

Fast Group Centrality Algorithms for Large Graphs

Henning Meyerhenke, Humboldt-Universität zu Berlin, Germany

Identifying important vertices is a fundamental task in network analysis. To this end, several vertex centrality measures have been introduced. Many of them can be generalized naturally to groups of vertices. This leads to the optimization problem to find the group of vertices with highest group centrality score. For several group centrality measures this problem has been shown to be NP-hard. Existing approximation algorithms fail to scale to large real-world networks. In this talk, we therefore address this lack of scalability from two directions.

First, in the context of group closeness centrality, we present new local search heuristics to compute highly central groups of vertices. These heuristics perform exchanges between vertices within the group and the rest of the graph until they reach a local optimum. Experimental results show that this strategy is two orders of magnitude faster than the state-of-the-art greedy method, while achieving up to 99.4% of the greedy method's solution quality.

Second, we introduce GED-Walk, a novel centrality metric inspired by Katz centrality. Similarly to Katz centrality, GED-Walk takes walks of any length into account, and gives greater importance to shorter walks than longer walks. We describe efficient algorithms to compute the GED-Walk score of a given group and to approximate the group of vertices with highest GED-Walk centrality. Our maximization algorithms keep lower and upper bounds of the marginal gain of every vertex to the GED-Walk score; then, they iteratively add to the (initially empty) group the vertex with highest marginal gain until the desired group size is reached.

Experimental results show that GED-Walk improves the precision of popular graph mining applications, and that our algorithm for maximizing it (in approximation) is two orders of magnitude faster than group betweenness approximation and, for group sizes up to 100 vertices, one to two orders of magnitude faster than greedy group closeness approximation. For real-world graphs with tens of millions of edges, our algorithms for GED-Walk maximization typically need less than one minute.

Our algorithms are implemented on top of NetworKit, a growing open-source toolkit for large-scale network analysis which we portray briefly, too.

Based on recent joint work with E. Angriman and A. van der Grinten (IEEE Big-Data 2019, <https://arxiv.org/abs/1911.03360>) as well as E. Angriman, A. van der Grinten, A. Bojchevski, D. Zügner, and S. Günnemann (SIAM ALENEX 2020, <https://arxiv.org/abs/1910.13874>).