechDefense 2025 work "A Hybrid NLP-Epidemiological Model for Fake News Network: BigBird-Assisted SEIZ Modelling with Approximate Bayesian Computation" is directly related to the dissertation because it connects BigBird-assisted fake-news identification with SEIZ modelling and ABC inference [89]. The IEEE Women in Engineering special-session record also documents that this poster received a Best Poster award [188]. This source is used as evidence of dissertation-related publication activity, while the mathematical ABC procedure is developed independently and supported by the broader ABC literature [24], [25], [104], [107], [180].

- *ABC for Network Spreading Processes*

Dutta, Mira, and Onnela proposed a Bayesian inference framework for spreading processes on networks [76]. Their work is important for this thesis because it demonstrates that ABC can be applied when the spreading process occurs on a known network and the available observations are incomplete snapshots of node states. The target of inference is not only the spreading parameter but also the source node of the process. This is directly analogous to fake-news diffusion, where the researcher may observe which users shared a claim at several time points but may not observe all exposures or private interactions.

The corresponding fake-news interpretation is as follows: if a temporal Twitter/X network is represented as $G\_T = (V, E\_T)$, and each node is observed as susceptible, exposed, infected, skeptic, or recovered at selected time points, ABC can be used to estimate the parameters that most plausibly generated the observed cascade. The source node can also be inferred probabilistically, which is useful for identifying the earliest influential spreaders or suspicious seed accounts.

Dutta et al. also provide the ABCpy framework for high-performance Approximate Bayesian Computation [107]. For this dissertation, ABCpy-style modularity is relevant because preprocessing, simulation, distance computation and posterior sampling can be treated as separate auditable components. For fake-news dynamics, the analogous control problem is intervention planning: warning labels, visibility reduction, debunking messages or influencer-based correction can be represented by increasing delta or gamma, or by reducing beta.

- *ABC on Twitter Information Cascades*

Carrignon, Bentley, and Ruck used ABC on Twitter data documenting the spread of true and false information [102]. Their work is useful because it applies ABC to online information cascades rather than biological disease alone. They compare simulated cultural-transmission models with observed cascade distributions, showing how ABC can be used to evaluate model plausibility when the full likelihood of online transmission is difficult to calculate.

For the proposed thesis, this supports the use of cascade-level summary statistics such as final cascade size, retweet distribution, and temporal growth pattern. While Carrignon et al. focus on cultural-transmission models, the same simulation-comparison logic can be applied to SIR and SEIZ fake-news models when the goal is to estimate parameters from observed temporal network cascades.

Supplementary ABC-SMC Model-Comparison Equations

Model-comparison formulation: SIR versus SEIZ

Let the observed temporal fake-news cascade be denoted by D_{obs}. Two candidate models are considered:

$$\mathcal{M} = \{M_1 = \text{SIR}, M_2 = \text{SEIZ}\} \tag{G.6}$$

For the SIR model, the parameter vector can be written as:

$$\theta_{\mathrm{SIR}} = (\beta, \gamma, I_0) \tag{G.7}$$

where beta is the transmission rate, gamma is the recovery or forgetting rate, and I_0 is the initial number of active spreaders. For the SEIZ model, the parameter vector is richer:

$$\theta_{\mathrm{SEIZ}} = (\beta, b, p, l, \rho, \varepsilon, \delta, S_0, E_0, I_0, Z_0) \tag{G.8}$$

Here beta is the contact rate between susceptible users and spreaders, b is the contact rate between susceptible users and skeptics, p is the probability of immediate adoption, l is the probability of becoming skeptical, rho is the exposed-spreader interaction rate, epsilon is the transition rate from exposed to spreader, and delta is the debunking or correction rate. The initial state variables S_0, E_0, I_0, and Z_0 describe the starting configuration of the cascade.

Bayesian model comparison treats the model itself as an unknown variable. The exact posterior model probability can be expressed as:

$$P(M \mid D_{\mathrm{obs}}) \propto P(M) \int P(D_{\mathrm{obs}} \mid \theta_M, M) p(\theta_M \mid M)\, d\theta_M \tag{G.9}$$

However, the likelihood term P(D_{obs} | θ_M, M) is usually unavailable for fake-news cascades because many relevant events are hidden: users may read content without reacting, share through private channels, delete posts, or be exposed through recommendation systems. ABC replaces this unavailable likelihood with simulation. The approximate posterior model probability can be written as:

$$P_{\mathrm{ABC}}(M \mid D_{\mathrm{obs}}) \propto P(M) \int \Pr\{\rho(S_{\mathrm{sim}}, S_{\mathrm{obs}}) \leq \varepsilon \mid \theta_M, M\} p(\theta_M \mid M)\, d\theta_M \tag{G.10}$$

In Equation (G.10), S(D) is a vector of summary statistics, rho is a distance function, and epsilon is the tolerance threshold. Under ABC-SMC, the tolerance threshold is progressively reduced across populations, so later accepted particles represent increasingly close matches to the observed cascade [106], [104].

