Abstract

The underlying theme of this thesis is the detection of dense regions of an undirected graph $G = (V, E)$. A dense region is a subset of the nodes $V' \subset V$ with many edges between nodes in V' and only few edges to nodes in $V \backslash V'$. The identification of these regions is helpful for solving the graph partitioning problem (GPP) and related clustering tasks. The GPP asks for a partition of V into k equally sized subdomains (clusters) such that the number of inter-cluster edges is minimized. Applications that involve problems related to GPP are numerous; they include parallel numerical simulations, network analysis, circuit design, and gene analysis in bioinformatics.

GPP and all relevant formulations of related partitioning problems are $\mathcal{NP}$-hard, so that no polynomial-time algorithms for their optimal solution are known. State-of-the-art graph partitioning libraries employ local node-exchanging heuristics within a multilevel framework and yield good solutions in very short time. However, the computed partitions do not necessarily meet the requirements of all users. This includes the choice of the appropriate objective function and the shape of the computed subdomains. Furthermore, due to their sequential nature, the most popular partitioning heuristics are difficult to parallelize, which is necessary for their efficient use as load balancers in parallel applications. For graph clustering problems, where the cluster sizes do not need to be balanced, there is no method which is both highly efficient, delivers high-quality results in many diverse applications, and is theoretically well understood.

To overcome these drawbacks, we introduce the disturbed diffusion scheme FOS/C. It is capable of distinguishing dense from sparse graph regions, which we explain by its relation to random walks. The combination of FOS/C with the k-means related framework Bubble yields the iterative and inherently parallel (re)partitioning/clustering algorithm Bubble-FOS/C. In our theoretical investigations on FOS/C and Bubble-FOS/C, we examine the random walk relation and its connection to the pseudoinverse of the input graph's Laplacian matrix. Amongst others, the derived results lead to an enhanced solution process of FOS/C and to a proof that Bubble-FOS/C converges to a local optimum which can be characterized by a potential function.

Since Bubble-FOS/C requires the solution of many linear systems, we construct an efficient algebraic multigrid solver, whose graph hierarchy is simultaneously used for a multilevel improvement process of the partitions. Despite the fact that our algebraic multigrid approach is significantly faster than previous implementations, the running time of Bubble-FOS/C is still very high. Thus, its very good solution quality experienced in graph partitioning experiments can hardly be exploited in practice. Further

acceleration approaches are discussed, but they are either not always successful or very complicated to implement.

That is why we develop in a next step a much faster and easier method for the improvement of partitions. This method is based on a different disturbed diffusive process, which is restricted to local areas of the graph and also contains a high degree of parallelism. By coupling this new technique with Bubble-FOS/C in a multilevel framework based on two different hierarchy construction methods, we obtain our new heuristic DibaP for (re)partitioning and clustering graphs. Compared to Bubble-FOS/C, DibaP shows a considerable acceleration, while retaining the positive properties of the slower algorithm. Extensive experiments with popular benchmark graphs show an extremely good behavior for partitioning graphs stemming from numerical simulations. DibaP computes consistently better results than the state-of-the-art libraries METIS and Jostle. Moreover, with our new algorithm, we have improved a large number of the best known partitions of six widely used benchmark graphs. In the related problems of load balancing by repartitioning and graph clustering, DibaP also improves the solution quality of state-of-the-art programs in many cases.

Insofar, our work consists of practical and theoretical advances concerning graph (re)partitioning and graph clustering, achieved by the development of new successful heuristic algorithms and the theoretical analysis of some important properties of these algorithms.