Title: Scalable Algorithm Design and Performance Analysis for Graph Motifs Discovery

Author list: Md Abdul Motaleb Faysal, Shaikh Arifuzzaman (Ph.D. Advisor)

Abstract: Community discovery is a prominent graph application with usages in classifying groups in social and business networks, finding similar proteins, detecting anomalous behavior in the cyber-security domain, finding critical points/entities in rumor propagation or infectious disease spreading, etc. State-of-the-art techniques for community discovery suffer from inadequate scalability, poor performance, and methodological inaccuracy. Scalable algorithms are required to process the massive networks available nowadays. The goal of this doctoral dissertation is to develop novel parallel algorithms and propose a high-performance computing architecture design for overcoming performance limitations in existing community detection approaches for processing massive social and biological data. There are two major types of community discovery; i) global community discovery where the entire network is broken down into disjoint communities of vertices, ii) goal-oriented/local community discovery where the goal is to find the communities of a query entity. In this doctoral research, we delve into both categories (global and local) of community detection problems, identify two outstanding sequential approaches, one from each category, and design, and implement parallel algorithm solutions for both of the sequential approaches. Our parallel algorithm design for global community discovery achieves up to 25X speedup compared to the original sequential approach using hybrid memory parallelism. Additionally, based on our observation from extensive benchmarking and performance analysis of software hash operations in the major compute kernel of our global community discovery, we propose a generalized accelerator design for hash accumulation and simulate the architecture design to attain up to 5.6X speedup using the accelerator. For the local/goal-oriented community discovery approach, we design a scalable novel algorithm based on shared memory parallelism that demonstrates 20X to 55X speedup for graphs with hundreds of millions to billions of edges.