Posterior model probability and parameter estimation

After the final ABC-SMC population is obtained, posterior parameter distributions are estimated from the accepted and weighted particles. The posterior probability of a candidate model can be approximated as:

$$P_{\mathrm{ABC}}(M = m \mid D_{\mathrm{obs}}) \approx \textstyle\sum_i w_i\, \mathbf{1}(M_i = m) \tag{G.11}$$

where w_i is the normalized weight of particle i and I(M_i = m) is an indicator that equals one if the particle belongs to model m. This expression allows the researcher to estimate whether the observed fake-news cascade is better explained by a simple SIR process or by the more detailed SEIZ process. In addition, the posterior particles provide credible intervals for parameters such as beta, gamma, epsilon, l, and delta. Such intervals are important because fake-news diffusion data are noisy and partially observed, so single-point parameter estimates can create a false sense of precision [90], [102].

Summary statistics for model comparison

For the SIR-SEIZ comparison, the summary statistics should capture both the temporal shape and the structural properties of the cascade. A suitable vector is:

$$S(D) = [C_T, t_{\mathrm{peak}}, r_{\max}, d_c, b_c, z_{\mathrm{share}}, \bar{p}_f] \tag{G.12}$$

Here C_T is the final cascade size, t_{peak} is the time to peak spreading, $r_{\max}$ is the maximum spreading rate, d_c is cascade depth, b_c is cascade breadth, z_{share} is the observed or approximated skeptical/non-sharing proportion, and $\bar{p}_f$ is the average Transformer-based fake-news probability. The final term links the content-detection layer to the diffusion-inference layer. A possible distance function is:

$$\rho(S_{\text{obs}}, S_{\text{sim}}) = \sum_k w_k \left|S_k^{\text{obs}} - S_k^{\text{sim}}\right| \tag{G.13}$$

The weights w_1...,w_5 should be chosen according to the purpose of the analysis. For early-warning systems, the time to peak and maximum spreading rate may be more important. For damage estimation, final cascade size and cascade depth may receive higher weights. For comparing SIR and SEIZ, the skeptical or non-sharing statistic is especially informative because SIR does not explicitly model skeptical users, whereas SEIZ does.

Appendix G.2. Supplementary Motivation for Likelihood-Free Inference and Implementation Context

Table G.1. Motivation for likelihood-free inference in fake-news epidemic modelling.

Observed difficulty	Meaning in fake-news cascades	Why ABC is useful
Hidden exposure	Users may see a post without reposting, commenting, or liking.	Simulation can include unobserved exposure states even if only reposts are observed.
Noisy observations	Bots, duplicates, missing timestamps, and deleted posts distort the cascade.	Distance functions can compare robust summaries rather than exact event histories.
Unknown belief state	A repost does not necessarily prove belief, and silence does not prove rejection.	Latent states such as E and Z can be inferred probabilistically.
Temporal dependence	The probability of later shares depends on earlier shares and network order.	Stochastic simulations can reproduce time-dependent cascades.
Intractable likelihood	The full probability of the cascade is not available in closed form.	ABC avoids explicit likelihood calculation.

Appendix G.3. ABC summary statistics and distance-function specification

The following supplementary material expands the technical audit trail for the ABC and ABC-SMC inference layer. It is intentionally placed in Appendix G so that Sections 3.5 and 3.6 remain results-focused and are not repeated. The tables and equations below document the extended summary-statistic vector, the standardized distance construction, posterior credible intervals and the component-level computation of the baseline diffusion-risk functional.

$$S_{\text{ext}}(D_{\text{obs}}) = \left(C_T, I_{\text{peak}}, t_{\text{peak}}, \text{AUC}(I_{\text{cum}}), I_7, I_{30}, t_{90}\right) \tag{G.14}$$

In Equation (G.14), C_T is the final cumulative cascade size, I_{peak} is the maximum new-spreader count, t_{peak} is the time to peak, $\text{AUC}(I_{\text{cum}})$ is the area under the cumulative infected curve, I_7 and I_{30} are the early-spreader counts during the first 7 and 30 days, and t_{90} is the time required to reach 90% of the final cascade size.

Table G.2 extends the compact ABC target evidence reported in Section 3.6.1. It is not a second interpretation of the SIR/SEIZ results; it records the technical summary vector used to compare observed and simulated fake-news cascade trajectories.

Table G.2. Extended ABC summary-statistic vector used for posterior inference.

