Location-Aware Social Network Recommendation via Temporal Graph Networks

Ziyi Zhang
zyzhang@tamu.edu
Texas A&M University
College Station, Texas, USA

Diya Li
diya.li@tamu.edu
Texas A&M University
College Station, Texas, USA

Zhenlei Song
songzl8_950606@tamu.edu
Texas A&M University
College Station, Texas, USA

Nick Duffield
duffieldng@tamu.edu
Texas A&M University
College Station, Texas, USA

Zhe Zhang
zhezhang@tamu.edu
Texas A&M University
College Station, Texas, USA

ABSTRACT

In the data-driven era, recommendations have become indispensable across various systems. Graphs, as versatile data structures, shine at abstracting complex systems. Many real-world scenarios effortlessly translate into graphs, representing individuals and their relationships as nodes and edges. Link prediction, a cornerstone of recommendations, excels in forecasting future network connections based on current structures. Its applications span diverse domains, including social networks, biological networks, and network security. Previous studies have leveraged classification algorithms like logistic regression and random forest, often complemented by node embedding techniques, yielding impressive results in addressing the challenge of link prediction. Today's dynamic networks continually reshape connections, introducing new links and nodes while removing others. Furthermore, the inclusion of location information associated with nodes provides a new opportunity. Adapting models to this dynamism necessitates capturing spatial and temporal dependencies for sustained effectiveness. In this paper, we undertake a comprehensive evaluation of various algorithms for link prediction. Subsequently, we further enriched the continuous-time dynamic graph networks by incorporating essential location information. This strategic enhancement results in a remarkable performance improvement, highlighting the crucial role of location-based temporal data in improving recommendations. It emphasizes the untapped potential of location and temporal information in refining user recommendations within interconnected networks.

CCS CONCEPTS

• **Information systems → Geographic information systems**; • **Computing methodologies → Neural networks**.

KEYWORDS

link prediction, recommendation system, dynamic graph, geographic information system

1 INTRODUCTION

In an increasingly interconnected world, social networking platforms have revolutionized the way we communicate, share, and connect. These digital ecosystems are vast repositories of human interaction, capturing the essence of our relationships, interests, and preferences. Yet, as the volume of location-based data in these networks grows exponentially, the quest for more intelligent and location-aware recommendation systems becomes paramount.

Graphs are highly versatile data structures that effectively represent numerous real-world applications, where individuals are depicted as nodes and their relationships are denoted as edges [9]. Their ubiquity has led to significant advancements in various fields, including biological networks [1] and network security [6]. In the realm of graph-related research, numerous tasks continue to challenge researchers, with areas like link prediction and community detection standing out prominently. Link prediction involves the prediction of potential future connections based on presently observed links, while community detection focuses on clustering nodes into similar groups within a network. These tasks remain active areas of investigation, contributing to the ongoing evolution of graph-based methodologies.

One especially compelling application of link prediction is in recommendation systems, which encompasses areas such as friend recommendations and item recommendations. While traditional recommendation systems often rely on collaborative filtering methods that utilize a user-item matrix derived from users' past activities, these approaches can encounter challenges related to sparsity and scalability [4]. In contrast, a link prediction-based approach for recommendation systems focuses on operating within a smaller network neighborhood. This approach presents a scalable solution to address the scalability issues encountered by traditional recommendation systems [9].

The evolution of link prediction-based recommendation systems is ongoing. Simultaneously, as sensor technology and data storage

capabilities continue to advance, an increasing volume of spatiotemporal data becomes accessible for analysis [10]. This development opens up a new avenue for researchers to enhance recommendation accuracy by incorporating temporal dependence and geographic location information. In fact, a substantial amount of effort has been dedicated to the development of dynamic graph-based methods specifically tailored to handle temporal information. In the early stages of developing dynamic graph models, the main focus was on discrete-time dynamic graphs. These methods typically involved taking snapshots of the graph at different points in time and then applying static analysis techniques to each snapshot. Recently, there has been a shift towards exploring continuous-time dynamic graphs [8]. This means that instead of just taking snapshots, we consider how the graph changes continuously over time. This development is relatively new and represents a more flexible and detailed way of modeling dynamic graphs.

