

Lecture 11. Community detection

In class we saw:

- A community is characterized by dense set of edges, denser than the neighborhood. That is, nodes in a community have more edges inside the community, few edges outside.
- Hardness of finding cliques and communities.
- Girvan-Newman algorithm from last class is a type of community detection.
- A definition of edge-density of a subgraph.
- Modularity as a measure of clustering.
- Correlation clustering.

Finding cliques and communities. Finding graph cliques (which are maximally dense subgraphs) is hard, therefore community detection is hard in the most general sense. However, this does not stop us from finding communities in more relaxed definitions. Which is what we do in the practical algorithms.

Subgraphs of highest edge density. We saw that finding the one most dense subgraph can be done in $O(n^2m)$ time. This is impractical in large graphs, and also finding just 1 community is not useful.

We saw an approximation algorithm for finding community containing a given set X . This is useful for example when we have knowledge of some members of a group but not all. For example, when we know a few friends belong to a club, but are looking for the whole club in the network.

Strong communities. The edge density notion does not enforce that a node should have fewer communities. For a definition that incorporates this idea, see *strong communities* in [1], Section 3.2.

References

- [1] David Kempe. Structure and dynamics of information in networks, lecture notes. Technical report, U.S.C, 2011.