AMRITA VISHWA
VIDYAPEETHAM

FAKE NEWS DETECTION

SOCIAL NETWORK ANALYTICS

2025

BARSHAN MONDAL | MAHIMA REMESH NAIR | NANDANA PRAVEEN

AGENDA

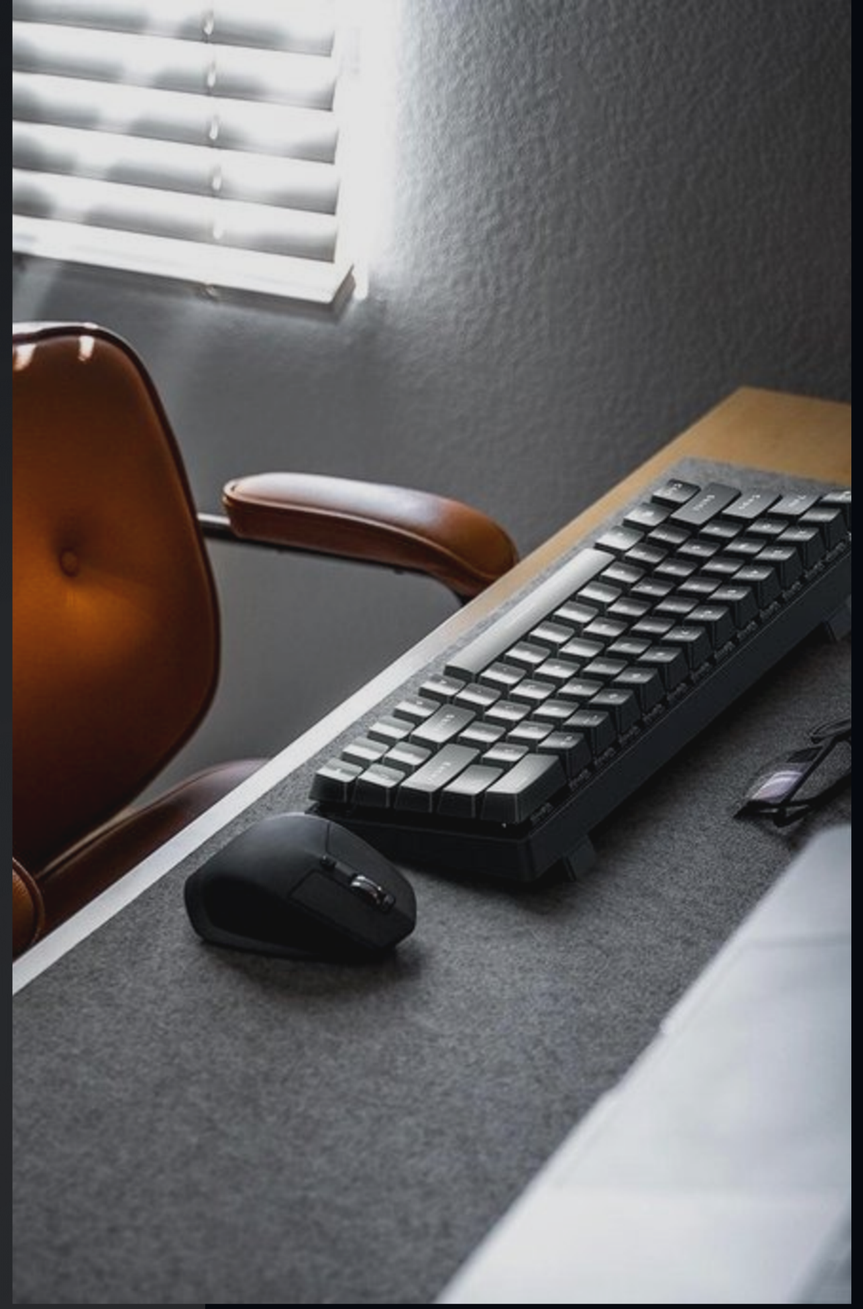
Our content today is divided into these parts. Each part will be described with examples.

CONTEXT OF WORK

“In the age of information, ignorance is a choice.”
- Donny Miller

In the digital age, misinformation and fake news have become widespread due to the vast reach and rapid dissemination through social media platforms. Traditional detection techniques often focus solely on textual analysis. However, they miss the relational and contextual cues that influence how misinformation spreads.