Continuous-time dynamic graph-based methods like TGNs [8] and GraphSAGE [3] have demonstrated impressive performance by effectively capturing continuous temporal information. In contrast, methods such as Node2Vec [2] and GAE [5] primarily operate on discrete or static graphs, which limits their ability to handle continuous changes in the data. However, the integration of geographic location information, a critical factor in graph analysis [11], has been a noteworthy challenge. These continuous-time dynamic graph-based models excel in capturing temporal dynamics but may not inherently account for location information. This presents an intriguing opportunity to explore how to seamlessly integrate location information into mainstream continuous-time dynamic graph-based models, thereby enhancing their effectiveness and relevance in real-world recommendation system.

In this paper, we introduced an extension of temporal graph networks, which we refer to as Location-Aware Temporal Graph Networks (Location-Aware TGNs). These networks are designed not only to capture continuous temporal information but also to incorporate location information. Through experimentation on a real-world social network, our results highlight the promise of location-based continuous-time dynamic networks in addressing link prediction challenges and enhancing recommendation systems.

2 PROBLEM FORMULATION

Define a dynamic graph as $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$, which evolves over discrete time intervals. In this representation, $\mathcal{V}$ represents nodes, potentially corresponding to entities like users or researchers. Importantly, $\mathcal{V}$ can exhibit dynamic characteristics, including attributes such as location and interests. Meanwhile, $\mathcal{E}_t$ signifies the temporal connections between nodes at time t, with each edge bearing a timestamp. By selectively removing designated edges, denoted as $\mathcal{E}_i \in \mathcal{E}_t$, from $\mathcal{G}_t$, we derive a new connected graph, denominated as $\mathcal{G}_{t-1} = (\mathcal{V}, \mathcal{E}_t - \mathcal{E}_i)$, positioned at time $t - 1$. The primary objective of link prediction-based recommendation is to forecast the probability of $\mathcal{E}_i$ resurfacing in $\mathcal{G}_t$ given the structural information of $\mathcal{G}_{t-1}$.

3 METHODOLOGY

In the context of addressing the location-aware social network recommendation challenge using temporal graph networks, our
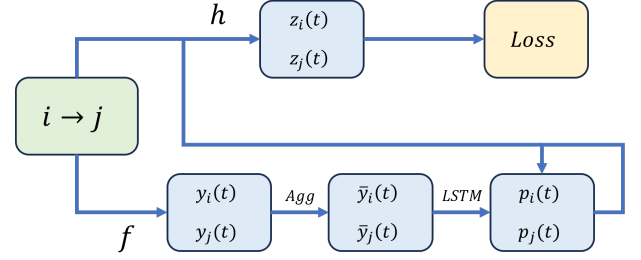

Figure 1: Location-Aware Temporal Graph Networks Workflow.

work takes a comprehensive approach. We begin by modeling the temporal graph as a sequence of n time-stamped events, denoted as $\mathcal{G} = \{\mathcal{S}_{t1}, \mathcal{S}_{t2}, \ldots, \mathcal{S}_{tn}\}$. These events capture the addition or modification of nodes and interactions between pairs of nodes at distinct time points $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$. The two fundamental event types of $\mathcal{S}$ are as follows:

1. **Location-Aware Node-Wise Event:** Represented by loc_i, where i signifies the node index, and the 2-D vector loc represents a location with latitude and longitude coordinates. If node i has not been encountered previously, this event creates it with its current location; otherwise, it updates the location information.

2. **Interaction Event:** Represented by an edge $\mathbf{e}_{ij}(t)$ connecting nodes i and j. In this context, $\mathbf{e}_{ij}(t)$ is treated as a vector that characterizes the interaction between nodes i and j. For instance, it can capture shared research interests between node i and node j in a collaboration social network.

