University of Sunderland

Faculty of Technology
Department of Computer Science

PROM02 – MSc Dissertation
MSc Data Science

Student Name: Chan Kin Lok Gerald
Supervisor Name: Cheung King Hong
Second Marker Name: Jiang Ming

Analysing the Spread of Misinformation in Social Networks Using Random Graph Models

Predictive Modelling and Mitigation Strategies for Enhancing Network Resilience

March 2025

Declaration

I declare the following:

(1) that the material contained in this dissertation is the end result of my own work and that due acknowledgement has been given in the bibliography and references to **ALL** sources be they printed, electronic or personal.

(2) the Word Count of this Dissertation is 12605 words.

(3) that unless this dissertation has been confirmed as confidential, I agree to an entire electronic copy or sections of the dissertation to being placed on the eLearning Portal, if deemed appropriate, to allow future students the opportunity to see examples of past dissertations. I understand that if displayed on the eLearning Portal it would be made available for no longer than five years and that students would be able to print off copies or download.

(4) I agree to my dissertation being submitted to a plagiarism detection service, where it will be stored in a database and compared against work submitted from this or any other Department or from other institutions using the service.

In the event of the service detecting a high degree of similarity between content within the service this will be reported back to my supervisor and second marker, who may decide to undertake further investigation that may ultimately lead to disciplinary actions, should instances of plagiarism be detected.

(5) I have read the University of Sunderland Policy Statement on Ethics in Research and I confirm that ethical issues have been considered, evaluated and appropriately addressed in this research.

SIGNED:

DATE: 20/03/2025

Abstract

In the contemporary digital age, social networks serve as prolific platforms for the fast dissemination of disinformation, presenting considerable threats to public discourse and societal health. This dissertation, entitled "Analysing the Spread of Misinformation in Social Networks Using Random Graph Models: Predictive Modelling and Mitigation Strategies for Enhancing Network Resilience," examines the dynamics of misinformation dissemination through the lens of random graph theory. The study synthesizes real data from social media platforms with theoretical modelling, grounded in an extensive survey of literature on network analysis, misinformation dynamics, and graph theory. The study utilizes a multi-phase technique encompassing data collecting, preprocessing, and the development of representative social network models. The study employs comprehensive simulations with random graph models to pinpoint significant network flaws and key nodes that enable the fast dissemination of disinformation. Advanced prediction models are subsequently created utilizing machine learning approaches to anticipate disinformation patterns across diverse network settings. The results indicate that network designs and high-centrality nodes substantially influence the dissemination of misinformation, implying that focused intervention measures might effectively alleviate these impacts. The research offers pragmatic insights and concrete suggestions for governments, social media platforms, and cybersecurity experts. This study illustrates the capacity of random graph models to forecast and manage the dissemination of misinformation, therefore aiding in the creation of more robust digital communication networks and fostering a safer, more informed online ecosystem.

Contents

1 Introduction

1.1 Background

In the contemporary digital age, social networks have transformed the methods of information dissemination and consumption. Nonetheless, this fast dissemination of knowledge has also led to the widespread proliferation of disinformation. False or false information sent through these networks can have extensive repercussions—eroding public trust, dividing communities, and perhaps affecting socio-political stability. The intricacy of social networks, defined by numerous linkages among members, necessitates strong analytical frameworks. Random graph models provide a robust theoretical framework to elucidate the structural characteristics of these networks, yielding insights into the mechanisms and routes by which disinformation disseminates.

1.2 Aims

This study seeks to examine and simulate the dissemination of disinformation inside social networks using the application of random graph theory. The primary objective is to create predictive models and efficient mitigation measures that improve network resilience against the spread of misinformation.

1.3 Objectives

To accomplish this objective, the study will:

- Perform extensive literature research to create a theoretical framework on disinformation dynamics, network analysis, and random graph models.
- Gather and preprocess empirical social network data to develop representative graph models.
- Utilize random graph models to replicate the dissemination of disinformation and pinpoint critical network weaknesses.
- Utilize machine learning methodologies to construct predictive models for anticipating disinformation trends.
- Formulate and assess ways to effectively mitigate the dissemination of disinformation.

1.4 Research Approach

The study employs a multi-phase methodology that combines theoretical and empirical approaches. A comprehensive literature assessment will be conducted to situate the research within the current academic debate on disinformation and network theory. This will be succeeded by data collecting and preparation, during which social network data are refined to construct precise network models. The study's essence is in employing random graph models to simulate misinformation dissemination, alongside predictive modelling utilizing machine learning techniques. The project will ultimately devise and assess intervention techniques, measuring their efficacy using simulation and comparative analysis.

1.5 Structure of the Report

The dissertation has six major chapters:

- Chapter 1: Introduction — delineating the context, aims, objectives, research methodology, and a summary of the report's organization.
- Chapter 2: Research Review — offering a critical assessment of the current research regarding misinformation, social network analysis, and random graph models.
- Chapter 3: Practical Research Methodology — outlining data collection, preprocessing, model development, and experimental design.
- Chapter 4: Results Analysis — showing the outcomes from the simulations and forecasting models.
- Chapter 5: Project Evaluation and Reflection — evaluating the study results, contemplating the methodology utilized, and addressing ethical, legal, societal, and professional issues.
- Chapter 6: Conclusions and Recommendations — encapsulating the principal results and suggesting avenues for further study and practical implementations.

1.6 Ethical, Social, Professional, Legal and Security Considerations

This study is executed with a robust dedication to ethical research methodologies. All social network data utilized in the study will be anonymised and processed in accordance with pertinent data protection legislation, including GDPR. The study examines the ethical ramifications of investigating disinformation by guaranteeing transparency in the

analysis and responsible utilization of the generated models. Furthermore, professional standards are upheld by stringent methodological design and comprehensive assessment, while possible social implications are meticulously assessed to avert inadvertent harm.

2 Literature Review

This chapter rigorously examines the scholarly literature pertinent to the study and forecasting of disinformation dissemination in social networks. It delineates the theoretical underpinnings of social network analysis, graph theory, and random graph models, while also addressing predictive modelling and mitigation measures. This review delineates the research issue and identifies the gaps that the current study aims to solve.

2.1 Misinformation in Social Networks

The fast propagation of information through social media platforms has concurrently led to an increase in the distribution of disinformation. Initial studies, like Al-Rawi (2020) and Vosoughi et al. (2018), investigate the dynamics of misinformation—its origins, dissemination methods, and effects on public perception. These studies indicate that misinformation frequently disseminates more rapidly and reaches more audiences than confirmed information. Nevertheless, whereas these studies offer significant insights into user behaviour and engagement, they frequently depend on empirical evidence lacking a solid theoretical framework to forecast dissemination trends.

2.2 Social Network Analysis and Graph Theory

Comprehending the foundational framework of social networks is crucial for modelling information dissemination. Freeman's (1978) foundational research on centrality measurements has facilitated the identification of major influencers inside networks. Subsequent research by Barabási (2009) and Watts and Strogatz (1998) presents the notions of scale-free and small-world networks, respectively, which are particularly pertinent for analysing intricate social connections. These network models include significant structural characteristics; nonetheless, they can inadequately depict the stochastic dynamics of misinformation dissemination.

2.3 Random Graph Models

Random graph models provide a probabilistic framework for simulating the evolution and dynamics of network topologies. Newman’s (2003) extensive review elucidates the adaptation of random graphs for modelling real-world networks. Moreover, the epidemic spreading models analysed by Pastor-Satorras and Vespignani (2001) illustrate the threshold-free behaviour in scale-free networks, which is directly relevant to the

dissemination of disinformation. While these models offer a robust theoretical foundation, their predictive efficacy is constrained without the incorporation of actual data, a deficiency rectified in the present effort.

2.4 Predictive Modelling and Mitigation Strategies

To tackle the issue of reducing disinformation, studies like Budak et al. (2011) and Kempe et al. (2003) provide algorithmic strategies to curtail dissemination and enhance influence minimization. These studies highlight the identification of pivotal nodes and intervention measures that might mitigate the dissemination of erroneous information. Although these technologies have potential, they frequently necessitate more refining when used in extensive, dynamic social networks. The amalgamation of machine learning methodologies with random graph simulations, as executed in the project's final code version, signifies an innovative strategy for addressing this disparity.

2.5 Critical Evaluation and Research Gaps

Notwithstanding the comprehensive study on disinformation dynamics and network modelling, some gaps persist. A multitude of studies has concentrated either on the empirical examination of disinformation or on theoretical models that exhibit a deficiency in predictive resilience. An integrated method that amalgamates the advantages of random graph models with sophisticated predictive algorithms is essential for enhancing the forecasting and mitigation of misinformation dissemination. The examined literature emphasizes the promise of this approach, while also revealing the shortcomings of current approaches in accurately depicting the intricate dynamics of social networks.

2.6 Conclusion

The literature demonstrates a strong basis in social network analysis and random graph theory, with developing methodologies in predictive modelling and mitigation techniques. This review affirms the significance of employing random graph models for analysing the dissemination of disinformation while also pinpointing essential gaps that the present work seeks to address. The insights obtained here justify the research question: "How can random graph models be employed to analyse and forecast the dissemination of misinformation in social networks, and what strategies can effectively curtail this dissemination?" The next chapters will elucidate the actual research technique and the incorporation of the final code implementation to tackle these problems.

3 Practical Research Methodology

This chapter offers a comprehensive elucidation of the methodologies employed to examine and predict the dissemination of disinformation inside social networks. The technique comprises several interrelated parts. The steps encompass data gathering and pre-processing, categorization using Natural Language Processing (NLP) models, network design and random graph modelling, SIR simulation for misinformation spread, sensitivity analysis, and sophisticated network analytics. Every phase is certain to be automated, reproducible, and adequately documented due to the technological foundation constituted by the final iteration of the project code.

3.1 Data Acquisition and Pre-processing

This phase guarantees the systematic collection, cleansing, and preparation of raw data from diverse sources for analysis. It establishes the groundwork for all ensuing modelling and simulation activities.

3.1.1 Data Sources and Collection

The study employs four datasets sourced from FakeNewsNet:

- gossipcop_fake.csv
- gossipcop_real.csv
- politifact_fake.csv
- politifact_real.csv

Each dataset comprises essential metadata, encompassing unique article identifiers, news URLs, article titles, and tweet identifiers. The justification for using these datasets is to get a varied and representative collection of news items that includes both false and authentic news. Employing several datasets not only amplifies the total data amount but also augments variability, which is essential for developing effective classification models.

```
GossipCop Fake Columns: ['id', 'news_url', 'title', 'tweet_ids']
GossipCop Real Columns: ['id', 'news_url', 'title', 'tweet_ids']
PolitiFact Fake Columns: ['id', 'news_url', 'title', 'tweet_ids']
PolitiFact Real Columns: ['id', 'news_url', 'title', 'tweet_ids']
```

Fig. 1 Column Headers of the GossipCop and PolitiFact Datasets (Fake and Real)

3.1.2 Data Collection Process

The data is imported into Python using the pandas module, a proficient tool for handling tabular data. The Pandas Data Frame format is employed because to its ability for rapid data processing, integration, and purification. Importing these CSV files into Data Frames ensures that the data is systematically arranged, hence easing subsequent pre-processing and integration.

3.1.3 Data Cleaning and Text Normalization

Raw text data sometimes includes superfluous parts that might hide significant trends. To prepare the text for analysis, a rigorous cleaning procedure will be implemented.

- Elimination of HTML Tags: News stories frequently include HTML code that does not enhance literary significance. All HTML tags are eliminated using regular expressions.
- Removal of Non-Alphabetic Characters: All characters that are not letters (including numbers, punctuation, or symbols) are eliminated to concentrate on the words themselves.
- Case Transformation: All text is shown in lowercase. This ensures that identical words in varying situations are regarded as the same entity, hence standardizing the input.
- Stop word Removal: Frequently occurring terms that lack substantial meaning, such as “the,” “is,” and “and,” are eliminated utilizing NLTK’s stop word corpus. Eliminating these words diminishes noise and amplifies the signal from more significant keywords.
- Tokenization and Reassembly: The sanitized text is segmented into individual lexemes, purged of stop words, and subsequently reconstructed into a cohesive, refined string.

These measures guarantee that the input for our models is of superior quality and that the data utilized for feature extraction is as devoid of noise as feasible.

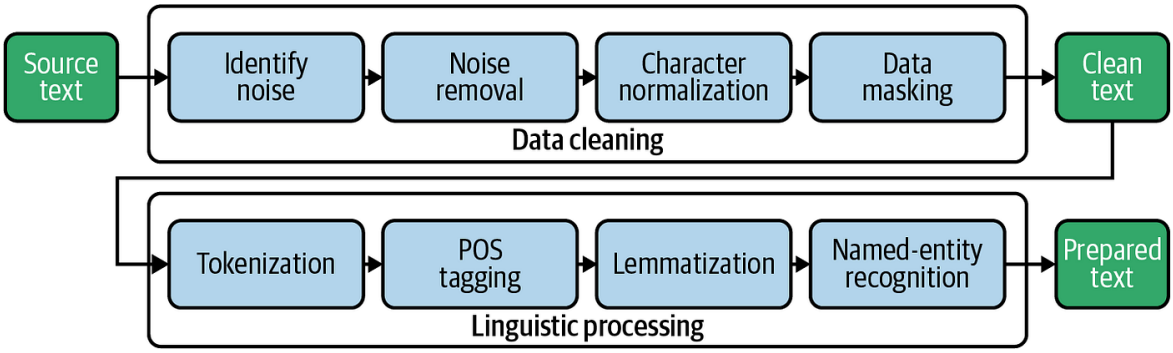


Fig. 2 Data Cleaning and Linguistic Processing Pipeline

3.1.4 Label Assignment and Data Consolidation

Subsequent to text cleansing, each item is allocated a binary designation:

- False information is designated as '1'.
- Authentic news is designated as '0'.

Upon completion of labelling, the four datasets are consolidated into a singular, comprehensive Data Frame. Throughout this procedure, any records with absent or null text values are eliminated to guarantee the dataset utilized for analysis is robust. The resulting consolidated dataset, including approximately 23,000 records, functions as the input for the ensuing classification and network modelling stages.

3.2 Classification Models for Misinformation Detection

The last phase in the process is constructing models capable of autonomously categorizing news stories as either fraudulent or authentic. This section delineates two separate methodologies: one utilizing classic machine learning (TF-IDF with Logistic Regression) and the other employing a contemporary deep learning technique (fine-tuned BERT). Both methodologies are thoroughly elucidated to convey the foundational concepts and to clarify their use in this project.

3.2.1 Baseline Model: TF-IDF Vectorization and Logistic Regression

3.2.1.1 TF-IDF Vectorization:

TF-IDF denotes Term Frequency–Inverse Document Frequency. This is a key approach in natural language processing that transforms text into numerical characteristics, indicating the significance of words within a document and across a corpus of texts.

- Term Frequency (TF): This metric quantifies the frequency of a word's occurrence in a document in relation to the document's overall word count. An elevated frequency signifies that the term holds significance inside that particular source.
- Inverse Document Frequency (IDF): This element assesses the uniqueness or rarity of a term within the whole dataset. Frequently occurring terms in several documents are assigned a diminished weight. IDF is often calculated as the logarithm of the total document count divided by the number of documents that include the term.
- TF-IDF Score: The product of TF and IDF yields a score that indicates a word's significance inside a specific text and its distinctiveness throughout the collection. This project restricts the feature space to 5000 words to achieve a compromise between comprehensiveness and computational efficiency.

3.2.1.2 Logistic Regression:

Logistic Regression is a commonly employed approach for binary classification tasks. It represents the likelihood that an input is associated with a specific class using the logistic function:

- The Logistic Function: The logistic function or sigmoid function is defined as $\sigma(z) = 1 / (1 + e^{(-z)})$, where z is a linear combination of input features. This function converts any input into a number ranging from 0 to 1, which may be read as a probability.
- Application in Classification: In the context of fake news identification, logistic regression computes the likelihood that an item is fraudulent (label '1') based on its TF-IDF characteristics. A threshold, often set at 0.5, is subsequently employed to determine the final classification.
- Model Training: The TF-IDF features are utilized to train the logistic regression model on 80% of the dataset, while the remaining 20% is allocated for testing. Performance is assessed using accuracy, precision, recall, and F1-score. This method offers a clear and comprehensible baseline for comparison with more intricate models.

3.2.2 Advanced Model: Fine-Tuned BERT

3.2.2.1 Overview of BERT:

BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model that utilizes the transformer architecture to comprehend the contextual meaning of words inside a text. BERT reads text bidirectionally, in contrast to typical models that read sequentially, therefore collecting subtle meanings.

3.2.2.2 Tokenization and Data Preparation:

The model starts the process by tokenizing the text with the "Bert-base-uncased" tokenizer, which divides the text into tokens (words or sub word units). A maximum sequence length of 128 tokens is implemented to maintain uniformity in input size. Padding and truncation ensure uniformity in text input length, which is essential for batch processing.

3.2.2.3 Fine-Tuning Process:

Subsequent to tokenization, the pre-trained BERT model undergoes fine-tuning for the false news classification objective. Fine-tuning encompasses:

- Fine-tuning the model on the designated dataset, modifying its pre-trained weights to optimize performance for the job.
- Configuring hyperparameters, including a learning rate of 2e-5, a batch size of 16, and conducting training for 3 epochs.
- Assessing assessment loss to guarantee adequate model convergence.

The refined model is anticipated to identify more intricate patterns in the text compared to the TF-IDF model, particularly nuanced language indicators that distinguish false news from authentic news.

3.2.2.4 Evaluation Metrics for BERT:

The performance of the fine-tuned BERT model, akin to the baseline, is evaluated by accuracy, precision, recall, and F1-score. The model's assessment outcomes are documented and juxtaposed with the baseline to evaluate enhancements.

3.3 Graph Construction and Random Graph Modelling

Following the classification of news stories, the subsequent phase involves constructing synthetic networks to model the dissemination of disinformation. This section elucidates

two prominent random graph models—Barabási–Albert (BA) and Erdős–Rényi (ER)—and their use in this work.

3.3.1 Background and Rationale:

A network, or graph, is a mathematical construct including nodes (or vertices) and edges (or connections). Within this research:

- Nodes: Denote distinct news pieces.
- Edges: Indicate possible encounters or routes for the dissemination of disinformation.

Random graph models facilitate the generation of synthetic networks that replicate the characteristics of actual social networks. This research employs two prominent models: the Barabási–Albert (BA) model and the Erdős–Rényi (ER) model.

3.3.2 Barabási–Albert (BA) Model:

3.3.2.1 Background and Concept:

The BA model relies on the notion of preference attachment, seen in several real-world networks, including the Internet and social media. In these networks, certain nodes (hubs) amass a significant number of connections, whilst the majority of nodes possess relatively few.

- Preferential Attachment: The addition of a new node to the network increases its likelihood of connecting to nodes that possess a high degree of connectivity. This results in a power-law distribution of node degrees, wherein the probability P(k) of a node possessing k connections diminishes as k escalates.
- Principal Variables:
  - Sample size: The aggregate quantity of nodes inside the network (e.g., 1000).
  - The quantity of edges that each new node forms upon joining the network (e.g., 3). This parameter affects the mean connectivity and the network's density.

