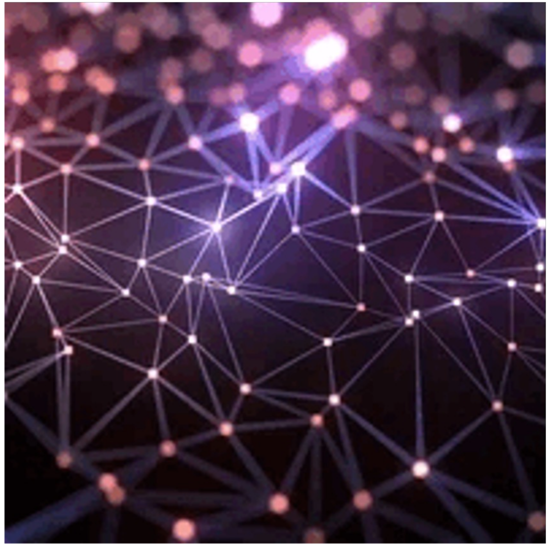
Exploring Time Granularity on Temporal Graphs for Dynamic Link Prediction in Real-world Networks

Xiangjian Jiang (xj265@cam.ac.uk)
Yanyi Pu (yp17@sheffield.ac.uk)

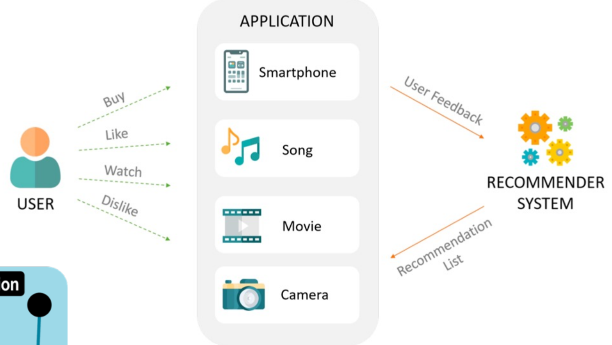
UNIVERSITY OF CAMBRIDGE
Department of Computer Science and Technology