This project operates at the intersection of Natural Language Processing, Graph Neural Networks, and Explainable AI, using a multi-modal approach to address the nuances of misinformation detection.



PROJECT GOALS

Elections

During the 2020 U.S. presidential election, widespread fake news about mail-in voting fraud was circulated, often appearing textually legitimate.

Security

These claims, when analyzed through metadata like speaker bias and repetition across partisan sources, revealed deeper patterns of misinformation.

Smart choice

This illustrates the need for models that go beyond content and incorporate relational indicators like source credibility, political alignment, and topical clustering.

PROJECT GOALS

The primary objectives are:

Develop a multi-modal GNN that integrates semantic embeddings, content features, and metadata.

Use community detection to enhance interpretability.

Construct a metadata-driven graph to capture inter-claim relationships.

Benchmark performance on a real-world dataset (PolitiFact) using stratified and temporal evaluation strategies.

RELATED WORK

We studied four key works:

BERT (2019)

Contextual language modeling; lacks relational data modeling.

CSI (2017)

Combines user and content data; no graph modeling.

Bi-GCN (2020)

Graph-based rumor detection; not generalizable.

MVAE (2019)

Multi-modal autoencoder; ignores graph structures.

How Our Work Compares with Related Work

Aspect	*Related Work*	*Our Approach*	*Key Points*
• Feature Integration	• Separate content or network analysis	• Unified BERT + metadata + graph structure	• Community-Aware Design: Explicit incorporation of community detection for enhanced interpretability
• Graph Construction	• Social networks or semantic only1	• Speaker credibility + political affiliation + topical similarity	• Metadata-Driven Graphs: Novel graph construction methodology based on news claim metadata rather than social connections
• Community Detection	• Limited interpretability focus	• Explicit community analysis for misinformation clusters	
• Evaluation	• Single metrics	• Comprehensive ablation studies + feature importance analysis	

FORMAL PROBLEM DEFINITION

MATHEMATICAL FORMULATION

Let G = (V, E, X, Y) represent a heterogeneous graph where :

- V: Set of news claim nodes
- E: Set of edges representing relationships between claims
- X ∈ R^(|V|×d): Node feature matrix
- Y ∈ {0,1}: Binary labels indicating fake (1) or real (0) news

OBJECTIVE FUNCTION

Learn a function f: (G, X) → Y that accurately classifies news claims while providing interpretable insights into the decision process

OPTIMIZATION GOAL

Minimize classification error while maximizing model interpretability through community-based analysis

Terms and Definitions

Core Concepts

- Multi-Modal Features: Integration of semantic (BERT), content (TF-IDF), and metadata features
- Heterogeneous Graph: Graph structure incorporating news claims as nodes with metadata-based similarity edges
- Community Detection: Algorithmic identification of densely connected node clusters representing misinformation patterns
- GraphSAGE: Graph neural network architecture using sampling and aggregation for scalable learning

Technical Terms

- Semantic Embeddings: 768-dimensional BERT representations capturing contextual meaning.
- Metadata Features: Speaker credibility scores, political affiliations, subject categories, temporal attributes.
- K-Nearest Neighbor Graph: Graph construction connecting each node to k=5 most similar neighbors based on cosine similarity.

Assumptions and Constraints

Key Assumptions

- Language Limitation: Current model focuses on English-language news claims
- Metadata Availability: Graph construction relies on consistent metadata across news sources
- Static Analysis: Framework analyzes claims at specific time points rather than temporal evolution
- Binary Classification: Simplified real/fake classification without nuanced credibility levels

Technical Constraints

- Computational Complexity: Community detection algorithms add processing overhead
- Graph Construction: Limited to metadata similarity, may miss other relevant relationships
- Scalability: Performance evaluation limited to single dataset domain

SOLUTION APPROACH

Multi-Modal Framework Design

Feature Extraction

- BERT-base-uncased for 768-dimensional semantic embeddings
- TF-IDF vectorization with 5000 max features, (1,2) n-gram range
- Metadata encoding for speaker, political, subject, and temporal features

Graph Construction

- K-nearest neighbor approach with k=5
- Cosine similarity on metadata feature vectors
- Heterogeneous graph with approximately $5|V|$ edges

