Meeting Notes

Thursday, March 5, 2020   2:03 PM

How to define the spatial-temporal career network?

The current career network only contains one type of node: (user, time-step) . Two users are connected based on the similarity (fuzzy matching score) between their institution names. In addition, each node has one or more labels that represent the general categories of their working institution.

To better incorporate the label information into the network, we discussed an alternative network definition: changing the graph at each time step into a bipartite graph of user nodes and institution nodes.

t=1992

User nodes   Institutions   (Institution labels)
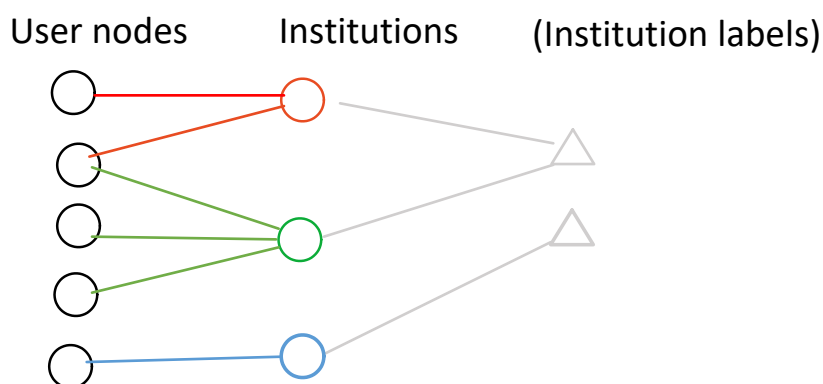
Users and institutions can have separated embeddings.

Additional modeling ideas:
- Represent supervisor/subordinary relationship between users as direct edges
- Modeling the hierarchical relationship between different institutions

Trajectory embedding

Trajectory embedding can be considered as a special kind of subgraph embedding.

Random walk-based methods (e.g. ST-Walk) consider the entire spatial-temporal graph

as a whole, such that random walks can be made spatially (within the same time step) and also temporally (going back in time). Each user node embedding will contain information of the user's career history. A trajectory embedding is the node embedding of the last node on the trajectory.

Other methods will model dynamic embedding of a user. i.e. the embedding of a user changes over time. A user trajectory can be represented as a sequence of node embeddings over time. In this way, we can easily predict future embedding of this user through an RNN-based model. The predicted embedding can be used to make inference on future links or attributes.

Novel Applications

1. **Missing link inference**: if a user wants to be working at a certain position/institution in the future, what intermediate steps should the user go through given his/her career history. This can be considered as an inverse problem to the career prediction problem.
2. **Pattern mining**: Learn repeated career trajectory patterns from all users. Such patterns will be useful for career prediction. It can also be used to solve the missing link inference problem.
3. **Anomaly detection**: Related to pattern mining is to detect unusual trajectory patterns. It may be interesting to uncover the cause of such anomaly.

Data Sources

1. In order to perform large-scale, meaningful analysis, we need to decide how to anonymize sensitive resume data.
2. In case of not having access to the Shenzhen resume dataset, we can consider similar problems on internet traffic networks.
3. Another public data alternative is the co-author network and scientific publication history.