

Dual-Force Community Evolution: Integrating Peer Influence and Interest Similarity in Social Networks

Bowen Guo

School of Engineering, Architecture and Information Technology, University of Queensland,
Brisbane, Australia

bowen.guo1@student.uq.edu.au

Abstract. Understanding how communities evolve requires accounting for both social influence and individual interests. In this paper, the Community Evolution Framework, which leverages the Louvain algorithm for initial community detection and extends it with a dynamic evolution mechanism was proposed. The framework integrates the peer influence and interest similarity through a tunable scoring function, allowing community memberships to be updated iteratively until convergence. To enhance interpretability, this paper further incorporates the large language models to generate thematic explanations of evolved communities based on interest distributions. The experiments conducted on a large-scale student social network dataset demonstrate that the proposed framework achieves strong modularity while capturing realistic membership reorganization patterns. The results highlight that balancing peer pressure and personal interests produces more cohesive yet adaptable communities. This work provides a quantitative and interpretable approach to studying the community evolution, with potential applications to broader the network science and the social computing tasks.

Keywords: Community Evolution, Louvain Algorithm, Social Network Analysis, Large Language Models (LLMs).

1. Introduction

In real life, individuals often face a dilemma between following social pressure and pursuing their own interests. On the one hand, peer influence strongly shapes behavior: people tend to follow what their friends watch, buy, or endorse, a mechanism that also explains the spread of risky habits such as smoking or drinking among adolescents. On the other hand, interest-driven preferences lead individuals toward communities that reflect their personal identities, such as music groups or volunteer circles. These two forces—peer pressure and interest alignment—are frequently in tension, yet both are crucial in determining how communities form, evolve, and stabilize.

Building on prior research, the Louvain algorithm [1] provides an efficient method for partitioning large networks into communities and has become a widely used foundation for structural analysis. A comprehensive survey by Fortunato [2] further highlighted the strengths and limitations of community detection methods, noting that modularity-based techniques, while effective, often struggle to capture overlapping and evolving structures. More recently, Rossetti and Cazabet [3] provided a systematic overview of community discovery in dynamic networks, emphasizing the importance of frameworks that account for temporal evolution and member reorganization. However, such static partitions cannot capture the dynamic nature of real communities [4]. Palla et al. [5] demonstrated that large groups remain stable only when membership is fluid, underscoring the need to model how individuals shift between communities over time [6]. Motivated by these insights, this study proposes a new framework for community evolution that explicitly incorporates both social influence and individual interests as driving factors, thereby capturing the dual forces that shape community change. In addition, this study leverages Large Language Models (LLMs) to analyze and interpret the evolution results, enhancing the explainability of community dynamics.

Further contributions in this area have examined how communities evolve over time. Hopcroft et al. [7] investigated methods for tracking community structures in large linked networks, highlighting the temporal nature of group stability. Asur et al. [8] proposed an event-based framework [6] to characterize community evolution through merging, splitting, and dissolution. Greene et al. [9, 10]

later extended this line of work by analyzing dynamic social networks, providing robust methodologies to monitor and evaluate community transitions. These studies collectively underscore the necessity of temporal perspectives in community analysis, which aligns closely with the goals of the present work.

To evaluate this approach, this study analyzes a dataset of 15,000 high school students collected from a well-known social platform between 2006 and 2009 [9]. The dataset contains both interest features (e.g., basketball, music, shopping) and demographic attributes (e.g., graduation year, gender, number of friends), offering a unique opportunity to study how structural ties and personal preferences jointly influence community membership and its evolution. This study first applies Louvain-based detection to partition the network into communities that reflect social connections. This study then extends this static view by introducing an evolution mechanism that models how individuals may realign their memberships in response to both peer influence and interest similarity.

2. Methods

2.1. Dataset Preparation

The dataset employed in this study is the publicly available Student Social Network dataset hosted on Kaggle. It contains approximately 15,000 high-school student profiles collected between 2006 and 2009. Each record consists of two categories of attributes [9]: 1) Interest features represented as binary indicators (e.g., basketball, music, dance, shopping), 2) Demographic attributes such as graduation year, gender, and number of friends.

To ensure the reliability of subsequent analysis, several preprocessing steps were performed. Records with missing gender information or implausible ages outside the 10–22 range were removed. Missing values in interest features were imputed with zeros to avoid distortions in similarity calculations. Each student was then assigned a unique identifier, and interest vectors were normalized using ℓ_2 – *normalized*, ensuring that all features contribute equally to similarity computation.