ABC summary statistic	COVID-19 Twitter/X fake-news claim	Hoaxy / OSoMeNet "Three million votes" claim
Observation window	2020-02-02 to 2020-12-31	2016-05-16 to 2016-09-26
Daily windows	334	134
Final I_{cum}	95,599	149,079
Peak I_{new}	5,892	4,660
Peak date	2020-07-30	2016-09-08
Time to peak (days)	179	115
Days to 90% final size	322	119
AUC of I_{cum}	18,197,133	8,712,522
First 7-day spreaders	16	10,289
First 30-day spreaders	1,922	15,225

$$d_{\text{curve}}(I_{\text{sim}}, I_{\text{obs}}) = \frac{\|I_{\text{sim}}(t) - I_{\text{obs}}(t)\|_2}{\|I_{\text{obs}}(t)\|_2} \quad \text{(G.15)}$$

$$d_{\text{summary}}(s_{\text{sim}}, s_{\text{obs}}) = \| z(s_{\text{sim}}) - z(s_{\text{obs}}) \|_2 \quad \text{(G.16)}$$

$$d_{\text{ABC}} = \alpha d_{\text{curve}} + (1-\alpha) d_{\text{summary}} \quad \text{(G.17)}$$

Equations (H.15)-(H.17) define the supplementary distance construction. The standardized summary distance prevents large-scale quantities such as final size and AUC from dominating smaller timing quantities, while the curve component remains the main target because the purpose is to reproduce the full cumulative spreading trajectory.

Appendix G.4. Posterior parameter distributions and credible intervals

This subsection gives the full posterior-distribution records behind the compact posterior summary reported in Section 3.6. The values should be read as uncertainty intervals conditional on the selected simulator, priors, tolerance schedule and summary statistics.

Tables H.3-H.5 supplement the compact posterior values in Chapter 3 by reporting distributional uncertainty. They should not be read as new model-fitting results; they are the audit trail behind the posterior estimates already summarized in Section 3.6.

Table G.3. ABC-SIR posterior parameter estimates and 95% credible intervals.

Dataset	Parameter	Mean	Median	SD	2.5%	97.5%
COVID-19	beta	0.19667	0.19515	0.01741	0.16666	0.23416
COVID-19	gamma	0.10249	0.10243	0.012122	0.082218	0.12762
COVID-19	R0	1.927	1.9223	0.093317	1.7733	2.1134
Hoaxy	beta	0.054983	0.055056	0.0016407	0.052106	0.057981
Hoaxy	gamma	0.0052291	0.005192	0.00085817	0.003772	0.0071409
Hoaxy	R0	10.775	10.572	1.6684	8.0378	14.195

Table G.4. ABC-SEIZ posterior parameter estimates for the COVID-19 Twitter/X fake-news cascade.

Parameter	Mean	Median	SD	2.5%	97.5%
beta	1.0968	1.0762	0.14112	0.84879	1.3861
b	0.1342	0.12899	0.030224	0.08448	0.19834
p	0.69873	0.69803	0.057028	0.5869	0.81925
l	0.092394	0.091512	0.040153	0.008891	0.1686
epsilon	0.010111	0.0099706	0.001704	0.0072491	0.013836

rho	0.00048268	0.00046762	0.00012845	0.00027831	0.00076659
gamma	0.27603	0.27106	0.060468	0.17865	0.41969
delta	0.39594	0.39921	0.070647	0.26002	0.5

Table G.5. ABC-SEIZ posterior parameter estimates for the Hoaxy / OSoMeNet fake-news cascade.

Parameter	Mean	Median	SD	2.5%	97.5%
beta	1.1855e-05	1e-05	3.2829e-06	1e-05	2.1785e-05
b	1.8852	2	0.1708	1.4281	2
p	0.012394	0.01184	0.0091162	0	0.031918
l	0.036275	0.035914	0.009719	0.017086	0.055573
epsilon	0.043851	0.042814	0.0083104	0.031086	0.063328
rho	1.1377e-05	1e-05	2.5604e-06	1e-05	1.9974e-05
gamma	0.0087209	0.0085732	0.0019061	0.0057081	0.012993
delta	0.0010334	0.0010149	0.00022608	0.00067631	0.0014442

Appendix G.5. Posterior predictive intervals and calibration diagnostics

Table G.6. ABC-SEIZ posterior predictive final-size intervals.

Dataset	Posterior model	Final tolerance	Relative L2	Final I_{cum}: observed / median / 95% interval
COVID-19 Twitter/X	ABC-SEIZ; 400 particles	0.0923	0.0463	95,599 / 84,921 / [72,269, 97,574]
Hoaxy / OSoMeNet	ABC-SEIZ; 400 particles	0.0635	0.0469	149,079 / 140,408 / [124,983, 155,833]

Table G.7 clarifies that ABC-SMC-SEIZ is not a different epidemic mechanism from ABC-SEIZ. It is a sequential calibration procedure applied to the same SEIZ state structure, with lower posterior predictive error under the same observed cumulative trajectories.

Table G.7. Supplementary posterior predictive calibration comparison between ABC-SEIZ and ABC-SMC-SEIZ.

Dataset	ABC-SEIZ error	ABC-SMC-SEIZ error	Error reduction	Interpretation
COVID-19 Twitter/X	Rel. L2 = 0.0463; RMSE = 2,836; MAE = 2,212	Rel. L2 = 0.0388; RMSE = 2,377; MAE = 1,900	Rel. L2 lower by 16.2%	ABC-SMC gives closer predictive calibration under the same SEIZ structure.
Hoaxy / OSoMeNet	Rel. L2 = 0.0469; RMSE = 3,805; MAE = 3,279	Rel. L2 = 0.0392; RMSE = 3,182; MAE = 2,657	Rel. L2 lower by 16.4%	ABC-SMC gives stronger sequential calibration for the smoother cascade.

