

Applications of Overlapping Community Detection

András Bóta, Miklós Krész, and András Pluhár

The analysis of networks is considered to be a traditional research field, which in the recent years has received a new impulse, thanks to the variety of available test databases and benchmarks. These graphs are so large in some cases, that only the fastest, near linear time algorithms have a chance of tackling with the given tasks. One of the popular approaches to network analysis, community detection has received a lot of attention in the recent years both from the point of theory and practical applications.

On the other hand, the concept of community doesn't have a strict definition. Thinking intuitively, one can imagine communities as subsets of nodes in a graph, in which the edge density is higher than between these sets. The traditional approach to community detection deals with disjoint communities. There exists, however another approach to communities: overlapping communities. In this case, the nodes in the graph can belong to multiple communities. Articles dealing with overlapping communities are scarce in the literature, and approach the subject mostly from the theoretical point of view. Just like traditional communities, overlapping communities also lack a strict definition. In most applications the overlaps between the communities can only happen at the edges of the given communities, another approach however allows greater overlaps, identifying a rich community structure in the original network. This can be an important source of information in itself, but the community graph built from the results can also form the basis of further analysis.

In most cases the data to be analyzed is in the form of unweighted graphs, but sometimes it is necessary to deal with the weighted counterparts, which can contain more information. Algorithms able to effectively utilize this additional information are also scarce both in literature and applications.

In this article we propose an improved version of an unpublished algorithm [1], which can detect overlapping communities in a weighted graph, and scales in quasi-linear time, thus allowing the analysis of large test databases. During this process the algorithm explores a dense community structure, which we use as the basis of further algorithms [2], like clustering, label propagation [3] and influence propagation [4]. We compare the results of this procedure with results from running these methods on the original, unmodified graph. Our test databases mostly come from the field of molecular biology, but we have tested our methods on graphs from different fields of science.

References

- [1] L. Csizmadia: Community detection in social graphs, Master's Thesis, Szeged, 2003. (In Hungarian: Közösségfelismerés ismeretségi gráfokban)
- [2] A. Pluhár: Fast algorithms for identifying communities in social graphs, Study, Szeged, 2002. (In Hungarian: Ismeretségi gráfok közösségeinek meghatározása gyors algoritmusokkal)
- [3] U. Raghavan, R. Albert, S. Kumura: Near linear time algorithm to detect community structures in large-scale networks, *Physical Review E* 76, 2007.
- [4] D. Kempe, J. Kleinberg,É. Tardos: Influential nodes in a diffusion model for social networks, *Proceedings of ICALP 2005*.