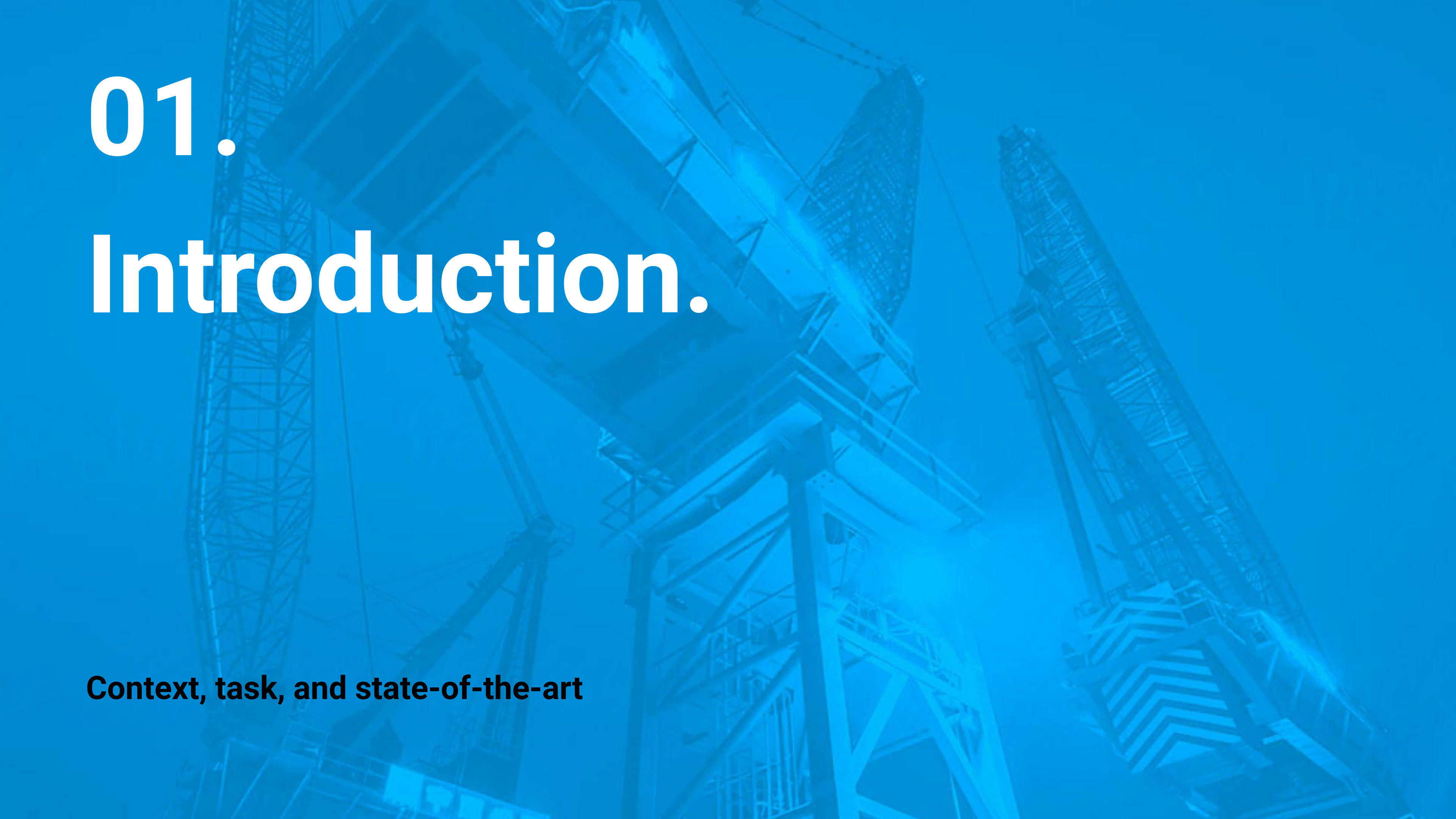
Temporal Graph Generation - An empirical study

