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 and 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]

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 $v_{i_{L+1}}$
    \[
    \{(v_{i_1}, v_{i_2}, t_{i_1}), (v_{i_2}, v_{i_3}, t_{i_2}), \ldots, (v_{i_L}, v_{i_{L+1}}, t_{i_L})\}
    \]
    where $t_{i_1} < t_{i_2} < \ldots < t_{i_L}$
  - 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 $w \in I^*_t(v)$ has probability of being selected given by
      \[
      \Pr(w) = \frac{\exp[-(\tau(w) - \tau(v))]}{\sum_{w' \in I^*_t(v)} \exp[-(\tau(w') - \tau(v))]} \]

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