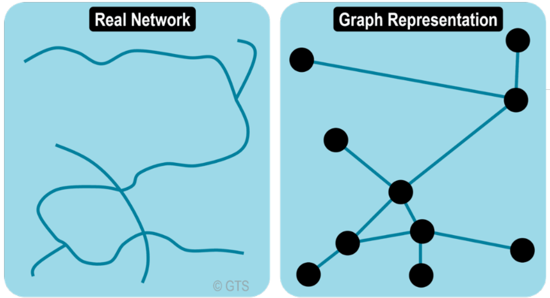
TGL Workshop @ NuralPS 2023

Dynamic Networks

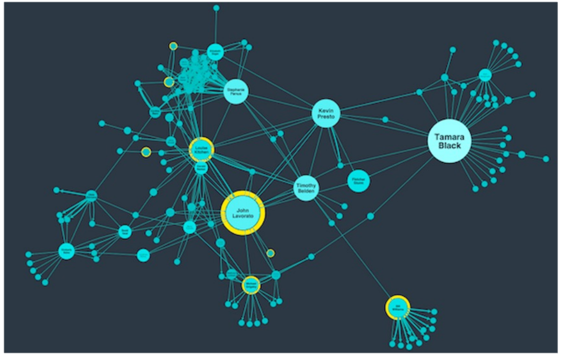


Recommendation systems





Transportations



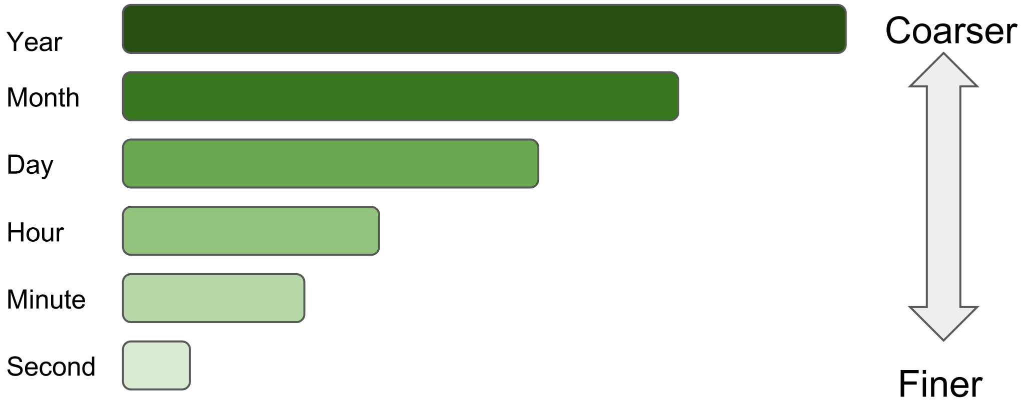
Social Networks

- Encode evolving connections & relationships in real-world scenarios
- Continuous (event-driven) vs Discrete (snapshots over fixed time intervals)

Time Granularity for Dynamic Graphs

Time Granularity:

- Time intervals at which dynamic graphs are observed or analyzed
- determine the level of temporal detail retained on the graph



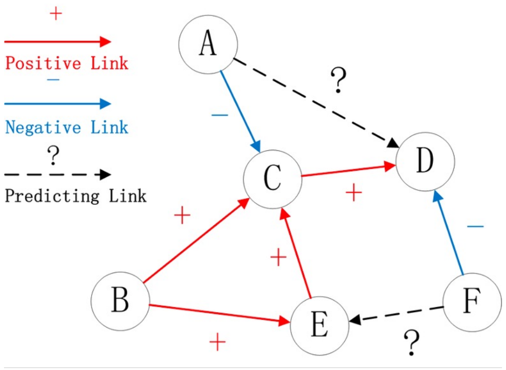
Importance for Graph Analysis:

- **Model Performance**
  - Coarse → lose information
  - Fine → introduce noise
- **Model Robustness**
  - Generalization
  - Sensitivity
- **Computational Efficiency**
  - # of training instances
  - Valuable insights
- **Transferability**
  - Across Domains
  - Across Tasks

Overview

- An early attempt to investigate the effect of Time Granularity on model performance
- No Prior Work
- Extension of (Poursafaei et al., 2022)[1] work

1. **One** Task → Dynamic Link Prediction
2. **Two** Types of Dynamic Graphs → Social & Interaction
3. **Three** Negative Sampling Strategies → Random, Historical, Inductive
4. **Four** Time Granularities: Second, Minute, Hour, Day
5. **Five** Models: EdgeBank_{tw}, EdgeBank_{∞}, JODIE, DyRep, TGN



Dynamic Link Prediction

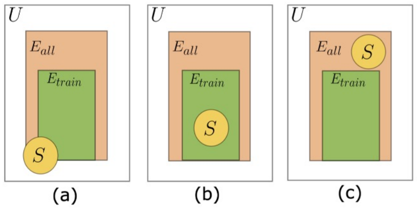


Figure 1: Negative edge sampling strategies during evaluation for dynamic link prediction; (a) random sampling, (b) historical sampling, (c) inductive sampling [21].

Negative Sampling Strategies

1. Poursafaei, Farimah, Shenyang Huang, Kellin Pelrine, and Reihaneh Rabbany. 'Towards Better Evaluation for Dynamic Link Prediction'. arXiv, 11 September 2022. https://doi.org/10.48550/arXiv.2207.10128.

Datasets

Table 1: Datasets Statistics with Associated Semantic Meanings

Dataset	Domain	Node	# of Nodes	Edge	Total Edges	Unique Edges	Unique Steps	Duration
Wikipedia [19]	Social	Editors & Wiki Pages	9,227	Editing Request	157,474	18,257	152,757	1 Month
Reddit [19]	Social	Uers & Posts	10,984	Posting Request	672,447	78,516	588,918	1 Month
MOOC [19]	Interaction	Students & Online Courses	7,144	Accessing a online course	411,749	178,443	345,600	1 Month
LastFM [19]	Interaction	Users & Songs	1,980	Listening a song	1,293,103	154,993	1,283,614	4 Years
Enron [39]	Social	Employees	184	Email communication	125,235	3,125	22,632	3 Years
Social Evo. [40]	Proximity	Students	74	Cellphone calls	2,099,519	4,486	565,932	1 Year
UCI [41]	Social	Students	1,899	Online Chats	59,835	20,296	58,911	196 Days

- Ubiquitously used datasets for Dynamic Graph Neural Networks (DGNNs)
- Directed edges lists recorded by Unix timestamp
- No Node/Edge Features

Dataset Split

Table 2: Datasets split by Day with split rate of 2/3-1/6-1/6 for training, validation, and testing.