3.3.2.2 Practical Implementation:

The BA network is built using NetworkX's ***barabasi_albert_graph*** function. Each node is assigned a fictitious or authentic label according to the relevant entry in the aggregated dataset. This not only establishes a realistic network topology but also incorporates the content properties directly into the model. The BA network, characterized by its

significant hubs, is especially beneficial for examining situations in which powerful nodes may swiftly disseminate disinformation.

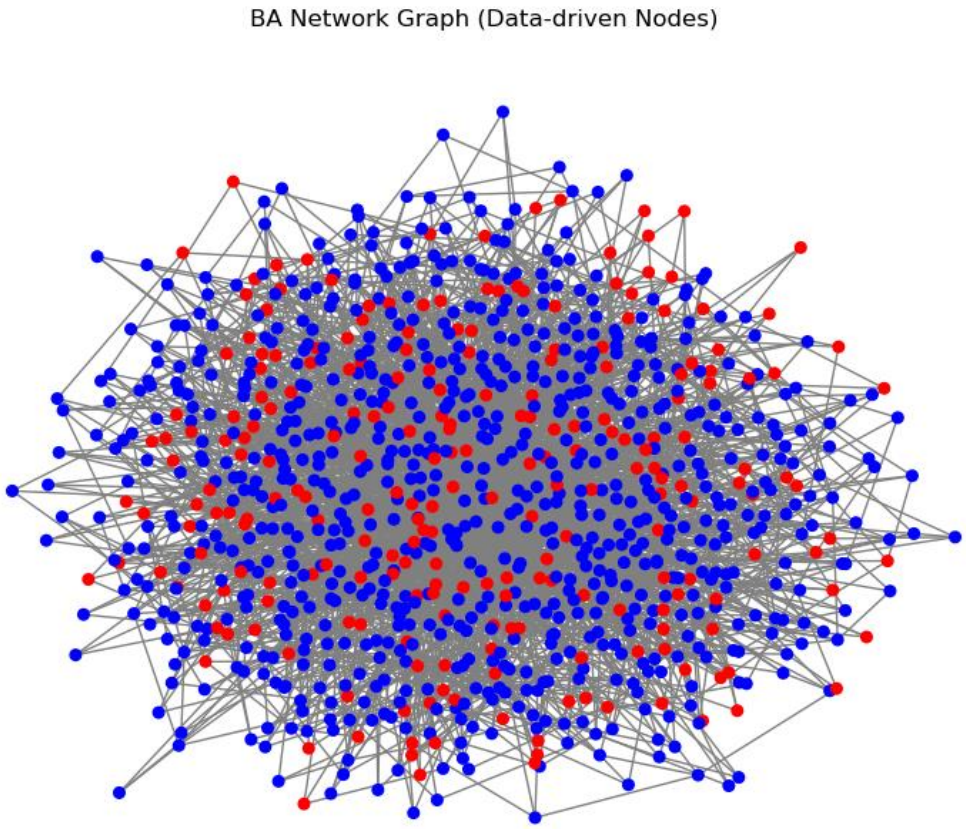


Fig. 3 BA Network Graph

3.3.3 Erdős–Rényi (ER) Model:

3.3.3.1 Background and Concept:

The ER model generates networks by randomly linking nodes with a predetermined frequency. In contrast to the BA model, which generates a scale-free network characterized by hubs, the ER model generally yields a more homogeneous network architecture in which most nodes possess a comparable number of connections.

- Random Connectivity: In an Erdős-Rényi network, each potential edge between any two nodes is established with probability p.
- Principal Variables:
  - Sample size: Identical to that in the BA model, ensuring equitable comparison.

- The probability that a specific pair of nodes is interconnected. In this study, er_p is established at around 6/(sample_size – 1) to get an average degree akin to that of the BA network.

3.3.3.2 Practical Implementation:

The Erdős-Rényi network is constructed with the ***erdos_renyi_graph*** tool from NetworkX. Nodes obtain identical fake/real labels as in the BA network, guaranteeing that both network models accurately represent the underlying dataset distribution. The ER model functions as a benchmark for comparison, facilitating the analysis of how randomness in connections affects the dissemination of disinformation.

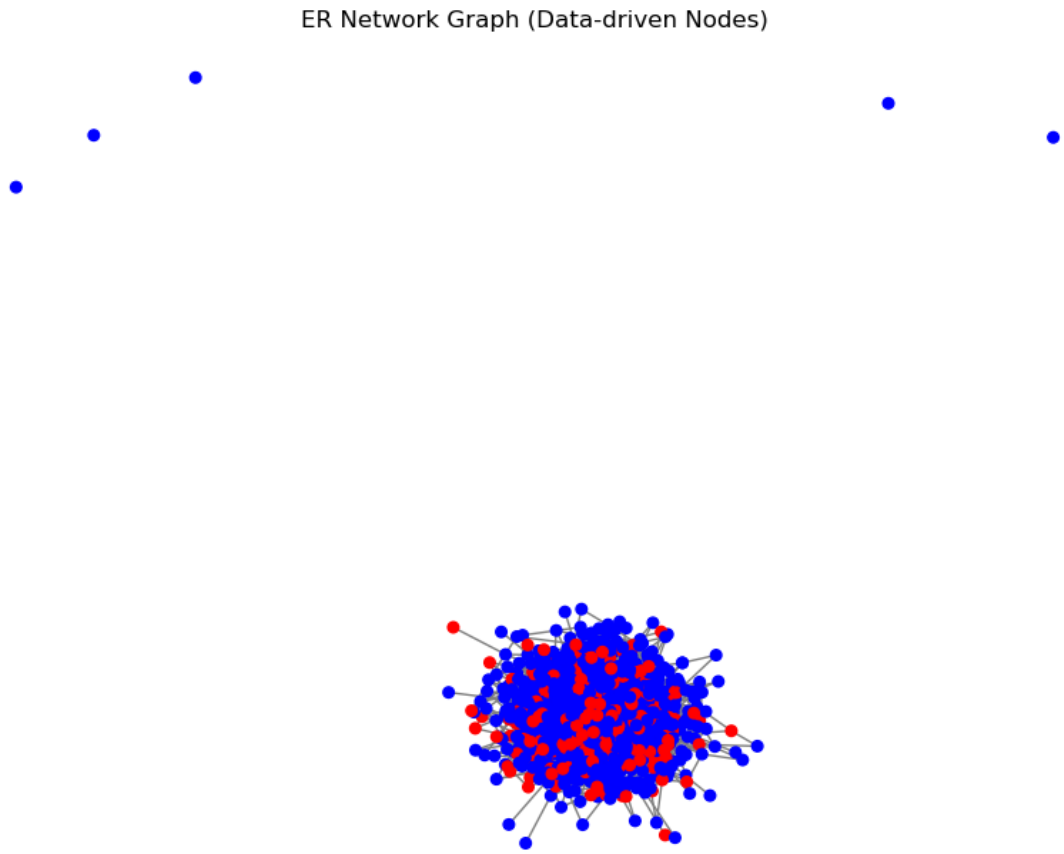


Fig. 4 ER Network Graph

3.3.4 Mathematical Foundations and Practical Implications

3.3.4.1 Mathematical Considerations:

- The preferential attachment of the BA model results in a power-law degree distribution, mathematically represented as P(k) ~ k^(-γ), with γ being a constant generally ranging from 2 to 3.
- The ER model adheres to a binomial (or Poisson for extensive networks) distribution, wherein the probability of a node possessing k connections is expressed as P(k) = (n-1 choose k) p^k (1-p)^(n-1-k).

3.3.4.2 Practical Implications:

The mathematical underpinnings elucidate that the BA network will possess prominent hubs capable of significantly accelerating the dissemination of disinformation, whereas the ER network typically propagates information more equally. This distinction is essential for assessing intervention methods in subsequent sections.

3.3.5 Linking Dataset Characteristics to the Network

An essential element of this research is the incorporation of dataset properties into the network model. The developed network does not depict direct user interactions but serves as a synthetic abstraction intended to encapsulate the dissemination potential of misinformation depending on article characteristics:

- Node Mapping: Each node in the network directly represents a news story in the dataset. The distribution of labels (false vs true) in the dataset establishes the initial state of the nodes, so mirroring the actual prevalence of misinformation in the real world.
- Attribute Integration: The results from the classification models (e.g., credibility scores or binary labels) are utilized to assign characteristics to nodes. These traits may then affect the dynamics of the simulation; for example, nodes designated as bogus news may possess an increased propensity to "infect" adjacent nodes in the SIR simulation.
- Synthetic Connectivity: The dataset lacks clear interconnections between articles; instead, network models (BA and ER) are utilized to simulate putative information dissemination paths. The selection of model parameters (e.g., ***ba_m*** in the BA model

and ***er_p*** in the ER model) is guided by the dataset's features, including the mean number of engagements or citations noted in pertinent literature.

- Data-Driven Calibration: The ratio of fabricated to authentic news in the dataset aids in fine-tuning the network. If the dataset reveals a significant prevalence of bogus news, the synthetic network will automatically have a greater sub-population of nodes disseminating disinformation, thereby affecting the overall simulation results.

This integration guarantees that the network model is not only an abstract mathematical entity but is intricately connected to the empirical characteristics of the dataset, hence augmenting the applicability of the simulation outcomes to actual misinformation dynamics.

3.4 SIR Simulation and Predictive Modelling

This phase uses the SIR model to mimic the dissemination of disinformation across synthetic networks. This section offers a comprehensive elucidation of the model's functionality and its relevance to this investigation.

3.4.1 SIR Simulation Overview

The SIR model is a traditional framework initially created for simulating infectious disease epidemics. It categorizes the population into three segments:

- Susceptible (S): Nodes that have yet to encounter the disinformation.
- Infected (I): Nodes actively disseminating disinformation.
- Recovered (R): Nodes that have stopped disseminating disinformation as a result of fact-checking, correction, or disengagement.

The adaptation of the SIR model simulates how disinformation may "infect" nodes within a network, offering a dynamic perspective on its temporal dissemination.

3.4.2 Detailed Implementation of the SIR Simulation

The simulation is executed by a function that incrementally modifies the state of each node in the network according to established probability. Presented above is a comprehensive account of the procedure:

3.4.2.1 Initialization:

All nodes within the network are first designated to the Susceptible (S) state. A limited fraction of nodes, as defined by the parameter **initial_infected**, is randomly selected to be

in the Infected (I) state. This first circumstance emulates the dissemination of disinformation.

3.4.2.2 Regulations for State Updates:

At every time interval:

- Each node in the Infected state strives to "contaminate" its adjacent nodes. A random number is produced for each neighbour in the Susceptible state. If this value is inferior to the infection probability, the neighbour becomes infected.
- Each sick node possesses an opportunity for recovery independently. A random number is created for each infected node, and if it is smaller than the recovery probability, the node changes to the Recovered (R) state.
- These updates are executed concurrently for each node in the network, encapsulating the dynamic interplay of dissemination and recuperation.

3.4.2.3 Documentation and Visualization:

Following each time step, the quantity of nodes in each state (S, I, R) is documented. The recorded state counts are subsequently utilized to create line graphs depicting the temporal variations in the populations of susceptible, infected, and recovered nodes. The visuals facilitate a comparison of the spread dynamics between the BA and ER network models.

3.4.3 Key Variables and Their Roles

- **initial_infected**: Quantity of nodes that are initially infected (e.g., 10). This option emulates the origin of disinformation.
- **infection_prob**: The likelihood that an infected node would effectively disseminate disinformation to a vulnerable neighbour (e.g., 0.2). This value indicates the "contagiousness" of the disinformation.
- **recovery_prob**: The likelihood that an infected node will recuperate at each time interval (e.g., 0.1). This simulates the impact of interventions like fact-checking.
- **Steps**: The total number of time intervals over which the simulation is conducted (e.g., 20), indicating the temporal scope throughout which the spread is monitored.

3.4.4 Interpreting SIR Simulation Outcomes

The simulation generates a time series depicting the quantities of susceptible, infected, and recovered nodes:

- A **rapid increase** in the quantity of infected nodes signifies a speedy epidemic, commonly observed in networks with prominent hubs (BA model).
- A **gradual infection curve** indicates a slower and more uniform transmission, as evidenced by the ER model.
- The **recovery** rate offers insights into the speed at which misinformation can be mitigated following the initiation of actions.

These results not only corroborate the simulation model but also establish a foundation for generating predictive insights. While the present emphasis is on simulation, the identified tendencies (including infection probability thresholds that result in substantial outbreaks) may inform future endeavours to develop prediction models for intervention measures.

3.5 Sensitivity Analysis

Sensitivity analysis is conducted to assess the impact of fluctuations in critical parameters on simulation results. This stage is essential for comprehending the model's robustness and dependability.

3.5.1 Methodology of Sensitivity Analysis

The emphasis of sensitivity analysis is on the **infection probability** (**infection_prob**):

- Parameter Variation: The infection probability is methodically altered from 0.1 to 0.5 in uniform increments. This range is selected to emulate varying degrees of disinformation "virulence."
- Multiple Iterations: For each infection probability number, the SIR simulation is executed several times (e.g., 5 runs) to reduce the influence of randomness inherent in the process. The data are subsequently averaged to get a reliable approximation.
- Key Outcome: The principal parameter assessed is the total count of recovered nodes, which acts as an indicator of the overall magnitude of the epidemic. An increased quantity of retrieved nodes signifies a broader dissemination of disinformation.

3.5.2 Visualizing and Interpreting Sensitivity Results

The mean final recovered numbers for each infection probability value are illustrated for both BA and ER networks:

- X-Axis: Values of infection likelihood.

- Y-Axis: Average final recovery count. Distinct plots for BA and ER networks are generated to demonstrate the impact of network topology on sensitivity.

This sensitivity study identifies important thresholds at which the dissemination of disinformation becomes disproportionately significant, hence guiding the development of suitable mitigation methods.

3.6 Advanced Network Metrics and Community Detection

Comprehending the network's structure is crucial for pinpointing pivotal nodes and sub-networks that affect the dissemination of disinformation. This section addresses the calculation of sophisticated centrality measures and the implementation of community discovery methods.

3.6.1 Centrality Metrics

Centrality measures assess the significance of nodes within a network.

- Betweenness Centrality:
  - Definition: Quantifies the occurrence of a node on the shortest pathways connecting pairs of nodes.
  - Interpretation: A high betweenness score signifies that a node acts as a conduit within the network. Such nodes can substantially influence the dissemination of information and, consequently, the proliferation of disinformation.
- Closeness Centrality:
  - Definition: Measures the proximity of a node to all other nodes by calculating the average of the shortest route distances.
  - Interpretation: Nodes exhibiting elevated proximity can swiftly engage with or exert influence over the whole network. They are likely to be the initial disseminators or recipients of disinformation.
- Eigenvector Centrality:
  - Definition: Evaluates a node's effect by assessing both the quantity and the quality of its connections. Nodes linked to other highly connected nodes receive elevated ratings.
  - Interpretation: This measure identifies nodes with numerous connections that are also linked to prominent nodes, rendering them potential super-spreaders.

3.6.2 Community Detection Using the Louvain Algorithm

Community detection organizes nodes into clusters characterized by denser intra-group connections compared to inter-group connections.

- Louvain Algorithm:
  - This approach enhances modularity, a statistic that quantifies the robustness of a network's partitioning into communities.
  - Principal Deliverables:
    - Number of Communities: Denotes the unique clusters recognized inside the network.
    - Modularity Score: An elevated score indicates a more robust and differentiated community structure.
- Practical Significance: Communities may function as echo chambers, wherein disinformation proliferates with limited external disruption. Recognizing these clusters is essential for formulating targeted actions that impede the dissemination of misinformation.

3.6.3 Visualization and Summary of Metrics

The calculated metrics and community structures are shown via network graphs:

- Node Colouring: Nodes are tinted according to their community association, facilitating the visual identification of clusters.
- Summary Tables: Average centrality values, community counts, and modularity scores are organized into tables for straightforward comparison between BA and ER networks.

This comprehensive research elucidates the influence of the network's architecture on the dissemination of disinformation and identifies pivotal nodes that may be targeted for mitigation efforts.

3.7 Code Implementation, Documentation, and Reproducibility

It is crucial that the process is replicable by other researchers. This section delineates the coding methodologies, documentation standards, and version control protocols utilized.

3.7.1 Modular Code Structure and Organization

The final code is structured into distinct, modular segments aligned with each methodological phase:

- Data Preprocessing Module: Comprises routines for text cleansing, label assignment, and dataset amalgamation.
- Classification Module: Utilizes TF-IDF vectorization, employs a Logistic Regression classifier, and incorporates fine-tuning procedures for the BERT model.
- Network Modelling Module: Contains functions for generating Barabási-Albert (BA) and Erdős-Rényi (ER) networks, as well as for assigning node characteristics according to the dataset.
- SIR Simulation Module: Comprises the simulation function that updates node states temporally and documents the history of the contagion's dissemination.
- Sensitivity Analysis Module: Comprises procedures for executing many simulations with differing infection probability and calculating the average outcomes.
- Metrics Module: Calculates sophisticated centrality metrics and implements the Louvain method for community discovery.

Every module is thoroughly documented with inline comments elucidating the purpose and operation of each function, hence enhancing code accessibility.

3.7.2 Parameter Logging and Version Control

- Parameter Logging: Essential parameters (e.g., *sample_size*, *ba_m*, *infection_prob*, *recovery_prob*, and *steps*) are recorded to facilitate the replication of simulations.
- Version Control: The code is preserved under version control, and all dependencies are recorded, guaranteeing that subsequent researchers may replicate the results.

3.8 Summary

This chapter has offered a comprehensive elucidation of the empirical research methods employed to examine and forecast the dissemination of misinformation in social networks. Commenced with rigorous data collection and cleansing protocols, guaranteeing that the dataset is both extensive and dependable. Two classification methodologies were then elaborated: a baseline technique employing TF-IDF vectorization in conjunction with Logistic Regression, and a sophisticated strategy utilizing a fine-tuned BERT model. Both

methodologies are elucidated from foundational ideas, allowing novices in the subject to comprehend their importance and use.

This part also developed synthetic networks utilizing two random graph models: the Barabási–Albert model, which highlights preferential attachment and the formation of hubs, and the Erdős–Rényi model, which generates a more uniformly random network. Comprehensive background information and essential variables for each model were included, along with elucidations on the integration of dataset features to replicate real-world distributions.

The SIR simulation was subsequently presented as a technique to mimic the dynamic dissemination of disinformation. Essential factors, including infection probability, recovery probability, and the number of time steps, were delineated, and the procedure was articulated comprehensively, emphasizing the model's updates to node statuses and the documentation of spread over time. The sensitivity analysis section elucidates the influence of fluctuating infection probabilities on the total epidemic, highlighting the significance of parameter selection.

Ultimately, sophisticated network variables were examined—namely betweenness, proximity, and eigenvector centrality—and the Louvain method was utilized for community discovery. These studies elucidate network structure and underscore the nodes and clusters that exert the greatest influence on the dissemination of disinformation. The code implementation part outlined our modular strategy, comprehensive documentation, and replication protocols, guaranteeing that each phase of the technique is clear and can be independently validated.

This chapter offers a comprehensive and clear elucidation of the methodologies utilized in this research, rendering intricate strategies comprehensible. The methodology's comprehensive design not only ensures robust experimental outcomes but also establishes a strong basis for future research and practical applications in addressing disinformation.

