Beyond Veracity: A Dynamic Framework for Modeling Adversarial Narratives in the Age of Misinformation

Paul Lowndes
ZeroTrust@NSHkr.com

January 1, 2025

Abstract

This paper presents the Dynamic Adversarial Narrative Network (DANN) framework, a novel approach to modeling the evolution and propagation of narratives in online spaces. We introduce mathematical formulations for analyzing narrative dynamics, incorporating veracity assessment, influence measurement, and reputational impact. The framework employs an ontology-free approach to knowledge representation and introduces ephemeral narrative graphs for dynamic analysis. Building upon Large Concept Models (LCMs), we present a modular architecture that enables multi-step reasoning and multi-source fusion. Through real-world case studies, we demonstrate how DANN can help identify and potentially mitigate harmful narrative patterns. We conclude by discussing ethical implications and safeguards against potential misuse of this technology.

1 Introduction

1.1 Background

The proliferation of online platforms has created unprecedented opportunities for narrative manipulation and targeted harassment campaigns. Traditional Multi-Agent Reinforcement Learning (MARL) approaches fail to capture the complex dynamics of these interactions, particularly when powerful actors leverage platform mechanics to amplify harmful narratives. This paper builds upon the work of Large Concept Models (LCMs) [**?**], integrating their capabilities with an ontology-free approach to knowledge representation that better captures the dynamic nature of online narratives.

1.2 Contributions

This paper makes the following contributions:

- A formal mathematical framework for modeling narrative dynamics in adversarial contexts
- Novel mechanisms for quantifying and tracking reputational damage
- Introduction of ephemeral narrative graphs for dynamic analysis
- A modular architecture supporting multi-step reasoning and multi-source fusion
- Practical strategies for detecting and mitigating coordinated manipulation

2 Framework Overview

2.1 Fundamental Spaces

Let $\mathcal{E}_G$ represent the global embedding space where:

$$\mathcal{E}_G = \{\mathbf{e} \in \mathbb{R}^d : \|\mathbf{e}\| \leq 1\} \tag{1}$$

For each agent a_i, we define a local embedding space $\mathcal{E}_i$ with mapping function ϕ_i:

$$\phi_i : \mathcal{E}_i \to \mathcal{E}_G \tag{2}$$

2.2 Ephemeral Narrative Graphs

We introduce query-specific narrative graphs $N_{i,Q}(t)$ for agent a_i at time t:

$$N_{i,Q}(t) = f_N(K_i(t), B_i(t), Q, \theta_i) \tag{3}$$

where Q represents the query or analysis context, and θ_i represents agent-specific parameters.

2.3 Knowledge and Belief Sets

For agent a_i, we define:

$$K_i(t) = \{\mathbf{e} \in \mathcal{E}_i : p_K(\mathbf{e}, t) > \tau_K\} \tag{4}$$

$$B_i(t) = \{\mathbf{e} \in \mathcal{E}_i : p_B(\mathbf{e}, t) > \tau_B\} \tag{5}$$

where p_K and p_B are probability functions for knowledge and belief respectively.

3 Enhanced Veracity Function

3.1 Multi-Source Fusion

We extend the veracity function to incorporate multiple information sources:

$$V(e, T, a_i, C, t) = \sum_{k=0}^{t} \lambda^{t-k} [\sum_{s \in S} w_s \cdot R(s) \cdot v_s(e, k)] \tag{6}$$

where S is the set of information sources, w_s is the source-specific weight, $R(s)$ is the reliability score, and v_s is the source-specific veracity assessment.

3.2 Source Reliability Assessment

The source reliability function incorporates multiple factors:

$$R(s) = \alpha H(s) + \beta E(s) + \gamma(1 - B(s)) + \delta \sum_{j \in J} \omega_j C_j(s) \tag{7}$$

where:

- $H(s)$: Historical accuracy
- $E(s)$: Domain expertise
- $B(s)$: Measured bias
- $C_j(s)$: Corroboration from independent source j

4 Multi-Step Reasoning Framework

4.1 Reasoning Pipeline

We implement a multi-step reasoning process:

Algorithm 1 Multi-Step Reasoning Process

1: Extract relevant information from sources
2: Construct ephemeral narrative graph
3: Perform entity disambiguation
4: Apply source credibility weights
5: Generate reasoning chain
6: Produce final analysis

4.2 Modular Architecture

The system is composed of independent modules:

- Information Extraction Module
- Graph Construction Module
- Entity Resolution Module
- Analysis Engine
- Verification Module

5 Implementation and Safeguards

[Previous sections on Implementation and Safeguards remain unchanged]

6 Dataset and Evaluation

[Previous sections on Dataset and Evaluation remain unchanged]

7 Limitations

- Computational complexity of full network analysis
- Challenges in ground truth determination
- Potential for system manipulation
- Privacy preservation concerns
- Scalability of ephemeral graph generation
- Reliability of source credibility assessment

8 Discussion and Future Work

8.1 Future Directions

- Integration with platform-specific monitoring tools
- Development of early warning systems
- Enhanced privacy-preserving mechanisms
- Improved temporal modeling capabilities
- Refinement of multi-source fusion techniques
- Optimization of ephemeral graph generation

9 Conclusion

The DANN framework provides a structured approach to understanding and potentially mitigating online narrative manipulation. The introduction of ephemeral narrative graphs, multi-source fusion, and modular architecture enhances its capability to handle complex, real-world scenarios. While powerful, it must be developed and deployed with careful consideration of ethical implications and potential misuse. Future work should focus on practical implementation strategies and robust safeguards.

References

[1] Large Concept Model. https://ai.meta.com/research/publications/large-concept-models-language-modeling-in-a-sentence-representation-space/