Multiscale Approaches for Network Compression

Combinatorial Scientific Computing for Petascale Simulations

Ilya Safro
Argonne National Laboratory
9700 S Cass ave, Argonne, IL 60439

Abstract

Finding a suitable compressed representation of large-scale networks/matrices/graphs has been intensively studied in both practical and theoretical branches of network analysis and data mining. In particular, the success of applying some of the recently proposed compression schemes strongly depends on the "compression-friendly" arrangement of network nodes. Usually, the goal of these arrangements is to order the nodes such that the endpoints of network links (edges) are located as close as possible. Doing so leads to a more compact representation of links and allows a better performance of compression schemes and network element access operations.

Recently, Chierichetti et al. formulated an NP-hard combinatorial optimization problem, namely, *the minimum logarithmic arrangement problem* (MLogA), that minimizes the gap encodings of edges stretched between their endpoints. This is achieved by ordering the network nodes and assigning to them unique integer values (ids) such that the endpoints of a link will obtain close values. Several other well-known optimization problems (such as the minimum linear arrangement and the minimum bandwidth) have goals similar to MLogA. In contrast, however, MLogA seeks a nearly optimal information-theoretical compressed encoding size for all network links as it minimizes the total number of bits to spend for this purpose. In practice, it often means that further compression algorithms will benefit if locality and similarity properties of a network will be reflected in the MLogA ordering. In this talk, we present a multiscale method for approximating a generalized *link-weighted* and *node-volumed* version of MLogA for *general* networks. The importance of a link-weighted property (when each link is assigned by a nonnegative weight) for this problem is twofold. First, the link weight can measure the significance of that link. For example, in many cases we have information regarding its access frequency, and we would prefer that frequently accessed links would be compressed better. Second, the multiscale algorithmic framework admits a natural aggregation of *weighted* links at different scales of the network structure. Similarly to the link weights, all nodes are assigned by nonnegative volumes that represent the length of their segments captured in the ordering of the network nodes.

Our multiscale algorithm is based on the algebraic multigrid methodology for linear ordering problems. In contrast to the existing approaches designed mostly for social networks, the main goal of this work is to provide a *generic* MLogA solver for general networks and matrices. Social networks (including big parts of web graphs) commonly consist of a combination of structural properties (such as degree distribution, small diameter, and expander-like topology) that enable fast greedy methods (usually based on some preferential ordering of graph traversal) to find arrangements for further high-quality compression. In real life, however, there occur many situations when the structure of a stored network (or a part of it) is complex and irregular (such as decision/detailed supply/infection spread networks and even parts of social networks that do not exhibit power law degree distribution), which creates many contradictions between greedy local decisions and solutions that consider a more global picture. We address this type of problem by introducing a multiscale solver. In the experimental part of this work we demonstrate significant improvement for minimization of logarithmic arrangement for various families of networks. In almost all cases, our solver exhibited better numerical results than previous best-known results.