EuroMLSys - April 2024

Lucas Ody

EURA NOVA

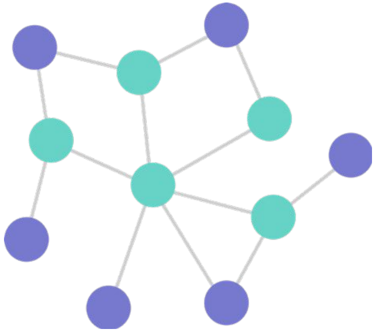
Summary



01.
Introduction.

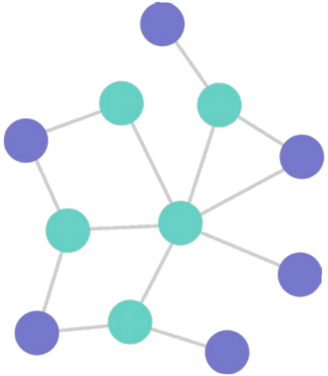
Context, task, and state-of-the-art

Problem Setting



Sensitive evolving real data

Temporal Graph generator



Evolving synthetic data

Scalability

Quality

Privacy

Temporal Graphs

A graph is defined as **$G = (A, X)$** where:

- **A** is a N by N **adjacency** matrix
- **X** is a N by D **feature** matrix.

A temporal graph is defined as a sequence of graphs **$G_T = \{G_i | i \in [0, T]\}$** where:

- **T** is the length of the considered time period

Graph generation

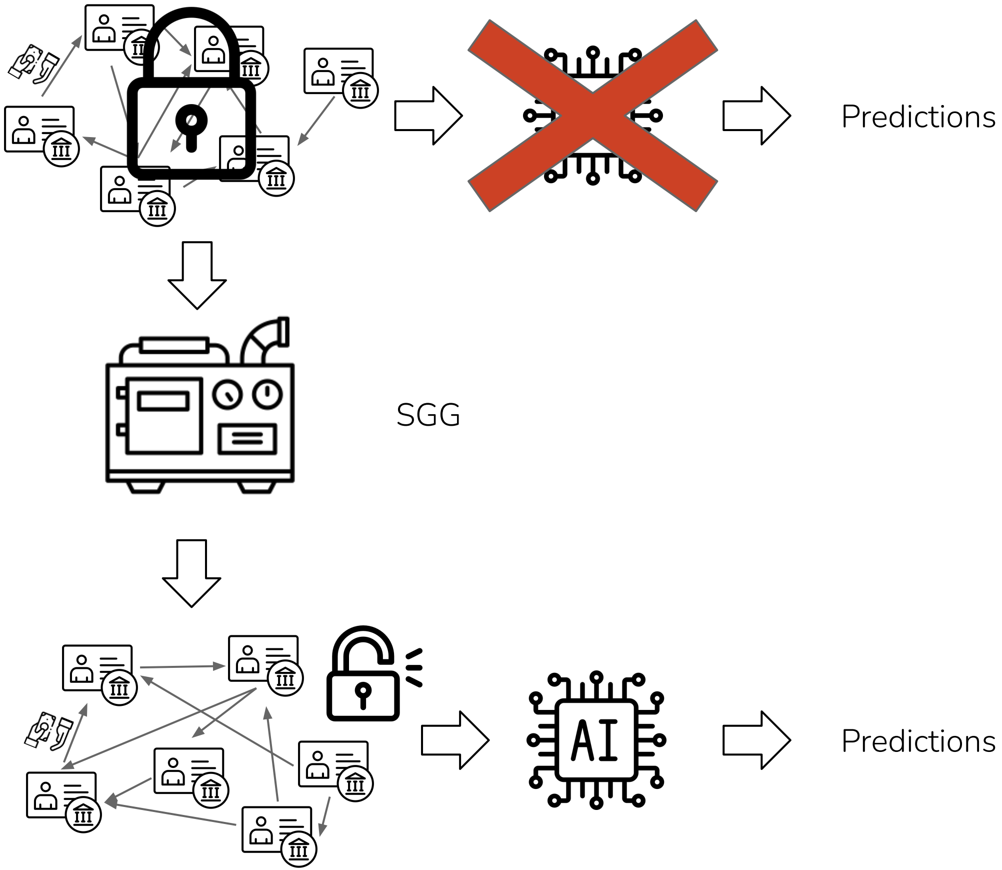
A graph is defined as **$G = (A, X)$** where:

- **A** is a N by N **adjacency** matrix
- **X** is a N by D **feature** matrix.