Appendix G.6. Supplementary computation of the Baseline Posterior Fake-News Diffusion-Risk Functional

The baseline risk functional is already defined in Equations (2.24)-(2.26) and discussed in Section 3.8. The supplementary material below does not redefine the method; it records how the reported component values are calculated from existing Chapter 3 outputs.

$$D_{30}^{\text{base}} = \frac{I_{\text{cum}}(t+30) - I_{\text{cum}}(t)}{N_c} \tag{G.18}$$

$$R_{30}^{\text{base}} = p_f(x_c) \cdot \left(1 + \lambda A_c(t)\right) \cdot D_{30}^{\text{base}} \tag{G.19}$$

Equation (G.18) gives the normalized 30-day posterior predictive growth component used in the demonstration, while Equation (G.19) expresses the corresponding baseline risk score when the amplification coefficient is set to the illustrative value $\lambda = 1$.

Table G.8 makes the risk-functional demonstration auditable. The values are horizon-specific and should not be interpreted as a universal ranking of platforms or misinformation domains.

Table G.8. Component-level computation of the baseline posterior diffusion-risk functional.

Empirical cascade	$p_f(x_c)$	$A_c(t)$	D_{30}^{base}	R_{30}^{base}	**Forecast validation**
COVID-19 Twitter/X fake-news claim	0.914	0.0616 = 5,892 / 95,599	0.0207 = (82,531 - 80,556) / 95,599	approx. 0.020	RMSE = 7,189; normalized RMSE = 0.075
Hoaxy / OSoMeNet political fake-news claim	0.963	0.0273 = 4,272 / 156,556	0.1980 = (144,564 - 113,565) / 156,556	approx. 0.196	RMSE = 4,019; normalized RMSE = 0.027

Appendix H. Additional Diffusion and Calibration Plots

Appendix H retains additional classification diagnostics, temporal-network visualizations, compartment-transition plots and posterior calibration figures that support Chapter 3 without overloading the main results chapter. Figures H.1-I.3 support the detection diagnostics in Section 3.3, Figures H.4-I.7 support the temporal-network discussion in Section 3.4, Figures H.8-I.9 support the compartment interpretation in Section 3.5, and Figures H.10-I.13 support the ABC/ABC-SMC calibration discussion in Section 3.6.

Figure H.1 complements the ISOT baseline discussion by showing the residual error pattern of the strongest classical model. It is retained as a supplementary diagnostic and should be read together with the RoBERTa comparison in Section 3.3.2.

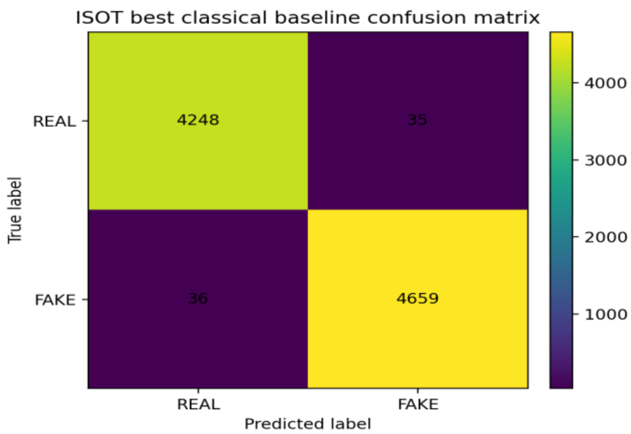

Figure H.1. Confusion matrix for the best-performing ISOT classical baseline (Linear SVM + TF-IDF).

Figure H.2 retains the standalone ROC visualization for the strongest ISOT Transformer setting. It complements Figure 3.3, where the ROC curve and confusion matrix are presented together in the main text.

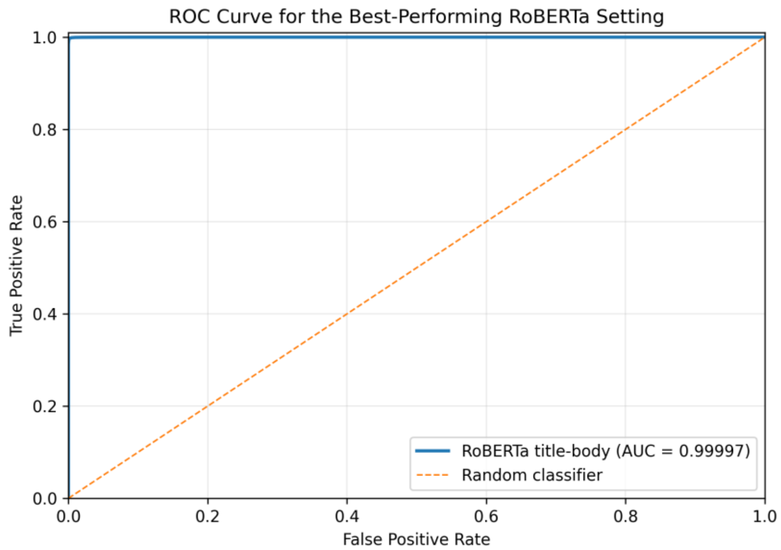

Figure H.2. Standalone ROC curve for the best-performing ISOT RoBERTa title-body setting.

