Community Detection on Graphs with Multiple Layered Community Structures

Selim Karaoglu

The main objective of this project is the analysis of community structures based on graphs that has multiple layers of communities. These layers of community structures are constructed by combination of the graphs that are devised on the same set of nodes with different edges. Each graph has their own community structures, and these structures are the ground truth for this experiment. The graph structures are then merged by unifying the edges of the graphs. The merged graph now contains multiple layered community structures for the nodes that are assigned to a community in the original graphs. Then the merged graph is analyzed by using overlapping - disjoint and nested community detection methods to recover information about the original graphs from the merged graph.

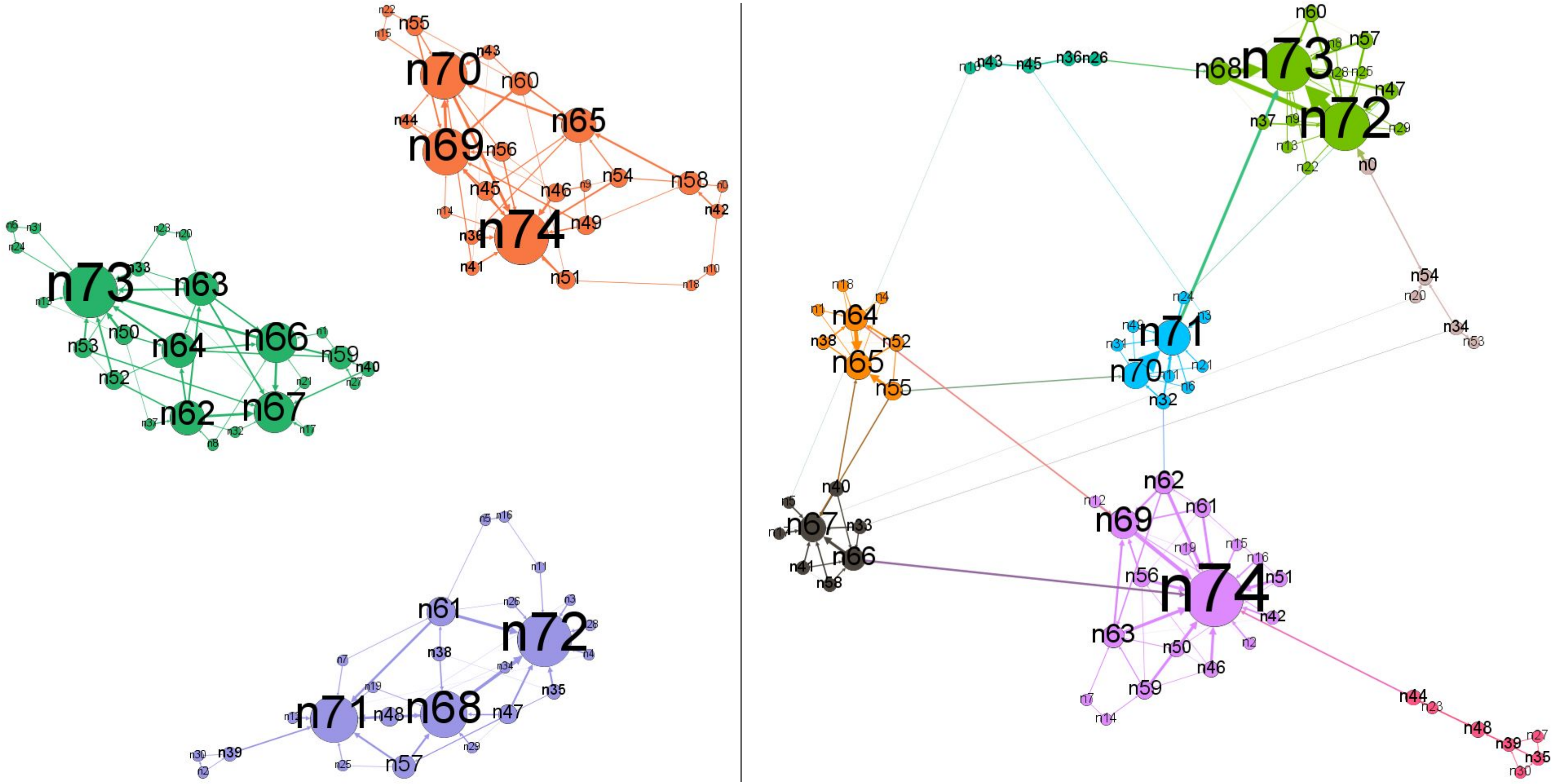
Background

- LFR Benchmark
- Igraph-R
- Python-igraph
- NetworkX
- Gephi

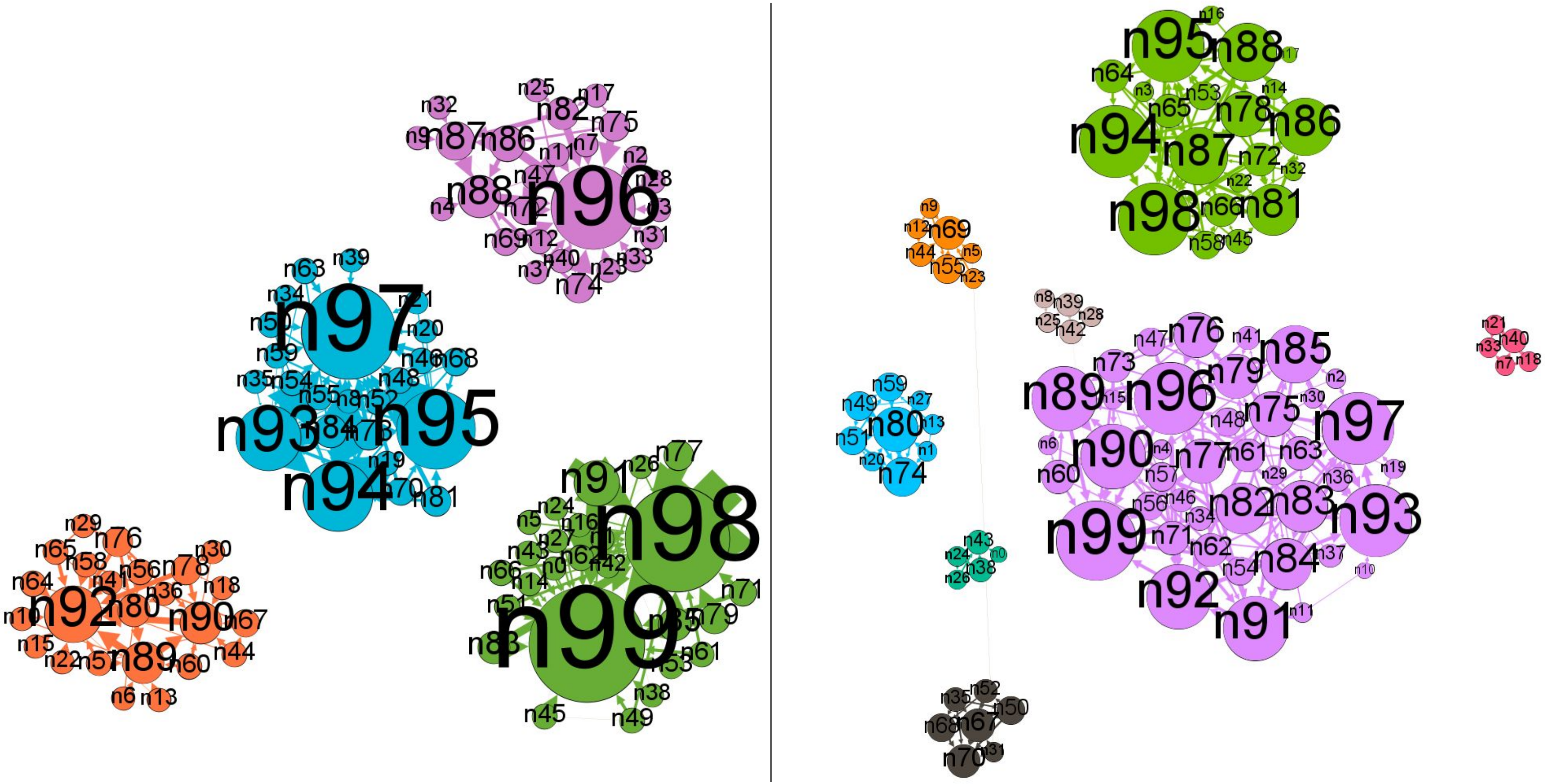
- Fast-Greedy
- Louvain
- Infomap
- Label Propagation
- Spinglass
- Walktrap
- Asynchronous Fluid
- Girvan-Newman
- Clique Percolation
- Link Communities

