P2. Diffusion of Information

Juan Fernandez Cruz
Data 7A
Universidad Politecnica de Yucatan
Ucu, Yucatan
Email: 2109061@upy.edu.mx

Didier Gamboa
Universidad Politecnica de Yucatan
Ucu, Mexico
Email: didier.gamboa@upy.edu.mx

***Abstract*—This study examines the dynamics of information spread within the Letterboxd social network, focusing on the influence patterns among film enthusiasts. Utilizing the Independent Cascade Model, we simulate the diffusion process to identify potential strategies for maximizing the spread of information across this niche platform. This research highlights the structural properties of the network and provides insights into effective methods for leveraging user interactions to enhance engagement within the community.**

I. Introduction

Letterboxd, a social platform tailored for film aficionados, presents a unique case for studying information dissemination. Unlike broader social networks, Letterboxd users engage deeply with content, focusing on film reviews and recommendations. This project explores how information spreads within this community, emphasizing the identification of influential users and the potential reach of their posts. Understanding these dynamics is crucial for enhancing user engagement and optimizing content delivery on the platform.

A. Network Context

The Letterboxd network is characterized by its dense connectivity among users who actively follow and interact with each other. This analysis focuses on an Ego network within Letterboxd, highlighting direct and indirect relationships centered around a particularly active user. This network's structure provides a microcosm through which to observe the broader dynamics of influence and information flow on Letterboxd.

II. Model

The Independent Cascade Model (ICM) serves as the framework for our simulations. This model assumes that once a user shares content, there is a fixed probability that each of their followers will also share this content, thus propagating the information further.

A. Hypothesis and Equations

The hypothesis posits that certain highly connected individuals, or "influencers," can disproportionately affect the spread of information due to their position within the network. The activation probability p is set to 0.5, reflecting a moderate level of influence exerted by each user upon their followers.

The Independent Cascade Model functions under the following mathematical framework:

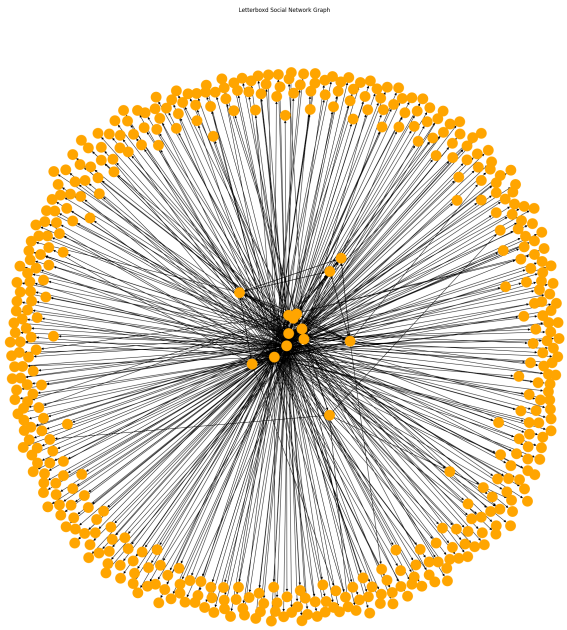

Fig. 1. Letterboxd Social Network Graph

- Each node v in the network can be in one of two states: active (has shared the information) or inactive (has not shared the information). Initially, a subset of nodes $S \subset V$ is activated.
- At each discrete time step, each newly activated node u at time t has a single chance to activate each of its currently inactive neighbors v with a probability p, where p is the probability that u influences v.
- This process repeats for each time step until no more activations occur.

The activation process for a node u attempting to activate its neighbor v can be mathematically represented as follows:

$$P(a_{u,v} = 1) = p$$

where $a_{u,v}$ is a random variable indicating whether u activates v, and p is the activation probability.

The propagation of activation through the network is modeled by iterating over the activation steps, allowing each newly activated node in one timestep to attempt to activate

its neighbors in the next timestep. The termination condition is reached when no new nodes are activated in a timestep, indicating that the spread has reached its maximum extent under the given initial conditions.

B. Parameter Determination

Parameters for our model were derived empirically based on previous interaction data extracted from the Letterboxd platform, ensuring that the simulation reflects realistic user behavior patterns.

