TREND: TempoRal Event and Node Dynamics for Graph Representation Learning

Zhihao Wen, Yuan Fang
School of Computing and Information Systems
Singapore Management University
Outline

- Introduction
- Methodology
- Experiment
- Conclusion
Issues of existing works

1. Take discrete snapshots
   - Dynamic network embedding: an extended approach for skip-gram based network embedding. IJCAI-2018.
   - DynGEM: Deep embedding method for dynamic graphs. arxiv-2018.
   - Dynamic network embedding by modeling triadic closure process. AAAI-2018.
   - Evolvegcn: Evolving graph convolutional networks for dynamic graphs. AAAI-2020.

2. Not inductive to new nodes
   - Embedding temporal network via neighborhood formation. KDD-2018.
   - Temporal network embedding with micro-and macro-dynamics. CIKM-2019.
   - Dynamic Heterogeneous Graph Embedding via Heterogeneous Hawkes Process. ECML-PKDD 2021.

3. Not model the exciting effects
   - DyRep: Learning representations over dynamic graphs. ICLR-2019.
   - Inductive representation learning on temporal graphs. ICLR-2020.
   - ......
Problem: temporal graph link prediction

Predict whether a link between $i$ & $j$ at $t$
Temporal point process and Hawkes Process

Point process models discrete sequential events, assuming that historical events can influence the current event.

Hawkes process is desirable for modeling temporal link formation!
For current event is influenced more by recent events, less by previous events.
Outline

- Introduction
- Methodology
- Experiment
- Conclusion
Proposed model: TREND
Hawkes Process on temporal graph

\[ \lambda_{i,j}(t) = \mu_{i,j}(t) + \sum_{(i,j',t') \in \mathcal{H}_i(t)} \gamma_{j'}(t') \kappa(t - t') \]

\[ + \sum_{(i',j,t') \in \mathcal{H}_j(t)} \gamma_{i'}(t') \kappa(t - t') \exp(-\delta(t - t')) \]

amount of excitement

base rate
Hawkes Process on temporal graph

\[ \lambda_{i,j}(t) = \mu_{i,j}(t) + \sum_{(i,j',t') \in \mathcal{H}_i(t)} \gamma_{j'}(t') \kappa(t-t') \]

\[ + \sum_{(i',j,t') \in \mathcal{H}_j(t)} \gamma_{i'}(t') \kappa(t-t') \]

\[ \lambda_{i,j}(t) = f(h^t_i, h^t_j) \]
Temporal GNN layer

\[
\begin{align*}
\mathbf{h}_i^{t,l} &= \sigma \left( \mathbf{h}_i^{t,l-1} \mathbf{W}_\text{self}^l + \sum_{(i,j',t') \in \mathcal{H}_i(t)} \kappa(t-t') \mathbf{h}_{j'}^{t',l-1} \mathbf{W}_\text{hist}^l \tilde{k}_i(t-t') \right) \\
& \quad \text{(self-information (for base intensity))} \\
& \quad \text{(historical neighbors’ information (for excitement by historical events))}
\end{align*}
\]
Modeling event dynamics

\[ \lambda_{i,j}(t) = f(h_i^t, h_j^t) = \text{FCL}_e((h_i^t - h_j^t)^2; \theta_e) \]
Event prior and adaptation

\[
\theta_{e}^{(i,j,t)} = \tau(\theta_{e}, h_{i}^{t} \parallel h_{j}^{t}; \theta_{\tau})
\]

\[
\lambda_{i,j}(t) = FCL_{e}((h_{i}^{t} - h_{j}^{t})^2; \theta_{e}^{(i,j,t)})
\]
Learnable transformation

\[ \theta_e^{(i,j,t)} = \tau(\theta_e, h_i^t \| h_j^t ; \theta_\tau) = (\alpha^{(i,j,t)} + 1) \odot \theta_e + \beta^{(i,j,t)} \]

\[ \alpha^{(i,j,t)} = \sigma((h_i^t \| h_j^t)W_\alpha + b_\alpha) \quad \beta^{(i,j,t)} = \sigma((h_i^t \| h_j^t)W_\beta + b_\beta) \]
Event loss

\[ L_e(i, j, t) = -\log(\lambda_{i,j}(t)) - Q \cdot \mathbb{E}_{k \sim P_n} \log(1 - \lambda_{i,k}(t)) \]
Modeling node dynamics

Predicted number of new events occurring on the node at $t$

$$\Delta \hat{N}_i(t) = FCL_n(h^t_i; \theta_n)$$
Overall loss

\[
\arg\min_\Theta \sum_{(i,j,t) \in \mathcal{I}^{tr}} L_e + \eta_1 L_n + \eta_2 (\|\alpha^{(i,j,t)}\|_2^2 + \|\beta^{(i,j,t)}\|_2^2)
\]