	Train		**Validation**		**Test**		**Total**	
Dataset	# of Days	# of Edges	# of Days	# of Edges	# of Days	# of Edges	# of Days	# of Edges
Wikipedia	20	99,701	5	26,697	5	26,359	30	152,767
Reddit	20	432,543	5	110,004	5	126,518	30	669,075
MOOC	20	216,364	5	65,815	5	63,421	30	345,610
LastFM	1,216	916,312	304	340,736	305	26,566	1,825	1,284,223
Enron	730	6,224	182	6,357	183	10,051	1,095	22,997
Social Evo.	160	268,758	40	136,849	40	160,325	240	566,012
UCI	130	55,202	32	2,402	34	1,307	196	58,977

- Prevents data leakage
- Remain same semantic meanings
- Promote fair cross-granularity comparisons

Benchmarks & DGNNs

- **EdgeBank$_{tw}$**:remembers edges from the short-term past
- **EdgeBank$_{\infty}$**: stores all observed edges in memory
- **JODIE:** a coupled recurrent neural network model
- **DyRep:** learn representations by capturing both topological and temporal dependencies
- **TGN** (Temporal Graph Networks): a generic, scalable and efficient framework to model dynamic graphs

JODIE[1]

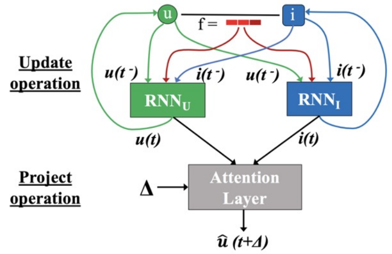


Figure 2: The *JODIE* model: After an interaction (u, i, t, f) between user u and item i, the dynamic embeddings of u and i are updated in the *update operation* with RNN_U and RNN_I, respectively. The *projection operation* predicts the user embedding at a future time $t + \Delta$.

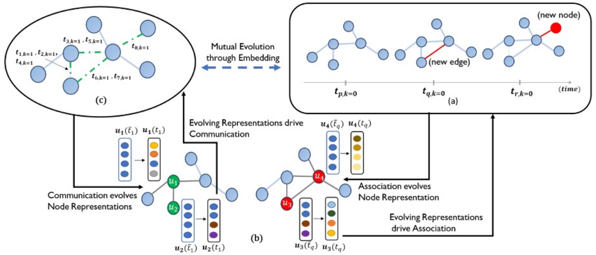


Figure 1: Evolution Through Mediation. **(a)** Association events (k=0) where the node or edge grows. **(c)** Communication Events (k=1) where nodes interact with each other. For both these processes, $t_{p,k=0} < (t_1, t_2, t_3, t_4, t_5)_{k=1} < t_{q,k=0} < (t_6, t_7)_{k=1} < t_{r,k=0}$. **(b)** Evolving Representations.

DyRep[2]

