Approximated and Randomized Betweenness Centrality

Introduction
Betweenness Centrality (BC) is a popular metric for identifying the most relevant nodes in a graph [Neuaman]. This measure basically consists of assigning a score to each vertex *v* (there is also a formulation on edges) based on how many times *v* occurs in the shortest paths of the graph.

The straightforward approach to calculate BC requires computing the all-pairs shortest path (APSP) among the vertices and storing their occurrences. Unfortunately, for the graphs of interest today this approach is not feasible, mainly due to the memory complexity of the APSP algorithm [Mikkel 1999]. In fact, the memory footprint required by this method is $O(n^2)$, which depends on the number of shortest paths between each pair of vertices.

In 2001, Brandes [Brandes 2001] re-formulated the BC algorithm in order to use a single shortest path iteratively. This approach basically computes for each vertex *v* of the graph its contribution to the BC score of the other vertices. Although this method solves the memory requirement issue, it does not allow the computation of the exact BC scores in graphs composed of millions, or even billions of vertices. Distributed memory techniques scale BC algorithms over multiple—typically thousands—parallel computing nodes equipped with accelerators, significantly reducing the time-to-solution that yet remains in the order of hours or days.

In this scenario, approximated techniques based on randomization achieve results by threading off the accuracy of the solution with performance. The most efficient one is based on a purely agnostic—in terms of knowledge of the topology of the graph—uniform sampling method, which guarantees a 2-alpha approximation of the solution [Bader 2007, van der Grinten 2020]. Other heuristics showed promising results on specific graphs [Geisberger 2008, Kourtellis 2013, Riondato 2018].

Project description
In this thesis, we propose to explore different aspects to design a more efficient approximated and randomized algorithm for the BC. The candidate will build his/her work on top of state-of-the-art parallel and distributed algorithms [Vella]. The project will achieve multiple intermediate goals, depending on the ability of the student and his/her graduate level. More specifically, the activity will focus on the following aspects:

- Survey on heuristics and other BC approximations [BSc, systems 50% - writing 50%]
- Benchmark new heuristics and analyze their quality according to graph characteristics. [BSc, theory 40% - coding 10% - systems 50%]
- Evaluating Approximated BC against other newer centrality measures, such as Expected Force, as pre-filter metrics for the influence maximization problem. [BSc/MSc: theory 40% - coding 30% - systems 30%]
- Design new heuristics, assess their quality and include them into the Approximated BC code. [Msc: theory 40% - coding 30% - systems 30%]

References

[Bader 2007] Bader, D.A., Kintali, S., Madduri, K. and Mihail, M., 2007, December. Approximating betweenness centrality. In International Workshop on Algorithms and Models for the Web-Graph (pp. 124-137). Springer, Berlin, Heidelberg.
[Brandes 2001] Brandes, U., 2001. A faster algorithm for betweenness centrality. Journal of mathematical sociology, 25(2), pp.163-177.
[Geisberger 2008] Geisberger, R., Sanders, P. and Schultes, D., 2008, January. Better approximation of betweenness centrality. In 2008 Proceedings of the Tenth Workshop on Algorithm Engineering and Experiments (ALENEX) (pp. 90-100). Society for Industrial and Applied Mathematics.
[Kourtellis 2013] Kourtellis, N., Alahakoon, T., Simha, R., Iamnitchi, A. and Tripathi, R., 2013. Identifying high betweenness centrality nodes in large social networks. Social Network Analysis and Mining, 3(4), pp.899-914.
[Mikkel 1999]. Mikkel, T. 1999: Undirected Single-Source Shortest Paths with Positive Integer Weights in Linear Time. Journal of the ACM, 46(3), pp. 362–94.
[Riondato 2018] Riondato, M. and Upfal, E., 2018. Abra: Approximating betweenness centrality in static and dynamic graphs with rademacher averages. ACM Transactions on Knowledge Discovery from Data (TKDD), 12(5), pp.1-38.
[Van der Grinten] Van der Grinten, A. and Meyerhenke, H., 2020, May. Scaling betweenness approximation to billions of edges by MPI-based adaptive sampling. In 2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS) (pp. 527-535).
[Vella] Vella, Tumeo, Bernaschi, Halappanavar. “Approximated Betweenness Centrality on Multi-GPU Systems”. Available on git https://github.com/fvella/axbc.