Figure H.3 supports the CoAID longer-news discussion in Section 3.3.2 by showing that the long-sequence BigBird-RoBERTa setting recovers most fake healthcare items despite strong class imbalance.

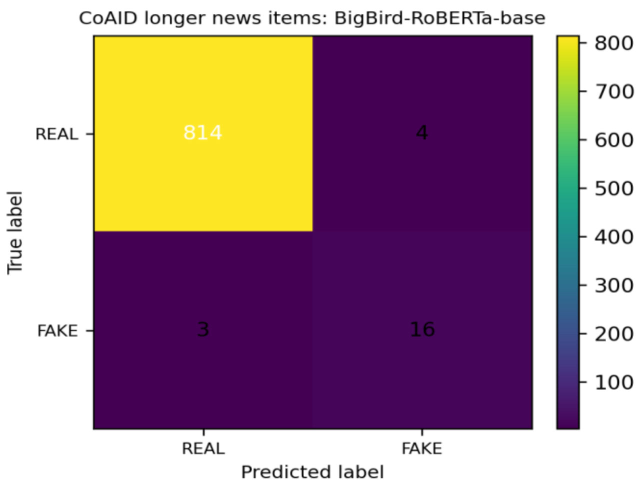

Figure H.3. CoAID BigBird-RoBERTa-base confusion matrix for longer news items.

Diffusion-oriented temporal-network visualizations

Figure H.4 supports Section 3.4.3 by showing the reduced core of the COVID-19 temporal network; the full model fitting relies on the complete $I_{\text{new}}(t)$ and $I_{\text{cum}}(t)$ trajectories, not only on this visualization.

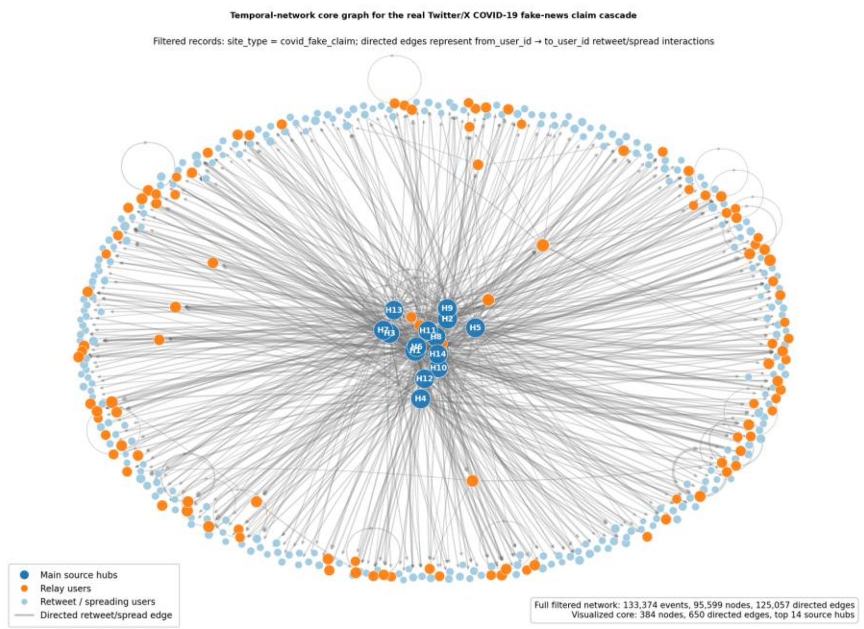

Figure H.4. Core temporal-network graph for the real Twitter/X COVID-19 fake-news claim cascade.

Figure H.5 supports Section 3.4.3 by showing the Hoaxy / OSoMeNet network core; it visualizes the hub-centred structure behind the large political fake-news cascade.