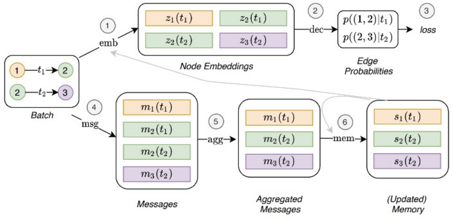


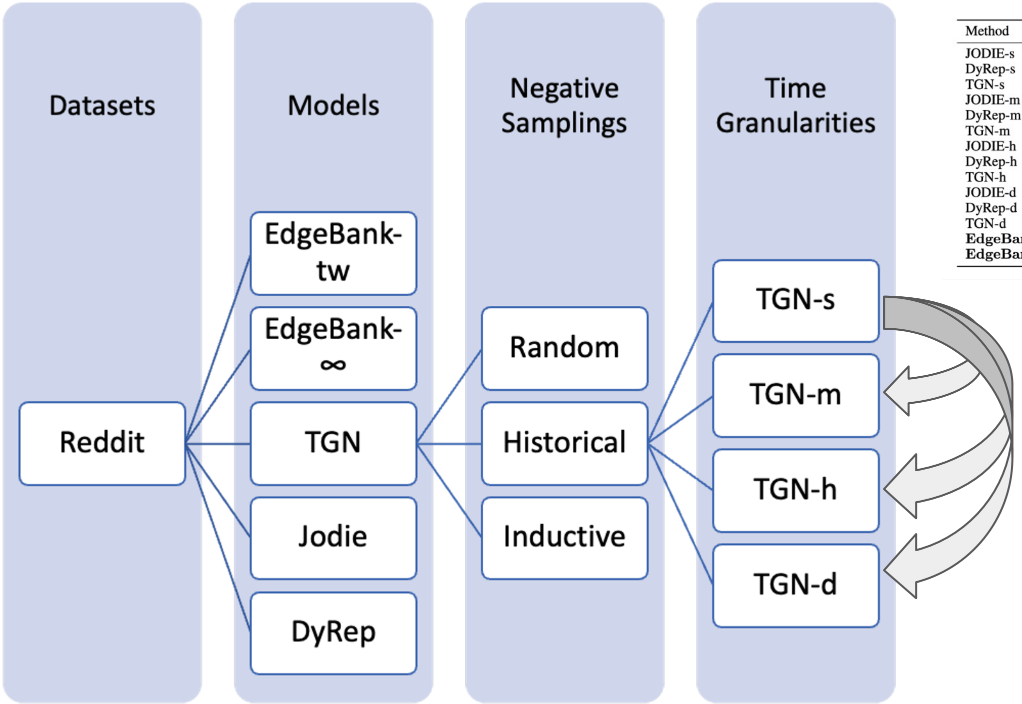
Figure 1: Computations performed by TGN on a batch of time-stamped interactions. *Top:* embeddings are produced by the embedding module using the temporal graph and the node's memory (1). The embeddings are then used to predict the batch interactions and compute the loss (2, 3). *Bottom:* these same interactions are used to update the memory (4, 5, 6). This is a simplified flow of operations which would prevent the training of all the modules in the bottom as they would not receiving a gradient. Section 3.2 explains how to change the flow of operations to solve this problem and figure 2 shows the complete diagram.

TGN[3]

1. Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pages 1269–1278, 2019. 4, 5
2. Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. Dyrep: Learning representations over dynamic graphs. In International Conference on Learning Representations, 2019. URL https://openreview.net/forum?id=HyePrhR5KX. 4, 5
3. Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. arXiv preprint arXiv:2006.10637, 2020. 2, 3, 4, 5