A temporal graph is defined as a sequence of graphs **$G_T = \{G_i | i \in [0, T]\}$** where:

- **T** is the length of the considered time period

Use cases:



02.
Evaluation

Comparison of models made possible

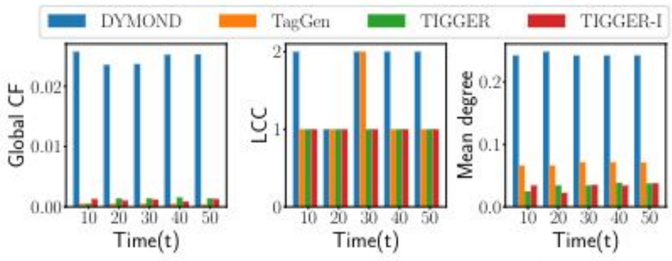
Current experiments:

Metrics for **Topological** quality

Scalability study not guaranteed

- What about **temporal** metrics?
- Downstream **utility**?
- **Privacy** leakage?

(3)



(2)

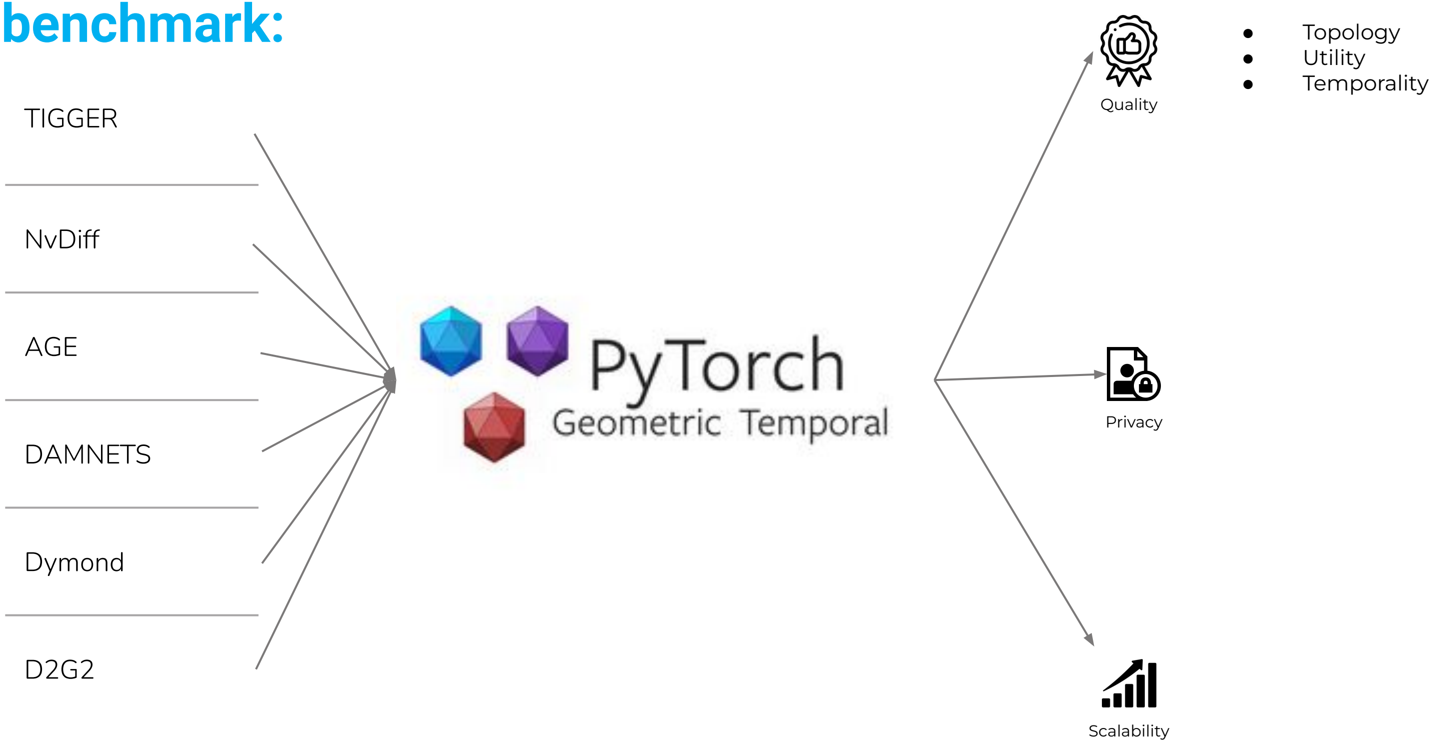
Table 1: The $\overline{\text{MMD}}$ on the B-A dataset for each network statistic. Lower is better.