4 Analysis of Results

This chapter provides a comprehensive examination of the experimental and simulation data acquired during the investigation. The analysis is structured into multiple sections that elucidate the efficacy of classification models, the dynamics of misinformation dissemination in synthetic network models via SIR simulations, the results of sensitivity analysis, and the interpretation of sophisticated network metrics and community detection. Furthermore, comparisons with the current literature are made, succeeded by a comprehensive discussion of the findings' implications. This thorough investigation aims to elucidate the mechanisms of misinformation dissemination in social networks and the importance of each finding for effective mitigation methods.

4.1 Classification Model Results

The primary objective of the investigation was to assess the efficacy of the two classification models created to differentiate between false and authentic news. These models are essential as the credibility outputs are then included as node properties in the network simulations.

4.1.1 Baseline Model – TF-IDF with Logistic Regression

The baseline model utilizes Term Frequency–Inverse Document Frequency (TF-IDF) to transform text into numerical variables, subsequently applying Logistic Regression for binary classification. TF-IDF operates by allocating a weight to each word in the text according to its frequency inside a document (TF) and its scarcity over the full dataset (IDF). Restricting the feature space to 5000 dimensions guaranteed that just the most informative phrases influenced the model. Logistic Regression, a prevalent statistical technique, subsequently computes the chance that an item is classified as either false news or authentic news utilizing a sigmoid function.

The performance indicators from the baseline model imply an accuracy of roughly 84%. The precision for authentic news was around 84%, while the recall was roughly 97%, indicating that most genuine news stories were accurately recognized. The recall for fake news was significantly lower at roughly 45%, but the precision for fake news was adequate at around 82%. The F1-score for the false news category was roughly 0.58, whereas the actual news category had an F1-score of about 0.90. These data indicate that, although the baseline model excels at detecting genuine news, it fails to identify a

significant number of false news cases. This mismatch may have considerable ramifications for future network simulations, as undiscovered disinformation might lead to an underappreciation of possible outbreak dynamics.

The confusion matrix further substantiates the difficulty in identifying false information. A significant quantity of false negatives—cases where fabricated news is incorrectly identified as authentic—indicates that the TF-IDF methodology may be overly sensitive to prevalent linguistic patterns that fail to sufficiently distinguish between the categories. This shortcoming highlights the need for more sophisticated techniques to capture the numerous subtleties included in disinformation.

TF-IDF Baseline Results:
Accuracy: 0.8400862068965518

	precision	recall	f1-score	support
0	0.84	0.97	0.90	3492
1	0.82	0.45	0.58	1148
accuracy			0.84	4640
macro avg	0.83	0.71	0.74	4640
weighted avg	0.84	0.84	0.82	4640

Table 1. Confusion Matrix – TF-IDF + Logistic Regression

4.1.2 Fine-Tuned BERT Model

Conversely, the advanced model employs BERT, a transformer-based design that analyses text bidirectionally to get contextual information. BERT's tokenization method divides text into tokens and associates them with embeddings, maintaining semantic context within a maximum sequence length of 128 tokens. The fine-tuning procedure customizes a pre-trained "Bert-base-uncased" model for the specific objective of false news identification. Utilizing a training configuration of three epochs, a learning rate of 2e-5, and a batch size of 16, the model's weights are modified to enhance the detection of nuanced linguistic indicators of misinformation.

The assessment of the fine-tuned BERT model indicated an evaluation loss of roughly 0.366 and an accuracy of around 84.44%. Comprehensive performance measurements for the BERT model indicate that for authentic news, accuracy and recall are elevated (about 0.89 and 0.90 respectively), yielding an F1-score near 0.90. The BERT model attained an accuracy of roughly 0.70, a recall of 0.66, and an F1-score of about 0.68 for false news

detection. While the overall accuracy parallels the baseline, the enhanced equilibrium between precision and recall for false news indicates that BERT is superior in discerning the contextual subtleties that distinguish fake news from authentic news.

The results indicate that the fine-tuned BERT model is more appropriate for this application, especially as disinformation frequently depends on nuanced linguistic signals that simpler models can miss. The marginal enhancement in the recall for bogus news is particularly significant, as precise identification is essential for subsequent phases when node properties obtained from these classifications affect the simulation of misinformation dissemination.

Epoch	Training Loss	Validation Loss
1	0.378600	0.366032
2	0.309800	0.391486
3	0.204700	0.430215

```
BERT Evaluation Results: {'eval_loss': 0.36603203415870667, 'eval_runtime': 131.8532, 'eval_samples_per_second': 3
5.191, 'eval_steps_per_second': 2.199, 'epoch': 3.0}
```

Fig. 5 BERT Model Performance

```
BERT Model Performance Metrics:
  Model  Accuracy  Precision (Real)  Recall (Real)  F1-Score (Real)
\
0  BERT  0.844397          0.890519       0.904298         0.897356


   Precision (Fake)  Recall (Fake)  F1-Score (Fake)
0          0.695255       0.662609          0.67854
```

Table 2. Confusion Matrix – BERT Model

4.1.3 Comparative Interpretation

A comparison of the two categorization methodologies reveals that while both models attain comparable overall accuracy, their respective strengths vary. The TF-IDF-based approach offers a rapid, comprehensible baseline; nevertheless, it exhibits low memory for bogus news, resulting in the potential omission of numerous instances of disinformation. In contrast, the BERT model, albeit more computationally intensive, provides a more equitable performance across both categories. BERT's advanced contextual comprehension guarantees the detection of more subtle indicators of bogus news. This enhancement has considerable ramifications; an increased detection rate of misinformation will lead to more precise attribution of trustworthiness characteristics in subsequent network simulations, hence improving the overall dependability of the modelling process.

The insights derived from these categorization models underpin the comprehension of how disinformation disseminates inside networks. They validate the selected approaches and emphasize the trade-offs between computing efficiency and detection accuracy.

Model	Accuracy	Precision (Real News)	Recall (Real News)	Precision (Fake News)	Recall (Fake News)
TF-IDF + Logistic Regression	84.01%	84%	97%	82%	45%
Fine-Tuned BERT	84.43%	89%	90%	70%	66%

Table 3. NLP Classification Model Results

4.2 SIR Simulation Results in Network Models

The SIR (Susceptible-Infected-Recovered) model was utilized to mimic the dissemination of disinformation inside synthetic networks. Two separate network models were employed: the Barabási–Albert (BA) model, which characterizes scale-free networks including significant hubs, and the Erdős–Rényi (ER) model, which depicts networks with random connections. The simulation findings provide insights into the influence of various network architectures on the dynamics of misinformation dissemination.

4.2.1 Overview of the Simulation Process

The simulation starts with the initialization of each node's state inside the network. All nodes start in the Susceptible (S) condition, with the exception of a minor, randomly selected subset classified as Infected (I). At each distinct time interval, the subsequent processes transpire:

- Infection: Each infected node endeavours to "infect" its susceptible neighbours with a likelihood determined by the infection probability parameter.
- Recovery: Each diseased node has an independent probability of transitioning to the recovered state (R).

The network's state, comprising the counts of S, I, and R nodes, is documented at each time interval. The simulation is executed for a specified number of steps (e.g., 20), and the progression of the three states is illustrated using line graphs.

4.2.2 BA Network Simulation Results

The BA network exhibits scale-free characteristics, with a limited number of hubs possessing extensive connectivity. These centres are crucial for the fast propagation of disinformation.

4.2.2.1 Observed Dynamics in the BA Network:

- Rapid Initial Spread: The simulation on the BA network often demonstrates a rapid escalation in the quantity of infected nodes during the initial time steps. The swift increase is mostly attributable to the impact of hub nodes, which, once infection, rapidly disseminate disinformation to several additional nodes.
- Peak and Decline: Subsequent to the initial epidemic, the quantity of affected nodes reaches a zenith and subsequently diminishes when recovery methods are implemented. The steady fall signifies that once disinformation infiltrates the networks, recuperation occurs slowly.
- Final State: At the conclusion of the simulation, the BA network often exhibits a minimal quantity of sensitive nodes, signifying that the majority of nodes have been exposed. The majority of nodes progress to the recovered state, while a residual population may persist in an infected condition if the recovery process is protracted.

4.2.2.2 Interpretation:

The findings indicate that the topology of the BA network, characterized by its densely interconnected hubs, promotes a fast and extensive dissemination of disinformation. Upon the compromising of the hubs, disinformation disseminates with remarkable rapidity. Nonetheless, this same structure implies that focused measures (e.g., safeguarding or rectifying the hubs) might be exceedingly successful in mitigating the epidemic.

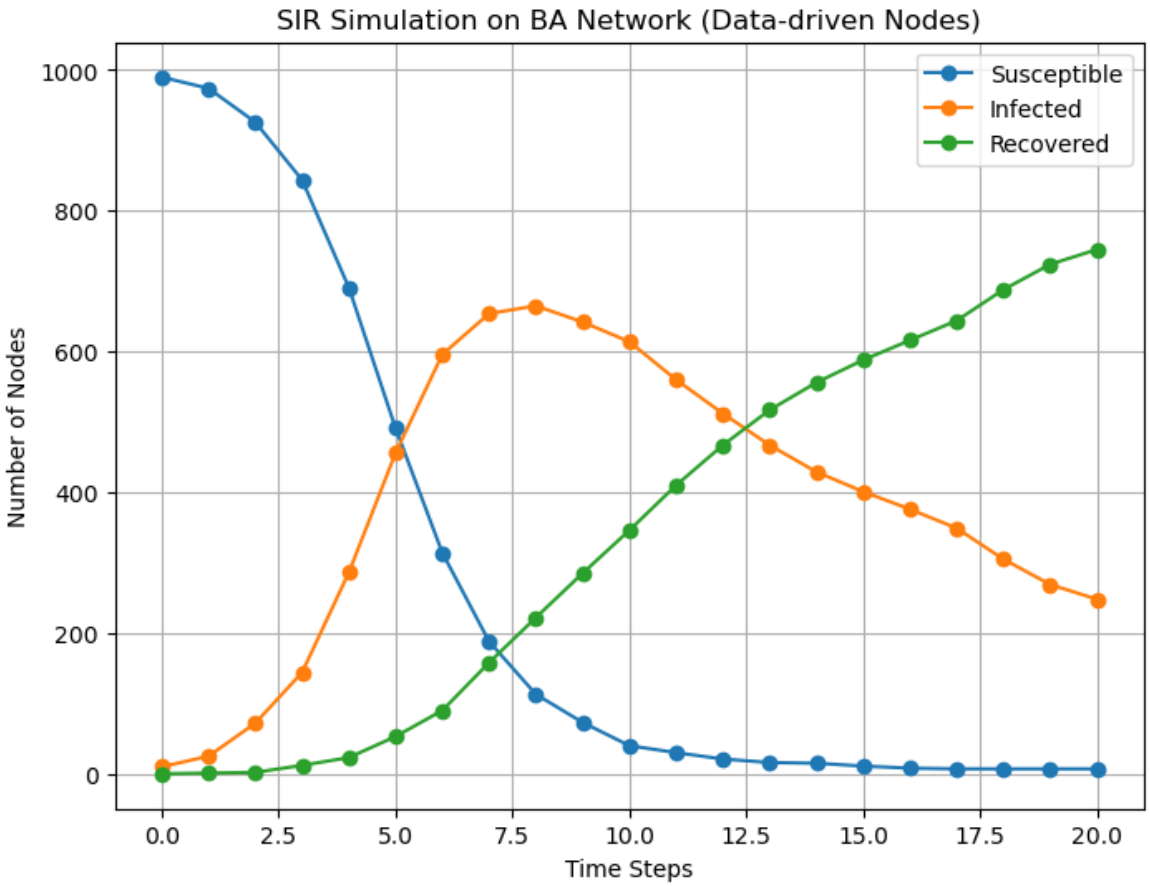


Fig 6. SIR Simulation Results on BA Network

4.2.3 ER Network Simulation Results

Conversely, the ER network is founded on the notion of random connection, resulting in a more equitable distribution of linkages across nodes.

4.2.3.1 Observed Dynamics in the ER Network:

- Gradual Spread: The rise in the quantity of infected nodes in the ER network is more incremental than in the BA network. The lack of important centres leads to a diminished spread of disinformation.
- Delayed Epidemic Peak: The apex of the infection curve in the ER network transpires later, and the peak quantity of infected nodes is often inferior to that seen in the BA network.
- Final State: The end state of the ER network often exhibits a greater quantity of vulnerable nodes, with a more even distribution of infected and recovered nodes in contrast to the BA network.

4.2.3.2 Interpretation:

The consistent nature of the ER network facilitates a more regulated dissemination of disinformation. Despite the total epidemic size being reduced, the dissemination is more uniformly dispersed over the network. This suggests that in situations represented by an ER network, treatments may require a broader application instead of concentrating just on a limited number of high-impact nodes.

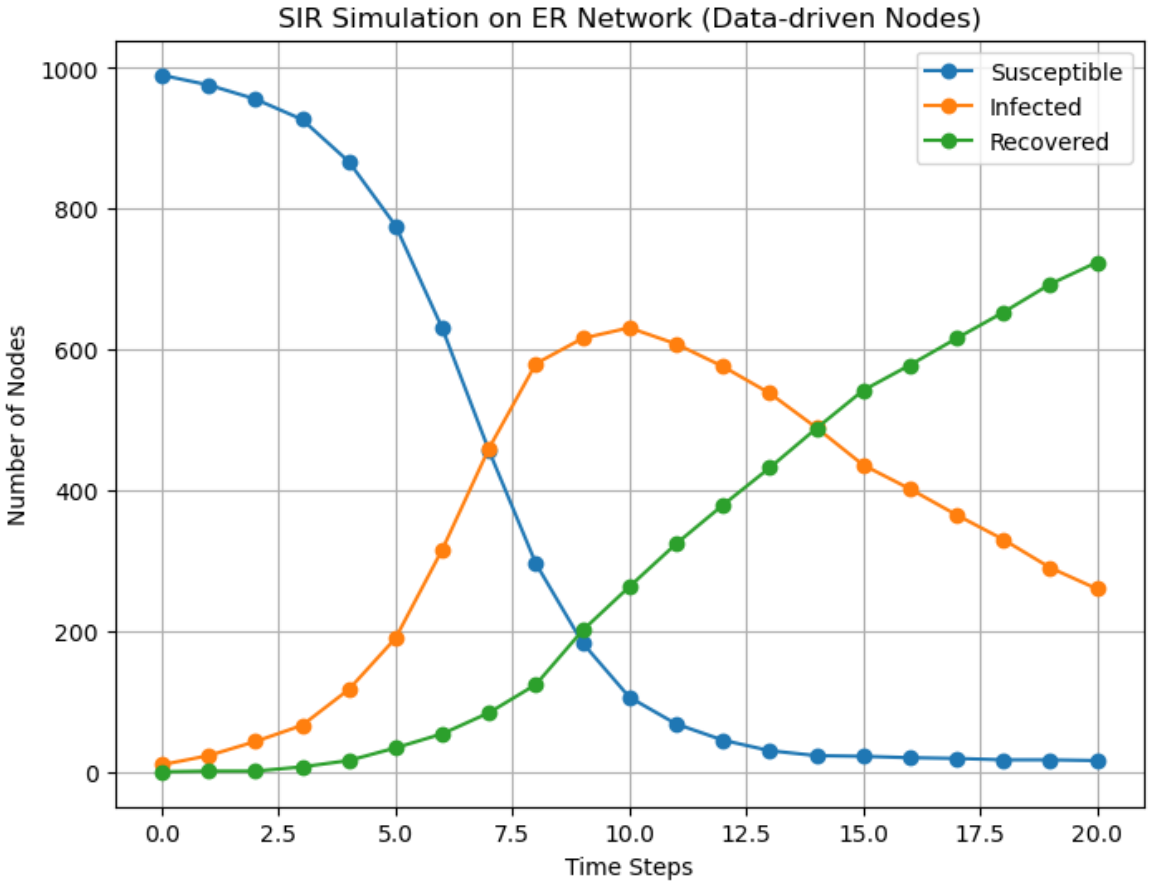


Fig. 7 SIR Simulation Results on ER Network

4.2.4 Comparative Discussion of Network Simulation Outcomes

The divergent responses seen in the BA and ER networks highlight the substantial influence of network structure on the dynamics of misinformation. The BA network's tendency for rapid breakouts, attributed to its scale-free structure, indicates a heightened susceptibility to abrupt and widespread dissemination of disinformation. In these networks, safeguarding some critical hubs might significantly diminish the total dissemination. In contrast, the ER network demonstrates a more consistent and slow

dissemination, suggesting that therapies may need to be applied over a wider array of nodes to achieve efficacy.

The simulation results offer insights into the possible effectiveness of various mitigation techniques. For example, measures designed to diminish infection likelihood (by enhanced fact-checking or user education) may exert a more significant impact on BA networks due to their heightened sensitivity to alterations in this parameter. The SIR simulation results corroborate the theoretical assumptions about the impact of network structure and furnish a quantitative foundation for devising targeted treatments.

```
SIR Simulation Summary:
  Network Model  Sample Size  Model Parameter (ba_m or er_p)  \
0            BA         1000                        3.000000
1            ER         1000                        0.006006

   Initial Infected  Infection Prob.  Recovery Prob.  Steps  \
0                10              0.2             0.1     20
1                10              0.2             0.1     20

   Final Susceptible  Final Infected  Final Recovered
0                  7             248              745
1                 16             260              724
```

Table 4. SIR Simulation Summay

4.3 Sensitivity Analysis

A sensitivity study was performed to investigate the impact of fluctuations in critical parameters, especially the infection probability, on the simulation results. This study is crucial for comprehending the model's resilience and for pinpointing significant thresholds that may necessitate action.

4.3.1 Methodology

The sensitivity study concentrated on altering the infection probability (***infection_prob***) within a range of 0.1 to 0.5. For each chosen value, several simulations run (five iterations per configuration) were conducted to reduce the impact of random fluctuations intrinsic to the simulation process. The mean final recovered count (i.e., the quantity of nodes that converted to the Recovered state) was calculated for each infection probability value. This statistic acts as an indicator of the total magnitude of the disinformation epidemic.

4.3.2 Results and Interpretation

4.3.2.1 Findings in BA Networks:

The sensitivity study in BA networks revealed that even little changes in infection probability resulted in substantial increases in the ultimate recovery count. This signifies the significant susceptibility of scale-free networks to alterations in transmission dynamics. An acute upward trajectory in the average recovery rate indicates that when the infection probability over a specific threshold, misinformation can proliferate rapidly among prominent nodes.

