*Proceedings of the 2025 Winter Simulation Conference*
*E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.*

# LEVERAGING USER EMBEDDINGS FOR IMPROVED INFORMATION DIFFUSION VIA AGENT-BASED MODELING

Xi Zhang[1], Chathika Gunaratne[1], Robert M. Patton[1], and Thomas Potok[1]

[1]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

## ABSTRACT

Understanding the cascade of behaviors influencing the actions of individuals in social networks is crucial to a multitude of application areas such as countering adversarial information operations. This paper presents a novel approach that leverages cosine similarity between user embeddings as the core mechanism of agent-based information diffusion modeling. We utilize SAGESim, a pure-Python agent-based modeling framework designed for distributed multi-GPU HPC systems, to simulate large-scale complex systems. Our methodology employs the Qwen3 embedding model to generate high-dimensional vector representations of social media users, capturing their behaviors and preferences. A cosine similarity-based influence mechanism, where agents with higher embedding similarity exhibit increased likelihood of information transmission, is evaluated. The framework enables scalable simulation of information diffusion by modeling individual agent interactions based on their semantic similarity rather than traditional network topology alone. Our approach demonstrates improved prediction accuracy by incorporating deep user representations into the agent-based modeling paradigm.

## 1 INTRODUCTION

Predicting how information spreads in social networks is a longstanding challenge with real-world impact on public health, information security operations, and marketing. Classical models such as the Independent Cascade (IC) and Linear Threshold (LT) treat diffusion as a contagion process where users probabilistically or conditionally influence their neighbors (Kempe, Kleinberg, and Tardos 2003), while more recent work has introduced temporal and neural approaches that infer diffusion paths from cascades (Gomez-Rodriguez, Leskovec, and Krause 2011) or model them using deep networks (Chen, Zhou, Zhang, Trajcevski, Zhong, and Zhang 2019). However, these approaches—whether classical or neural—primarily rely on network topology and assume static influence patterns, overlooking individual differences in user behavior and ignoring richer user-level semantics that may fundamentally shape influence and adoption dynamics. Furthermore, agent-based simulation provides the added benefit of traceability not easily derived from deep neural surrogates – where the adoption of a particular agent can be traced back and explained by the cascade of prior agent behaviors and interactions. Representation learning offers a promising alternative: user embeddings can encode nuanced behavioral traits and preferences. Prior studies have used such embeddings for feature extraction in predictive models, but rarely as active components in simulation environments. Similarity-based influence mechanisms—motivated by the principle of homophily have also been explored, though typically using simple features or static metadata. We argue that deep embedding-based similarity provides a more expressive basis for modeling influence. Our work integrates LLM-derived user embeddings into scalable agent-based simulations via cosine similarity-driven influence, offering a behavior-aware alternative to topology-based diffusion models.

## 2 METHODOLOGY

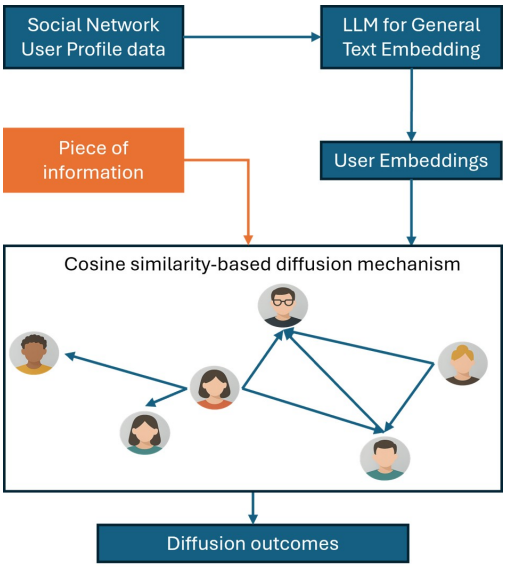


Figure 1: Cosine similarity-based information diffusion model implemented on SAGESim.

We evaluated our methodology on the Memetracker dataset (Leskovec, Backstrom, and Kleinberg 2009), which captures the temporal dynamics of phrase diffusion across mainstream news outlets and blogs. From this data, we constructed 1,417 high-quality quote cascades, encompassing 291,988 timestamped propagation events across 23,026 fully connected sources. To simulate this process, we employed SAGESim—a distributed agent-based modeling framework specifically designed for multi-GPU high-performance computing environments, including exascale systems like the Frontier and Summit supercomputers (Gunaratne, Zhang, Patton, Coletti, and Potok 2024). SAGESim retains key traits of classical ABMs, such as non-differentiable agent behavior and heterogeneous agent types, while enabling efficient large-scale simulation with full entity-level resolution and traceability.

Each information source is modeled as an agent, $i$, represented by a 1,024-dimensional semantic embedding vector, $x_i = \mathbf{F}(\mathrm{q}i)$, obtained by encoding the source's aggregated quote content, q$i$, using the Qwen3 transformer-based embedding model, $\mathbf{F}$. This embedding effectively captures the agent's preferences and topical focus reflected in the quotes. Rather than relying on static network topology, we introduce a cosine similarity-based influence mechanism. The similarity between agents $i$ and $j$ is given by $\mathbf{G}_{ij} = \cos(x_i, x_j)$, which captures the cosine similarity between the agents' semantic vectors. If this similarity value exceeds a threshold $\tau$, and $j$ is infected, the transmission probability to $i$ is $P(\mathrm{infect}_{j\to i}) = p \cdot \frac{\mathbf{G}_{ij}+1}{2}$, where $p$ is the base infection strength parameter. Evolutionary algorithms are employed to precisely calibrate parameter $p$ and $\tau$, enabling the model to reflect observed cascade dynamics accurately. This framework enables scalable simulation of embedding-based information diffusion, where cosine similarity between user representations drives influence dynamics through distributed GPU acceleration.

## ACKNOWLEDGEMENTS

This manuscript has been authored by UT-Battelle LLC under contract DE-AC0500OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (https://energy.gov/downloads/doe-public-access-plan). This research was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy.

## REFERENCES

Chen, X., F. Zhou, K. Zhang, G. Trajcevski, T. Zhong, and F. Zhang. 2019. "Information diffusion prediction via recurrent cascades convolution". *IEEE Transactions on Knowledge and Data Engineering* 32(2):368–380.

Gomez-Rodriguez, M., J. Leskovec, and A. Krause. 2011. "Inferring networks of diffusion and influence". In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1019–1028. ACM.

Gunaratne, Chathika and Zhang, Xi and Patton, Robert M and Coletti, Mark and Potok, Thomas 2024, Mar. "SAGESim: Scalable Agent-based GPU-Enabled Simulator".

Kempe, D., J. Kleinberg, and É. Tardos. 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*, 137–146. ACM.

Leskovec, J., L. Backstrom, and J. Kleinberg. 2009. "Meme-tracking and the dynamics of the news cycle". In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 497–506.