Model	Degree	Clustering	Spectral	Transitivity	Assortativity	Closeness
DYMOND	14.01	61.20	8.78	7.28	4.76	3.19
TagGen	16.33	16.55	2.29	2.06	23.95	0.10
AGE	15.08	25.15	9.45	3.42	6.37	2.36
DAMNETS	**$8e^{-3}$**	**0.78**	**0.14**	**0.01**	**0.01**	**$5e^{-6}$**

(1)

	Community-small				Ego-small			
	Deg.↓	Clus.↓	Orbit↓	NN↓	Deg.↓	Clus.↓	Orbit↓	NN↓
GraphRNN	0.080	0.120	0.040	0.997	0.090	0.220	0.003	**1.094**
GRAN	0.070	0.045	0.021	0.701	0.020	0.126	0.010	**1.095**
GraphCNF	**0.021**	0.141	0.044	0.876	0.011	**0.011**	**0.001**	**1.094**
EDP-GNN	0.053	0.144	0.026	0.789	0.052	0.093	0.007	**1.095**
GDSS	0.045	0.086	**0.007**	0.858	0.021	0.024	0.007	1.097
NVDiff	**0.021**	**0.035**	0.018	**0.688**	**0.005**	0.045	**0.001**	**1.092**

Our benchmark:



Quality

Utility

Run a Link Prediction Model, or Node Classification

Topology

- Compute topological property
- Compare distribution of properties

Time metrics

Compare the temporal correlations

Privacy

- Nearest Neighbour Distance Ratio, **NNDR**

d_1 = Closest Nodes Distance
d_2 = Second Closest Nodes Distance

NNDR = d_1/d_2

- Distances computed from the embedders of the utility downstream tasks

	DTW
Perturbation Percentage	*NNDR (mean ± std)*
Orig. data	0 ± 0
Pert. data (5%)	0.323 ± 0.1
Pert. data (10%)	0.397 ± 0.127
Pert. data (25%)	0.543 ± 0.166
Pert. data (50%)	0.694 ± 0.179
Pert. data (75%)	0.75 ± 0.162
Pert. data (90%)	0.784 ± 0.155

Scalability

Threshold-based

- Memory consumption can be estimated per model
- Training time can be measured per epoch, to estimate a full training time

03. Limitations

The lacking points & the future

Privacy

- Similarity-based metrics are not reliable:

(5)

On the Inadequacy of Similarity-based Privacy Metrics: Reconstruction Attacks against "Truly Anonymous Synthetic Data"

Georgi Ganev[1,2] and Emiliano De Cristofaro[3]
[1]University College London [2]Hazy [3]UC Riverside

- Difficult enough problem for tabular data
    - Room for **improvement**!

04.
Closing words

What's the best model?

- Choice of **Dataset is more important** than choice of model

	n. snapshots	*n. nodes*	*min n. edges*	*mean n. edges*	*max n. edges*	*Node features*
Insecta	39	165	4894	12677.64	21548	X
Twitter Tennis	120	1000	41	340.32	936	16
England Covid	53	129	836	1358.5	2158	8
Wiki small	50	1616	33	59.60	93	X
Bitcoin	190	3783	1	127.29	1171	X
Reddit	588915	10984	1	1.14	7	X
IMDB	28	150544	6822	21156.29	37040	X

(1)

References

(1) Xiaohui Chen, Yukun Li, Aonan Zhang, and Li-ping Liu. 2022. NVDiff: Graph Generation through the Diffusion of Node Vectors. (2022)

(2) Jase Clarkson, Mihai Cucuringu, Andrew Elliott, and Gesine Reinert. 2022. DAMNETS: A deep autoregressive model for generating Markovian network time series. In Learning on Graphs Conference.

(3) Shubham Gupta, Sahil Manchanda, Srikanta Bedathur, and Sayan Ranu. 2022. TIGGER: Scalable Generative Modelling for Temporal Interaction Graphs. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36. 6819–6828.

(4) Penghang Liu and Ahmet Erdem Sariyüce. 2023. Using Motif Transitions for Temporal Graph Generation. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 1501–1511.

(5) Georgi Ganev and Emiliano De Cristofaro, 2023, On the Inadequacy of Similarity-based Privacy Metrics: Reconstruction Attacks against "Truly Anonymous Synthetic Data''

(6) Benedek Rozemberczki and Paul Scherer and Yixuan He and George Panagopoulos and Alexander Riedel and Maria Astefanoaei and Oliver Kiss and Ferenc Beres and and Guzman Lopez and Nicolas Collignon and Rik Sarkar, 2021, PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models 4564–4573