2.2. Graph Construction and Baseline Community Detection

2.2.1 Interest Representation and Cosine Similarity

Cosine similarity has been widely adopted in clustering tasks as a robust measure of affinity between high-dimensional vectors [10]. In this paper students are represented as binary interest vectors, where each dimension corresponds to an interest tag. To compute pairwise affinity, each vector is ℓ_2 – *normalized*, ensuring comparability across individuals. Given the normalized featured matrix $\hat{\mathbf{X}}$, the pairwise cosine similarities are obtained through the product matrix $\mathbf{S} = \hat{\mathbf{X}} \hat{\mathbf{X}}^T$. This produces a dense similarity graph in which entries S_{ij} represent the similarity between students i and j .

2.2.2 Top-k Sparsification and Graph Construction

Inspired by prior applications of k-nearest-neighbor (kNN) sparsification in spectral clustering, where retaining only the most salient local neighbors helps preserve meaningful structure while avoiding excessive density [11]. To reduce computational overhead and emphasize the most relevant ties, a k -nearest-neighbor truncation is applied to each row of $\mathbf{S}$ with $k = 20$. Only the strongest neighbors are retained, while self-connections are removed. The data are formalized as an undirected weighted graph $G = (V, E)$ in which each vertex denotes a student and the weight of an edge indicates the similarity score between two students.

2.2.3 Baseline Community Detection

Initial communities are identified using the Louvain algorithm, which optimizes modularity by iteratively reallocating nodes. Louvain is widely adopted in large-scale network analysis due to its scalability and sensitivity to mesoscale structure [1]. The resulting partition $\{c_i\}_{i=1}^n$ serves as the baseline community assignment for subsequent evolution experiments.

2.2.4 Community Interest Centroids

For each detected community C , an interest centroid vector is computed in the original binary feature space (prior to normalization):

$$\mu_C = \frac{1}{|C|} \sum_{i \in C} X_i. \quad (1)$$

Each dimension of μ_C reflects the prevalence of the corresponding interest within the community [12]. These centroids capture prototypical interest patterns and are later used as references in the community evolution framework.

2.3. Community Evolution Framework

While the Louvain algorithm yields a static partition of the similarity graph, real communities are inherently dynamic. To capture membership shifts driven by both social influence and interest similarity, a Community Evolution framework is introduced.

2.3.1 Scoring function

For a student u and candidate community C , a composite score is defined as

$$\text{Score}(u, C) = \alpha \cdot \text{NeighborPro}(u, C) + \beta \cdot \text{Sim}(\mathbf{x}_u, \mu_C) \quad (2)$$

Where

- $\text{NeighborPro}(u, C)$ denotes the proportion of u 's neighbors (in the Top-20 similarity graph) currently assigned to community C , reflecting peer pressure;
- $\text{Sim}(\mathbf{x}_u, \mu_C)$ is the cosine similarity between u 's interest vector and the centroid of community C , capturing interest alignment;
- α and β are tunable parameters controlling the relative influence of social versus interest-driven factors [13], with $\alpha + \beta = 1$.

2.3.2 Update rule

At each iteration, every student is reassigned to the community that maximizes $\text{Score}(u, C)$. If the highest score corresponds to the current label, the assignment remains unchanged. Once all labels are updated, community centroids μ_C are recomputed based on the new memberships.

2.3.3 Iteration and convergence

The procedure is repeated until no further label changes occur, indicating convergence. In practice, convergence is typically achieved within a limited number of iterations. The resulting community structure reflects a balance between social conformity and personal interests, determined by the parameter setting (α, β) . More specifically.

- Higher values of α emphasize neighbor conformity, producing more stable but less flexible communities.
- Higher values of β emphasize interest similarity, allowing greater reorganization and cross-community migration.
- Balanced weights ($\alpha \approx \beta$) capture the interplay of both forces.

2.4. Measurement Metrics and Analysis

Community quality is assessed using modularity, a widely adopted metric in network science [2]. Modularity measures the extent to which edges are concentrated within communities compared to a random baseline. Higher values indicate stronger internal cohesion and clearer separation between groups [14]. The modularity of the baseline Louvain partition is compared with that of the evolved partitions obtained from the Community Evolution Framework under different (α, β) settings to evaluate structural stability and reorganization.

2.5. LLM-based Interpretability

To provide qualitative interpretations of the evolved communities, an LLM-based post-hoc procedure is employed. First, the largest communities are identified by membership size. For each selected community, frequencies of predefined interest tags are aggregated and the top-10 tags are extracted to form a compact profile. These profiles are then used to construct natural-language prompts supplied to a general-purpose large language model, which returns concise descriptions of each community’s salient themes and plausible contextual explanations. This interpretability step is strictly post-hoc [15, 16] and does not affect community assignments or quantitative evaluation. It offers practitioner-friendly summaries that complement modularity-based analysis and facilitates understanding of how interest prevalence patterns align with the final community structure [17].