In Summary

- Extract BERT + TF-IDF + Metadata features
- Construct graph using cosine similarity of metadata
- Train GraphSAGE with 2 layers
- Apply Louvain/Label Propagation for interpretability
- Evaluate on PolitiFact dataset

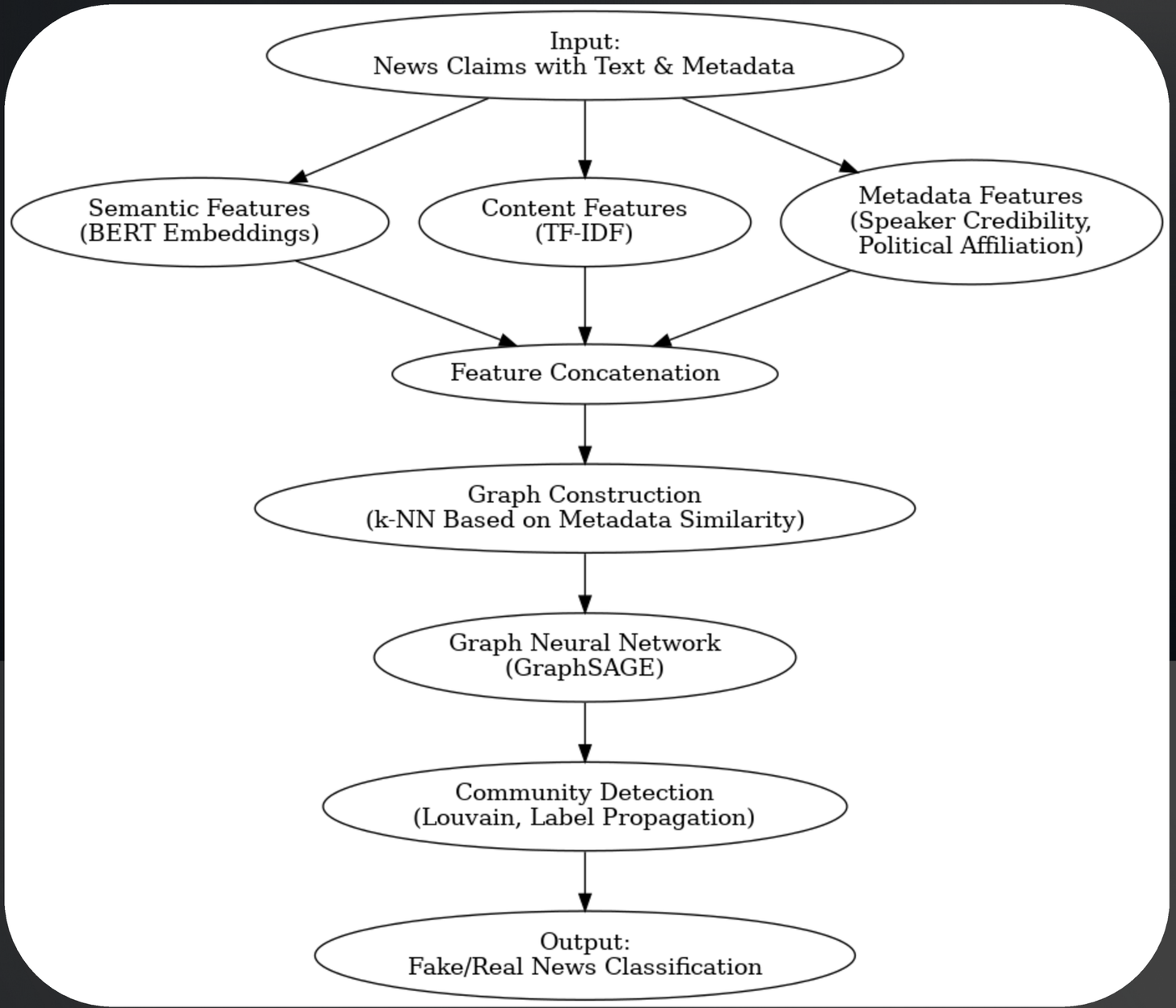
Neural Architecture

- Two-layer GraphSAGE with mean aggregation
- ReLU activation and dropout regularisation (p=0.5)
- Cross-entropy loss for binary classification

Community Analysis

- Louvain algorithm for modularity optimisation
- Label propagation for neighbour-based community assignment

BLOCK DIAGRAM



Problem Approach Overview

- Multi-Modal Graph Neural Network (GNN) for Fake News Detection.
- Integration of BERT-based semantic embeddings with metadata-driven structural features.
- Construction of a heterogeneous graph based on metadata (credibility, political affiliation, topical similarity).

