Community detection and k-cores decomposition

You are given three small **networks with sociological partitioning** of nodes.

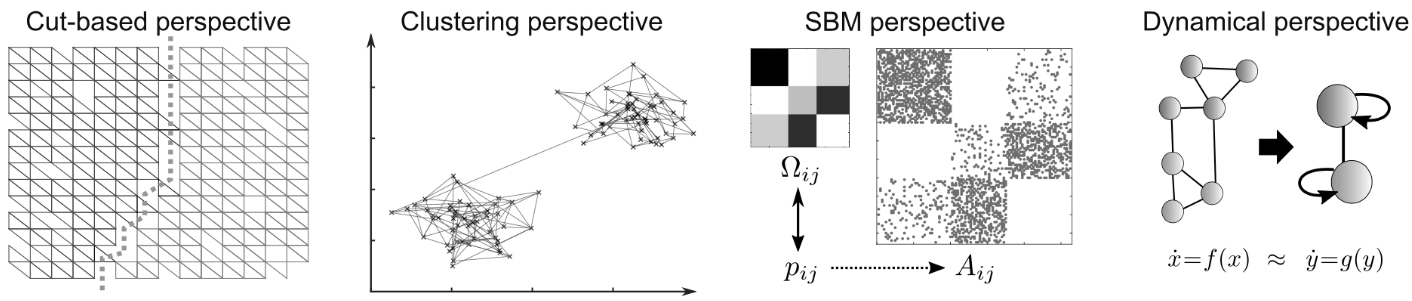
- Zachary’s karate club network (2 clusters)
- Davis’s southern women network (2 to 4 clusters)
- Lusseau’s bottlenose dolphins network (2 clusters)

Later you will be studying also selected larger **networks with node labels**.

- Game of Thrones character appearance network (characters)
- Human diseasome network by common symptoms (diseases)
- Conflicts and alliances between world nations (countries)
- Ingredients network by shared compounds (ingredients)

All networks are available in Pajek format.

Browse CDlib library for implementations of **community detection or graph partitioning** algorithms. Select an algorithm which you will be using for the exercises below. Some of the most popular algorithms are optimization of modularity known as the Leiden algorithm, map equation algorithm called Infomap, label propagation algorithm, hierarchical clustering based on edge betweenness, stochastic block models etc.



I. Small networks with known sociological partitioning

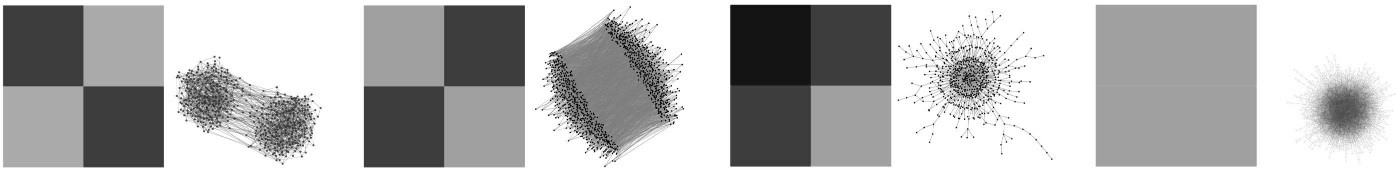
Apply the selected algorithm to **small social networks** above and test whether **communities coincide with sociological partitioning** of these networks. Ideally, you should apply the algorithm to each network multiple times and compare partitions with some standard measure like normalized mutual information or adjusted Rand index. Since these networks are very small, you can also visualize communities.

II. Larger networks with labels associated with nodes

Apply the selected algorithm to **larger real networks** above and test whether **communities provide good decomposition or abstraction** of the networks. Ideally, you should apply the algorithm to each network multiple times and examine the labels of the nodes in different communities. For simplicity, you may rather compute some standard quality function like modularity or examine only the nodes in the largest community.

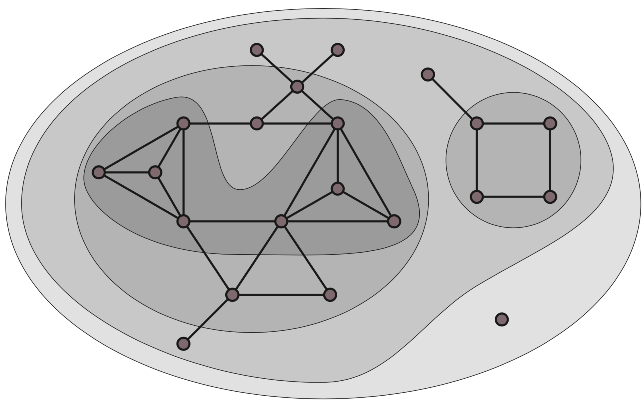
III. Random graphs with no mesoscopic structure

(tentative) Apply the selected algorithm also to **Erdös-Rényi random graphs** that should have **no community or other structure** and test whether the algorithm is able to detect this. You should apply the algorithm to random graphs with increasing average degree $\langle k \rangle$ and test for which values of $\langle k \rangle$ the algorithm doesn't reveal any structure.



IV. k-cores decomposition of real networks

1. Recall the following **algorithm for computing network k-cores** for a given k. Starting with the original network, iteratively remove nodes with degree less than k. When no such node remains, connected components of the resulting network are the k-cores.



2. *(tentative)* Implement the algorithm and compute all **k-cores of larger networks** above. For different values of k, print out the number of k-cores and the size of the largest one. Remember that k-core is always a subset of $k-1$-core. What is the maximum value of k denoted k_{max} for which there exists at least one k-core?

3. For each network, print out the **labels of nodes in k_{max}-core** and try to interpret the results.