3. Results and Discussion

3.1. Results

The Community Evolution Framework was systematically evaluated using modularity as the primary structural quality metric. Table 1 presents the modularity values for the baseline Louvain partition alongside those obtained under different (α, β) parameter settings. The baseline Louvain method achieved a modularity score of 0.7584, which serves as a reference point for subsequent comparisons. When the evolution mechanism was introduced, the framework demonstrated clear adaptability: parameter settings that heavily favored social influence ($\alpha = 0.9, \beta = 0.1$) produced stable but less flexible clusters, with modularity reaching 0.7620. Conversely, interest-driven configurations ($\alpha = 0.1, \beta = 0.9$) resulted in greater reorganization of community membership, albeit at the cost of lower modularity 0.7434. The balanced configuration ($\alpha = 0.5, \beta = 0.5$) achieved the highest modularity 0.7608, suggesting that the interplay of peer pressure and interest alignment yields communities that are both cohesive and adaptive.

Table 1. Modularity Comparison under different parameter settings

α	β	Modularity
0.1	0.9	0.7434
0.3	0.7	0.7572
0.5	0.5	0.7608
0.7	0.3	0.7620
0.9	0.1	0.7620

The evolution results of the Community Evolution Framework are illustrated in Fig. 1, where the visualization focuses on the top six communities, each represented by the 60 highest-degree nodes.

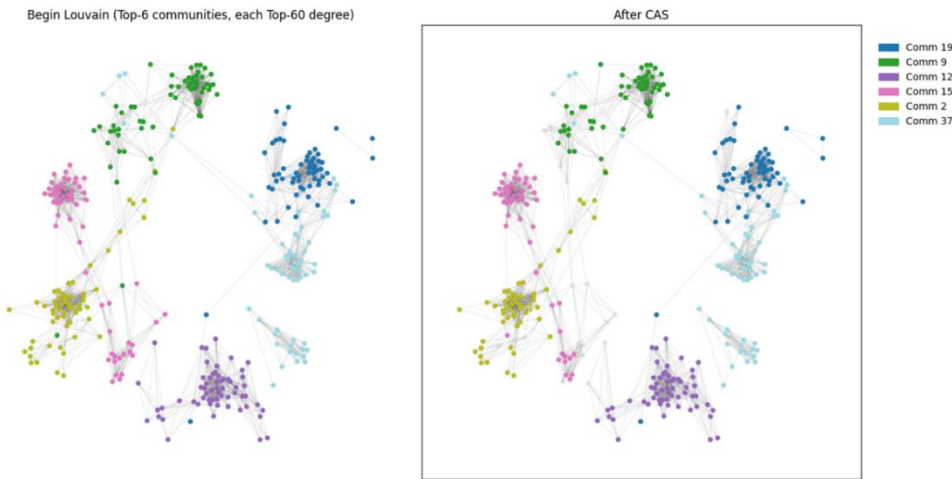


Figure 1. Community Evolution Result (Picture credit: Original)

To complement quantitative evaluation, LLM analysis was applied to the evolved communities. The model was prompted with the top interest distributions of major groups and asked to summarize their characteristics shown in Fig. 2.

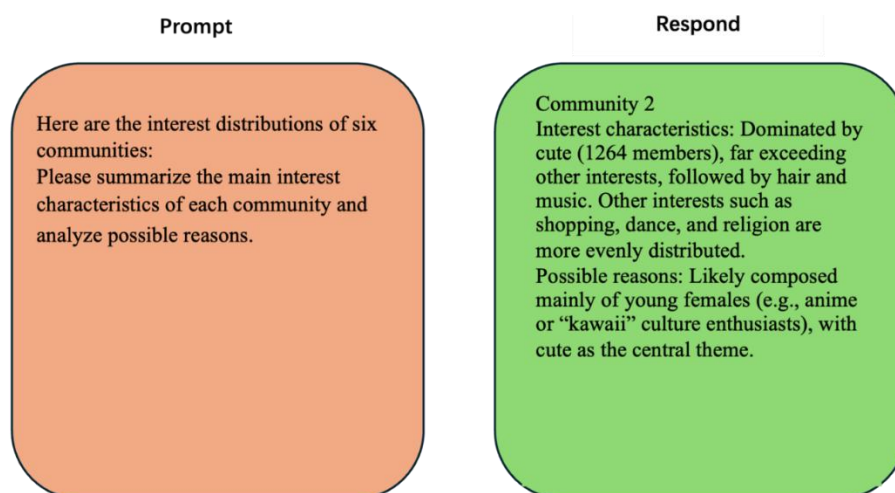


Figure 2. LLM interaction (Picture credit: Original)

3.2. Discussion

The results highlight the dual role of social influence and interest similarity in community evolution. High α values preserve stable clusters, while high β values encourage cross-community reassignment and reorganization. Balanced settings yield slightly higher modularity, indicating that considering both factors allows communities to remain cohesive while adapting to individual preferences.