Methodology

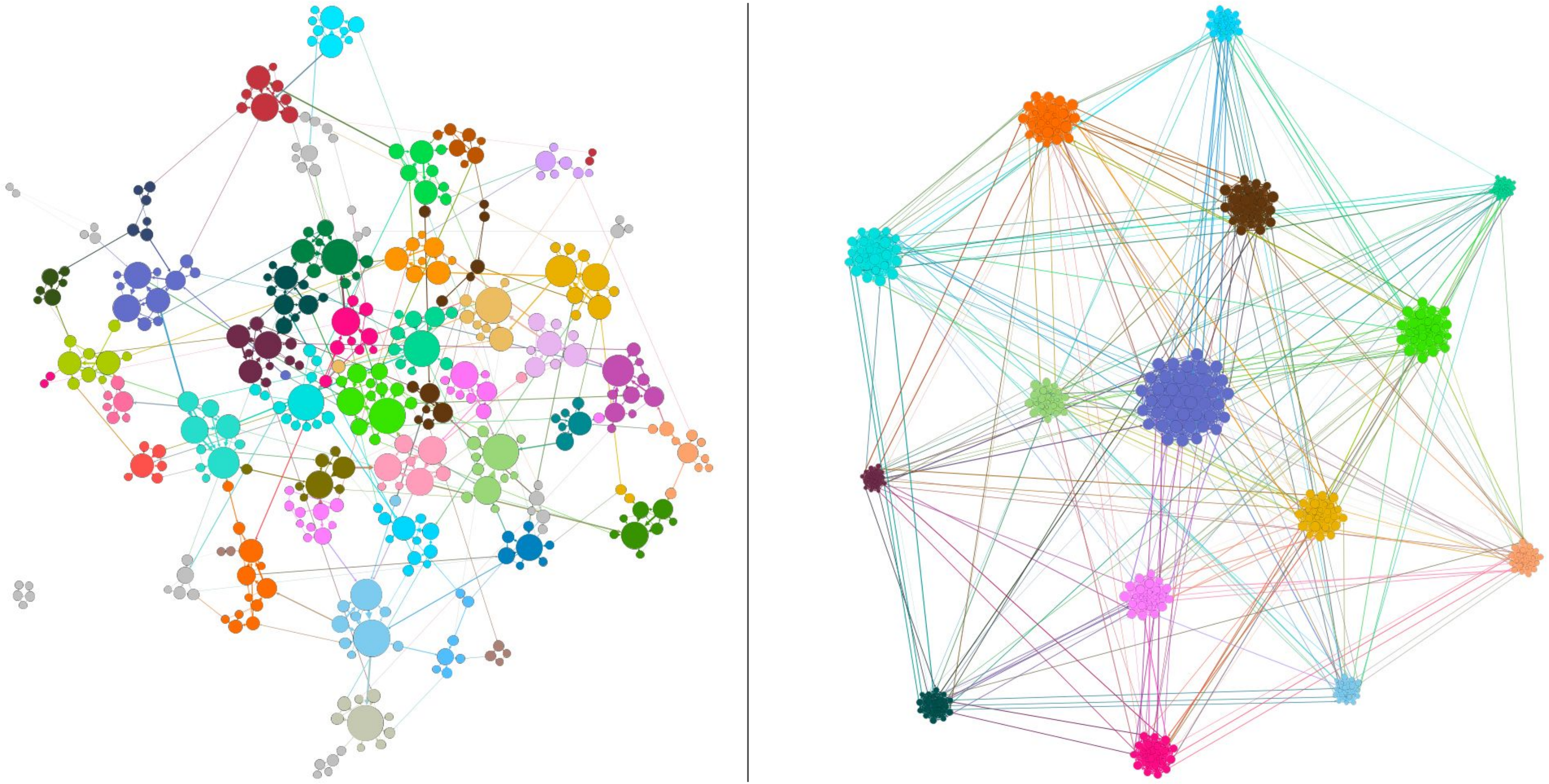
There are several combined graphs created and experimented in this project. All these combined graphs are based on two original graphs built on the same set of nodes. This set of nodes represent the small isolated group. This project’s focus is on small scale networks just like Zachary’s Karate Club and the graphs are all based on an isolated space with small number of nodes. Several different real network types are implemented to compare how different community structures affect the community detection methods' results.