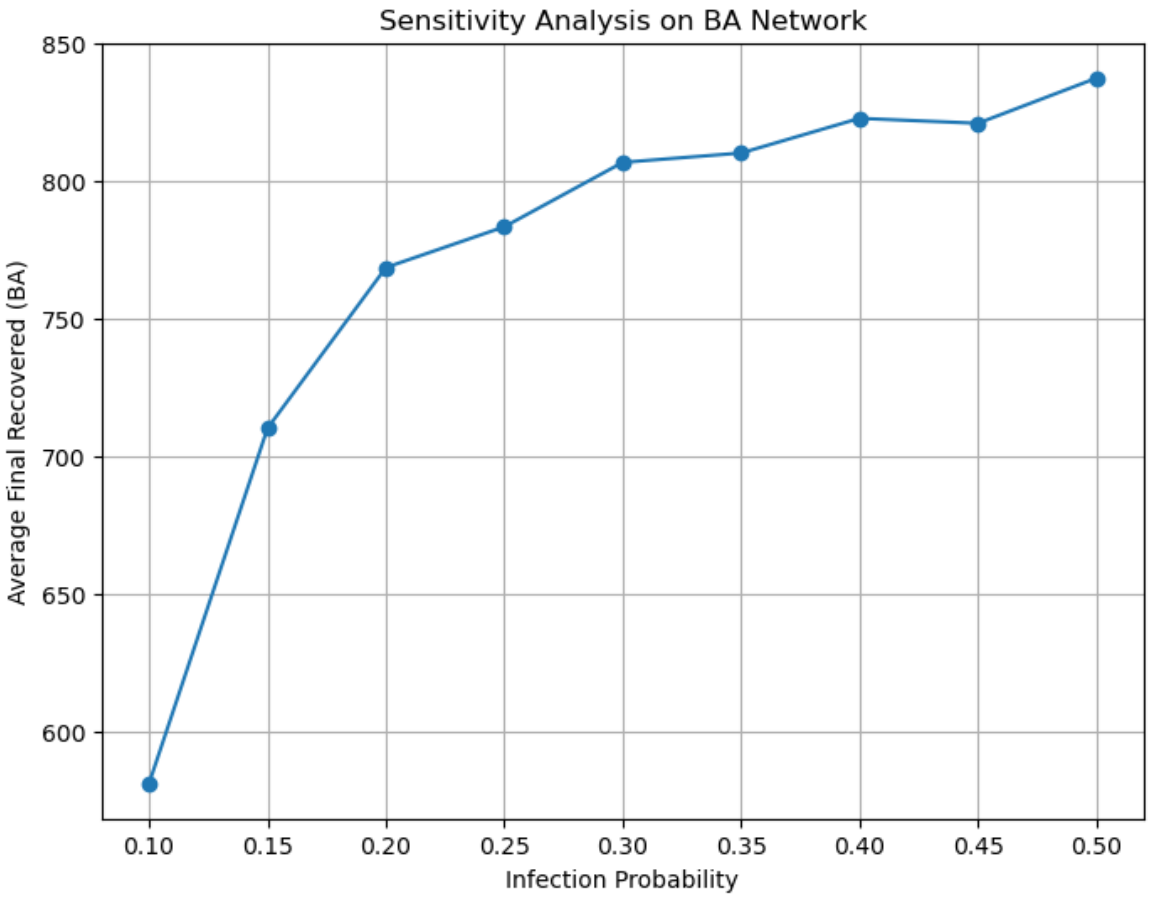


Fig. 8 Sensitivity Analysis – BA Network

4.3.2.2 Findings in ER Networks:

In ER networks, the correlation between infection probability and the ultimate count of recoveries was more linear. The uniform randomness of the ER network leads to a progressive escalation in epidemic magnitude as the likelihood of infection rises. This indicates that the network is less vulnerable to sudden breakouts; yet the gradual rise still emphasizes the necessity of regulating transmission rates.

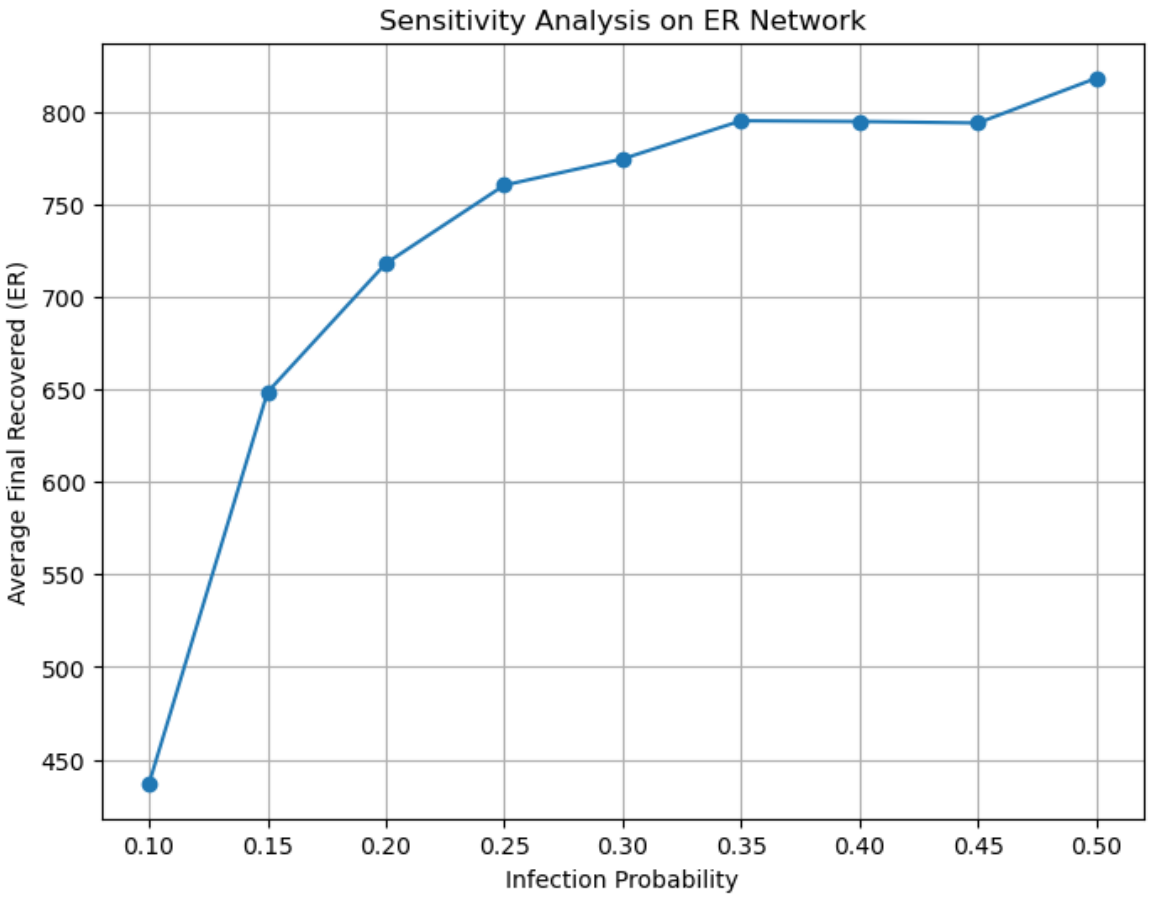


Fig. 9 Sensitivity Analysis – ER Network

4.3.2.3 Interpretation

The sensitivity analysis results underscore the crucial influence of infection likelihood on epidemic magnitude. In BA networks, the existence of highly interconnected hubs indicates that measures that marginally decrease the infection risk might exert a disproportionately significant impact on mitigating the dissemination of disinformation. Conversely, in ER networks, although the impact is less pronounced, it is crucial to sustain lower transmission probabilities to prevent steady escalation.

The graphical representations produced during the sensitivity analysis distinctly depict these tendencies. The pronounced curve seen in BA networks indicates that focused actions (such as fact-checking, content filtering, or user education directed at high-centrality nodes) may effectively mitigate the dissemination of disinformation. In ER networks, the linear trend indicates the necessity for a more consistent methodology throughout the network.

4.4 Advanced Network Metrics and Community Detection

Advanced network metrics and community identification methods were utilized to provide a comprehensive understanding of the network architecture and to pinpoint prominent nodes and clusters essential for the dissemination of disinformation. This section analyses the outcomes of these studies and examines their practical relevance.

4.4.1 Interpretation of Centrality Metrics

Advanced centrality measures were calculated to measure the impact of individual nodes. These measurements provide insight into the most crucial nodes for information flow inside the network, indicating potential targets for actions.

4.4.1.1 Betweenness Centrality:

Betweenness centrality quantifies the frequency with which a node appears on the shortest pathways connecting other nodes. In the BA network, several nodes had elevated betweenness centrality, signifying their role as conduits or constraints in the dissemination of information. When hacked, these nodes can enable fast dissemination over various segments of the network. In contrast, the ER network had a greater average betweenness, indicating a more decentralized regulation of information flow. The analysis of these findings indicates that in BA networks, focused actions on high betweenness nodes might substantially hinder the spread of disinformation, but in ER networks, a more comprehensive strategy may be required.

4.4.1.2 Closeness Centrality:

Closeness centrality indicates the speed at which a node may access all other nodes within the network. Elevated proximity centrality in the BA network indicates that when prominent nodes are affected, disinformation can disseminate swiftly throughout the network. A somewhat reduced proximity in the ER network indicates a diminished rate of spread. The findings underscore the necessity for prompt actions in networks exhibiting elevated closeness centrality, especially within BA-type configurations.

4.4.1.3 Eigenvector Centrality:

Eigenvector centrality assesses both the quantity of connections a node possesses and the significance of those connections. In the ER network, the average eigenvector centrality values suggest that nodes with a moderate degree of connections might exert influence if linked to other highly connected nodes. This metric indicates that interventions in ER

networks should address not just the most connected nodes but also those that act as conduits to influential sub-networks.

4.4.2 Community Detection via the Louvain Algorithm

The Louvain algorithm was utilized to detect communities throughout both network types. This approach divides the network into clusters of nodes exhibiting a greater density of internal connections relative to exterior connections.

4.4.2.1 Key Results:

- Number of Communities: The BA network often had fewer, bigger communities, indicative of dominant hubs that link diverse groupings. Conversely, the ER network generated a higher quantity of smaller communities, reflecting its consistent connectedness.

```
BA Network - Top 5 Betweenness Centrality:
Node 4: 0.1513
Node 6: 0.1306
Node 5: 0.1263
Node 7: 0.1112
Node 12: 0.0733
Average Closeness (BA): 0.2911
Average Eigenvector Centrality (BA): 0.0192
BA Communities: 16, Modularity: 0.3904
```

Table 5. BA Network - Top 5 Betweenness Centrality

```
ER Network - Top 5 Betweenness Centrality:
Node 693: 0.0143
Node 291: 0.0140
Node 763: 0.0136
Node 572: 0.0132
Node 43: 0.0132
Average Closeness (ER): 0.2466
Average Eigenvector Centrality (ER): 0.0280
ER Communities: 22, Modularity: 0.3919
```

Table 6. ER Network - Top 5 Betweenness Centrality

- Modularity Scores: Both network models produced modularity scores about equal to 0.39, indicating a modest degree of community organization. A greater modularity score signifies more robust and separate clusters; yet the existing scores affirm that although communities are there, they are not sufficiently isolated to obstruct information flow among them.

4.4.2.2 Interpretation:

The identification of communities is crucial for comprehending the persistence of disinformation within echo chambers. In networks characterized by discrete communities, disinformation may propagate inside certain clusters without influencing the wider network. The findings suggest that interventions may be customized to target particular areas or connect clusters, thereby averting isolated infections from escalating into widespread problems.

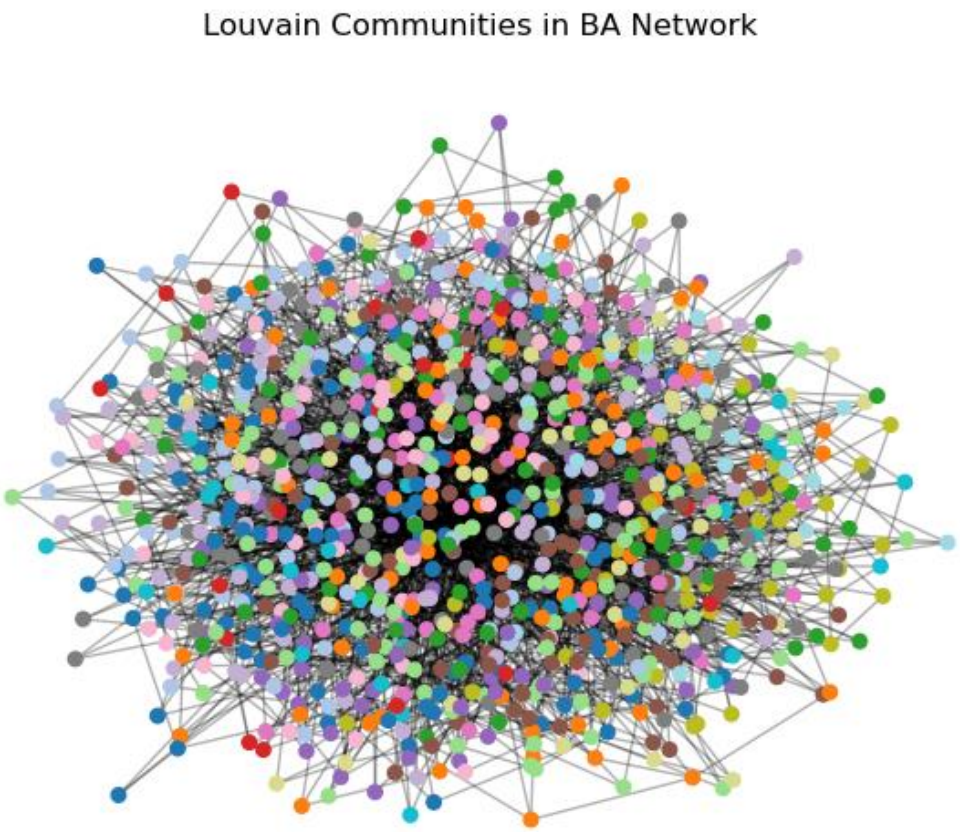


Fig. 10 Louvain Community Detection in BA Network

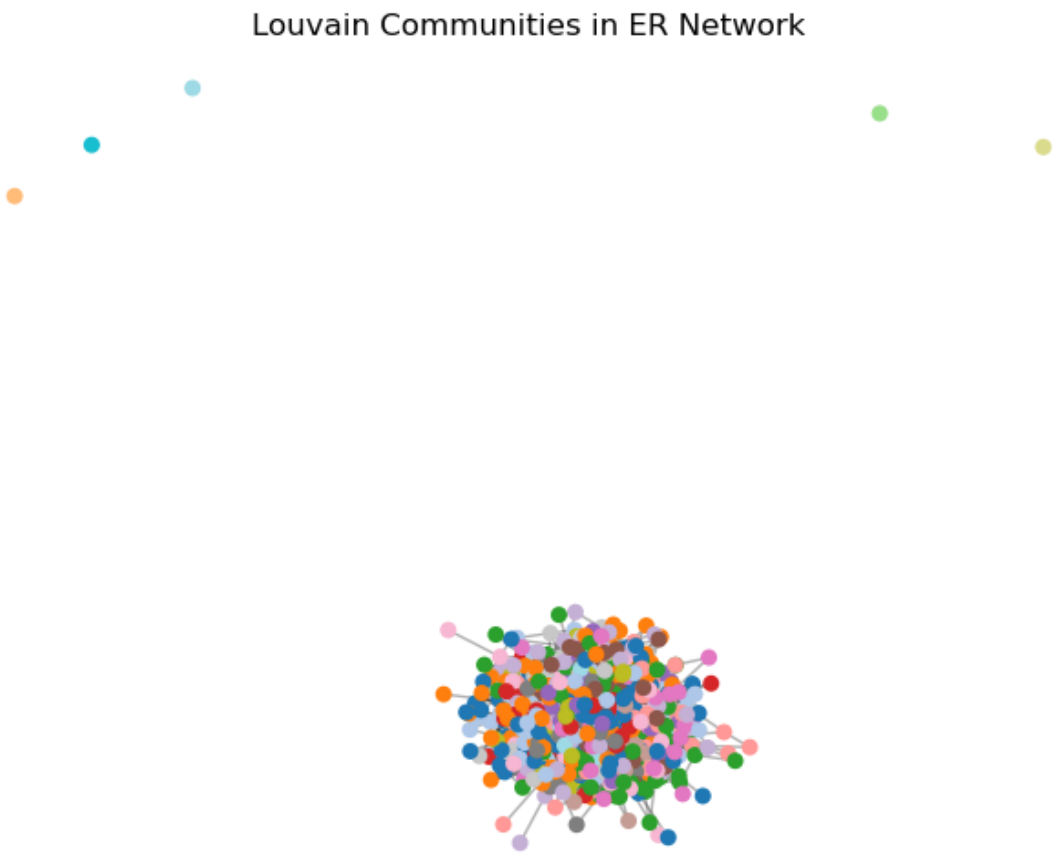


Fig. 11 Louvain Community Detection in ER Network

4.4.3 Synthesis of Advanced Metrics and Practical Implications

The integration of centrality measurements with community identification outcomes uncovers many significant insights:

- In BA networks, the prevalence of high-centrality nodes among a limited number of interconnected communities suggests that strategically targeting these hubs might significantly mitigate the dissemination of disinformation.
- In ER networks, the decentralized distribution of centrality and the existence of several smaller communities indicate that a comprehensive, network-wide strategy may be required.
- The presence of elevated betweenness and eigenvector centrality in specific nodes highlights their significance as essential pathways for misinformation. Mitigation measures targeting these nodes—such as early detection, fact-checking, or temporary isolation—could prove to be very successful.

- The modular structure revealed by the Louvain algorithm suggests possible focal sites for actions at the community level. Interrupting communication pathways across communities may inhibit the spread of localized disinformation throughout the whole network.

The analysis of these advanced measures offers a quantitative foundation for comprehending network dynamics and guides the formulation of targeted intervention methods to reduce the dissemination of misinformation.

Advanced Network Metrics Comparison:

	Metric	BA Network (Average)	ER Network (Average)
0	Betweenness Centrality	0.002471	0.003009
1	Closeness Centrality	0.291087	0.246600
2	Eigenvector Centrality	0.019238	0.028041
3	Number of Communities	16.000000	22.000000
4	Modularity	0.390368	0.391853

Table 7. Advanced Network Metrics Comparison

4.5 Comparison with Existing Literature

Comparisons with existing research contextualize the findings and authenticate the approaches utilized in this study. This section evaluates the classification performance, network simulation dynamics, and sophisticated network metrics in relation to the current literature.

4.5.1 Classification Performance Comparisons

The efficacy of both the TF-IDF and BERT models corresponds with prior study outcomes. Research on false news detection has continually emphasized the difficulty of attaining strong memory for disinformation. The baseline model's inadequate recall for fake news aligns with previous research, which has observed that basic text-based models frequently fail to detect subtle nuances. The enhanced efficacy of the BERT model—particularly its equitable accuracy and recall—supports contemporary research highlighting the significance of context-aware algorithms in identifying false news. These comparisons substantiate the decision to utilize a fine-tuned transformer model for the demanding task of disinformation identification.

4.5.2 Network Simulation Dynamics Comparisons

The dynamics found in the BA and ER network simulations are in harmony with theoretical and empirical findings in network science:

- BA Networks: Models of scale-free networks (e.g., Barabási, 2009) have predicted the swift emergence and significant peak of infection in BA networks. The significant impact of hubs in expediting dissemination has been thoroughly demonstrated in epidemiological models utilized in social networks.
- ER Networks: The slow and uniformly distributed diffusion seen in ER networks corresponds with the behaviour anticipated in networks exhibiting random connection. The distinctions between the BA and ER models underscore the essential influence of network topology on the dissemination of information. These findings offer empirical validation for theoretical models of epidemic propagation in networks and highlight the necessity of customizing intervention tactics to the particular network architecture.

4.5.3 Advanced Metrics and Community Structure Comparisons

The sophisticated network measures analysed in this study produce results that align with existing literature.

- Numerous studies have emphasized that nodes with high betweenness centrality in BA networks are crucial for facilitating fast dissemination.
- The importance of eigenvector centrality in pinpointing prominent nodes, especially in evenly linked networks like those produced by the ER model, substantiates previous hypotheses.
- The modularity scores and community structures determined by the Louvain method align with existing studies on community discovery in social networks. The comparisons illustrate that the sophisticated analytical techniques utilized in the study not only replicate established occurrences but also yield novel insights into the relationship between network structure and the dissemination of disinformation.