These findings align with prior work emphasizing the interplay of peer pressure and personal interests. Unlike static methods, the proposed Community Evolution Framework captures dynamic adjustments in membership, offering a more realistic representation of evolving communities.

The use of large language models adds an interpretability layer by identifying thematic orientations (e.g., lifestyle, music, religion), providing cultural insights beyond modularity. Integrating quantitative evaluation with LLM-based interpretation thus offers a more comprehensive framework for community analysis, with potential applications in broader network science domains.

4. Conclusion

This work proposed the Community Evolution Framework to capture dynamic changes in social networks by leveraging the Louvain algorithm as the baseline community detection method and extending it with an evolution mechanism that incorporates both peer influence and interest similarity. To enhance interpretability, this paper further employed large language models to generate thematic explanations of community structures. Experiments on a large-scale student social network dataset demonstrated that the proposed framework maintains strong modularity while providing richer insights into how communities reorganize over time. In the future, the further study plans to apply this framework to heterogeneous datasets and explore predictive modeling of community evolution across diverse online platforms.

References

- [1] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp.* 2008; 2008 (10): P10008. doi: 10.1088/1742-5468/2008/10/P10008.
- [2] Fortunato S. Community detection in graphs. *Phys Rep.* 2010; 486 (3-5): 75-174. doi: 10.1016/j.physrep.2009.11.002.

- [3] Rossetti G, Cazabet R. Community discovery in dynamic networks: A survey. *ACM Comput Surv.* 2019; 51 (2): 1-37. doi:10.1145/3172867.
- [4] Ahmad H, Sajid M, Mazhar F, Fuzail M. Mapping unseen connections: Graph clustering to expose user interaction patterns. *J Future Artif Intell Technol.* 2025; 1 (4): 474-96. doi: 10.62411/faith.3048-3719-77.
- [5] Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. *Nature.* 2007; 446 (7136): 664-7. doi: 10.1038/nature05670.
- [6] Anuar SHH, Rahman NA, Abdullah R, et al. Bird flock effect-based dynamic community detection: Unravelling network patterns over time. *Alex Eng J.* 2025; 112: 177-208. doi: 10.1016/j.aej.2024.10.097.
- [7] Hopcroft J, Khan O, Kulis B, Selman B. Tracking evolving communities in large linked networks. *Proc Natl Acad Sci U S A.* 2004; 101 Suppl 1: 5249-53. doi: 10.1073/pnas.0307750100.
- [8] Asur S, Parthasarathy S, Ucar D. An event-based framework for characterizing the evolutionary behavior of interaction graphs. *ACM Trans Knowl Discov Data.* 2009; 3 (4): 1-36. doi: 10.1145/1631162.1631164.
- [9] Students' social network profile clustering [Internet]. 2025 [cited 2025 Sep 20]. Available from: <https://www.kaggle.com/datasets/zabihullah18/students-social-network-profile-clustering>.
- [10] Huang A. Similarity measures for text document clustering. In: *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008)*, Christchurch, New Zealand 2008 Apr 14 (Vol. 4, pp. 9-56).
- [11] Zhou D, Bousquet O, Lal T, Weston J, Schölkopf B. Learning with local and global consistency. *Advances in neural information processing systems.* 2003; 16.
- [12] Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett.* 2010; 31 (8): 651-66. doi: 10.1016/j.patrec.2009.09.011.
- [13] RafieeFard J, Teimourpour B, Shaghrouzi M, Jami M. A comparison of centrality measures for community evolution discovery. In: *Proc 8th Int Conf Web Res (ICWR)*. Tehran, Iran: IEEE; 2022. p.183-8. doi: 10.1109/ICWR54782.2022.9786228.
- [14] Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci U S A.* 2006; 103 (23): 8577-82. doi: 10.1073/pnas.0601602103.
- [15] Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: An overview of interpretability of machine learning. In: *Proc IEEE 5th Int Conf Data Sci Adv Anal (DSAA)*. Turin, Italy: IEEE; 2018. p.80-9. doi: 10.1109/DSAA.2018.00018.
- [16] Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv: 1702.08608.* 2017. doi: 10.48550/arXiv.1702.08608.
- [17] Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth. In: *Proceedings of the ACM SIGKDD workshop on mining data semantics 2012 Aug 12 (pp. 1-8).*