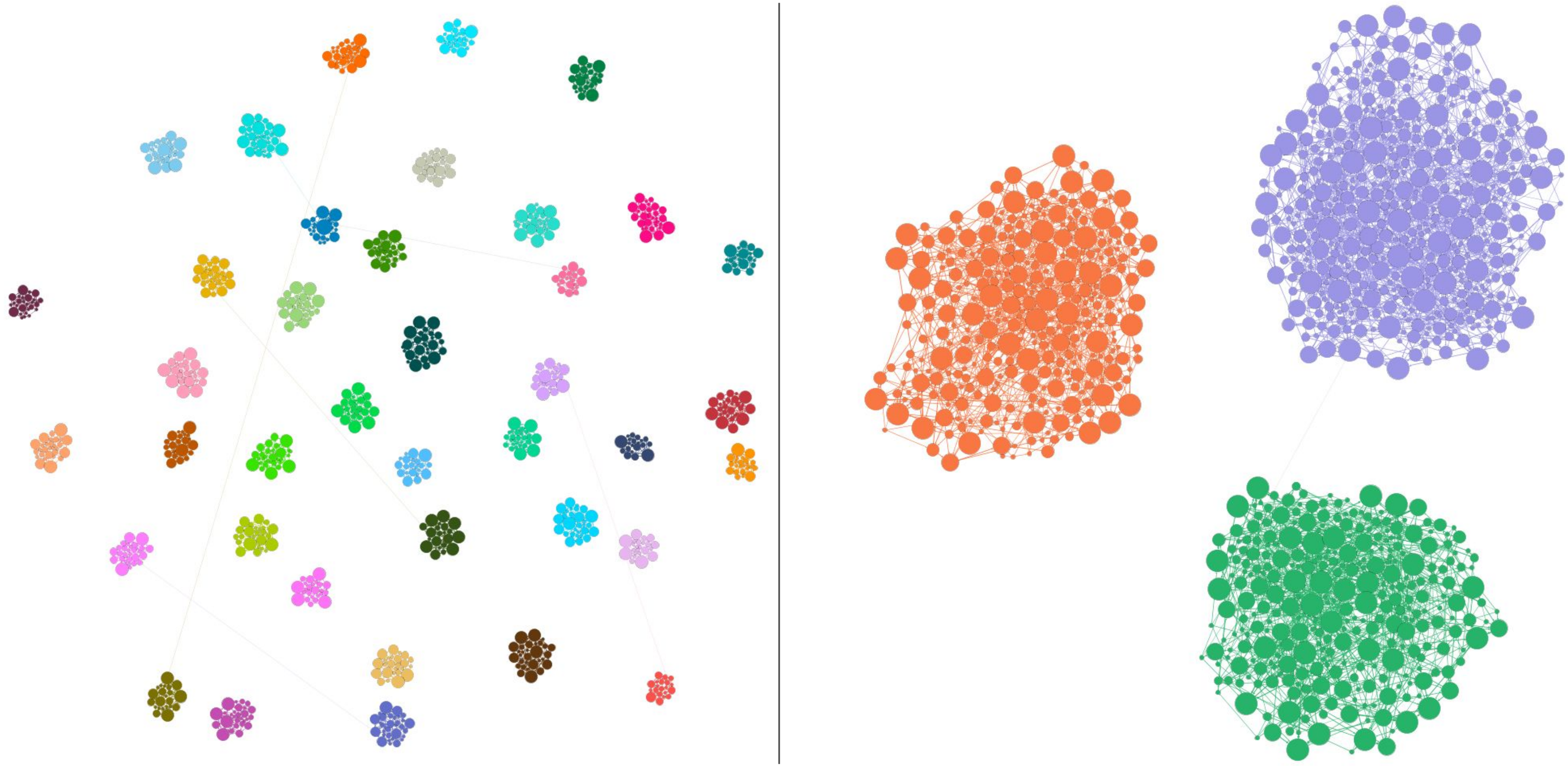
Test75 original graph community structures