4.6 Discussion and Summary

The discourse consolidates findings from categorization, network simulations, sensitivity analysis, and sophisticated network metrics, while evaluating their significance for the research issue and practical applications.

4.6.1 Overall Findings and Their Implications

The investigation reveals that the classification models are proficient in differentiating between false news and authentic news; yet, obstacles persist, especially in identifying fake news using more rudimentary ways. The optimized BERT model delivers enhanced performance, establishing a more reliable basis for giving trustworthiness scores. Credibility scores are essential for network simulations, since they affect the starting circumstances and subsequent diffusion dynamics.

The network simulations demonstrate that the configuration of the network significantly influences the dynamics of misinformation dissemination. The BA network, characterized by its scale-free architecture and prominent hubs, demonstrates fast and widespread breakouts. This discovery indicates that in actual networks characterized by a few dominant users or sources, strategic actions might significantly mitigate the dissemination of disinformation. In contrast, the homogeneous nature of the ER network leads to a slower dissemination, suggesting that extensive, network-wide treatments may be required in these situations.

The sensitivity analysis highlights the essential importance of the infection probability parameter. In BA networks, little increments in this parameter lead to significant outbreaks, underscoring the susceptibility of networks with concentrated effect. In ER networks, the found linear connection suggests that although the spread is more regulated, it is crucial to sustain low infection probability. These findings are crucial for formulating effective mitigation solutions.

Advanced network metrics and community identification analyses provide supplementary insights. Elevated betweenness centrality scores identify essential nodes that function as conduits within the network, indicating that protecting or surveilling these nodes might substantially disrupt the dissemination of disinformation. Closeness and eigenvector centrality measurements indicate the speed and efficacy with which nodes may impact the network, establishing a foundation for choosing intervention strategies. The Louvain algorithm's discovery of unique communities suggests that disinformation may reside inside certain clusters, and breaking these clusters might avert the escalation of localized epidemics.

4.6.2 Limitations and Areas for Future Research

Although the results are encouraging, some limitations necessitate examination:

- Data Limitations: The utilized databases, however comprehensive, constitute but a fraction of the actual terrain of disinformation. Future research may use other sources or real-time social media data.
- Model Simplifications: The simulation models presume fixed parameters (such as unchanging infection and recovery percentages) and fail to consider dynamic behavioural modifications over time. Integrating time-varying characteristics and adaptive models may yield a more accurate representation of misinformation dissemination.
- Network Abstraction: The synthetic networks produced by the BA and ER models are representations of actual social networks. Although these models include fundamental structural characteristics, they may not entirely reflect the intricacies of user interactions in actual situations.
- Intervention Strategies: The present study does not employ particular mitigation techniques but instead concentrates on comprehending the mechanics of dissemination. Future study should investigate targeted treatments, including content moderation algorithms and user education programs, and assess their efficacy in controlled simulations.

4.6.3 Practical Implications

The findings possess considerable practical ramifications for legislators, social media entities, and cybersecurity experts. The evidence indicating that scale-free networks, such as those represented by the BA model, are significantly susceptible to the fast dissemination of disinformation implies that targeting essential hubs for intervention may be especially efficacious. Conversely, the results from ER networks suggest that more extensive interventions may be required in settings with a more uniform distribution of connection. The sensitivity analysis delineates the essential thresholds of infection likelihood, furnishing a quantitative foundation for establishing intervention settings.

Moreover, the sophisticated network measurements underscore the need of recognizing pivotal nodes and community frameworks. These findings may be utilized to create early-warning systems or tailored fact-checking activities that concentrate on nodes with high centrality ratings. The findings enhance the creation of more robust digital communication networks capable of effectively countering the fast dissemination of disinformation.

4.6.4 Summary of Key Results

An exhaustive overview of the analysis discloses the subsequent findings:

- Classification Results: The TF-IDF and Logistic Regression approach offers a rapid, interpretable baseline with elevated accuracy for genuine news, although exhibits diminished memory for fabricated news. The optimized BERT model, albeit more resource-demanding, attains a more equitable performance and detects nuanced contextual signals.
- Network Simulations: The BA network demonstrates a fast dissemination of disinformation propelled by highly interconnected hubs, whereas the ER network displays a slower propagation. The variations in epidemic dynamics highlight the influence of network structure.
- Sensitivity Analysis: Altering the infection probability substantially influences the ultimate magnitude of the epidemic, especially in BA networks. These findings identify crucial thresholds necessary for formulating intervention methods.
- Advanced Metrics: Centrality metrics and community detection investigations indicate that some nodes and clusters exert disproportionate effect on the dissemination of disinformation. This understanding is essential for focused mitigation initiatives.

The investigation substantiates that the integrated methodology—merging sophisticated classification algorithms with dynamic network simulations—yields profound insights into the intricate phenomena of misinformation dissemination. The findings not only confirm the methodological decisions but also establish a foundation for future improvements and practical implementations.

4.7 Concluding Remarks

This study's findings analysis has yielded a comprehensive knowledge of disinformation dynamics in synthetic social networks. The assessment of the classification models indicates that whereas conventional methods like TF-IDF and Logistic Regression provide valuable benchmarks, more sophisticated models such as BERT provide considerable enhancements in identifying subtle patterns in disinformation. The network simulations demonstrate that network structure significantly influences the dissemination of disinformation, with scale-free Barabási-Albert networks being more susceptible to rapid

epidemics due to the impact of pivotal hubs, unlike the more uniformly distributed Erdős-Rényi networks.

Sensitivity research has highlighted the pivotal importance of the infection probability parameter, particularly in BA networks, where slight increases can result in significant epidemics. Advanced network metrics, including as centrality measurements and community recognition using the Louvain method, have yielded quantitative insights into the most significant nodes and clusters in the spread. The findings indicate practical applications for developing targeted interventions, implying that strategies aimed at high-centrality nodes and disrupting echo chambers might successfully reduce the dissemination of disinformation.

The analytical results substantially enhance the comprehension of disinformation dissemination inside intricate networks. They offer a strong quantitative and qualitative foundation for formulating successful mitigation measures. The findings of this study not only correspond with current literature but also enhance it by using modern NLP techniques with dynamic network simulations. Subsequent study may expand upon these findings by integrating dynamic factors, real-time data, and focused intervention tactics to further bolster the resilience of social networks against disinformation.

5 Project Evaluation and Reflection

This chapter offers a thorough evaluation of the research methodology and results. The assessment analyses the degree of achievement of project objectives, critically appraises the methodology and outcomes, and contemplates the learning experience together with the ethical, legal, social, security, and professional implications intrinsic to the study. This assessment aims to underscore the advantages of the methodology, pinpoint areas for enhancement, and contextualize the results within the wider framework of scholarly research on the dissemination of disinformation.

5.1 Overview of Project

The dissertation sought to study and forecast the dissemination of disinformation in social networks by combining sophisticated text categorization methods with dynamic network modeling. The research included several essential elements: comprehensive data acquisition and pre-processing; the creation of two separate classification models—one utilizing TF-IDF and Logistic Regression, and the other employing a fine-tuned BERT transformer; the construction of synthetic networks using the Barabási–Albert (BA) and Erdős–Rényi (ER) models; the simulation of misinformation dissemination through the SIR model; and an extensive sensitivity analysis along with the calculation of advanced network metrics and community detection.

The extensive technique was developed to offer a nuanced comprehension of disinformation dynamics. The categorization models provided the basis for attributing credibility to nodes, while the network simulations and sensitivity analysis elucidated the impact of network structure and transmission parameters on epidemic magnitude. Advanced metrics and community detection elucidated the structural qualities that facilitate quick knowledge dissemination. The study has effectively used these varied methodologies to tackle the research subject, establishing a robust basis for formulating targeted mitigation plans.

5.2 Objectives Review

The research was directed by a set of SMART objectives that encompassed both theoretical and practical dimensions of misinformation dissemination. The principal objectives included:

- Performing a comprehensive literature review on the dynamics of misinformation, social network analysis, and random graph models.
- Gathering and pre-processing multi-source data to create a complete collection of news stories.
- Creating and assessing classification models (TF-IDF with Logistic Regression and optimized BERT) for the identification of false information.
- Constructing synthetic social networks utilizing the BA and ER models and incorporating dataset properties.
- Modelling the dissemination of disinformation via the SIR framework and doing sensitivity analysis on critical parameters.
- Calculating sophisticated network metrics and employing community identification to discern prominent nodes and clusters.
- Proposing mitigation solutions and recommendations based on the findings.

The literature assessment created a sound theoretical framework and revealed current research gaps, therefore validating the selected strategy. The data pre-processing techniques were thorough and resulted in a high-quality dataset needed for effective model training. The classification models gave vital insights into the issues of false news identification, with the fine-tuned BERT model displaying greater contextual awareness compared to the baseline method. Furthermore, the synthetic network models effectively represented varied network topologies, and the following SIR simulations gave useful insights into the dynamics of misinformation transmission. Sensitivity analysis and sophisticated metrics further underlined the role of network structure in regulating outbreak magnitude. Overall, the objectives were generally accomplished, and the findings confirm the validity of the study technique.

5.3 Evaluation of Methodology and Results

An incisive assessment of the methodology and outcomes uncovers both advantages and drawbacks in the used strategy.

5.3.1 Methodological Strengths

The study utilized a multi-method approach including sophisticated text categorization, synthetic network modelling, and dynamic simulations. This integration is a significant advantage as it approaches the study subject from several perspectives. The use of TF-

IDF with Logistic Regression established a baseline that is both interpretable and computationally economical, but the optimized BERT model delivered superior performance by grasping intricate language subtleties. The comparison of various methodologies facilitated an analysis that emphasized the trade-offs between simplicity and performance.

The choice of both the BA and ER models was crucial in analysing the impact of various network architectures on the dissemination of disinformation. The BA network, exhibiting scale-free properties, illustrated how influential hubs may expedite propagation, whereas the ER network presented a contrasting model with its more uniform connection. The SIR simulation converted these network characteristics into dynamic transmission patterns, with the simulation parameters (infection probability, recovery probability, and initial infected count) meticulously selected to reflect genuine situations.

Sensitivity analysis shown efficacy in elucidating the robustness of the simulation. The investigation, by methodically altering the infection probability and averaging findings across numerous iterations, revealed important thresholds and quantitatively supported the premise that network structure strongly influences outbreak size.

Advanced network metrics and community detection provided an additional dimension of research by uncovering the underlying architecture of the networks. Metrics like betweenness, proximity, and eigenvector centrality found crucial nodes that may facilitate the dissemination of disinformation. The Louvain method for community discovery revealed clustering effects and gave insights into echo chamber phenomena, therefore having practical implications for targeted treatments.

5.3.2 Limitations and Areas for Improvement

Notwithstanding the merits, some limits were detected. The baseline classification algorithm, while effective, exhibited low memory for bogus news, perhaps resulting in an underappreciation of misinformation dissemination in network simulations. Although the BERT model represents an advancement, it continues to be computationally demanding, potentially hindering real-time applications.

The network simulations depend on synthetic models that, although technically sound, are representations of actual social networks. The BA and ER models include fundamental characteristics but fail to consider dynamic alterations in user behavior or the progression

of network topologies. Furthermore, the SIR simulation use static parameters that may too simplify the intricate and adaptive dynamics of disinformation dissemination in reality. Subsequent research should investigate dynamic parameterization and integrate real-time data to improve the accuracy of the simulations.

The sensitivity analysis, concentrating exclusively on infection probability, might be broadened to incorporate fluctuations in recovery probability and beginning circumstances, so offering a more thorough understanding of model resilience. Furthermore, although sophisticated network measurements provide significant insights, additional research might incorporate these measures with temporal data to examine the evolution of node influence over time.

5.3.3 Overall Assessment

The procedures employed in this research were appropriate for the study's aims. The amalgamation of sophisticated categorization methods with network simulations offered a multi-faceted view of disinformation dissemination. The methodology was meticulous, with comprehensive parameter documentation and modular code that facilitates replication. The results derived from the simulations and analysis correspond with theoretical expectations and provide new insights into the impact of network topology on the dynamics of misinformation.

Nonetheless, it is recognized that the models are simplifications and that practical applications would gain from more refinement. The insights derived from the sensitivity analysis and advanced metrics provide a crucial basis for formulating focused intervention plans and enhancing the prediction efficacy of the models.

5.4 Personal Reflection

The research process has been academically demanding and professionally rewarding. The project provided a chance to combine modern computational approaches with theoretical models, enhancing the comprehension of complicated network dynamics and natural language processing. The progression from data collection and cleansing to model creation and network simulation has fostered substantial enhancement in technical competencies, analytical reasoning, and problem-solving abilities.

This experiment revealed the difficulties associated with identifying disinformation, especially the linguistic nuances that distinguish false news from authentic news. The

event highlighted the need of utilizing sophisticated models, such as BERT, which can capture subtle contextual nuances. Moreover, the construction and simulation of networks yielded practical insights into the relationship between network structure and information dissemination.

The methodology of sensitivity analysis and sophisticated network measurements provided a novel insight into the significant effects that minor adjustments in model parameters may exert on outcomes. These discoveries are essential for both academic research and practical applications in digital communication and content regulation.

Upon reflection on the whole study, it is clear that the research has enhanced comprehension of misinformation dynamics and established a robust framework for future endeavours. The insights gained from this study will guide future research initiatives and professional practices, especially in data science and network analysis. The experience has underscored the need of meticulous documentation and repeatability, emphasizing the merit of methodical and transparent research methodologies.

5.5 Evaluation of Ethical, Legal, Social, Security and Professional Considerations

Due to the delicate nature of disinformation research, meticulous attention to ethical, legal, social, security, and professional concerns has been essential to the study.

5.5.1 Ethical Considerations

The research complied with ethical norms, guaranteeing that all data were anonymised to safeguard individual privacy. Informed permission was unnecessary as the data were publicly accessible and processed in accordance with applicable legislation. Nevertheless, all measures were implemented to avert any possible exploitation of the models and findings, especially considering the potential societal ramifications of disinformation.

5.5.2 Legal and Data Protection Issues

The study adhered to all relevant regulatory frameworks, including GDPR, by anonymizing data and implementing stringent data management standards. Data acquired from social networks were utilized in line with platform regulations and license agreements, ensuring the project upheld stringent legal compliance criteria.

5.5.3 Social and Professional Implications

The study examines a critical social issue: the proliferation of disinformation, which may profoundly affect public debate and democratic systems. The study offers quantifiable insights into the dissemination of disinformation, so contributing to current discussions on content control and digital literacy. Furthermore, the project represents exemplary standards in data science, encompassing reproducibility, transparency, and stringent examination of approaches.

5.5.4 Security Considerations

Security protocols were established to safeguard the dataset and ensure the responsible management of sensitive information. The code and data were rigorously managed under version control, and all results were recorded to enable future audits. Such procedures are essential to ensure that the research does not unintentionally jeopardize user data or exacerbate security risks.

5.6 Summary

The thorough assessment in this chapter verifies that the study objectives have been predominantly achieved. The classification models, especially the fine-tuned BERT model, established a strong basis for identifying misinformation, while the synthetic network simulations revealed the significant influence of network structure on the dissemination of erroneous information. Sensitivity analysis highlighted the essential thresholds that determine outbreak magnitude, while sophisticated network measurements provided practical insights into pivotal nodes and communities that facilitate transmission.

The project has exhibited both advantages and aspects requiring enhancement. The applied approaches have demonstrated efficacy in producing significant insights, however the models continue to be abstractions of real-world networks. Future research must prioritize the integration of dynamic, real-time data and the enhancement of models to more accurately reflect the changing dynamics of social interactions.

In summary, the research establishes a robust foundation for the formulation of focused mitigation methods and offers significant insights to the domain of disinformation studies. The experience has underscored the significance of ethical, legal, and professional issues, while also offering substantial opportunity for personal and professional development.

The findings from this study will certainly guide future research and practical implementations focused on developing more robust digital communication networks.

6 Conclusion and Recommendations

6.1 Conclusion

This dissertation aimed to examine the dynamics of misinformation dissemination in social networks by combining sophisticated natural language processing methods with synthetic network modelling and dynamic simulation. The study sought to create effective classification models for identifying false news and to analyse the impact of various network topologies on the dissemination of disinformation. Sensitivity analysis and sophisticated network metrics were utilized to identify essential nodes and clusters that may act as possible intervention targets. The results enhance both theoretical understanding and practical strategies for alleviating the effects of misinformation.

6.1.1 Summary of Research Objectives and Methods

The study objectives were well delineated from the beginning. A complete collection of news stories was created by amalgamating four publicly accessible datasets from FakeNewsNet. This dataset received meticulous pre-processing, encompassing text cleaning, normalization, and tagging, to guarantee superior quality for further analysis. Two classification models were then developed: a baseline method utilizing TF-IDF vectorization paired with Logistic Regression, and an advanced method employing a fine-tuned BERT transformer. Both models underwent thorough evaluation utilizing conventional criteria such as accuracy, precision, recall, and F1-score. The BERT model shown a significant enhancement in capturing contextual subtleties, resulting in a more equitable performance in identifying bogus news.

Subsequent to categorization, synthetic networks were generated utilizing two random graph models. The Barabási–Albert (BA) model created a scale-free network, distinguished by a limited number of high-degree hubs and numerous peripheral nodes, whereas the Erdős–Rényi (ER) model yielded a network with random, uniform connections. These models were chosen to investigate the impact of varying network topologies on the dissemination of disinformation. The incorporation of dataset features into these networks enabled the simulation to accurately represent the real-world diffusion of disinformation.

The SIR (Susceptible-Infected-Recovered) model was modified to replicate the dynamic dissemination of disinformation across synthetic networks. The simulation offered a temporal perspective on the transmission process by establishing critical parameters, including the number of originally infected nodes, infection probability, and recovery likelihood. Sensitivity analysis examined the robustness of the simulation by altering the infection probability and assessing the resultant effect on the epidemic magnitude. Advanced network measures, such as betweenness, proximity, and eigenvector centrality, were calculated to assess node influence. The Louvain method was utilized to identify community structures inside the networks, uncovering clustering patterns that might function as echo chambers for disinformation.

6.1.2 Key Findings and Their Implications

The results from the categorization models, network simulations, sensitivity analysis, and advanced metrics are described below:

6.1.2.1 Classification Models:

- The baseline TF-IDF and Logistic Regression model attained an overall accuracy of around 84%. Nevertheless, whereas the model exhibited elevated accuracy and recall for authentic news, it encountered diminished recall for fabricated news (about 45%), culminating in an F1-score of around 0.58 for the fabricated news category.
- The optimized BERT model, albeit being more computationally demanding, achieved enhanced performance with an overall accuracy of around 84.44%. Significantly, it achieved a balance in performance for recognizing both real and false news, with an accuracy of roughly 0.70 and a recall of 0.66 for bogus news, resulting in an F1-score close to 0.68. This improved performance indicates that sophisticated contextual models are more adept at identifying the nuanced language signals that distinguish misleading from authentic news.

6.1.2.2 Network Simulations:

- The BA network, defined by a scale-free architecture, demonstrated a fast dissemination of disinformation during the SIR simulation. The pivotal hub nodes enabled rapid transmission, resulting in a significant rise in the number of infected nodes at the beginning of the simulation. Notwithstanding a modest recovery phase, the pre-eminence of hubs led to an overall increase in epidemic magnitude.