\((\theta_g, \theta_e, \theta_\tau, \theta_n)\)
Outline

- Introduction
- Methodology
- Experiment
- Conclusion
# Statistics of datasets

| Dataset         | CollegeMsg | cit-HepTh | Wikipedia | Taobao   |
|-----------------|------------|-----------|-----------|----------|
| # Events        | 59,835     | 51,315    | 157,474   | 4,294,000|
| # Nodes         | 1,899      | 7,577     | 8,227     | 1,818,851|
| # Node features |            |           |           |          |
| Multi-edge?     | Yes        | No        | Yes       | Yes      |
| New nodes in testing | 22.79% | 100%      | 7.26%     | 23.46%   |

[Image 1](https://www.shutterstock.com/th/video/clip-21752794-message-network-icon-link-connection-technology-loop)

[Image 2](https://www.researchgate.net/publication/297894915)

[Image 3](https://www.dreamstime.com/concept-e-commerce-shopping-web-icons-line-style-mobile-shop-digital-marketing-bank-card-gifts-digital-concept-e-commerce-image159818445)
Performance comparison with baselines

In each column, the best result is **bolded** and the runner-up is underlined. Improvement by TREND is calculated relative to the best baseline. "-" indicates no result obtained due to out of memory issue; * indicates that our model significantly outperforms the best baseline based on two-tail \( t \)-test (\( p < 0.05 \)).

|                | CollegeMsg |         |         |         |         |         |         |
|----------------|------------|---------|---------|---------|---------|---------|---------|
|                | Accuracy   | F1      | Accuracy| F1      | Accuracy| F1      | Accuracy|
| DeepWalk       | 66.54±5.36 | 67.86±5.86 | 51.55±0.90 | 50.39±0.98 | 65.12±0.94 | 64.25±1.32 | 53.59±0.18 | 56.67±0.12 |
| Node2vec       | 65.82±4.12 | 69.10±3.50 | 65.68±1.90 | 66.13±2.15 | 75.52±0.58 | 75.61±0.52 | 52.74±0.33 | 54.86±0.32 |
| VGAE           | 65.82±5.68 | 68.73±4.49 | 66.79±2.58 | 67.27±2.84 | 66.35±1.48 | 68.04±1.18 | 55.97±0.22 | 59.80±0.16 |
| GAE            | 62.54±5.11 | 66.97±3.22 | 69.52±1.10 | 70.28±1.33 | 68.70±1.34 | 69.74±1.43 | 58.13±0.15 | 61.40±0.07 |
| GraphSAGE      | 58.91±3.67 | 60.45±4.22 | 70.72±1.96 | 71.27±2.41 | 72.32±1.25 | 73.39±1.25 | 60.74±0.18 | 61.61±0.20 |
| CTDNE          | 62.55±3.67 | 65.56±2.34 | 49.42±1.86 | 44.23±3.92 | 60.99±1.26 | 62.71±1.49 | 51.64±0.32 | 43.99±0.38 |
| EvolveGCN      | 63.27±4.42 | 65.44±4.72 | 61.57±1.53 | 62.42±1.54 | 71.20±0.88 | 73.43±0.51 | -         | -         |
| GraphSAGE+T    | 69.09±4.91 | 69.41±5.45 | 67.80±1.27 | 69.12±1.12 | 57.93±0.53 | 63.41±0.91 | 67.05±0.23 | 67.69±0.17 |
| TGAT           | 58.18±4.78 | 57.23±7.57 | 78.02±1.93 | 78.52±1.61 | 76.45±0.91 | 76.99±1.16 | 70.07±0.59 | 71.31±0.18 |
| HTNE           | 73.82±5.36 | 74.24±5.36 | 66.70±1.80 | 67.47±1.16 | 77.88±1.56 | 78.09±1.40 | 59.03±0.17 | 60.34±0.19 |
| MMDNE          | 73.82±5.36 | 74.10±3.70 | 66.28±3.87 | 66.70±3.39 | 79.76±0.89 | 79.87±0.95 | 58.24±0.10 | 59.04±0.16 |
| TREND          | 74.55±1.95 | 75.64±2.09 | **80.37**±2.08 | **81.13**±1.92 | **83.75**±1.19 | **83.86**±1.24 | **78.56**±0.17 | **80.67**±0.15 |
| (improv.)      | (+0.99%)   | (+1.89%) | (+3.01%) | (+3.32%) | (+5.00%) | (+4.99%) | (+12.11%) | (+13.12%) |
Ablation study
Outline

- Introduction
- Methodology
- Experiment
- Conclusion
Conclusion

• Conclusion:
  
  • Studied the problem of temporal graph representation learning and temporal link prediction.
  
  • Proposed TREND, a novel framework driven by event and node dynamics on a Hawkes process-based GNN.
  
  • Conduct extensive experiments on four real-world graph datasets and demonstrated the superior performance of TREND.
THANK YOU FOR YOUR ATTENTION

Paper, code, data...  
www.yfang.site

We are also hiring! Multiple positions open.  
www.yfang.site/hiring