Test100 original graph community structures



Test400 original graph community structures



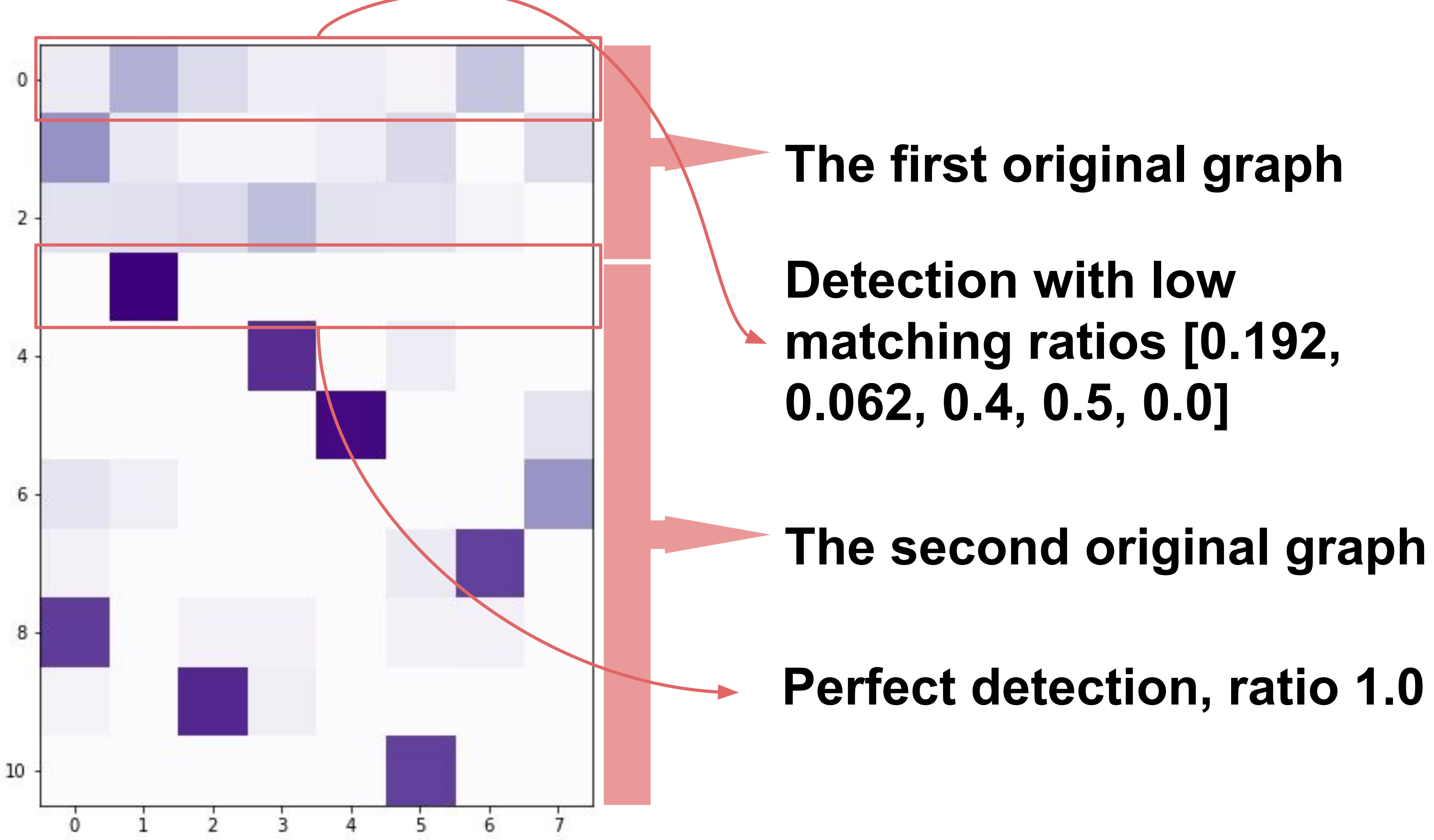
Test1000 original graph community structures