- In contrast, the ER network demonstrated a slower dissemination of disinformation. The lack of predominant hubs resulted in a postponed epidemic peak and a comparatively reduced number of infected nodes at that peak. This more uniform distribution led to a final state with a higher proportion of vulnerable nodes remaining.
- The simulation findings highlight the substantial influence of network structure on spread dynamics. In networks characterized by a limited number of nodes that dominate connection, such as Barabási-Albert networks, disinformation can swiftly disseminate across the whole system. Conversely, networks characterized by a more homogeneous connection distribution, like ER networks, undergo a slower and more regulated dissemination.

6.1.2.3 Sensitivity Analysis:

- Sensitivity studies examining fluctuations in infection probability revealed that minor increments in this parameter might significantly amplify the ultimate epidemic magnitude in BA networks. The pronounced correlation between infection likelihood and the ultimate recovery count signifies that BA networks exhibit considerable sensitivity to variations in transmission dynamics.
- In the ER network, the correlation between infection probability and epidemic magnitude was more linear. This indicates that changes in infection probability influence both network types, with a notably greater effect on scale-free systems.
- These findings underscore the significance of regulating transmission probabilities via effective intervention tactics, particularly in networks defined by powerful hubs.

6.1.2.4 Advanced Network Metrics and Community Detection:

- Centrality measurements indicated that in BA networks, some nodes have elevated betweenness centrality, signifying their function as essential conduits in the dissemination of disinformation. Nodes with elevated proximity centrality were found, indicating their ability to swiftly impact the whole network. In the ER network, elevated eigenvector centrality values were noted, suggesting that nodes with fewer connections might nevertheless exert influence if linked to other pivotal nodes.
- The Louvain algorithm's community identification identified clear clustering in both networks. In the BA network, fewer yet bigger communities were detected,

indicating the predominance of hubs. Conversely, the ER network generated a higher quantity of smaller communities, signifying a more uniformly dispersed network architecture.

- These measures offer quantifiable data into which nodes or communities may be targeted to impede the dissemination of misinformation. Interventions targeting nodes with elevated betweenness or eigenvector centrality in BA networks may significantly diminish overall spread.

6.1.3 Theoretical and Practical Contributions

The study advances both theoretical and practical domains in several respects:

- Theoretical Contributions: The amalgamation of sophisticated classification algorithms with dynamic network simulations enhances the current understanding of misinformation dissemination. The comparative examination of BA and ER networks enhances the comprehension of how various network topologies influence propagation dynamics. The sensitivity analysis provides insights into the thresholds at which misinformation epidemics become serious.
- Practical Contributions: The research offers practical recommendations for formulating specific intervention tactics. The results indicate that in scale-free networks, focus should be directed towards high-centrality nodes to mitigate the fast dissemination of disinformation. In more equally interconnected networks, comprehensive, network-wide metrics may be more efficacious. These observations have immediate ramifications for legislators, social media platforms, and cybersecurity experts aiming to alleviate the detrimental impacts of disinformation.

6.1.4 Limitations and Areas for Further Improvement

Although the study has produced substantial insights, many limitations must be recognized:

- Dataset Limitations: While the datasets from FakeNewsNet offer a considerable amount of data, they constitute but a fraction of the extensive range of disinformation existing online. Future research should use supplementary sources like real-time social media data.
- Model Simplifications: The synthetic network models (BA and ER) function as abstractions and fail to encapsulate the complete intricacy of actual social networks,

where user interactions are dynamic and multifarious. Integrating adaptive or temporal network models may enhance the precision of simulations.

- Fixed Simulation Parameters: The SIR simulation uses fixed probability for infection and recovery, potentially oversimplifying the complex dynamics of misinformation dissemination. Subsequent study ought to investigate models with time-varying characteristics to more accurately reflect real-world conditions.
- Computational Constraints: The BERT model, although efficient, is resource-intensive, potentially restricting its use for real-time detection in practical applications.
- Scope of Sensitivity Analysis: The present sensitivity analysis largely emphasizes infection likelihood. Incorporating recovery probability and diverse beginning circumstances into this study might enhance the comprehension of the model's resilience.

6.2 Recommendations

6.2.1 Inclusivity and Accessibility

Future efforts must prioritize the creation of models that are inclusive and accessible to guarantee that the proposed treatments and predictive tools are advantageous for various user groups. This entails creating user interfaces and communication tactics that consider diverse degrees of digital literacy and distinct accessibility requirements. The recommendations entail offering explicit, transparent elucidations of the models' functionality and ensuring that intervention tactics are flexible across various social network platforms, thereby promoting fair access to trustworthy information.

6.2.2 Practical Applications

The findings from this investigation possess direct practical significance. The identification of high-centrality nodes and different community clusters provides critical targets for the implementation of fact-checking mechanisms and user education initiatives. Social media networks may incorporate these findings to establish real-time monitoring systems that identify possible disinformation outbreaks. Additionally, practical applications may involve utilizing automated algorithms that exploit these network indicators to prioritize content moderation initiatives, so improving the overall resilience of digital communication networks against disinformation.

6.2.3 Further Research

Subsequent study should focus on enhancing the categorization and network simulation models based on the existing findings. This may entail including continuous credibility scores rather than binary labels to offer a more comprehensive comprehension of misinformation dissemination. Furthermore, subsequent research should include real-world dynamic data, including temporal fluctuations in user interactions and changing network architecture, to reconcile the disparity between synthetic and authentic social networks. Extending sensitivity analysis to include other characteristics such as recovery likelihood and initial infection levels would provide further insights into the factors that promote or impede the spread of misinformation.

6.2.4 Alternative Approaches

The present work has effectively utilized random graph models and SIR simulations; nevertheless, investigating alternate methodologies may further improve comprehension and mitigation measures. For instance, employing agent-based modelling might better accurately represent individual user behaviours and interactions. Hybrid models integrating deterministic and stochastic techniques may offer a more resilient framework for forecasting the dissemination of disinformation. Assessing these alternate approaches alongside the current strategy will ascertain the most efficacious strategies for intervention and prevention, so enhancing the overall defence against digital disinformation.

7 Reference List

1. Agrawal, D. and El Abbadi, A. (2012) ‘Mitigating the Spread of Rumors in Social Networks’, IEEE Transactions on Knowledge and Data Engineering, 24(1), pp. 1–10.
2. Albert, R. and Barabási, A.L. (2002) ‘Statistical Mechanics of Complex Networks’, Reviews of Modern Physics, 74(1), pp. 47–97.
3. Al-Rawi, A. (2020) ‘Misinformation and Coronavirus: A Conceptual Framework’, Social Media & Society, 6(3), pp. 1–13.
4. Allcott, H. and Gentzkow, M. (2017) ‘Social Media and Fake News in the 2016 Election’, Journal of Economic Perspectives, 31(2), pp. 211–236.
5. Barabási, A.L. (2009) ‘Scale-Free Networks: A Decade and Beyond’, Science, 325(5939), pp. 412–413.
6. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.U. (2006) ‘Complex Networks: Structure and Dynamics’, Physics Reports, 424(4–5), pp. 175–308.
7. Bovet, A. and Makse, H.A. (2019) ‘Influence of Fake News in Twitter During the 2016 US Presidential Election’, Nature Communications, 10, p. 7.
8. Budak, C., Agrawal, D. and El Abbadi, A. (2011) ‘Limiting the Spread of Misinformation in Social Networks’, in Proceedings of the 20th International Conference on World Wide Web (WWW), pp. 665–674.
9. Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2019) ‘BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding’, in Proceedings of NAACL-HLT, pp. 4171–4186.
10. Erdős, P. and Rényi, A. (1959) ‘On Random Graphs I’, Publicationes Mathematicae, 6, pp. 290–297.
11. FakeNewsNet (n.d.) FakeNewsNet. Available at: http://github.com/KaiDMML/FakeNewsNet/tree/master (Accessed: 20 November 2024).
12. Freeman, L.C. (1978) ‘Centrality in Social Networks: Conceptual Clarification’, Social Networks, 1(3), pp. 215–239.
13. Granovetter, M.S. (1973) ‘The Strength of Weak Ties’, American Journal of Sociology, 78(6), pp. 1360–1380.
14. Hutto, C.J. and Gilbert, E. (2014) ‘VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text’, in Proceedings of the International Conference on Weblogs and Social Media (ICWSM).

15. Jin, F., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, C. and Ramakrishnan, N. (2013) ‘Epidemiological Modeling of News and Rumors on Twitter’, in Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, pp. 609–618.
16. Kempe, D., Kleinberg, J. and Tardos, É. (2003) ‘Maximizing the Spread of Influence through a Social Network’, in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146.
17. Kumar, S., West, R. and Leskovec, J. (2016) ‘Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes’, in Proceedings of the 25th International Conference on World Wide Web, pp. 591–600.
18. Lazer, D.M.J., Baum, M.A., Benkler, Y., et al. (2018) ‘The Science of Fake News’, Science, 359(6380), pp. 1094–1096.
19. Li, F. and Li, T. (2019) ‘Detecting Fake News on Social Media: A Data Mining Perspective’, IEEE Access, 7, pp. 141343–141359.
20. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J. (2013) ‘Efficient Estimation of Word Representations in Vector Space’, in Proceedings of ICLR
21. Newman, M.E.J. (2003) ‘The Structure and Function of Complex Networks’, SIAM Review, 45(2), pp. 167–256.
22. Newman, M. (2010) Networks: An Introduction. Oxford: Oxford University Press.
23. Pastor-Satorras, R. and Vespignani, A. (2001) ‘Epidemic Spreading in Scale-Free Networks’, Physical Review Letters, 86(14), pp. 3200–3203.
24. Pennycook, G. and Rand, D.G. (2018) ‘The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings’, Management Science, 64(11), pp. 4944–4957.
25. Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) ‘“Why Should I Trust You?”: Explaining the Predictions of Any Classifier’, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144.
26. Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A. and Menczer, F. (2018) ‘The Spread of Low-Credibility Content by Social Bots’, Nature Communications, 9, p. 4787.
27. Shu, K., Mahudeswaran, D., Wang, S., Lee, D. and Liu, H. (2018) ‘FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media’, arXiv preprint. arXiv:1809.01286.

28. Shu, K., Mahudeswaran, D. and Liu, H. (2020) ‘Fakenewschallenge: A Benchmark Dataset for Fake News Detection’, in Proceedings of the 29th ACM International Conference on Information and Knowledge Management, pp. 1709–1712.
29. Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. (2017) ‘Fake News Detection on Social Media: A Data Mining Perspective’, ACM SIGKDD Explorations Newsletter, 19(1), pp. 22–36.
30. Shu, K., Wang, S. and Liu, H. (2017) ‘Exploiting Tri-Relationship for Fake News Detection’, arXiv preprint. arXiv:1712.07709.
31. Traag, V.A., Van Dooren, P. and Nesterov, Y. (2011) ‘Narrow Scope for Resolution-limit-free Community Detection’, Physical Review E, 84(1), p. 016114.
32. Vosoughi, S., Roy, D. and Aral, S. (2018) ‘The Spread of True and False News Online’, Science, 359(6380), pp. 1146–1151.
33. Wang, Y., McKee, M., Torbica, A. and Stuckler, D. (2019) ‘Systematic Literature Review on the Spread of Misinformation on Social Media’, Journal of Medical Internet Research, 21(8), p. e13607.
34. Watts, D.J. and Strogatz, S.H. (1998) ‘Collective Dynamics of “Small-World” Networks’, Nature, 393(6684), pp. 440–442.
35. Weng, L., Menczer, F. and Ahn, Y.Y. (2013) ‘Virality Prediction and Community Structure in Social Networks’, Scientific Reports, 3, p. 2522.
36. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Brew, J. (2020) ‘Transformers: State-of-the-Art Natural Language Processing’, in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45.
37. Zhou, X. and Zafarani, R. (2020) ‘A Survey of Fake News: Fundamental Theories, Detection Methods and Opportunities’, ACM Computing Surveys, 53(5), pp. 1–40.

Research Proposal

1. Project Title

Analyzing the Spread of Misinformation in Social Networks Using Random Graph Models

Predictive Modelling and Mitigation Strategies for Enhancing Network Resilience

2. Background and Rationale

In recent years, the fast proliferation of disinformation on social media has become a substantial societal concern, affecting public perception, political debate, and health effects. The intrinsic complexity of social networks, characterized by extensive linkages and dynamic user behaviours, necessitates robust analytical frameworks to represent these interactions. Random graph models provide as a robust mathematical framework for simulating network topologies and may be integrated with prediction and intervention tactics. This study proposal expands upon the expanding literature about disinformation dynamics, network theory, and sophisticated Natural Language Processing (NLP) methodologies for identifying bogus news.

3. Research Aim

The primary objective of this research is to examine and model the dissemination of disinformation inside social networks using the application of random graph theory. The research aims to create prediction models and focused mitigation measures that improve network resilience against the fast spread of misinformation.

4. Research Objectives

- To do exhaustive literature research on the dissemination of disinformation, social network analysis, and random graph models, therefore creating the theoretical framework for the study.
- To gather and preprocess multi-source data from FakeNewsNet, encompassing datasets such as gossipcop_fake, gossipcop_real, politifact_fake, and politifact_real, to create a comprehensive dataset.
- To construct and assess classification models, comprising a baseline TF-IDF with Logistic Regression and a sophisticated fine-tuned BERT model, for the detection of false news.

- To create synthetic social networks utilizing Barabási–Albert (BA) and Erdős–Rényi (ER) random graph models, including features derived from classification results.
- To analyse the dissemination of disinformation across these networks utilizing a SIR (Susceptible-Infected-Recovered) framework and to conduct sensitivity studies on critical parameters.
- To calculate sophisticated network metrics and implement community detection techniques to find prominent nodes and clusters that may function as essential intervention sites.
- To provide actionable mitigation methods derived from the simulation results and to delineate suggestions for subsequent study and implementation.

5. Research Question

In what ways may random graph models facilitate the analysis and prediction of disinformation dissemination within social networks, and which measures might successfully curtail this dissemination?

6. Methodology

The research adopts a mixed-methods approach that combines quantitative data analysis, machine learning, and network simulation:

- Data Acquisition and Pre-processing: Four datasets from FakeNewsNet will be amalgamated and sanitized. The text data will be standardized and categorized (fake vs. real) to form a cohesive dataset.
- Classification Models: Two methodologies will be employed: a baseline TF-IDF combined with a Logistic Regression model and a fine-tuned BERT model for the classification of news articles. The outputs will elucidate node properties in the network simulations.
- Network Modelling: Synthetic networks will be developed utilizing the Barabási-Albert and Erdős-Rényi models. The BA model will imitate scale-free networks, whereas the ER model will depict uniformly random connections.
- SIR Simulation: An SIR model will be utilized for both network types to mimic the temporal dissemination of disinformation. Critical metrics, like infection and recovery probabilities, will be altered to evaluate the network's response.
- Sensitivity Analysis and Advanced Metrics: The sensitivity analysis will examine the effects of differing infection probabilities. Advanced network

measures, including betweenness, closeness, and eigenvector centrality, with community recognition via the Louvain method, will be utilized to find prominent nodes and sub-network structures.

7. Timeline and Deliverables

The project will be executed within a timeframe delineated by certain milestones and deliverables:

- Phase 1 (Planning and Literature Review): To be completed by November 2024, including a finished study proposal and literature review report.
- Phase 2 (Data Collection and Pre-processing): Finalize by the conclusion of November 2024, ensuring a pristine and organized dataset.
- Phase 3 (Model Development): Construct classification models and scripts for network development by mid-December 2024.
- Phase 4 (Simulation and Analysis): Execute SIR simulations and sensitivity analyses from January to February 2025, accompanied with comprehensive analytical results.
- Phase 5 (Reporting and Review): Complete the dissertation report, encompassing assessment and suggestions, by March 2025.

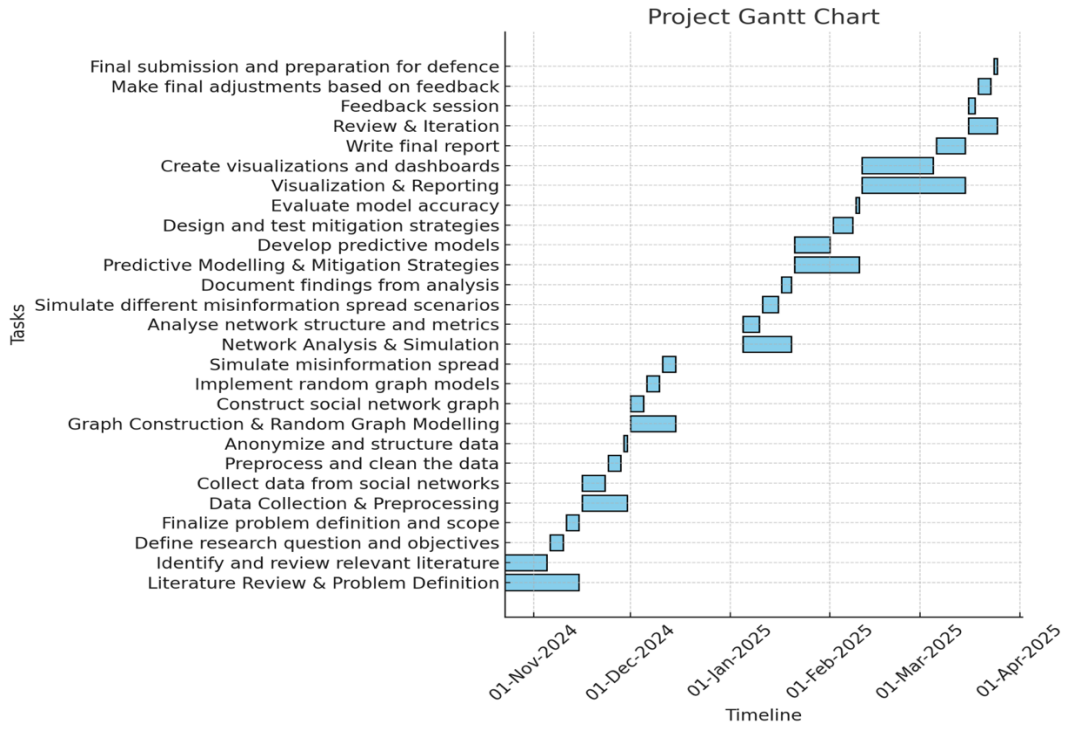


Table 8. Project Schedule and Milestones

8. Ethical and Legal Considerations

This study will rigorously comply with ethical standards and legal obligations. All utilized data will be anonymized to safeguard personal information, and data management will adhere to GDPR and other pertinent data protection requirements. Ethical permission will be obtained as required, and all procedures will be meticulously documented to guarantee ethical research methods.

9. Conclusion

This study proposal delineates a comprehensive approach to tackle the intricate problem of misinformation dissemination in social networks. The work seeks to deliver practical insights for anticipating and minimizing the effects of disinformation by merging advanced NLP classification algorithms with random graph modelling and dynamic simulations. The suggested methodology enhances scholarly dialogue and provides actionable solutions for adaptation by social media platforms and policymakers.