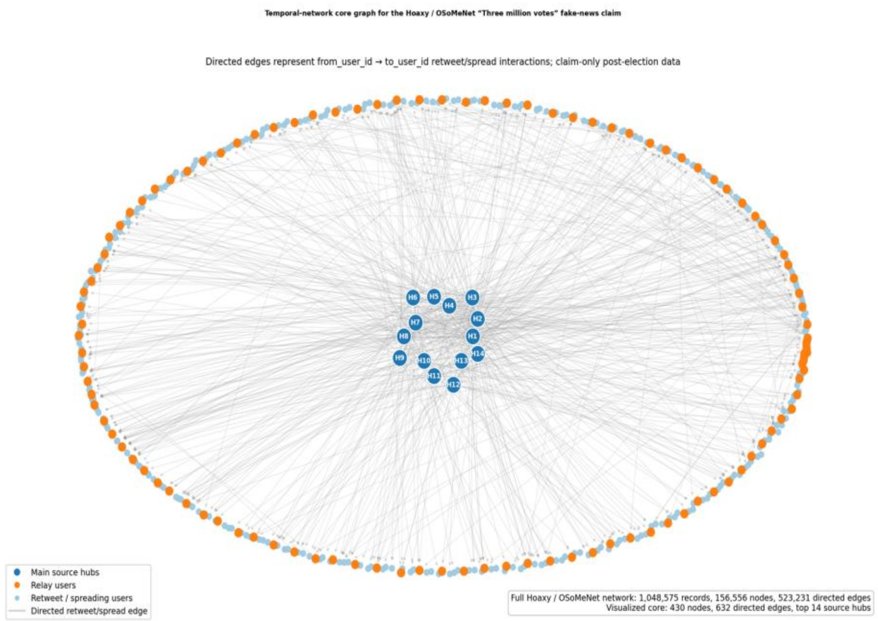

Figure H.5. Temporal-network core graph for the Hoaxy / OSoMeNet fake-news claim “Three million votes in the Presidential Election were cast by illegal aliens.”

Figure H.6 documents the auxiliary Hoaxy-imported social cascade and remains supplementary because this case is not used as a principal ABC/ABC-SMC inference target.

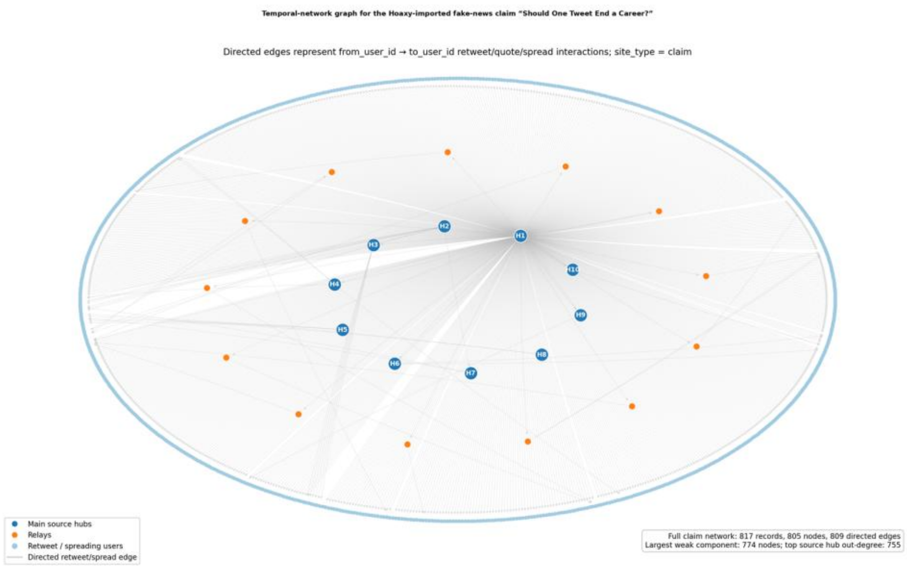

Figure H.6. Temporal-network graph for the Hoaxy-imported fake-news claim "Should One Tweet End a Career?".

Figure H.7 documents the synthetic Albanian validation cascade and should be interpreted only as pipeline validation rather than empirical Twitter/X or Hoaxy evidence.

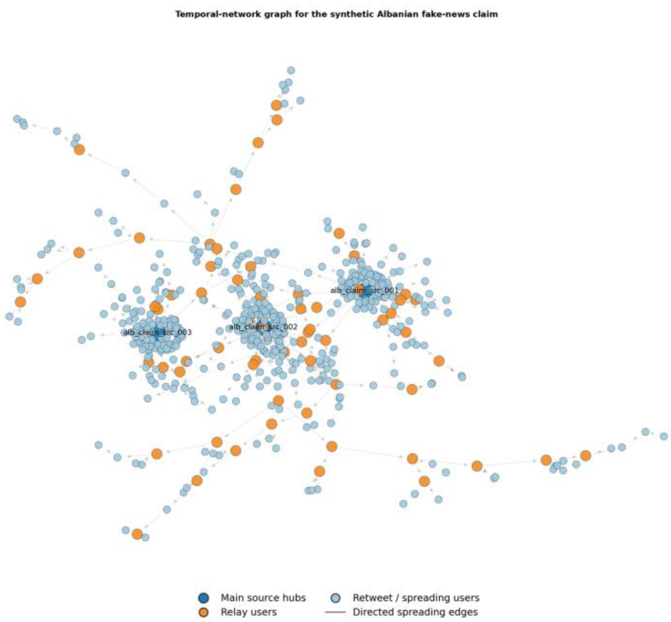

Figure H.7. Temporal-network graph for the synthetic Albanian fake-news claim "Shqipëria do të presë 100 mijë palestinezë nga Gaza."

SIR/SEIZ fitting and compartment-transition graphics

Figure H.8 supports the SIR/SEIZ comparison in Section 3.5 by retaining the Hoaxy SIR compartment decomposition outside the main chapter.

