



DISSERTATION DEFENSE



DI JIN

Graph Summarization Meets Representation Learning for Scalable Feature Summarization: Methods & Applications

Monday, November 1, 2021

1:30 – 3:30pm

[Virtual](#)

ABSTRACT: Graphs are ubiquitous as they naturally capture interactions between entities. Recently, graph representation learning has gained significant popularity thanks to its superiority in downstream machine learning (ML) tasks, such as friend recommendation and anomaly detection. Specifically, node representation learning (embedding) aims to find a dense vector of rich latent features per entity for these tasks. However, these representations with fixed dimensions come with computational and storage challenges for large real-world graphs, and the “black-box” nature of the latent features impedes interpretability. Graph summarization aims to describe the original graph with concise and interpretable representation, but it is often lossy and trades off space and performance in ML tasks.

In this thesis, we bridge the two lines of research, node embedding and graph summarization by introducing scalable methods for generating summaries of latent or non-latent node features that achieve the state-of-the-art performance on ML tasks while requiring significantly reduced storage and supporting interpretability. Specifically, we introduce latent network summarization that summarizes the graph structural features in static networks as latent node embeddings for storage and query efficiency, and extend this idea to incorporate temporal proximity as the temporal summaries in continuous-time dynamic networks. We also perform an extensive systematic study of temporal summaries and show that they capture the graph structure and temporal dependency at least as well as recently-proposed dynamic embedding approaches, while being less complex and easy-to-understand. Finally, we summarize the non-latent graph features by modeling feature importance as the high-level knowledge that can be used for graph analysis and transfer learning. We demonstrate the effectiveness, scalability, and space efficiency of our methods on industrial applications such as entity linkage and professional role inference.

CHAIR: Prof. Danai Koutra