Experiment

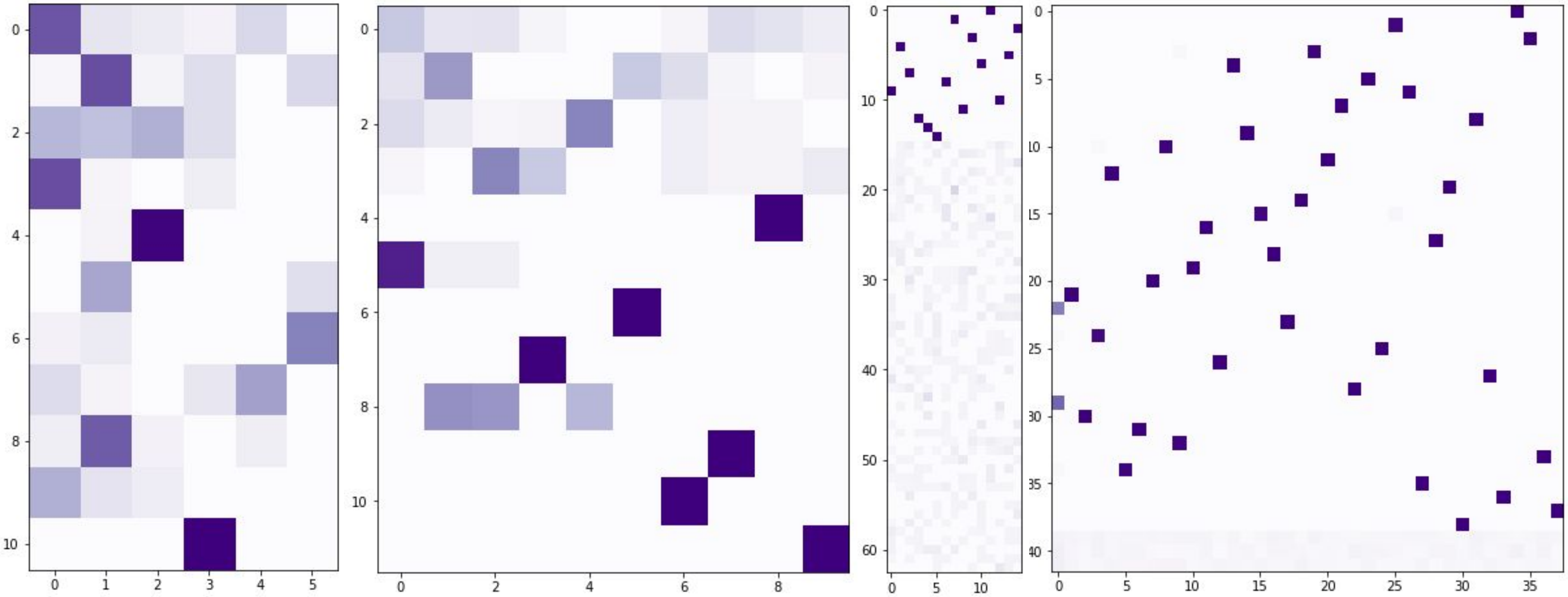
Each graph is inspected with the same disjoint, overlapping and nested community detection methods.

The results of community detection algorithms are then compared to the ground truth communities with sequential matching.