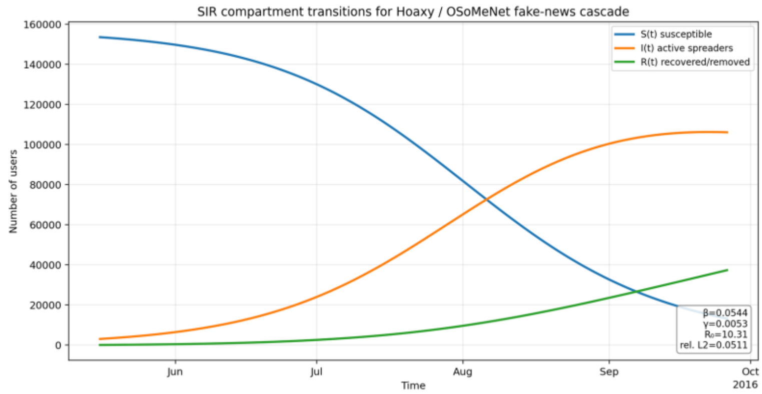

Figure H.8. Supplementary SIR compartment-transition curves for the Hoaxy / OSoMeNet "Three million votes" fake-news claim cascade.

Figure H.9 supports the SEIZ interpretation in Section 3.5 by retaining the Hoaxy exposed and skeptic-state decomposition outside the main chapter.

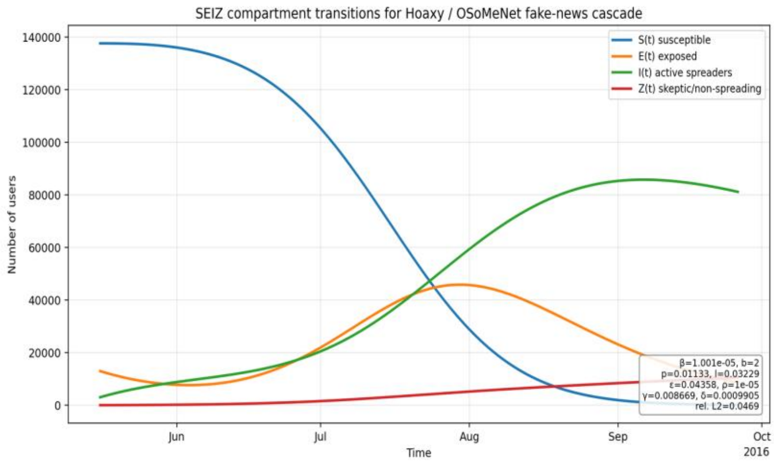

Figure H.9. Supplementary SEIZ compartment-transition curves for the Hoaxy / OSoMeNet "Three million votes" fake-news claim cascade.

ABC, ABC-SMC, posterior predictive and forecast calibration graphics

Figure H.10 supports the posterior predictive validation in Section 3.6.2 by providing the standalone COVID-19 ABC-SEIZ fit.

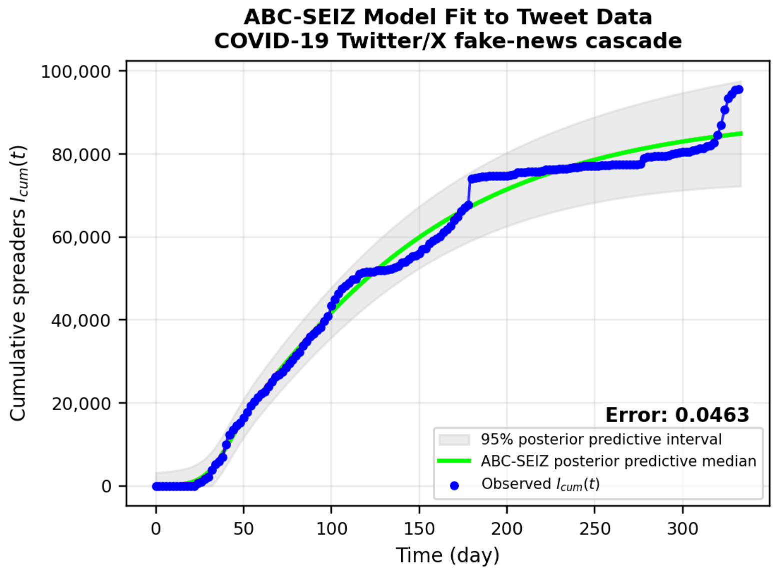

Figure H.10. Supplementary ABC-SEIZ posterior predictive fit for the COVID-19 real Twitter/X fake-news cascade.

Figure H.11 supports the posterior predictive validation in Section 3.6.2 by providing the standalone Hoaxy ABC-SEIZ fit.