After introducing these event types, we shift our focus to exploring the effective utilization of location information associated with nodes and edge features within a temporal graph network, and the mechanisms by which this information propagates throughout the network. To commence this investigation, let's consider an interaction event that involves nodes i and j. In this context, we define two essential update functions as follows:

$$y_i(t) = f_{ij}(p_i(t^-), p_j(t^-), \Delta t, e_{ij}(t))$$
$$y_j(t) = f_{ji}(p_j(t^-), p_i(t^-), \Delta t, e_{ij}(t))$$

$$(1)$$

Similarly, for location-aware node-wise event loc_i, we update it using another function for the involved node as follows:

$$y_i(t) = f_i(p_i(t^-), t, loc_i(t)) \tag{2}$$

Here, $p_i(t^-)$ signifies the historical information associated with node i, while f represents a trainable neural network. In scenarios involving multiple events, which is a common occurrence in social networks where a node may concurrently update its current location and engage with another node within the same time interval, an information aggregator comes into play. For instance, consider a collaboration social network where a faculty member might secure a new position in a different location and simultaneously interact with new collaborators. In such cases, an information aggregator can be employed:

$$\bar{y}_i(t) = \mathbf{Aggregator}(y_i(t_1), \ldots, y_i(t_b)) \tag{3}$$

In Equation (3), the **Aggregator** can be implemented using one of two strategies. The first strategy involves computing the mean

of the values, while the second strategy entails retaining the most recent value.

Simultaneously, we employ a Long Short-Term Memory (LSTM) mechanism to continuously update $p_i(t)$, as utilized in Equation (1) and Equation (2), in order to preserve and incorporate historical information:

$$p_i(t) = \text{LSTM}(\bar{y}_i(t), p_i(t^-)) \qquad (4)$$

In the final step, to produce the embedding $z_i(t)$ for node i at time t, we apply the subsequent embedding formulation:

$$z_i(t) = \sum_{j \in \mathcal{N}_i^{[0,t]}} h(p_i(t), p_j(t), e_{ij}, loc_i(t), loc_j(t)) \qquad (5)$$

In Equation 5, h represents a neural network, while $\mathcal{N}_i^{[0,t]}$ denotes the neighborhood of node i within the time interval $[0, t]$. The workflow of Location-Aware TGNs is depicted in Figure 1.

4 EXPERIMENTS

4.1 Dataset

Scholars@TAMU (accessible at https://scholars.library.tamu.edu/vivo/) is a valuable platform that helps faculty members and organizations at Texas A&M University (TAMU) showcase their expertise. It gathers information from various sources, including TAMU's systems, public research data (like grants and publications), and authoritative references. This data is then used to create individual profiles that faculty members can edit to accurately represent their skills and knowledge. For our research, we collected 13k samples from **Scholars@TAMU** over several decades.

This dataset contains a variety of features, which are detailed in Table 1. These features serve as the basis for our subsequent analyses and investigations. In the construction of our graph-based network, we establish connections between authors and their coauthors who share the same publication, utilizing the publication's abstract as edge features. Additionally, We include location information as attributes linked to each author, which provide details about the author's educational institution and current workplace, thereby serving as node features. It's important to note that our network is dynamic in nature, and we segment it by publication year, denoted as t, in accordance with the problem statement's requirements. The depicted process outlined in Figure 2.
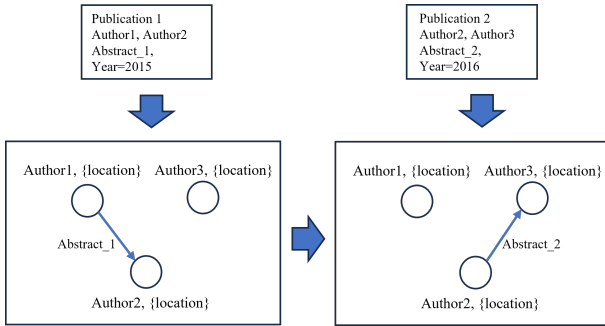

Figure 2: Illustration of Data Processing Diagram: Dynamic Graphs for Year 2015 (Left) and Year 2016 (Right).

