Dynamic Community Detection in Bipartite Networks (master's thesis)

The most popular community detection method, `Louvain` [1], produces arbitrarily bad results. Recently, the `Leiden` algorithm [2] outclassed `Louvain` and attracted considerable attention after publication in Nature. It promises performance and quality gains for static, unipartite networks, and the potential applications range from protein-protein interaction networks to customer behavior analysis in product co-purchasing networks. However, many of these applications evolve dynamically and have bipartite symmetry. While both bipartite [3] and dynamic [4] versions for `Louvain` exist, their adoption to `Leiden` will be crucial for the practical usability of the algorithm. In this thesis, you will develop a bipartite, dynamic version of the `Leiden` algorithm. We will publish the results, and the method will improve a production system of a sporting goods manufacturer.

Challenges

Both methods heuristically optimize the modularity, a measure for clustering quality for unipartite networks. However, a singleton clustering, where every node has its own community, already has high modularity in bipartite networks. As a remedy, one could project the bipartite graph to its unipartite equivalent, compromising on higher memory requirements. Alternatively, your algorithm could optimize a bipartite version of the modularity [5], which might be more difficult to implement.

While a dynamic bipartite version of `Louvain` exists [6], its ideas will have to be aligned with the `Leiden` algorithm. We will have to ensure that an incremental cluster update conserves the convergence guarantees that set `Leiden` apart from its predecessor and demonstrate your algorithm's performance on a benchmark dataset [7].

Profile and Outlook

If you bring

- an interest in the foundation of modern AI and are curious about the underlying mathematics
- previous experience implementing advanced algorithms, preferably in C++

This thesis offers

- the opportunity to make a practical impact in the field of e-commerce, bringing you in a good position for a career as, e.g. backend developer
- the potential for a solid publication, increasing your academic visibility in the growing field of AI

References

[1] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech.*, vol. 2008, no. 10, p. P10008, Oct. 2008.

[2] V. A. Traag, L. Waltman, and N. J. van Eck, "From louvain to leiden: Guaranteeing well-connected communities," *Sci Rep*, vol. 9, no. 1, p. 5233, Mar. 2019.

[3] T. Alzahrani and K. J. Horadam, "Community detection in bipartite networks: Algorithms and case studies," in *Complex systems and networks*, J. Lü, X. Yu, G. Chen, and W. Yu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 25–50.

[4] N. Dakiche, F. Benbouzid-Si Tayeb, Y. Slimani, and K. Benatchba, "Tracking community evolution in social networks: A survey," *Information Processing & Management*, vol. 56, no. 3, pp. 1084–1102, May 2019.

[5] M. J. Barber, "Modularity and community detection in bipartite networks," *Phys. Rev. E*, vol. 76, no. 6, p. 066102, Dec. 2007.

[6] T. Hecking, L. Steinert, T. Göhnert, and H. U. Hoppe, "Incremental clustering of dynamic bipartite networks," in *2014 european network intelligence conference*, 2014, pp. 9–16.

[7] J. Leskovec and A. Krevl, "SNAP datasets: Stanford large network dataset collection." Jun-2014.