Code

Analyzing the Spread of Misinformation in Social Networks

Using Random Graph Models, SIR Simulation, and Classification Models¶

Overview:
This notebook integrates data from four FakeNewsNet datasets to analyze the spread of misinformation. It covers:

1. Data loading and preprocessing (with fake/real labeling)
2. Classification models (TF-IDF + Logistic Regression and fine-tuned BERT)
3. Synthetic network modeling using data-driven nodes for both Barabási–Albert (BA) and Erdős–Rényi (ER) models
4. SIR simulations on these networks with network graph outputs
5. Sensitivity analysis on infection probability
6. Advanced network metrics and community detection (Louvain)

Future Work:
Extend sensitivity analysis to other parameters, incorporate node weights from classification scores, and compare results with real social network data.

1. Import Libraries and Setup

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import networkx as nx
import random
import re
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords

# For classification and NLP
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score
from transformers import AutoTokenizer, AutoModelForSequenceClassification, Trainer, TrainingArguments
from datasets import Dataset

# For community detection
import community as community_louvain

# Set random seeds for reproducibility
np.random.seed(42)
random.seed(42)
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/peipei/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

2. Data Loading and Preprocessing

I load four FakeNewsNet datasets, clean the text data, and assign a label:

- Fake news = 1
- Real news = 0

The datasets are then combined into a single DataFrame.

```
# Load datasets from FakeNewsNet/Dataset
gossipcop_fake = pd.read_csv('FakeNewsNet/Dataset/gossipcop_fake.csv')
gossipcop_real = pd.read_csv('FakeNewsNet/Dataset/gossipcop_real.csv')
politifact_fake = pd.read_csv('FakeNewsNet/Dataset/politifact_fake.csv')
politifact_real = pd.read_csv('FakeNewsNet/Dataset/politifact_real.csv')

# Display columns for inspection
print("GossipCop Fake Columns:", gossipcop_fake.columns.tolist())
print("GossipCop Real Columns:", gossipcop_real.columns.tolist())
print("PolitiFact Fake Columns:", politifact_fake.columns.tolist())
print("PolitiFact Real Columns:", politifact_real.columns.tolist())
```

```
GossipCop Fake Columns: ['id', 'news_url', 'title', 'tweet_ids']
GossipCop Real Columns: ['id', 'news_url', 'title', 'tweet_ids']
PolitiFact Fake Columns: ['id', 'news_url', 'title', 'tweet_ids']
PolitiFact Real Columns: ['id', 'news_url', 'title', 'tweet_ids']
```

```
# Define text cleaning function
stop_words = set(stopwords.words('english'))

def clean_text(text):
    if pd.isna(text):
        return ""
    text = re.sub(r'<.*?>', '', text)        # Remove HTML tags
    text = re.sub(r'[^a-zA-Z\s]', '', text)    # Remove non-alphabet characters
    text = text.lower()                    # Convert to lowercase
    tokens = text.split()
    tokens = [word for word in tokens if word not in stop_words]
    return ' '.join(tokens)

def apply_cleaning(df):
    # Check for 'title' or 'content' column
    if 'title' in df.columns:
        source_col = 'title'
    elif 'content' in df.columns:
        source_col = 'content'
    else:
        print("No text column found!")
        return df
    df['cleaned_text'] = df[source_col].apply(clean_text)
    return df

# Clean each dataset
gossipcop_fake = apply_cleaning(gossipcop_fake)
gossipcop_real = apply_cleaning(gossipcop_real)
politifact_fake = apply_cleaning(politifact_fake)
politifact_real = apply_cleaning(politifact_real)

# Assign labels
gossipcop_fake['label'] = 1
politifact_fake['label'] = 1
gossipcop_real['label'] = 0
politifact_real['label'] = 0

# Combine all datasets and drop missing texts
```

```
combined_df = pd.concat([gossipcop_fake, gossipcop_real, politifact_fake, politifact_real], ignore_index=True)
combined_df = combined_df.dropna(subset=['cleaned_text'])
print("Combined dataset shape:", combined_df.shape)
```

```
Combined dataset shape: (23196, 6)
```

3. Classification Models

I build two models to classify fake vs. real news:

1. **TF-IDF + Logistic Regression (Baseline)**
2. **Fine-Tuned BERT**

The performance of both models is printed for comparison.

```
# Prepare data for classification
X = combined_df['cleaned_text']
y = combined_df['label']

# TF-IDF Vectorization and Logistic Regression
vectorizer = TfidfVectorizer(max_features=5000)
X_tfidf = vectorizer.fit_transform(X)
X_train, X_test, y_train, y_test = train_test_split(X_tfidf, y, test_size=0.2, random_state=42)

clf = LogisticRegression(max_iter=1000)
clf.fit(X_train, y_train)
y_pred_tfidf = clf.predict(X_test)

print("TF-IDF Baseline Results:")
print("Accuracy:", accuracy_score(y_test, y_pred_tfidf))
print(classification_report(y_test, y_pred_tfidf))
```

```
TF-IDF Baseline Results:
Accuracy: 0.8400862068965518
              precision    recall  f1-score   support

           0       0.84      0.97      0.90      3492
           1       0.82      0.45      0.58      1148

    accuracy                           0.84      4640
   macro avg       0.83      0.71      0.74      4640
weighted avg       0.84      0.84      0.82      4640
```

```
# BERT Fine-Tuning
# BERT Fine-Tuning
hf_dataset = Dataset.from_pandas(combined_df[['cleaned_text', 'label']])
split_dataset = hf_dataset.train_test_split(test_size=0.2, seed=42)

tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")

def tokenize_function(example):
    return tokenizer(example["cleaned_text"], truncation=True, padding="max_length", max_length=128)

tokenized_datasets = split_dataset.map(tokenize_function, batched=True)
cols_to_remove = [col for col in ["cleaned_text", "__index_level_0__"] if col in
tokenized_datasets["train"].column_names]
tokenized_datasets = tokenized_datasets.remove_columns(cols_to_remove)
tokenized_datasets = tokenized_datasets.rename_column("label", "labels")
tokenized_datasets.set_format("torch")

model = AutoModelForSequenceClassification.from_pretrained("bert-base-uncased", num_labels=2)

training_args = TrainingArguments(
```

```
    output_dir="./results",
    evaluation_strategy="epoch",
    save_strategy="epoch",
    learning_rate=2e-5,
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,
    num_train_epochs=3,
    weight_decay=0.01,
    logging_dir='./logs',
    logging_steps=50,
    load_best_model_at_end=True,
)

trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_datasets["train"],
    eval_dataset=tokenized_datasets["test"],
    tokenizer=tokenizer,
)

trainer.train()
bert_results = trainer.evaluate()
print("BERT Evaluation Results:", bert_results)
```

```
Map:   0%|          | 0/18556 [00:00<?, ? examples/s]
Map:   0%|          | 0/4640 [00:00<?, ? examples/s]
Some weights of BertForSequenceClassification were not initialized from
the model checkpoint at bert-base-uncased and are newly initialized: ['c
lassifier.bias', 'classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to
 use it for predictions and inference.
/Users/peipei/anaconda3/lib/python3.11/site-packages/transformers/traini
ng_args.py:1594: FutureWarning: `evaluation_strategy` is deprecated and
will be removed in version 4.46 of  Transformers. Use `eval_strategy` i
nstead
  warnings.warn(
/var/folders/cn/wqzk89gx2hv061vhqvfb94yw0000gn/T/ipykernel_1203/14308690
92.py:32: FutureWarning: `tokenizer` is deprecated and will be removed i
n version 5.0.0 for `Trainer.__init__`. Use `processing_class` instead.
  trainer = Trainer(
```

[3480/3480 1:37:12, Epoch 3/3]

Epoch	Training Loss	Validation Loss
1	0.383100	0.364423
2	0.312000	0.386554
3	0.215200	0.441053

[290/290 02:10]

```
BERT Evaluation Results: {'eval_loss': 0.36442306637763977, 'eval_runtim
e': 130.6968, 'eval_samples_per_second': 35.502, 'eval_steps_per_second
': 2.219, 'epoch': 3.0}
```

```
# Get predictions from the BERT model on the test dataset
bert_preds_output = trainer.predict(tokenized_datasets["test"])
bert_preds = np.argmax(bert_preds_output.predictions, axis=1)
y_test_bert = tokenized_datasets["test"]["labels"]

# Compute the overall accuracy and detailed classification metrics
accuracy = accuracy_score(y_test_bert, bert_preds)
report = classification_report(y_test_bert, bert_preds, output_dict=True)

# Extract performance metrics for class '0' (Real) and class '1' (Fake)
precision_real = report['0']['precision']
recall_real = report['0']['recall']
f1_real = report['0']['f1-score']

precision_fake = report['1']['precision']
recall_fake = report['1']['recall']
f1_fake = report['1']['f1-score']

# Create a DataFrame to neatly display the results
bert_performance = pd.DataFrame({
    'Model': ['BERT'],
    'Accuracy': [accuracy],
    'Precision (Real)': [precision_real],
    'Recall (Real)': [recall_real],
    'F1-Score (Real)': [f1_real],
    'Precision (Fake)': [precision_fake],
    'Recall (Fake)': [recall_fake],
    'F1-Score (Fake)': [f1_fake]
})

print("BERT Model Performance Metrics:")
print(bert_performance)
```

```
BERT Model Performance Metrics:
  Model  Accuracy  Precision (Real)  Recall (Real)  F1-Score (Real)
\
0  BERT  0.844397          0.890519       0.904298         0.897356


   Precision (Fake)  Recall (Fake)  F1-Score (Fake)
0          0.695255       0.662609          0.67854
```

4. Network Modeling and SIR Simulation Using Data Attributes

I sample 1,000 nodes from the combined dataset to represent articles (nodes). Each node is assigned its fake/real label.
I then generate synthetic networks using:

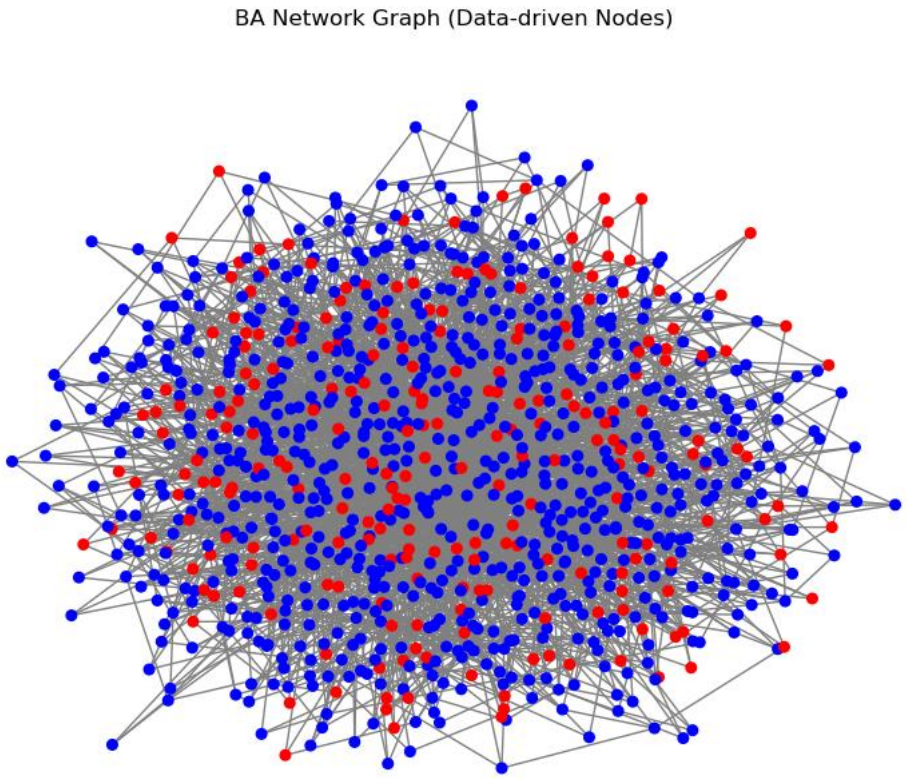
- **Barabási–Albert (BA) Model**
- **Erdős–Rényi (ER) Model**

For each network, I output the network graph and run an SIR simulation.

4.1 BA Network: Graph Generation, Visualization, and SIR Simulation

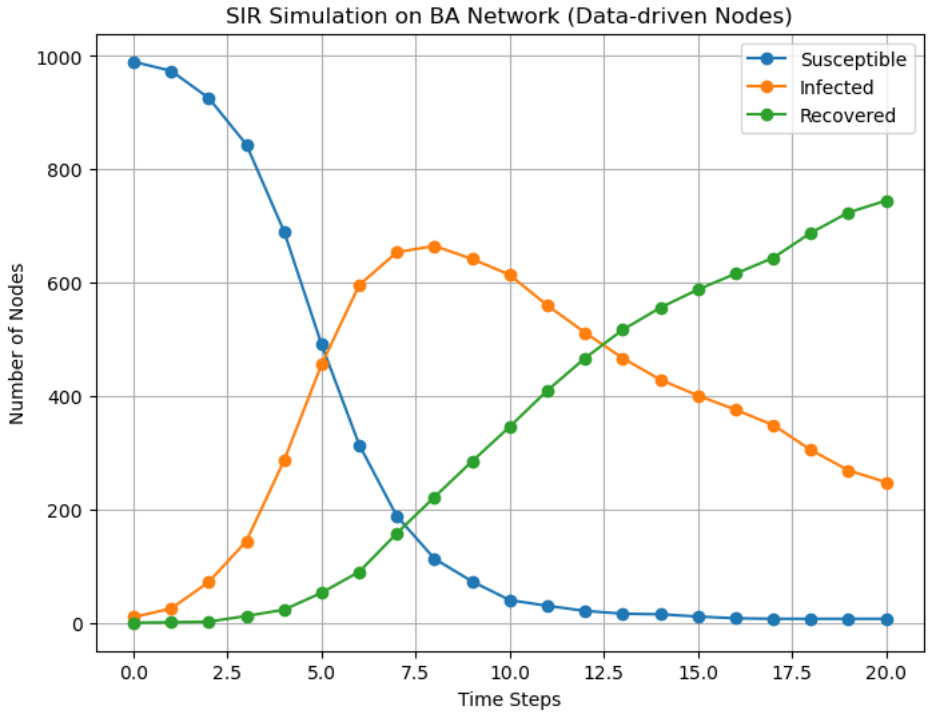
```
# Sample nodes from the combined dataset
sample_size = 1000
combined_sample = combined_df.sample(n=sample_size, random_state=42).reset_index(drop=True)

# Generate a BA network with sample_size nodes
ba_m = 3  # Each new node attaches to 3 existing nodes
G_ba = nx.barabasi_albert_graph(sample_size, ba_m, seed=42)

# Assign node attributes from the combined sample
```

```
for i in range(sample_size):
    G_ba.nodes[i]['label'] = combined_sample.loc[i, 'label']

# Visualize the BA network graph
plt.figure(figsize=(8,6))
pos_ba = nx.spring_layout(G_ba, seed=42)
node_colors_ba = ['red' if G_ba.nodes[node]['label'] == 1 else 'blue' for node in G_ba.nodes()]
nx.draw(G_ba, pos_ba, node_color=node_colors_ba, node_size=30, edge_color='gray', with_labels=False)
plt.title("BA Network Graph (Data-driven Nodes)")
plt.axis('off')
plt.show()
```



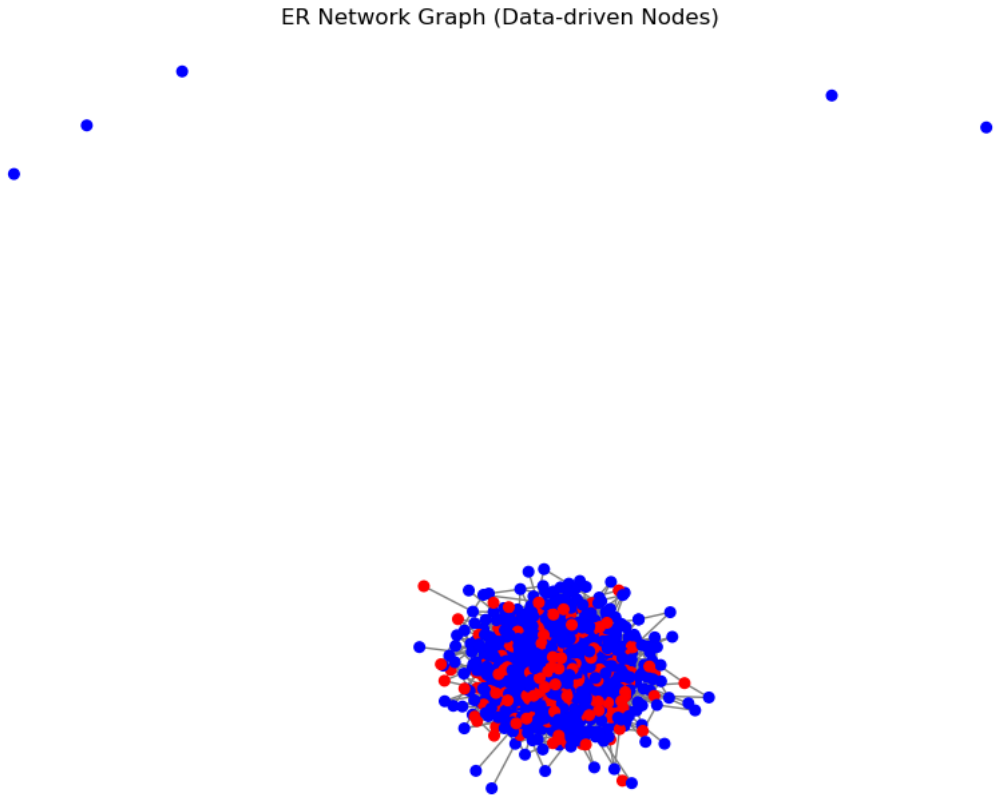

```
# Define SIR simulation function
def simulate_SIR(G, initial_infected=10, infection_prob=0.2, recovery_prob=0.1, steps=20):
    """
    Simulate the SIR model on graph G.
    States: 0 = Susceptible, 1 = Infected, 2 = Recovered
    Returns final state and history dictionary with counts for S, I, R at each time step.
    """
    state = {node: 0 for node in G.nodes()}
    initial_nodes = random.sample(list(G.nodes()), initial_infected)
    for node in initial_nodes:
        state[node] = 1
    history = {'S': [sum(1 for s in state.values() if s == 0)],
               'I': [sum(1 for s in state.values() if s == 1)],
               'R': [sum(1 for s in state.values() if s == 2)]}
    for t in range(steps):
        new_state = state.copy()
        for node in G.nodes():
            if state[node] == 1:
                for neighbor in G.neighbors(node):
                    if state[neighbor] == 0 and random.random() < infection_prob:
                        new_state[neighbor] = 1
                if random.random() < recovery_prob:
                    new_state[node] = 2
        state = new_state.copy()
        history['S'].append(sum(1 for s in state.values() if s == 0))
        history['I'].append(sum(1 for s in state.values() if s == 1))
        history['R'].append(sum(1 for s in state.values() if s == 2))
```

```
    return state, history

# SIR simulation parameters
steps = 20
initial_infected = 10
infection_prob = 0.2
recovery_prob = 0.1

# Run SIR simulation on BA network
state_ba, history_ba = simulate_SIR(G_ba, initial_infected, infection_prob, recovery_prob, steps)

# Plot SIR simulation results for BA network
plt.figure(figsize=(8,6))
plt.plot(history_ba['S'], label='Susceptible', marker='o')
plt.plot(history_ba['I'], label='Infected', marker='o')
plt.plot(history_ba['R'], label='Recovered', marker='o')
plt.xlabel("Time Steps")
plt.ylabel("Number of Nodes")
plt.title("SIR Simulation on BA Network (Data-driven Nodes)")
plt.legend()
plt.grid(True)
plt.show()
```