Experimental Design

7 Datasets X **5** Methods X **3** Random Seeds X **4** Time Granularities = **420** Models



(a) Random Sampling

Method	Wikipedia	Reddit	MOOC	LastFM	Enron	Social Evo.	UCI	Rank
JODIE-s	0.992 (0.001)	0.999 (0.000)	0.857 (0.015)	0.999 (0.000)	0.930 (0.005)	0.996 (0.000)	0.994 (0.001)	3
DyRep-s	0.681 (0.010)	0.566 (0.000)	0.606 (0.005)	0.505 (0.009)	0.643 (0.011)	0.679 (0.053)	0.833 (0.001)	12
TGN-s	0.979 (0.001)	0.961 (0.000)	0.789 (0.012)	0.694 (0.013)	0.808 (0.019)	0.923 (0.001)	0.901 (0.006)	6
JODIE-m	0.993 (0.000)	0.999 (0.000)	0.865 (0.007)	0.999 (0.000)	0.926 (0.002)	0.995 (0.001)	0.995 (0.000)	1
DyRep-m	0.677 (0.011)	0.561 (0.000)	0.588 (0.010)	0.495 (0.005)	0.634 (0.019)	0.696 (0.015)	0.837 (0.008)	13
TGN-m	0.978 (0.001)	0.955 (0.000)	0.805 (0.029)	0.695 (0.016)	0.807 (0.020)	0.869 (0.007)	0.903 (0.039)	5
JODIE-h	0.992 (0.001)	0.996 (0.000)	0.845 (0.018)	0.997 (0.000)	0.923 (0.017)	0.914 (0.000)	0.995 (0.001)	2
DyRep-h	0.640 (0.011)	0.496 (0.000)	0.590 (0.012)	0.503 (0.000)	0.593 (0.057)	0.701 (0.000)	0.829 (0.016)	14
TGN-h	0.960 (0.001)	0.928 (0.000)	0.688 (0.000)	0.684 (0.000)	0.764 (0.027)	0.562 (0.000)	0.813 (0.023)	8
JODIE-d	0.976 (0.005)	0.886 (0.000)	0.623 (0.000)	0.928 (0.000)	0.920 (0.004)	0.633 (0.000)	0.994 (0.001)	4
DyRep-d	0.582 (0.032)	0.462 (0.000)	0.591 (0.000)	0.505 (0.000)	0.649 (0.006)	0.513 (0.000)	0.835 (0.009)	11
TGN-d	0.944 (0.005)	0.923 (0.000)	0.570 (0.000)	0.602 (0.000)	0.795 (0.016)	0.595 (0.000)	0.836 (0.024)	9
EdgeBank$_{tw}$	0.888 (0.000)	0.924 (0.000)	0.607 (0.000)	0.840 (0.000)	0.867 (0.000)	0.600 (0.000)	0.733 (0.000)	7
EdgeBank$_{\infty}$	0.911 (0.000)	0.954 (0.000)	0.548 (0.000)	0.827 (0.000)	0.858 (0.000)	0.538 (0.000)	0.749 (0.000)	10

AU-ROC Scores

Standard Deviation

Evaluation Metrics
- AU-ROC
- Average Precision (AP)

Experimental Results

4 granularities

2 evaluation metrics

Table 6: AU-ROC of dynamic link prediction on the "second" granularity data across three negative sampling strategies. Note that we report the mean AU-ROC over three runs with the standard deviations in parenthesis, and the rank is computed by averaging the ranks over all datasets.

(a) Random Sampling

3 negative sampling strategies

Method	Wikipedia	Reddit	MOOC	LastFM	Enron	Social Evo.	UCI	Rank
JODIE-s	0.992 (0.001)	0.999 (0.000)	0.857 (0.015)	0.999 (0.000)	0.930 (0.005)	0.996 (0.000)	0.994 (0.001)	3
DyRep-s	0.681 (0.010)	0.566 (0.000)	0.606 (0.005)	0.505 (0.009)	0.643 (0.011)	0.679 (0.053)	0.833 (0.001)	12
TGN-s	0.979 (0.001)	0.961 (0.000)	0.789 (0.012)	0.694 (0.013)	0.808 (0.019)	0.923 (0.001)	0.901 (0.006)	6
JODIE-m	0.993 (0.000)	0.999 (0.000)	0.865 (0.007)	0.999 (0.000)	0.926 (0.002)	0.995 (0.001)	0.995 (0.000)	1
DyRep-m	0.677 (0.011)	0.561 (0.000)	0.588 (0.010)	0.495 (0.005)	0.634 (0.019)	0.696 (0.015)	0.837 (0.008)	13
TGN-m	0.978 (0.001)	0.955 (0.000)	0.805 (0.029)	0.695 (0.016)	0.807 (0.020)	0.869 (0.007)	0.903 (0.039)	5
JODIE-h	0.992 (0.001)	0.996 (0.000)	0.845 (0.018)	0.997 (0.000)	0.923 (0.017)	0.914 (0.000)	0.995 (0.001)	2
DyRep-h	0.640 (0.011)	0.496 (0.000)	0.590 (0.012)	0.503 (0.000)	0.593 (0.057)	0.701 (0.000)	0.829 (0.016)	14
TGN-h	0.960 (0.001)	0.928 (0.000)	0.688 (0.000)	0.684 (0.000)	0.764 (0.027)	0.562 (0.000)	0.813 (0.023)	8
JODIE-d	0.976 (0.005)	0.886 (0.000)	0.623 (0.000)	0.928 (0.000)	0.920 (0.004)	0.633 (0.000)	0.994 (0.001)	4
DyRep-d	0.582 (0.032)	0.462 (0.000)	0.591 (0.000)	0.505 (0.000)	0.649 (0.006)	0.513 (0.000)	0.835 (0.009)	11
TGN-d	0.944 (0.005)	0.923 (0.000)	0.570 (0.000)	0.602 (0.000)	0.795 (0.016)	0.595 (0.000)	0.836 (0.024)	9
EdgeBank$_{tw}$	0.888 (0.000)	0.924 (0.000)	0.607 (0.000)	0.840 (0.000)	0.867 (0.000)	0.600 (0.000)	0.733 (0.000)	7
EdgeBank$_{\infty}$	0.911 (0.000)	0.954 (0.000)	0.548 (0.000)	0.827 (0.000)	0.858 (0.000)	0.538 (0.000)	0.749 (0.000)	10

