Don't Thread on Me: Predicting Online Community Health with Temporal GNNs

David Lupea
Department of Statistics
Stanford University
dlupea@stanford.edu

Kalyani Limaye
ICME
Stanford University
limayk@stanford.edu

Vedant Sahu
ICME
Stanford University
vsahu@stanford.edu

1 Introduction

Online communities such Stack Exchange sites, Reddit subreddits, and Discord servers constantly wrestle with structural decisions. Should a thriving community split into focused subcommunities? Should two inactive communities with overlapping interests merge? When should a community narrow its scope to improve quality? Right now, these decisions happen through moderator intuition, community votes, and ad hoc heuristics rather than a systematic methodology [Zhu et al., 2014, Kiene et al., 2016]. We propose a temporal graph neural network (GNN) framework that predicts community health trajectories under alternative structural configurations (e.g., split, merge, scope changes). This framework enables platform designers to test "what if" scenarios and make evidence-based decisions about community structure rather than relying on intuition.

2 Problem Setup

We demonstrate our framework on Stack Exchange, a network of 173 Q&A communities spanning diverse topics from programming to poker to cooking [Stack Exchange Inc., 2025]. Each site faces structural decisions: Stack Overflow (30M+ users) debates splitting by specialization, while small sites like Bioacoustics (<5K users) consider merging with related communities. This diversity in scale and maturity provides an ideal testbed for learning generalizable patterns about community structure optimization.

Dynamic, hierarchical representation. Each site at month t is represented as a two-level graph:

- **Level 1 (Dense topic graph)**: Nodes are tags/topics (e.g., Python, Machine Learning, Database); edges encode topic co-occurrence weighted by frequency. Topic features include popularity, growth rate, diversity (entropy), answer quality, and difficulty.
- **Level 2 (User-topic bipartite)**: Users connect to topics with contribution-weighted edges (questions, answers, votes). User features include reputation, tenure, activity level, expertise entropy, and retention indicators.

Prediction task. Given a community's 12-month graph sequence (months $t-11$ to t), we predict four engagement metrics at month $t+6$: questions per day, answer rate (percentage with accepted answers), user retention (fraction who remain active), and new user growth rate.

Structure comparison. To recommend optimal structures, we encode candidate interventions as graph transformations and predict outcomes under each:

- **Split**: Partition topics via clustering (e.g., "python-web" vs. "python-data") and reassign users by contribution weights to produce two child sites.
- **Merge**: Unify two small, overlapping sites by combining topic sets and user bases.
- **Narrow scope**: Prune peripheral topics/users with low engagement to reduce fragmentation.
- **Broaden scope**: Import related topics/users from neighboring communities.

For each alternative, we predict outcomes and rank by a weighted health score combining the four metrics. The recommended intervention is the one with highest predicted health.

3 Data

We use the **Stack Exchange Data Dump** available via Archive.org [Stack Exchange Inc., 2024], which provides historical, longitudinal records of all Q&A communities across the network. The dataset covers 173 distinct Q&A sites, spanning activity from 2008 to 2024. This rich temporal coverage enables learning community evolution patterns across diverse scales and maturity levels. Each site is released as a set of relational tables with a common schema, including: *Posts*, *Tags*, *Users*, *Comments*, *Votes*, *Badges*, and *Post History*.

4 Methodology

4.1 Model Architecture

Stage 1: Spatial encoding. Aggregate user features to topics; apply multi-layer *Graph Attention Networks (GAT)* [Veličković et al., 2018] on the dense topic graph; propagate enriched topic representations back to users to obtain user embeddings capturing both direct activity and topic-mediated relations.

Stage 2: Temporal modeling. Process 12-month graph sequences (months $t - 11$ to t) with an LSTM + temporal attention to emphasize salient periods (growth spurts, decline signals). Output is a community trajectory representation.

Stage 3: Multi-task prediction. Predict engagement metrics at $t + 6$: questions per day, answer rate, user retention, and new user growth. Separate prediction heads with joint training to leverage shared representations.

4.2 Training Strategy

Supervised learning on historical trajectories where each example consists of 12 months of graph history predicting metrics 6 months ahead. Temporal split: train on 2008–2020, validate on 2021–2022, test on 2023–2024. Multi-task loss using MSE for continuous metrics.

5 Experiments and Evaluation

Baselines. We compare against three baselines: (1) linear extrapolation from recent trends, (2) ARIMA time series forecasting, and (3) standard GCN without temporal modeling [Kipf and Welling, 2017] to isolate the benefit of temporal dynamics.

