Location-Aware Social Network Recommendation via Temporal Graph Networks

Ziyi Zhang, Diya Li, Zhenlei Song, Nick Duffield, and Zhe Zhang

Texas A&M University, College Station, TX

November 13, 2023

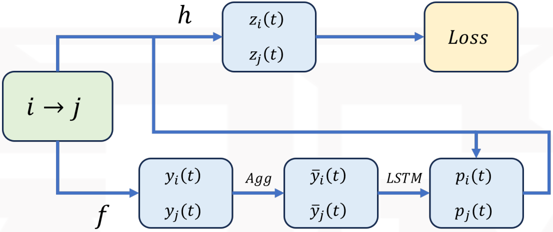
Introduction

1. In the data-driven era, recommendations have become indispensable across various systems.
2. Link prediction, a cornerstone of recommendations, excels in forecasting future network connections based on current structures.
3. Today's dynamic networks continually reshape connections, introducing new links and nodes while removing others.
4. The inclusion of location information associated with nodes provides a new opportunity.
5. Adapting models to this dynamism necessitates capturing spatial and temporal dependencies for sustained effectiveness.

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 $\mathcal{G}_{t-1}$.

Methodology



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.

1. An **interaction event** that involves nodes i and j. In this context, we define two essentials update functions as follows:

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

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

2. A **location-aware node-wise event** loc_i, we update it using:

$$y_i(t) = f_i(p_i(t^-), t, loc_i(t))$$

3. An information **aggregator** can be employed:

$$\overline{y}_i(t) = \textbf{Aggregator}(y_i(t_1), ..., y_i(t_b))$$

4. Updating $p_i(t)$ using a **Long Short-Term Memory** mechanism:

$$p_i(t) = \mathsf{LSTM}(\overline{y}_i(t), p_i(t^-))$$

5. Produce the embedding $z_i(t)$ for node i at time t:

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

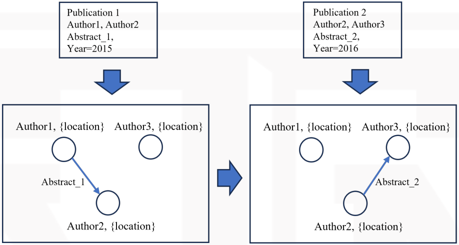
Dataset

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.

Scholars@TAMU 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.

Data Pre-processing



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.

Evaluation Metrics

1. **AUC-ROC** (Area Under the Curve - Receiver Operating Characteristic) and **AP** (Average Precision) are used to evaluate our link prediction model.
2. **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.
3. **AP** summarizes precision-recall performance by calculating the weighted mean of precision achieved at each threshold, emphasizing precision at different recall levels.

Main Reuslts

Evaluation of Location-Aware TGNs Performance on Scholars@TAMU nad its Comparative Performance.

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

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	$\mathbf{0.8725} \pm \mathbf{0.0133}$

Discussion

1. 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.
2. This distinction arises from Location-Aware TGNs' ability to differentiate between discrete-time dynamic graphs and continuous-time dynamic graphs.
3. 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.
4. 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.

Thank You!