- 24 tables in total ➡️ 2 tables of ranking

More results are available in the main manuscript.

Experimental Results

Table 2: Average rank of AU-ROC on dynamic link prediction for different time granularities over three negative sampling strategies. Note that the top three methods are coloured by **First**, **Second** and **Third** respectively. Note that the absolute difference between any two given methods can be determined by calculating the difference in their numerical scores in Appendix B.

Granularity	**Second**			**Minute**			**Hour**			**Day**		
NS	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu
JODIE-s	**3**	11	9	**2**	11	9	**3**	11	10	4	12	11
DyRep-s	12	7	6	11	7	6	14	7	6	14	7	5
TGN-s	6	**2**	**1**	5	**1**	**1**	6	5	**3**	9	5	4
JODIE-m	**1**	12	14	**1**	12	12	**2**	13	14	**1**	13	12
DyRep-m	13	9	8	12	8	7	13	8	7	12	9	8
TGN-m	5	**1**	**2**	4	**2**	**2**	7	5	4	7	4	**3**
JODIE-h	**2**	14	11	**3**	14	14	**1**	14	13	**2**	14	14
DyRep-h	14	8	7	13	6	5	11	6	5	13	8	6
TGN-h	8	5	**3**	7	5	**3**	5	**1**	**1**	6	**3**	**2**
JODIE-d	4	13	12	6	13	13	4	12	11	**3**	11	13
DyRep-d	11	6	5	14	9	8	12	9	8	11	6	7
TGN-d	9	4	4	10	4	4	8	**2**	**2**	5	**1**	**1**
EdgeBank$_{tw}$	7	**3**	13	8	**3**	11	9	**3**	12	8	**2**	10
EdgeBank$_{\infty}$	10	10	10	9	10	10	10	10	9	10	10	9

- 24 tables in total ➡️ 2 tables of ranking

More results are available in the main manuscript.

Experimental Results

- Intuitions
  - On **fine** granularity test data, **fine models >> coarse models**.
  - On **coarse** granularity test data, **fine models ≥ coarse models**.
- Results
  - **Finer granularity ≠ Better Performance under Random Negative Sampling**
    - On **second** granularity, JODIE-m/h outperform JODIE-s.
    - On **minute** granularity, JODIE-m outperforms JODIE-s.
    - On **hour** granularity, JODIE-h outperforms JODIE-s/m.

Granularity	Second			Minute			Hour			Day		
NS	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu
JODIE-s	3	11	9	2	11	9	3	11	10	4	12	11
JODIE-m	1	12	14	1	12	12	2	13	14	1	13	12
JODIE-h	2	14	11	3	14	14	1	14	13	2	14	14
JODIE-d	4	13	12	6	13	13	4	12	11	3	11	13

More results are available in the main manuscript.

Experimental Results

- Intuitions
  - On **fine** granularity test data, **fine models >> coarse models**.
  - On **coarse** granularity test data, **fine models ≥ coarse models**.
- Results
  - **Finer granularity ≠ Better Performance under alternative negative samplings**
    - On **second** granularity, TGN-m outperforms TGN-s.
    - On **hour** granularity, TGN-h/d outperforms TGN-s/m.
    - On **day** granularity, TGN-d outperforms TGN-s/m/h.

Granularity	Second			Minute			Hour			Day		
NS	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu
TGN-s	6	2	1	5	1	1	6	5	3	9	5	4
TGN-m	5	1	2	4	2	2	7	5	4	7	4	3
TGN-h	8	5	3	7	5	3	5	1	1	6	3	2
TGN-d	9	4	4	10	4	4	8	2	2	5	1	1

More results are available in the main manuscript.