FEATURE EXTRACTION

- Semantic Features: 768-dim BERT embeddings.
- Content Features: TF-IDF vectors (5000 max features, bigrams, frequency thresholds).
- Metadata Features:

1. Speaker credibility.
2. Political affiliation (one-hot encoded).
3. Subject category.
4. Temporal features (publication date, day).

Graph Construction

- k-NN Graph constructed with k=5 using cosine similarity over metadata features.

- Each node (news claim) connected to its 5 most similar neighbors.Metadata used includes:

1. Speaker credibility,
2. Political affiliation,
3. Subject category,
4. Temporal features.

- Edges represent contextual similarity between claims, not text similarity.

- Helps the model capture structural relationships crucial for fake news detection.

GNN ARCHITECTURE

- 2-layer GraphSAGE with mean aggregation.
- Community-aware feature fusion via Louvain algorithm.
- Dropout applied (p=0.5); cross-entropy loss function for binary classification.

COMMUNITY DETECTION

- Louvain and Label Propagation algorithms applied for community detection.
- Helps in identifying clusters of related fake news claims.
- Louvain method optimizes modularity to find dense communities.
- Label Propagation assigns labels based on neighbor majority voting.
- Reveals patterns and structures in the misinformation network.
- Enhances interpretability by showing how fake news spreads within communities

ANALYSIS OF SOLUTION

ADVANTAGES

- Fusion of content and structure enhances detection capability.
- Community detection improves explainability of misinformation patterns.
- Robust across temporal splits.

CHALLENGES

- Metadata availability inconsistency.
- Alignment between community structure and true labels remains low (ARI/NMI).
- Generalization to multilingual or cross-domain datasets requires further research.

SOFTWARE, LIBRARIES USED

- Python 3.9+
- PyTorch Geometric (PyG)
- HuggingFace Transformers (BERT)
- Scikit-learn (for TF-IDF, evaluation)
- NetworkX (Graph operations)
- Community (Louvain algorithm)
- Matplotlib, Seaborn (Visualization)

IMPLEMENTATION STATISTICS

- Training Time: ~2.5 hours (GPU-enabled environment).
- Louvain Community Detection: < 5 minutes.
- Label Propagation: ~15 minutes.
- Dataset Size: 41,054 news claims.
- Graph Edges: Approx. 5|V| edges (k=5).

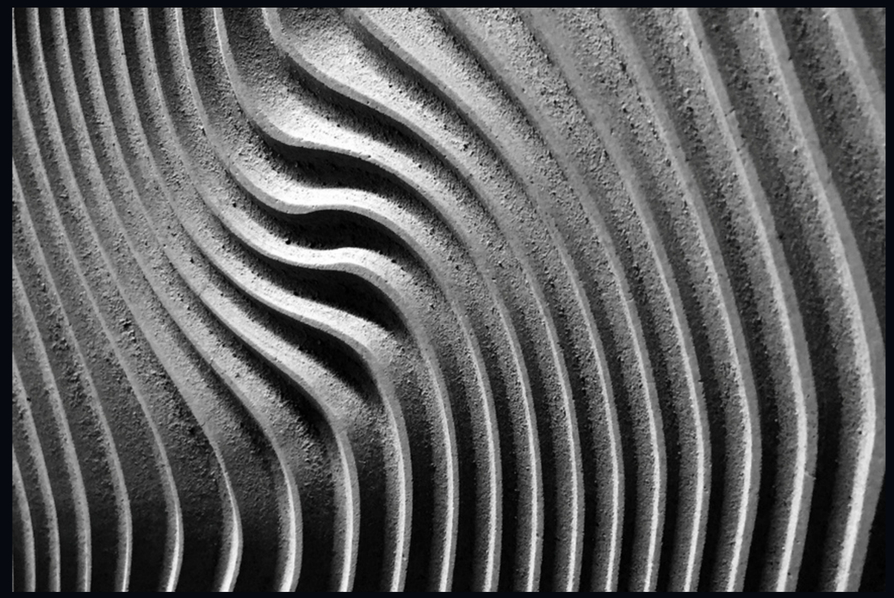
EXPERIMENTAL SETUP

Dataset: PolitiFact Fake News Dataset

★ **Split:**
- Stratified split.
- Temporal split (training: past data; testing: future data).

