



# Embedding-Based Node Clustering in Temporal Interaction Networks

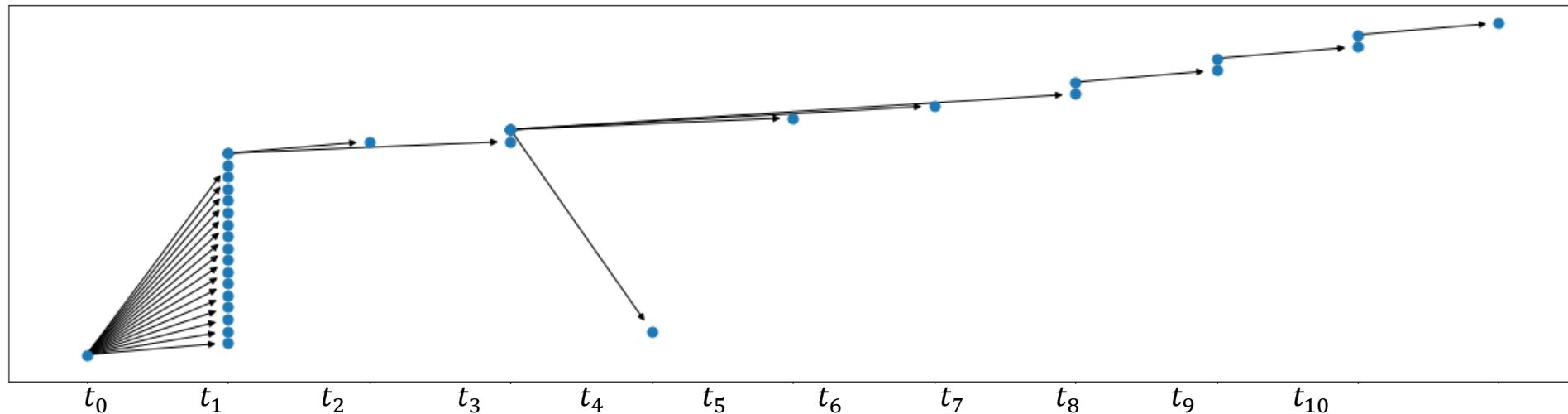
Adriana M. Ortiz Aquino  
Computing/DSSI  
Mentor: Goran Konjevod





# Temporal Data

- Radoslaw email dataset [1]
- It contains 167 nodes and 82.9K edges, where an edge represents an email exchange between two employees and each edge has a *time* attribute which corresponds to the timestamp the email was sent (57K unique emails)
  - Period covered is from January 1, 2010 to September 30, 2010
- It is also accompanied by the graph representing organizational hierarchy showing who supervises whom (without positions) [2]



[1] Datasets, K. O. N. E. C. T. (2015). The koblenz network collection. [http://konect.uni-koblenz.de/networks/radoslaw\\_email](http://konect.uni-koblenz.de/networks/radoslaw_email)

[2] Michalski, Radosław, 2020, "Manufacturing company email metadata and corporate hierarchy", <https://doi.org/10.7910/DVN/6Z3CGX>, Harvard Dataverse, V2, UNF:6:/mJDDYEUXZ0pb1bvKiKttA== [fileUNF]



# Time-Respecting Node Embedding

- Continuous-Time Dynamic Network Embedding (CTDNE) [3] learns a time dependent network representation for a temporal interaction network  $G = (V, E_T, \tau)$ 
  - Learns a temporal embedding by searching over the space of temporal random walks that obey time
  - Example: A random walk from node  $v_{i_1}$  to  $v_{i_{L+1}}$ 
$$\{(v_{i_1}, v_{i_2}, t_{i_1}), (v_{i_2}, v_{i_3}, t_{i_2}), \dots, (v_{i_L}, v_{i_{L+1}}, t_{i_L})\}$$
where  $t_{i_1} < t_{i_2} < \dots < t_{i_L}$
- Walks are biased towards edges that appear closer in time, i.e., the walks represent a (possible) chain of emails in a week
  - This is achieved using an exponential bias where given an arbitrary edge  $e = (u, v, t)$ , each temporal neighbor  $w \in \Gamma_t(v)$  has probability of being selected given by