4.2 Data Preprocessing

In our data preprocessing pipeline, we carefully process the features to ensure their suitability for subsequent analysis. Specifically, the 'author_id' and 'publication_id' samples are subjected to expansion and uniqueness operations. Meanwhile, the 'publication_year' samples are normalized by scaling them to fall within the range defined by their maximum and minimum values. In the case of 'abstract' features, we employ doc2vec for vectorization, thereby capturing the semantic essence of text. The 'location' features are uniquely managed by representing them as 2-dimensional vectors that incorporate longitude and latitude information.

4.3 Evaluation Metrics

AUC-ROC (Area Under the Curve - Receiver Operating Characteristic) and **AP** (Average Precision) are used to evaluate our link prediction model. AUC-ROC measures overall classification performance, capturing the area under the ROC curve which plots True Positive Rate against False Positive Rate. Higher scores indicate better discrimination ability. AP summarizes precision-recall performance by calculating the weighted mean of precision achieved at each threshold, emphasizing precision at different recall levels.

4.4 Main Results

We firstly conducted Location-Aware Temporal Graph Networks analysis on real-world social network data from Scholars@TAMU. Then we assessed the outcomes using AUC-ROC and AP metrics. The results for both the training and test datasets are presented in Table 2. We observe that Location-Aware TGNs perform effectively on the real-world dataset. However, the absence of exceptional results, as reported in [8], may be attributed to variations in data preprocessing and data quality. Subsequently, we employed Node2Vec [2] as node embedding method, a popular technique in machine learning and network analysis. Node2Vec is designed to learn low-dimensional vector embeddings for nodes in a graph while preserving its structural information. In Node2Vec-related experiments, we configured the hyper-parameters with a walk length of 16 and 50 walks. After generating the node embeddings, we proceeded to apply four distinct classification algorithms: Logistic Regression, Random Forest, XGBoost, and LightGBM. To assess the effectiveness of these classification algorithms in the context of link prediction, we employed the AUC-ROC score as our evaluation metric. Additionally, we selected two other state-of-the-art approaches, GraphSAGE and CTDNE [7], as baseline methods for comparison. The comprehensive results are presented in Table 3.

Based on the aforementioned comparison, it is evident that Location-Aware TGNs outperforms the other methods, demonstrating the best performance. This notable achievement can be attributed to the thoughtful network design, the incorporation of diverse edge features, and the inclusion of location information.

5 DISCUSSION

From the results obtained by applying Location-Aware TGNs to real-world social networks, it becomes evident that Location-Aware TGNs represent an advanced framework. This distinction arises from its ability to differentiate between discrete-time dynamic

Table 1: Description of Collected Features in Scholars@TAMU Dataset

Name	Description
'author_id'	An *author_id* is a distinct and anonymous identifier assigned to each faculty member within the dataset.
'publication_id'	A *publication_id* is a unique and anonymous identifier assigned to each publication within the dataset.
'publication_year'	The *publication_year* denotes the specific year in which an article was published or made publicly accessible.
'abstract'	The *abstract*, linked to the *publication_id*, provides a concise summary of each distinct publication within the dataset.
'location'	The *location* attribute contains information about each author's educational institution and current workplace.

Table 2: Evaluation of Location-Aware TGNs Performance on Scholars@TAMU.

Metric	Value
AUC-ROC	0.8615 ± 0.0125
Average Precision	0.8482 ± 0.0129
AUC-ROC (Test)	0.8725 ± 0.0133
Average Precision (Test)	0.8712 ± 0.0126

Table 3: Comparative Performance of Node2Vec-Based Methods and Location-Aware TGNs.

