Paper Summary

David Miller
CIS 5930: Social Network Mining

February 14, 2018

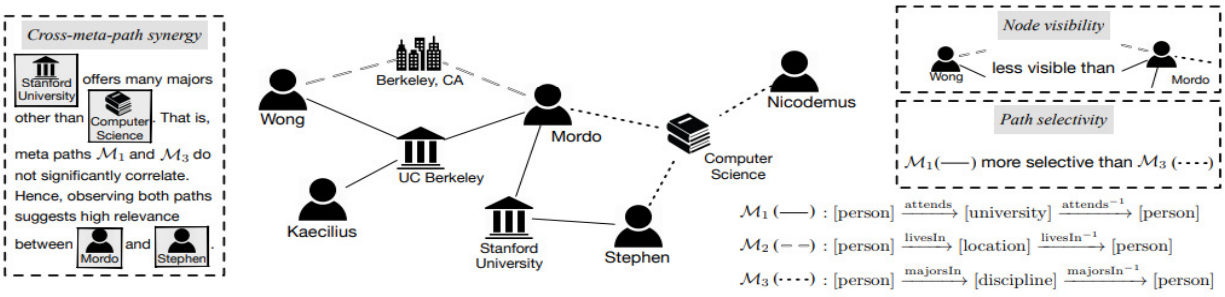

Figure 1

A heterogeneous information network (HIN) is a network with varying types of objects that are interconnected. An example is shown un figure 1 where person, places, universities, and disciplines are in the same network. This networks are mathematically described by a graph $G = (V, E)$. The Path-based Relevance Probabilistic (PReP), in a nutshell, consists of two major parts: (i) inferring model parameters by finding the maximum a posteriori (MAP) estimate to fit the input HIN, and (ii) deriving relevance score between any node pair based on the learned model [1]. The two parts of the algorithm can be described as

1. **The Inference Algorithm:** Input the path counts P and the parameters. Initialize the parameters ρ, Φ, and Θ. While there hasnt been convergence, update the parameter η by equation 10 in the paper. While η has not converged, update ρ_u by equation 11 in the paper for all u. Update Φ with equation 13 in paper and update θ with equation 12 in the paper.

2. The relevance measure is given by

$$r(s) = \sum_{t=1}^{T} \frac{P_{st}}{\rho_u \rho_\nu \eta_t \sum_{k-1}^{K} \phi_{sk} \theta_{kt}} + (1 - \beta) \sum_{k=1}^{K} \log \phi_{sk}$$

   where the relevance is defined for all pairs of nodes.

In terms of future work, the author alludes to a couple things in the paper

1. Sampling algorithm design to efficiently calculate marginal likelihood.

2. Exploration of defining relevance from the proposed PReP model with marginal likelihood.

3. Study on the cases with label information and meta-path selection.

** I do not have a complete understanding of this paper so my strengths, weaknesses, and questions wont be too in depth **

Three strengths I found with the paper are

1. The results look better than current algorithms.
2. Real world applications of the algorithm lend itself to commercialization.
3. Algorithm can be parallelized, making it efficient.

Three weaknesses I found with the paper are

1. The paper is not an easy read for those who lack a strong statistical background, making it hard for people to understand the research.
2. I honestly can not determine other weaknesses since I had a hard time reading the paper.

Questions for the reader

1. I do not really have questions since the paper is unclear to me so I am hoping to understand the paper more in today's talk.

References

[1] Yu Shi, Po-Wei Chan, Honglei Zhuang, Huan Gui, Jiawei Han *PReP: Path-Based Relevance from a Probabilistic Perspective in Heterogeneous Information Networks*, KDD17, August 13-17, 2017, Halifax, NS, Canada.