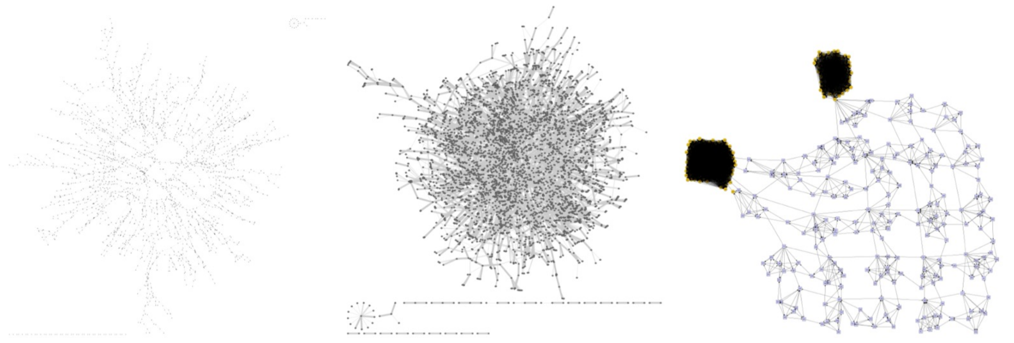
Experimental Results

- Intuitions
  - On **fine** granularity test data, **fine models >> coarse models**.
  - On **coarse** granularity test data, **fine models ≥ coarse models**.
- Results
  - **Long-term dependency is important for dynamic link prediction in real-world networks**
    - JODIE has a significant decline in performance under challenging negative sampling.
    - DyRep consistently achieve normal performance
    - TGN stably achieves competitive performance across all datasets, granularities and negative sampling strategies.

Granularity	**Second**			**Minute**			**Hour**			**Day**		
NS	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu	Rand	Hist	Indu
JODIE-s	3	11	9	2	11	9	3	11	10	4	12	11
DyRep-s	12	7	6	11	7	6	14	7	6	14	7	5
TGN-s	6	2	1	5	1	1	6	5	3	9	5	4
EdgeBank$_{tw}$	7	3	13	8	3	11	9	3	12	8	2	10
EdgeBank$_{\infty}$	10	10	10	9	10	10	10	10	9	10	10	9

More results are available in the main manuscript.

Existing Benchmark Limitations



(a) (b) (c)

Figure 2: An example of "hairball" graph due to repetitive edge additions and aggregation. (a) Original Wikipedia graph used in our experiment (no edge repetition); (b) The "hairball" visualisation of the Wikipedia graph under our edge aggregation method; (c) A synthetic example of a globally sparse but locally dense graph, containing multiple "black holes". (a) and (b) are visualised using the Backbone layout [44] in Visone [45] without edge sparsification. The width of the edge indicates the number of communications between two designated edges. (c) is visualised using the Organic layout [46] in yEd [47].

- Transductive Edges: **no** edge deletion included
- Transductive Nodes: **no** node addition/deletion included
- "Hairballs" and "Black Holes": **globally sparse but locally dense** graph

More discussions are available in the main manuscript.

Discussion

- Takeaways
  - We introduce **a novel data-splitting approach** that allows fair comparison across different time granularities **without data leakage issues**.
  - We empirically investigate the effect of time granularity on dynamic link prediction task, and the results suggest that:
    - **Finer granularity does not guarantee better performance due to potential noise.**
    - **Long-term dependency is significant for link prediction in real-world scenarios.**
  - We provide an insightful discussion on the **inherent limitations of existing benchmarks** from the perspective of data properties.
- Future Work
  - **Inductive Link Prediction:** explore the dynamic graphs where the **node addition/deletion** happens across test/valid/test set
  - **Learnable Time Granularity:** design models that can learn temporal information from **different time granularities inherently** without manual specifications
  - **Novel Timestamp Aggregation:** aggregate **both links and timestamps**, which might cause the fundamental change to the graph properties, e.g. multigraph

THANK YOU

UNIVERSITY OF CAMBRIDGE
Department of Computer Science and Technology

Xiangjian Jiang (xj265@cam.ac.uk)
Yanyi Pu (yp17@sheffield.ac.uk)