Methods	AUC-ROC Score
Node2Vec + Logistic Regression	0.5132 ± 0.0210
Node2Vec + Random Forest	0.5671 ± 0.0171
Node2Vec + XGBoost	0.5812 ± 0.0142
Node2Vec + LightGBM	0.6273 ± 0.0194
CTDNE	0.8069 ± 0.0412
GraphSAGE	0.8117 ± 0.0342
Location-Aware TGNs	**0.8725 ± 0.0133**

graphs and continuous-time dynamic graphs. Discrete-time dynamic graphs consist of sequences of static graph snapshots taken at specific time intervals, while continuous-time dynamic graphs are a broader concept, encompassing timed lists of events. These events may involve edge additions or deletions, node insertions, and node-to-edge feature transformations.

Most of the current research and applications predominantly revolve around discrete-time dynamic graphs. However, it is essential to recognize that, in real-world social networks, continuous-time dynamic graphs often align more naturally with the underlying dynamics of the systems being studied. Simultaneously, the incorporation of location information holds paramount importance. One of the key advantages of Location-Aware TGNs is their proficiency in managing dynamic location data, which ensures they are not constrained by the rapid changes in locations. Consequently, the advancement of Location-Aware TGNs carries significant potential for recommendations and can pave the way for broader and more impactful applications, including practical implementations in fields like transportation and urban planning, epidemiology, and environmental monitoring.

6 CONCLUSION

In summary, we have introduced Location-Aware TGNs, an extension of the temporal graph network framework that integrates dynamic location information. Location-Aware TGNs offer several advantages, such as efficient memory utilization, incorporation of dynamic location information and continuous temporal information, and the ability to handle spatio-temporal dependencies within temporal networks. Consequently, the adoption of Location-Aware TGNs resulted in a significant boost, achieving an improvement in AUC-ROC score. The outcomes derived from Location-Aware TGNs not only benefit researchers at TAMU in pinpointing potential collaborators with precision but also establish an advanced framework for tackling recommendation tasks enriched with dynamic location information.

7 ACKNOWLEDGEMENTS

REFERENCES

[1] Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. 2017. Protein interface prediction using graph convolutional networks. *Advances in neural information processing systems* 30 (2017).

[2] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.

[3] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584* (2017).

[4] Zan Huang, Xin Li, and Hsinchun Chen. 2005. Link prediction approach to collaborative filtering. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*. 141–142.

[5] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).

[6] Igor Kotenko and Mikhail Stepashkin. 2006. Attack graph based evaluation of network security. In *Communications and Multimedia Security: 10th IFIP TC-6 TC-11 International Conference, CMS 2006, Heraklion, Crete, Greece, October 19-21, 2006. Proceedings 10*. Springer, 216–227.

[7] Giang Hoang Nguyen, John Boaz Lee, Ryan A Rossi, Nesreen K Ahmed, Eunyee Koh, and Sungchul Kim. 2018. Continuous-time dynamic network embeddings. In *Companion proceedings of the the web conference 2018*. 969–976.

[8] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. 2020. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637* (2020).

[9] Emir Alaattin Yilmaz, Selim Balcisoy, and Burcin Bozkaya. 2023. A link prediction-based recommendation system using transactional data. *Scientific Reports* 13, 1 (2023), 6905.

[10] Ziyi Zhang, Diya Li, Zhe Zhang, and Nicholas Duffield. 2021. A time-series clustering algorithm for analyzing the changes of mobility pattern caused by COVID-19. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop On Animal Movement Ecology And Human Mobility*. 13–17.

[11] Zhe Zhang, Zhangyang Wang, Angela Li, Xinyue Ye, E Lynn Usery, and Diya Li. 2021. An AI-based Spatial Knowledge Graph for Enhancing Spatial Data and Knowledge Search and Discovery. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Searching and Mining Large Collections of Geospatial Data*. 13–17.