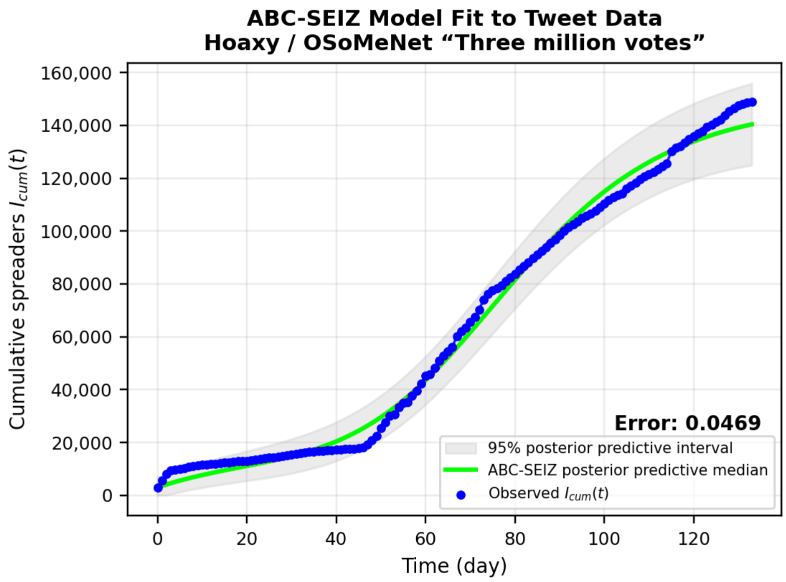

Figure H.11. Supplementary ABC-SEIZ posterior predictive fit for the Hoaxy / OSoMeNet "Three million votes" fake-news cascade.

Figure H.12 supports the ABC-SMC calibration discussion in Section 3.6.3 by retaining the COVID-19 comparative posterior predictive plot.

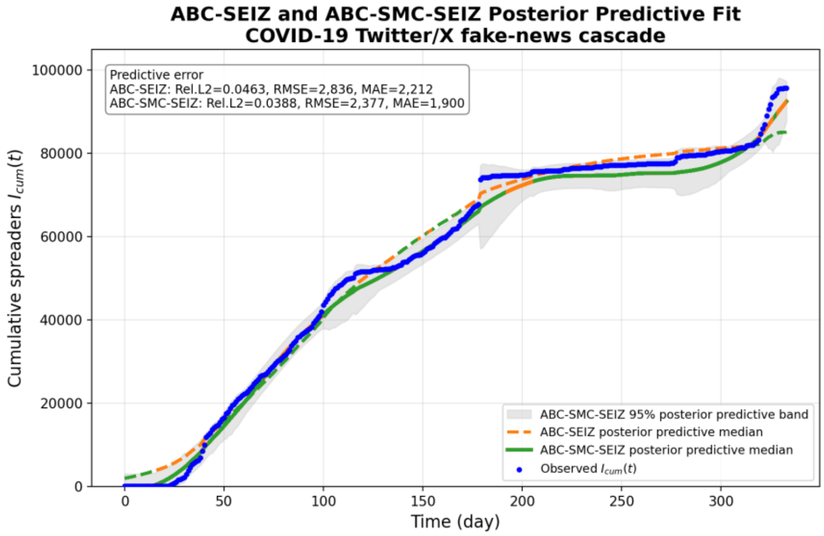

Figure H.12. Supplementary comparative ABC-SEIZ and ABC-SMC-SEIZ posterior predictive fit for the COVID-19 real Twitter/X fake-news claim cascade.

Figure H.13 supports the ABC-SMC calibration discussion in Section 3.6.3 by retaining the Hoaxy comparative posterior predictive plot.

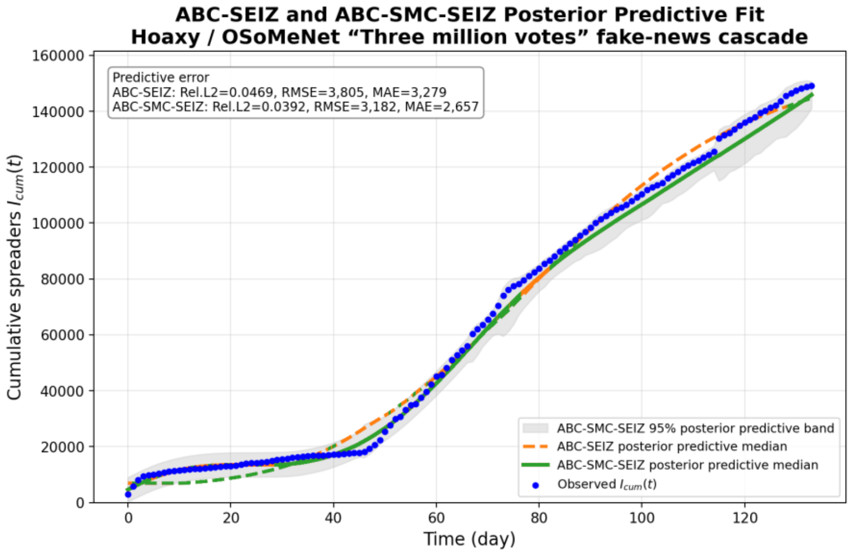

Figure H.13. Supplementary comparative ABC-SEIZ and ABC-SMC-SEIZ posterior predictive fit for the Hoaxy / OSoMeNet "Three million votes" fake-news claim cascade.