Evaluation tasks:

- **Trajectory prediction:** Predict engagement metrics 6 months ahead, stratified by community size (small < 500, medium $500 - 2000$, large > 2000 users). Metrics: MAE, RMSE, R^2.
- **Ablation studies:** Compare temporal window lengths (3, 6, 12 months) to determine optimal historical context for prediction.

References

P. Aragón, V. Gómez, D. García, and A. Kaltenbrunner. Identifying engagement patterns in stack exchange communities. *EPJ Data Science*, 12(1):43, 2023.

W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin. Graph neural networks for social recommendation. In *Proceedings of the 28th International Conference on World Wide Web*, pages 417–426, 2019.

W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

C. Kiene, A. Monroy-Hernández, and B. M. Hill. Surviving an "eternal september": How an online community managed a surge of newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1152–1156, 2016.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Stack Exchange Inc. Stack exchange data dump, 2024. URL `https://archive.org/details/stackexchange`. Accessed: 2025-10-21.

Stack Exchange Inc. About stack exchange, 2025. URL `https://stackexchange.com/about`. Accessed: 2025-10-21.

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

H. Zhu, R. E. Kraut, Y.-C. Wang, and A. Kittur. Identifying shared leadership in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3431–3440, 2014.

A Related Work

Graph Neural Networks (GNNs) have become a powerful tool for modeling social networks. Fan et al. [2019] introduced GraphRec, a framework that integrates user-user social graphs with user-item interactions for social recommendation, addressing challenges of sparsity and heterogeneity . More generally, Hamilton et al. [2017] proposed GraphSAGE, an inductive approach that learns node embeddings by aggregating neighborhood information, enabling scalability and generalization to unseen nodes. These methods illustrate how GNNs can capture both relational structure and evolving user behavior, forming the methodological basis for our temporal community modeling.

There is also extensive research on the dynamics of online communities. Aragón et al. [2023] analyzed long-term engagement patterns in Stack Exchange communities, showing how user retention and participation trajectories shape site health. Earlier studies of governance and community resilience highlight how leadership structures and newcomer surges impact sustainability [Zhu et al., 2014, Kiene et al., 2016]. Together, these works motivate our focus on predicting community health under structural interventions using temporally-aware GNNs.

B Prediction Task

Given the temporal graph sequence $\mathcal{S}_t$, we predict a vector of engagement metrics at time $t+6$:

$$\hat{\mathbf{y}}_{t+6} = f_\theta(\mathcal{S}_t) = \begin{bmatrix} \widehat{\text{QPD}} \\ \widehat{\text{AnsRate}} \\ \widehat{\text{Retention}} \\ \widehat{\text{NewUserGrowth}} \end{bmatrix}_{t+6}$$

where:

- $\widehat{\text{QPD}}$: predicted questions per day
- $\widehat{\text{AnsRate}}$: predicted answer rate (fraction of questions with accepted answers)
- $\widehat{\text{Retention}}$: predicted retention rate (fraction of month-t active users still active at $t+6$)
- $\widehat{\text{NewUserGrowth}}$: predicted new user growth rate relative to month t

and f_θ is our temporal GNN model parameterized by θ.

C Structure Comparison

For a given community at time t, we evaluate each transformation $T \in \mathcal{T}$:

$$\hat{\mathbf{y}}_{t+6}^{(T)} = f_\theta(\mathcal{S}_t^{(T)})$$

where $\mathcal{S}_t^{(T)}$ is the 12-month sequence with the final graph $\mathcal{G}_t$ replaced by $T(\mathcal{G}_t)$.

We compute a composite health score:

$$H(\hat{\mathbf{y}}_{t+6}^{(T)}) = \mathbf{w}^\top \hat{\mathbf{y}}_{t+6}^{(T)} = w_1 \cdot \widehat{\text{QPD}}^{(T)} + w_2 \cdot \widehat{\text{AnsRate}}^{(T)} + w_3 \cdot \widehat{\text{Retention}}^{(T)} + w_4 \cdot \widehat{\text{NewUserGrowth}}^{(T)}$$

where $\mathbf{w} \in \mathbb{R}^4$ is a weight vector with $\|\mathbf{w}\|_1 = 1$ (we use uniform weights $w_i = 0.25$ unless domain knowledge suggests otherwise).

The recommended intervention is:

$$T^* = \arg\max_{T \in \mathcal{T}} H(\hat{\mathbf{y}}_{t+6}^{(T)})$$