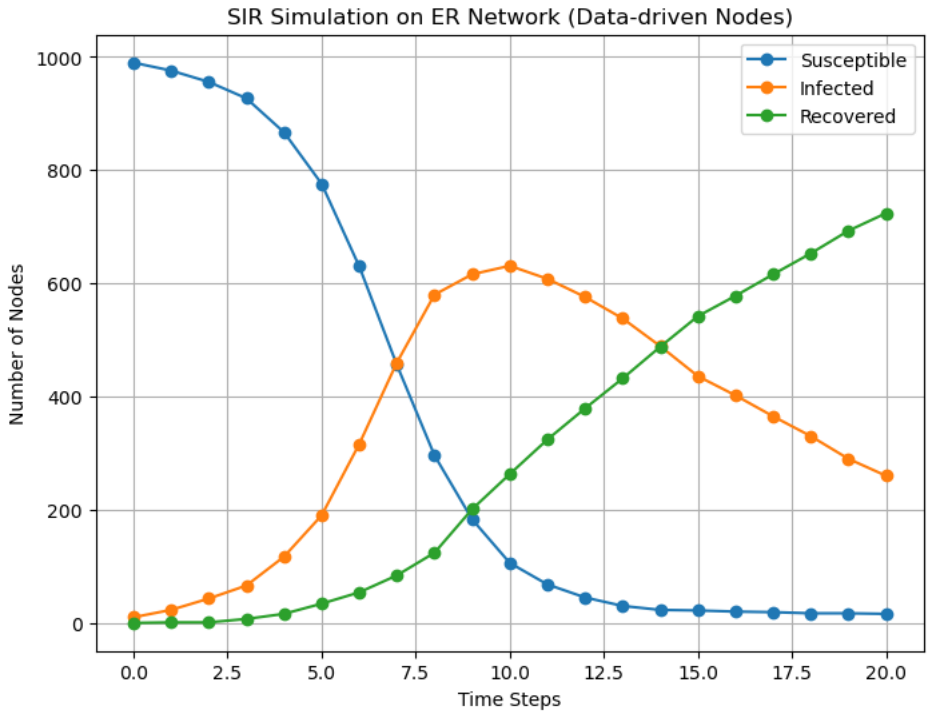


4.2 ER Network: Graph Generation, Visualization, and SIR Simulation

```
# Generate an ER network using the same sample
er_p = 6 / (sample_size - 1)  # set probability for average degree around 6
G_er = nx.erdos_renyi_graph(sample_size, er_p, seed=42)

# Assign node attributes from the combined sample
for i in range(sample_size):
    G_er.nodes[i]['label'] = combined_sample.loc[i, 'label']

# Visualize the ER network graph
plt.figure(figsize=(8,6))
pos_er = nx.spring_layout(G_er, seed=42)
node_colors_er = ['red' if G_er.nodes[node]['label'] == 1 else 'blue' for node in G_er.nodes()]
nx.draw(G_er, pos_er, node_color=node_colors_er, node_size=30, edge_color='gray', with_labels=False)
plt.title("ER Network Graph (Data-driven Nodes)")
plt.axis('off')
plt.show()
```




```
# Run SIR simulation on ER network
state_er, history_er = simulate_SIR(G_er, initial_infected, infection_prob, recovery_prob, steps)

# Plot SIR simulation results for ER network
plt.figure(figsize=(8,6))
plt.plot(history_er['S'], label='Susceptible', marker='o')
plt.plot(history_er['I'], label='Infected', marker='o')
plt.plot(history_er['R'], label='Recovered', marker='o')
plt.xlabel("Time Steps")
plt.ylabel("Number of Nodes")
plt.title("SIR Simulation on ER Network (Data-driven Nodes)")
plt.legend()
plt.grid(True)
plt.show()
```



4.3 SIR Simulation Summary¶

```
# Define simulation parameters
sample_size = 1000
ba_m = 3
er_p = 6 / (sample_size - 1)  # Calculated for ER network to achieve an average degree around 6
initial_infected = 10
infection_prob = 0.2
recovery_prob = 0.1
steps = 20

# Retrieve final simulation values from the history dictionaries
final_S_ba = history_ba['S'][-1]
final_I_ba = history_ba['I'][-1]
final_R_ba = history_ba['R'][-1]

final_S_er = history_er['S'][-1]
final_I_er = history_er['I'][-1]
final_R_er = history_er['R'][-1]

# Create a summary DataFrame
summary_data = {
    "Network Model": ["BA", "ER"],
    "Sample Size": [sample_size, sample_size],
    "Model Parameter (ba_m or er_p)": [ba_m, er_p],
    "Initial Infected": [initial_infected, initial_infected],
    "Infection Prob.": [infection_prob, infection_prob],
    "Recovery Prob.": [recovery_prob, recovery_prob],
    "Steps": [steps, steps],
    "Final Susceptible": [final_S_ba, final_S_er],
    "Final Infected": [final_I_ba, final_I_er],
    "Final Recovered": [final_R_ba, final_R_er]
}

summary_df = pd.DataFrame(summary_data)

print("SIR Simulation Summary:")
print(summary_df)
```

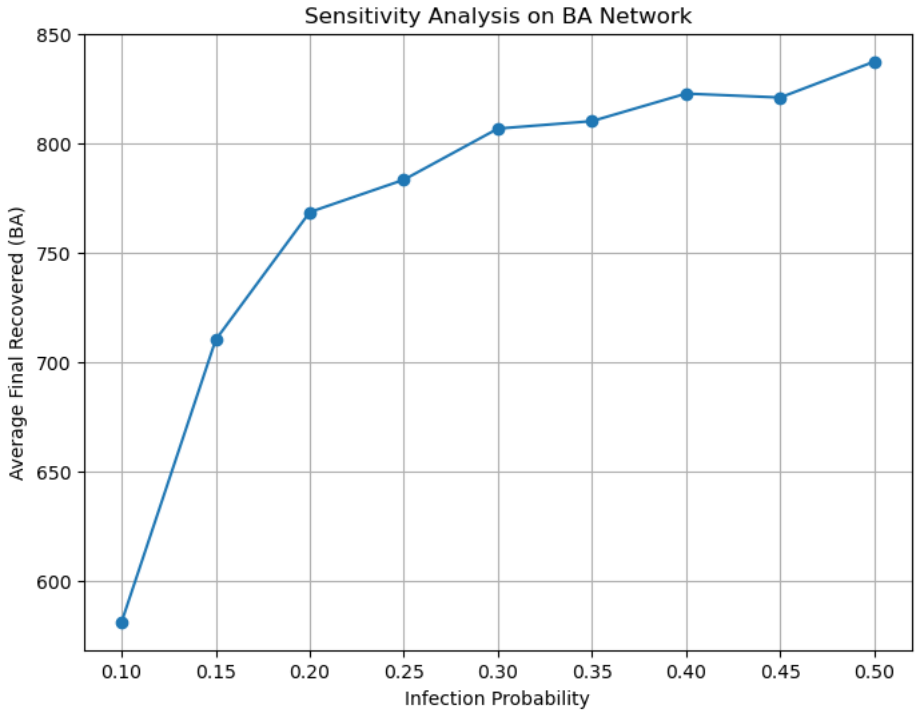
```
SIR Simulation Summary:
  Network Model  Sample Size  Model Parameter (ba_m or er_p)  \
0            BA         1000                        3.000000
```

```
1             ER         1000                          0.006006

   Initial Infected  Infection Prob.  Recovery Prob.  Steps  \
0                10              0.2             0.1     20
1                10              0.2             0.1     20

   Final Susceptible  Final Infected  Final Recovered
0                  7             248              745
1                 16             260              724
```

5. Sensitivity Analysis

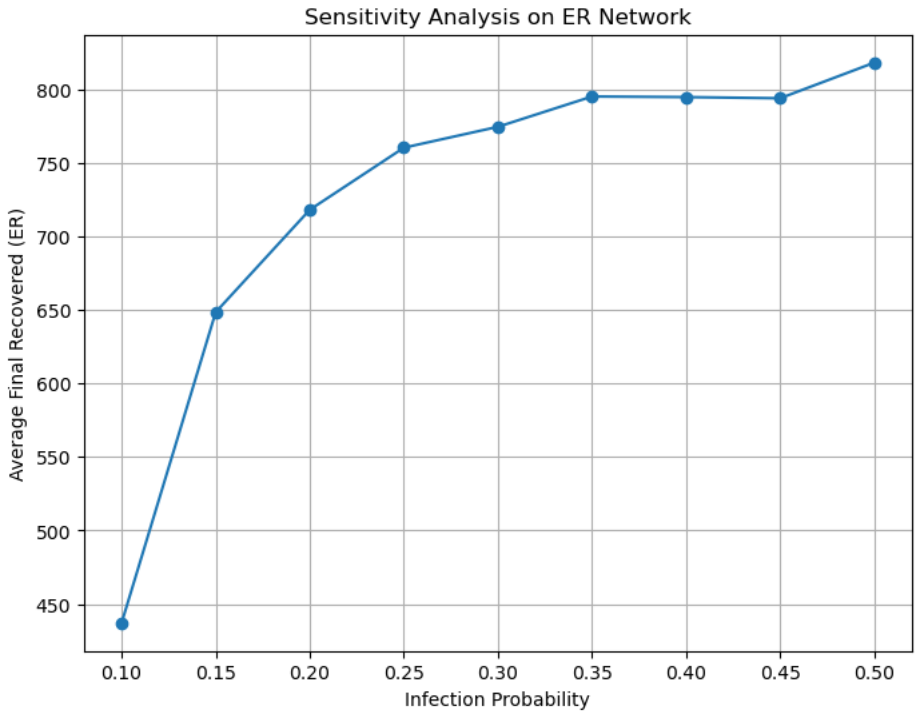
I explore how varying the infection probability affects the final epidemic size (measured as the average final recovered count) for both BA and ER networks.

```
# Sensitivity Analysis for BA Network
infection_probs = np.linspace(0.1, 0.5, 9)
runs = 5
final_recovered_avg_ba = []

for ip in infection_probs:
    recovered_list = []
    for _ in range(runs):
        _, history = simulate_SIR(G_ba, initial_infected, ip, recovery_prob, steps)
        recovered_list.append(history['R'][-1])
    final_recovered_avg_ba.append(np.mean(recovered_list))

plt.figure(figsize=(8,6))
plt.plot(infection_probs, final_recovered_avg_ba, marker='o')
plt.xlabel("Infection Probability")
plt.ylabel("Average Final Recovered (BA)")
plt.title("Sensitivity Analysis on BA Network")
plt.grid(True)
plt.show()
```




```
# Sensitivity Analysis for ER Network
final_recovered_avg_er = []

for ip in infection_probs:
    recovered_list = []
```

```
    for _ in range(runs):
        _, history = simulate_SIR(G_er, initial_infected, ip, recovery_prob, steps)
        recovered_list.append(history['R'][-1])
    final_recovered_avg_er.append(np.mean(recovered_list))

plt.figure(figsize=(8,6))
plt.plot(infection_probs, final_recovered_avg_er, marker='o')
plt.xlabel("Infection Probability")
plt.ylabel("Average Final Recovered (ER)")
plt.title("Sensitivity Analysis on ER Network")
plt.grid(True)
plt.show()
```



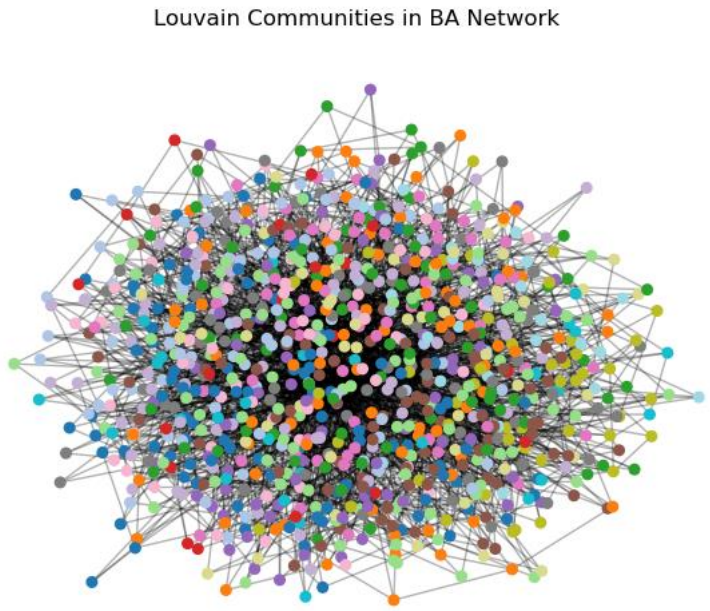
6. Advanced Network Metrics and Community Detection

I compute advanced network metrics (betweenness, closeness, eigenvector centrality) and perform community detection using the Louvain algorithm for both BA and ER networks.

```
# Advanced Metrics for BA Network
betweenness_ba = nx.betweenness_centrality(G_ba)
closeness_ba = nx.closeness_centrality(G_ba)
eigenvector_ba = nx.eigenvector_centrality(G_ba)

print("BA Network - Top 5 Betweenness Centrality:")
for node, val in sorted(betweenness_ba.items(), key=lambda x: x[1], reverse=True)[:5]:
    print(f"Node {node}: {val:.4f}")
print(f"Average Closeness (BA): {np.mean(list(closeness_ba.values())):.4f}")
print(f"Average Eigenvector Centrality (BA): {np.mean(list(eigenvector_ba.values())):.4f}")

# Louvain Community Detection on BA Network
partition_ba = community_louvain.best_partition(G_ba)
num_communities_ba = len(set(partition_ba.values()))
modularity_ba = community_louvain.modularity(partition_ba, G_ba)
print(f"BA Communities: {num_communities_ba}, Modularity: {modularity_ba:.4f}")

# Visualize BA communities
colors_ba = [partition_ba[node] for node in G_ba.nodes()]
pos_ba = nx.spring_layout(G_ba, seed=42)
plt.figure(figsize=(8,6))
nx.draw_networkx_nodes(G_ba, pos_ba, node_size=30, node_color=colors_ba, cmap=plt.cm.tab20)
nx.draw_networkx_edges(G_ba, pos_ba, alpha=0.3)
plt.title("Louvain Communities in BA Network")
```

```
plt.axis('off')
plt.show()
```

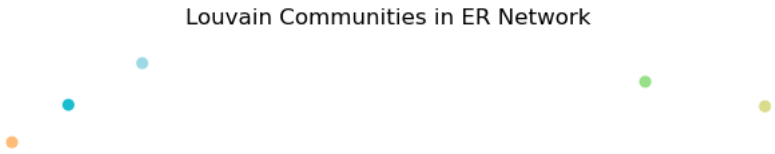
```
BA Network - Top 5 Betweenness Centrality:
Node 4: 0.1513
Node 6: 0.1306
Node 5: 0.1263
Node 7: 0.1112
Node 12: 0.0733
Average Closeness (BA): 0.2911
Average Eigenvector Centrality (BA): 0.0192
BA Communities: 16, Modularity: 0.3904
```




```
# Advanced Metrics for ER Network
betweenness_er = nx.betweenness_centrality(G_er)
closeness_er = nx.closeness_centrality(G_er)
eigenvector_er = nx.eigenvector_centrality(G_er)

print("ER Network - Top 5 Betweenness Centrality:")
for node, val in sorted(betweenness_er.items(), key=lambda x: x[1], reverse=True)[:5]:
    print(f"Node {node}: {val:.4f}")
print(f"Average Closeness (ER): {np.mean(list(closeness_er.values())):.4f}")
print(f"Average Eigenvector Centrality (ER): {np.mean(list(eigenvector_er.values())):.4f}")

# Louvain Community Detection on ER Network
partition_er = community_louvain.best_partition(G_er)
num_communities_er = len(set(partition_er.values()))
modularity_er = community_louvain.modularity(partition_er, G_er)
print(f"ER Communities: {num_communities_er}, Modularity: {modularity_er:.4f}")

# Visualize ER communities
colors_er = [partition_er[node] for node in G_er.nodes()]
pos_er = nx.spring_layout(G_er, seed=42)
plt.figure(figsize=(8,6))
nx.draw_networkx_nodes(G_er, pos_er, node_size=30, node_color=colors_er, cmap=plt.cm.tab20)
nx.draw_networkx_edges(G_er, pos_er, alpha=0.3)
plt.title("Louvain Communities in ER Network")
plt.axis('off')
plt.show()
```

```
ER Network - Top 5 Betweenness Centrality:
Node 693: 0.0143
Node 291: 0.0140
```

```
Node 763: 0.0136
Node 572: 0.0132
Node 43: 0.0132
Average Closeness (ER): 0.2466
Average Eigenvector Centrality (ER): 0.0280
ER Communities: 22, Modularity: 0.3919
```



```
# Calculate average metrics for BA network
avg_betweenness_ba = np.mean(list(betweenness_ba.values()))
avg_closeness_ba = np.mean(list(closeness_ba.values()))
avg_eigen_ba = np.mean(list(eigenvector_ba.values()))
num_communities_ba = len(set(partition_ba.values()))
modularity_ba_val = modularity_ba

# Calculate average metrics for ER network
avg_betweenness_er = np.mean(list(betweenness_er.values()))
avg_closeness_er = np.mean(list(closeness_er.values()))
avg_eigen_er = np.mean(list(eigenvector_er.values()))
num_communities_er = len(set(partition_er.values()))
modularity_er_val = modularity_er

# Create a summary DataFrame
metrics_summary = pd.DataFrame({
    "Metric": ["Betweenness Centrality", "Closeness Centrality", "Eigenvector Centrality",
               "Number of Communities", "Modularity"],
    "BA Network (Average)": [avg_betweenness_ba, avg_closeness_ba, avg_eigen_ba, num_communities_ba,
modularity_ba_val],
    "ER Network (Average)": [avg_betweenness_er, avg_closeness_er, avg_eigen_er, num_communities_er,
modularity_er_val]
})

print("Advanced Network Metrics Comparison:")
print(metrics_summary)
```

```
Advanced Network Metrics Comparison:
                  Metric  BA Network (Average)  ER Network (Average)
0  Betweenness Centrality              0.002471              0.003009
1    Closeness Centrality              0.291087              0.246600
2  Eigenvector Centrality              0.019238              0.028041
3   Number of Communities             16.000000             22.000000
```

4	Modularity	0.390368	0.391853

7. Conclusions and Future Work

In this notebook I have:

- Loaded and preprocessed four FakeNewsNet datasets.
- Combined them and assigned fake/real labels.
- Built classification models (TF-IDF baseline and fine-tuned BERT) for fake news detection.
- Generated synthetic networks (BA and ER) with data-driven node attributes.
- Visualized network graphs and simulated misinformation spread using an SIR model.
- Performed sensitivity analysis on infection probability.
- Computed advanced network metrics and performed community detection using the Louvain algorithm.

Future Work:

- Extend sensitivity analysis to other parameters (recovery probability, initial infected count).
- Incorporate classification outcomes as node weights for more nuanced simulations.
- Compare synthetic results with real-world social network interaction data.
- Explore targeted mitigation strategies based on network centrality and community structure.