COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Dissertation Defense

Mark Heimann

Unsupervised Structural Embedding Methods for Efficient Collective Network Mining

Friday, July 10, 2020
12:00 – 2:00 pm
umich.zoom.us/j/8456614642

ABSTRACT: How can we align accounts of the same user across social networks? Can we identify the professional role of an email user from their patterns of communication? Can we predict the medical effects of chemical compounds from their atomic network structure? Many problems in graph data mining, including all of the above, are defined on multiple networks. The central element to all of these problems is cross-network comparison, whether at the level of individual nodes or entities in the network or at the level of entire networks themselves. To perform this comparison meaningfully, we must describe the entities in each network expressively in terms of patterns that generalize across the networks. Moreover, because the networks in question are often very large, our techniques must be computationally efficient.

In this thesis, we propose scalable unsupervised methods that embed nodes in vector space by mapping nodes with similar structural roles in their respective networks, even if they come from different networks, to similar parts of the embedding space. We perform network alignment by matching nodes across two or more networks based on the similarity of their embeddings, and refine this process by reinforcing the consistency of each node's alignment with those of its neighbors. By characterizing the distribution of node embeddings in a graph, we develop graph-level feature vectors that are highly effective for graph classification. With principled sparsification and randomized approximation techniques, we make all our methods computationally efficient and able to scale to graphs with millions of nodes or edges. We demonstrate the effectiveness of structural node embeddings on industry-scale applications, and propose an extensive set of embedding evaluation techniques that lay the groundwork for further methodological development and application.

Chair: Prof. Danai Koutra