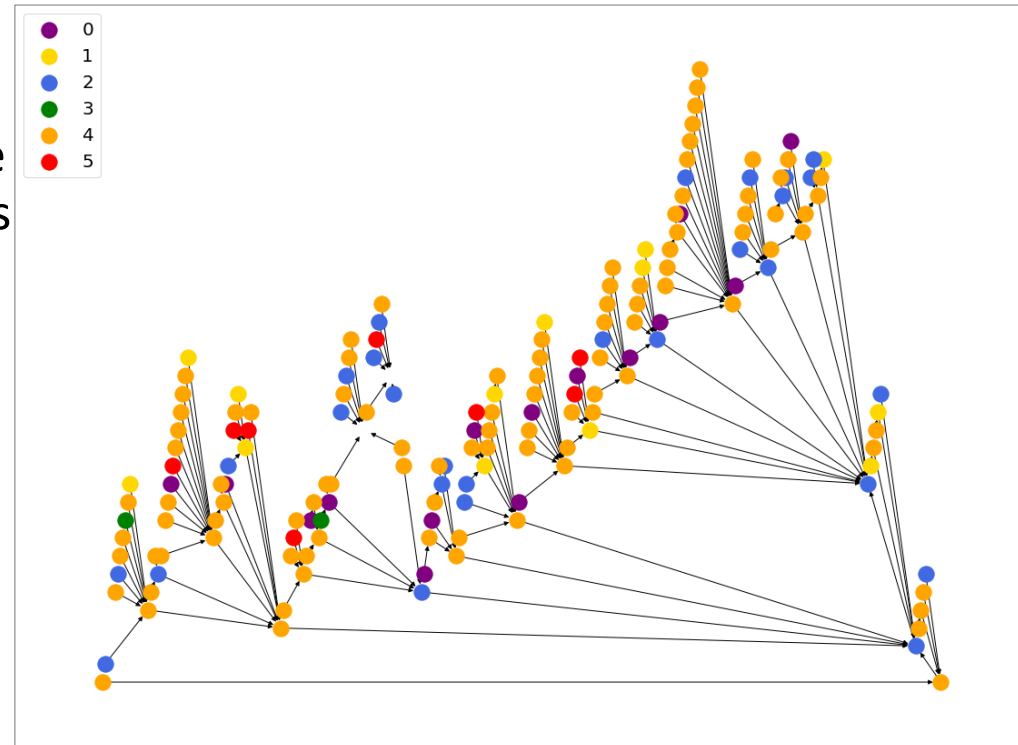
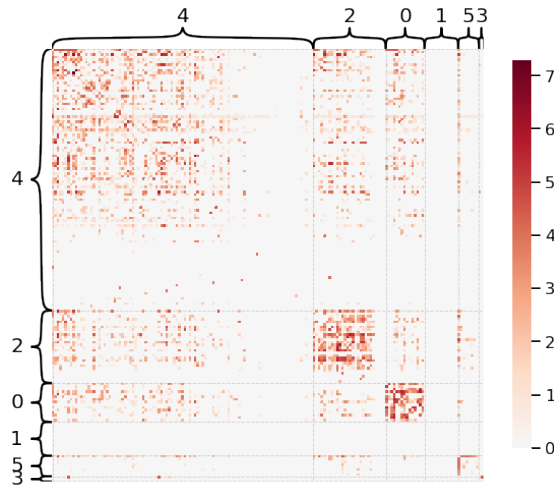
$$\Pr(w) = \frac{\exp[-(\tau(w) - \tau(v))]}{\sum_{w' \in \Gamma_t(v)} \exp[-(\tau(w') - \tau(v))]}$$

[3] Nguyen, G. H., Lee, J. B., Rossi, R. A., Ahmed, N. K., Koh, E., & Kim, S. (2018, April). Continuous-time dynamic network embeddings. In *Companion Proceedings of the The Web Conference 2018* (pp. 969-976).



# Clustering

- Used Gaussian Mixture Model [4] to cluster the temporal embedding obtained from CTDNE
- The clusters assign nodes based on structural equivalence [5] which implies that the organization of the nodes is based on hierarchical roles in the network



[4] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.

[5] Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5), 75-174.



**Disclaimer**

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.