TRENDS: 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 by modeling triadic closure process. AAAI-2018.

2. Not inductive to new nodes
   Dynamic Heterogeneous Graph Embedding via Heterogeneous Hawkes Process. ECML-PKDD 2021.

3. Not model the exciting effects
   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

\[
\lambda_{i,j}(t) = \mu_{i,j}(t) + \sum_{(i,j',t') \in \mathcal{H}_i(t)} \gamma_{j'}(t') \kappa(t - t') + \sum_{(i',j,t') \in \mathcal{H}_j(t)} \gamma_{i'}(t') \kappa(t - t') \exp(-\delta(t - t'))
\]
Hawkes Process on temporal graph

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

\[ \lambda_{i,j}(t) = f(h_i^t, h_j^t) \]
Temporal GNN layer

\[ h_{i}^{t,l} = \sigma \left( h_{i}^{t,l-1} W_{self}^{l} + \sum_{(i',j',t') \in \mathcal{H}_i(t)} h_{j'}^{t',l-1} W_{hist}^{l} \hat{k}_i(t-t') \right) \]

- \( h_{i}^{t,l} \) represents the node representation at time \( t \) and layer \( l \).
- \( \sigma \) is the activation function.
- \( W_{self}^{l} \) and \( W_{hist}^{l} \) are the weights for self- and historical neighbors' information.
- \( \mathcal{H}_i(t) \) is the set of historical neighbors of node \( i \) at time \( t \).
- \( \hat{k}_i(t-t') \) is a function that assigns weights to historical neighbors based on the time difference.

The diagram illustrates the input graph and the node representations at time \( t \).
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

\[ \theta_{e(i,j,t)} = \tau(\theta_e, h_i^t || h_j^t; \theta_{\tau}) \]

\[ \lambda_{i,j}(t) = FCL_e((h_i^t - h_j^t)^2; \theta_{e(i,j,t)}) \]

Event prior

Event condition

Event input

Event-specific transfer

Scaling

Shifting
Learnable transformation

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

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

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

- a vector of ones
- element-wise multiplication
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 $t$

$$\Delta \hat{N}_i(t) = \text{FCL}_n(h^t_i; \theta_n)$$

$h^t_i$  $\theta_n$  $\Delta \hat{N}_i(t)$  Node loss $L_n$
Overall loss

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

where \((\theta_g, \theta_e, \theta, \theta_n)\) are the parameters to be optimized.
Outline

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CollegeMsg</th>
<th>cit-HepTh</th>
<th>Wikipedia</th>
<th>Taobao</th>
</tr>
</thead>
<tbody>
<tr>
<td># Events</td>
<td>59,835</td>
<td>51,315</td>
<td>157,474</td>
<td>4,294,000</td>
</tr>
<tr>
<td># Nodes</td>
<td>1,899</td>
<td>7,577</td>
<td>8,227</td>
<td>1,818,851</td>
</tr>
<tr>
<td># Node features</td>
<td>–</td>
<td>128</td>
<td>172</td>
<td>128</td>
</tr>
<tr>
<td>Multi-edge?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>New nodes in testing</td>
<td>22.79%</td>
<td>100%</td>
<td>7.26%</td>
<td>23.46%</td>
</tr>
</tbody>
</table>

[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)

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 ($p < 0.05$).

<table>
<thead>
<tr>
<th></th>
<th>CollegeMsg</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Taobao</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>F1</td>
<td>Accuracy</td>
<td>F1</td>
<td>Accuracy</td>
<td>F1</td>
</tr>
<tr>
<td>DeepWalk</td>
<td>66.54±5.36</td>
<td>67.86±5.86</td>
<td>51.55±0.90</td>
<td>50.39±0.98</td>
<td>65.12±0.94</td>
<td>64.25±1.32</td>
</tr>
<tr>
<td>Node2vec</td>
<td>65.82±4.12</td>
<td>69.10±3.50</td>
<td>65.68±1.90</td>
<td>66.13±2.15</td>
<td>75.52±0.58</td>
<td>75.61±0.52</td>
</tr>
<tr>
<td>VGEA</td>
<td>65.82±5.68</td>
<td>68.73±4.49</td>
<td>66.79±2.58</td>
<td>67.27±2.84</td>
<td>66.35±1.48</td>
<td>68.04±1.18</td>
</tr>
<tr>
<td>GAE</td>
<td>62.54±5.11</td>
<td>66.97±3.22</td>
<td>69.52±1.10</td>
<td>70.28±1.33</td>
<td>68.70±1.34</td>
<td>69.74±1.43</td>
</tr>
<tr>
<td>GraphSAGE</td>
<td>58.91±3.67</td>
<td>60.45±4.22</td>
<td>70.72±1.96</td>
<td>71.27±2.41</td>
<td>72.32±1.25</td>
<td>73.39±1.25</td>
</tr>
<tr>
<td>CTDNE</td>
<td>62.55±3.67</td>
<td>65.56±2.34</td>
<td>49.42±1.86</td>
<td>44.23±3.92</td>
<td>60.99±1.26</td>
<td>62.71±1.49</td>
</tr>
<tr>
<td>EvolveGCN</td>
<td>63.27±4.42</td>
<td>65.44±4.72</td>
<td>61.57±1.53</td>
<td>62.42±1.54</td>
<td>71.20±0.88</td>
<td>73.43±0.51</td>
</tr>
<tr>
<td>GraphSAGE+T</td>
<td>69.09±4.91</td>
<td>69.41±5.45</td>
<td>67.80±1.27</td>
<td>69.12±1.12</td>
<td>57.93±0.53</td>
<td>63.41±0.91</td>
</tr>
<tr>
<td>TGAT</td>
<td>58.18±4.78</td>
<td>57.23±7.57</td>
<td>78.02±1.93</td>
<td>78.52±1.61</td>
<td>76.45±0.91</td>
<td>76.99±1.16</td>
</tr>
<tr>
<td>HTNE</td>
<td>73.82±5.36</td>
<td>74.24±5.36</td>
<td>66.70±1.80</td>
<td>67.47±1.16</td>
<td>77.88±1.56</td>
<td>78.09±1.40</td>
</tr>
<tr>
<td>MMDNE</td>
<td>73.82±5.36</td>
<td>74.10±3.70</td>
<td>66.28±3.87</td>
<td>66.70±3.39</td>
<td>79.76±0.89</td>
<td>79.87±0.95</td>
</tr>
<tr>
<td>TREND (improv.)</td>
<td>74.55±1.95</td>
<td>75.64±2.09</td>
<td><strong>80.37</strong>±2.08</td>
<td><strong>81.13</strong>±1.92</td>
<td><strong>83.75</strong>±1.19</td>
<td><strong>83.86</strong>±1.24</td>
</tr>
<tr>
<td></td>
<td>(+0.99%)</td>
<td>(+1.89%)</td>
<td>(+3.01%)</td>
<td>(+3.32%)</td>
<td>(+5.00%)</td>
<td>(+4.99%)</td>
</tr>
</tbody>
</table>
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