

Discovering motifs or structural patterns such as communities is a prominent graph application used for 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, lack of performance, and methodological inaccuracy. The goal of this doctoral dissertation is to develop novel parallel algorithms and propose high-performance computing architecture design for overcoming performance limitations in existing community detection approaches for processing massive social and biological data. In this doctoral research, we delve into two major categories of community discovery problems: i) global community discovery and ii) local community discovery, identify two outstanding sequential approaches, one from each category, and design and implement parallel algorithm solutions for both 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, we run extensive benchmarking and performance analysis of software hash accumulation on two prominent community discovery approaches. Based on our observation, we propose a generalized accelerator design for hash accumulation, and we conduct simulations of that proposed architecture design to attain up to 5.6X speedup. For the local/goal-oriented community discovery approach, we design a novel parallel algorithm based on shared memory parallelism that demonstrates 20X to 55X speedup for graphs with hundreds of millions to billions of edges.