★ **Baseline Comparisons:**
- TF-IDF + SVM.
- BERT + Linear Classifier.
- TF-IDF + GraphSAGE.
- Metadata + GraphSAGE.

★ **Evaluation on GPU for GNN training.**



DATASET & ITS RELEVANCE

PolitiFact Dataset:

- 41,054 labeled news claims (Fake/Real).
- Rich metadata (speaker, party, topic, time).

Relevance:

- Real-world political fake news claims.
- Temporal spread and speaker network available.
- Suitable for graph-based modeling and temporal generalization analysis

METRICS USED + JUSTIFICATION

CLASSIFICATION METRICS:

1

- Accuracy.
- Precision.
- Recall.
- F1-Score.
- AUC-ROC.

COMMUNITY DETECTION METRICS:

2

- Adjusted Rand Index (ARI).
- Normalized Mutual Information (NMI).

JUSTIFICATION

3

- Accuracy & F1-score for balanced performance measure.
- ARI & NMI to evaluate clustering alignment with ground truth.

RESULTS

	Method	Accuracy	Precision	Recall	F1-Score
0	TF-IDF + SVM	62.3	0.618	0.634	0.626
1	BERT + Linear Classifier	71.2	0.706	0.721	0.713
2	TF-IDF + GraphSAGE	66.7	0.664	0.671	0.667
3	Metadata + GraphSAGE	58.9	0.583	0.595	0.589
4	Our Method	76.9	0.714	0.778	0.744

Classification Performance

- Our Method outperforms all baselines with highest accuracy (76.9%) and F1-Score (0.744).
- BERT + Linear Classifier shows strong performance but lacks structural feature integration.
- TF-IDF + GraphSAGE improves over TF-IDF + SVM by leveraging graph structure.
- Metadata + GraphSAGE alone performs poorly, indicating content features are essential.
- Multi-modal fusion (Our Method) captures both semantic and structural information effectively.

Community Detection

- Louvain detected fewer communities (237), representing coarse-grained clusters.
- Label Propagation identified more fine-grained communities (3710).
- NMI score is higher for Label Propagation (0.1743), indicating better information retention.
- ARI scores are low for both, showing limited alignment with ground truth labels.
- Graph construction captures some structural patterns, but further improvement needed for community-ground truth alignment.

	Algorithm	Communities	ARI	NMI
0	Louvain	237	0.0013	0.0661
1	Label Propagation	3710	0.0014	0.1743

RESULTS INTERPRETATION

- Multi-modal fusion significantly outperforms unimodal baselines.
- BERT semantic embeddings + metadata structural features are highly complementary.
- Community detection reveals hidden misinformation clusters.
- Model remains robust in temporal generalization scenarios.
- Structural features like speaker credibility and political affiliation play a crucial role.

CONCLUSION

- Proposed a novel Multi-Modal GNN for fake news detection.
- Achieved 76.9% accuracy on PolitiFact dataset.
- Outperformed traditional and unimodal GNN models.
- Enabled explainable detection through community analysis.
- Demonstrated temporal generalizability and metadata usefulness.

POTENTIAL FUTURE WORK

- Extend to multilingual fake news detection.
- Incorporate heterogeneous GNNs for multipl node/edge types.
- Develop attention-based edge weighting.
- Improve community-ground truth alignment.
- Evaluate across other misinformation domains (health, science).

REFERENCES

- S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
- J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
- N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in Proc. ACM CIKM, 2017, pp. 797–806.
- T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in Proc. AAAI, vol. 34, no. 1, pp. 549–556, 2020.
- D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: Multimodal variational autoencoder for fake news detection," in Proc. WWW, 2019, pp. 2915–2921.

THANK YOU