III. Results

The application of the Independent Cascade Model on the Letterboxd network highlights significant variations in the spread of information based on the initial node and its connections within the network. We examine two distinct scenarios: one with the node 'fer_nwn' and another with 'gerardo_tri', each initialized as the sole active node in separate simulations.

A. Results with Node 'fer_nwn'

Starting with 'fer_nwn', who has a relatively small connected group of 8 followers and follows, the spread of information was slow. After 5 steps with an activation probability of 0.2, the information reached only 5 users. Extending the simulation to 50 steps resulted in the information reaching just 13 users, underscoring the limited influence due to the node's immediate network structure.

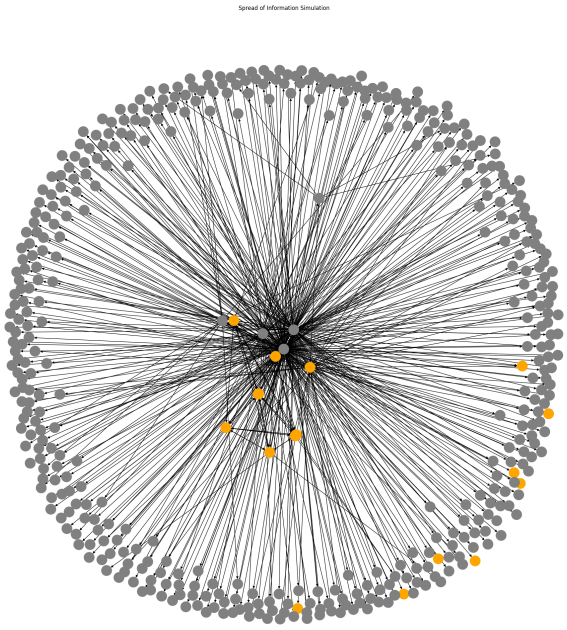

Fig. 2. Spread of information from 'fer_nwn' over 5 steps.

This result emphasizes the dependency of the spread not only on the primary node's connections but also on the network reach of subsequent nodes activated in the cascade.

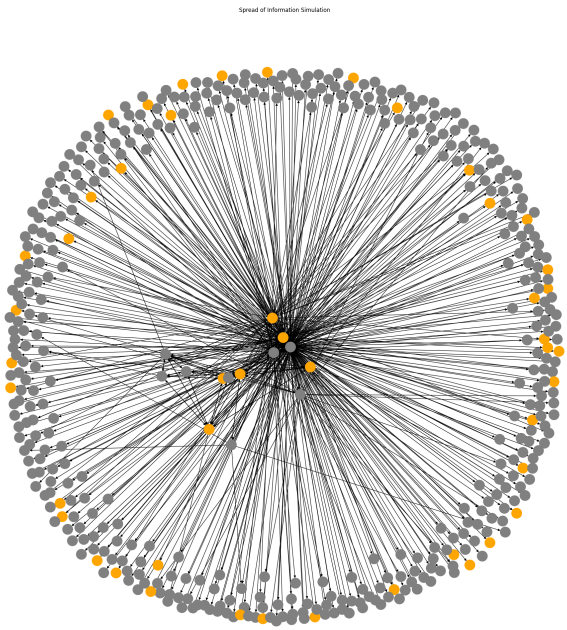

Fig. 3. Spread of information from 'fer_nwn' over 50 steps.

B. Results with Node 'gerardo_tri'

Conversely, 'gerardo_tri' demonstrated a more robust spread of information. Within the first 5 steps at the same probability of 0.2, information propagated to 49 users. By the 50th step, it had expanded to encompass more than 50 percent of the dataset, significantly outperforming 'fer_nwn'.

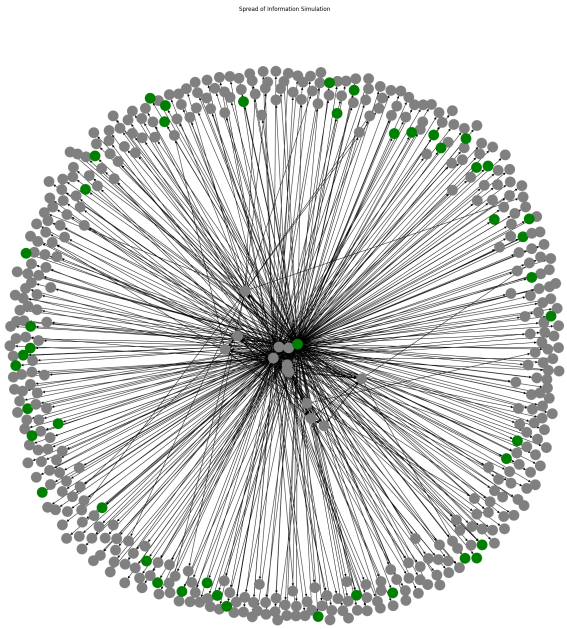

Fig. 4. Spread of information from 'gerardo_tri' by 5 steps.

