

# STRUCTURE DETECTION IN GRAPHS AND HYPERGRAPHS

Abstract

by Xu Liu, Ph.D.  
Washington State University  
July 2021

Co-Chairs: Assefaw H. Gebremedhin and Andrew Lumsdaine

Understanding relationships among entities in large-scale systems is a fundamental operation in data science. Such systems can be modeled as graphs if the relationships are pairwise or as hypergraphs if the relationships are multi-way. Structures in graphs and hypergraphs such as communities or connected components and metrics such as centrality indices encapsulate relationships among entities. For community detection and connected components decomposition, in graphs and hypergraphs, as the sizes of the systems continue to increase, we generally need fast algorithms to enable large-scale real-time processing. The first contribution of this thesis is to develop linear-time algorithms that can be used to detect communities and connected components in graphs. We achieve this by developing a new framework based on Direction-optimizing Label Propagation Algorithm (DOLPA). We parallelize DOLPA in both shared-memory and distributed-memory. Since community detection is NP-hard,

it is difficult for community detection algorithms to achieve both high quality of solution and fast convergence at the same time. As the second contribution of this thesis, DOLPA provides a trade-off to accommodate different community detection scenarios: accurate detection versus fast detection.

The other two contributions of the thesis target hypergraphs. Hypergraphs are increasingly becoming important in data science applications due to their capability for robust data representation. However, their computational complexity can be high. To reduce this complexity, lower-dimensional approximations of hypergraphs to compute relevant structures are needed. Line graphs, in particular, high order  $s$ -line graphs for  $s > 1$  of hypergraphs are a critical tool for this purpose, since they capture important hypergraph properties. Obtaining high order  $s$ -line graphs is highly compute-intensive and irregular. Until recently, no efficient algorithm has been proposed to compute  $s$ -line graphs for  $s \geq 2$ . The third contribution of this thesis is to develop a scalable framework for  $s$ -line graph computation and  $s$ -metrics computation for arbitrary  $s$  values. The framework is applied for hypergraph structure detection in real-world systems. The last contribution of the thesis is to design algorithms to speedup  $s$ -line graph computation. The new algorithms are orders of magnitude faster than a naive algorithm and a sparse general matrix-matrix multiplication (SpGEMM) implementation with filtration.