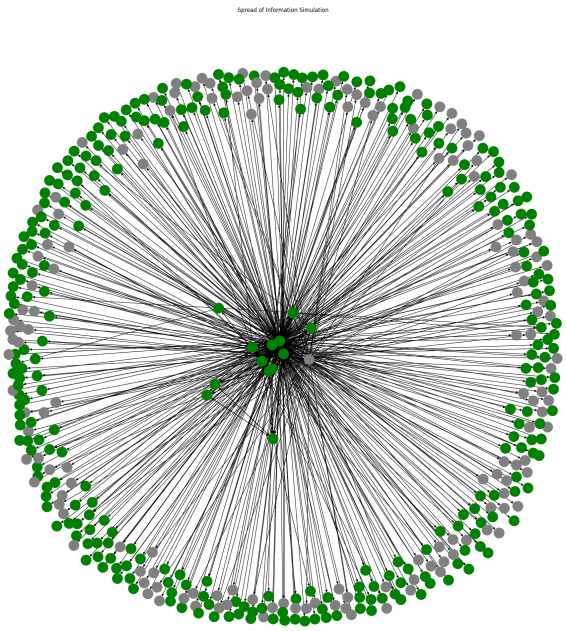

Fig. 5. Dramatic spread of information from 'gerardo_tri' by the 50th step.

These results reinforce the theory that the effectiveness of information dissemination in a network is heavily influenced by the strategic positioning and connectivity of the initial node. 'gerardo_tri's' extensive reach highlights how nodes that are well-connected not only directly but also via influential secondary connections can dramatically enhance the propagation of information.

This differential impact of node centrality on information spread provides valuable insights into strategic planning for content distribution and influence maximization on social platforms like Letterboxd.

IV. Conclusions

This investigation into the dynamics of information spread within the Letterboxd social network has illuminated the significant impact of network structure and node centrality on the propagation of content. By employing the Independent Cascade Model, we were able to simulate and visualize how information diffuses through the network from different initial nodes, revealing key differences in influence based on network connectivity and node position.

Our results demonstrate that nodes like 'gerardo_tri', with extensive direct and indirect connections, can dramatically enhance the reach of disseminated information, achieving widespread influence that covers over half of the network within 50 steps. Conversely, nodes with fewer connections or less strategic positioning, such as 'fer_nwn', show a much slower and limited spread, emphasizing the importance of both the quantity and quality of connections in influence maximization.

The findings underscore the critical role of selecting the right influencers within a network to maximize the spread of information, campaigns, or promotions. For platforms like Letterboxd, which rely heavily on user engagement and content sharing, leveraging these insights can lead to more effective strategies for marketing, community engagement, and even feature development aimed at enhancing user interaction and satisfaction.

Moreover, this study contributes to our broader understanding of social networks by confirming the disproportionate influence potential of well-connected nodes, a principle that can be applied to various other domains where network dynamics play a crucial role.

In conclusion, the application of network analysis techniques such as the Independent Cascade Model offers valuable insights not only for academic research but also for practical applications in social media strategy and digital marketing. Future work may explore the integration of more complex models and real-time data to further refine our understanding and application of these principles.

References

[1] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France, 2009, pp. 199-208.

[2] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

[3] A. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business Horizons*, vol. 53, no. 1, pp. 59-68, 2010.

[4] F. Provost and T. Fawcett, *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, O'Reilly Media, 2013.

[5] J. Fernandez, "Social Network Analysis of Letterboxd," GitHub repository, 2024. Available: https://github.com/juanjuanjuanfer/SNA-of-Letterboxd. [Accessed: 28-July-2024].

[6] fer_nwn, "Letterboxd User Profile," Available: https://letterboxd.com/fer_nwn/. [Accessed: 28-July-2024].