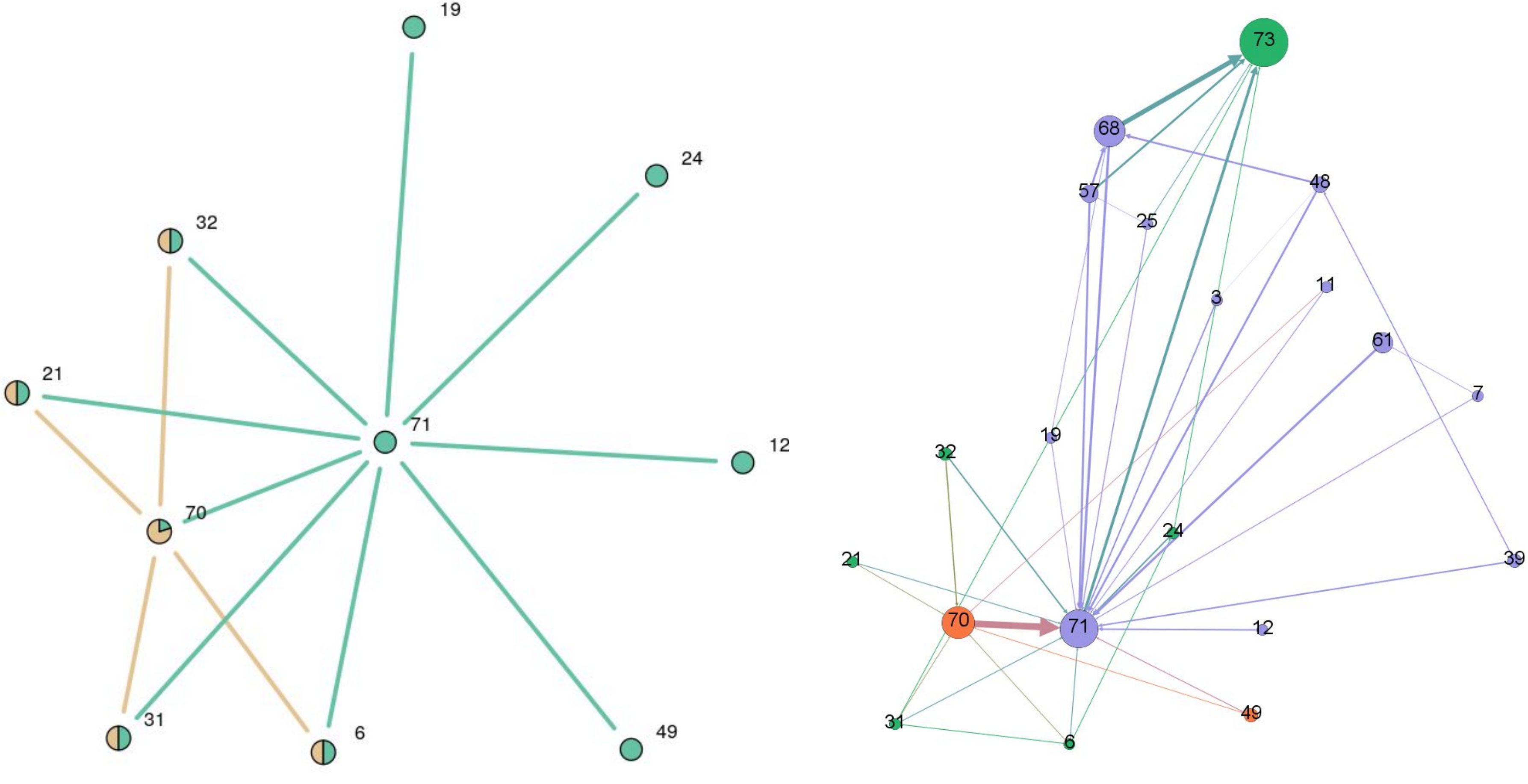
To compare the results with ground truth values, sequential matching ratios are employed.


The first original graph
Detection with low matching ratios [0.192, 0.062, 0.4, 0.5, 0.0]
The second original graph
Perfect detection, ratio 1.0

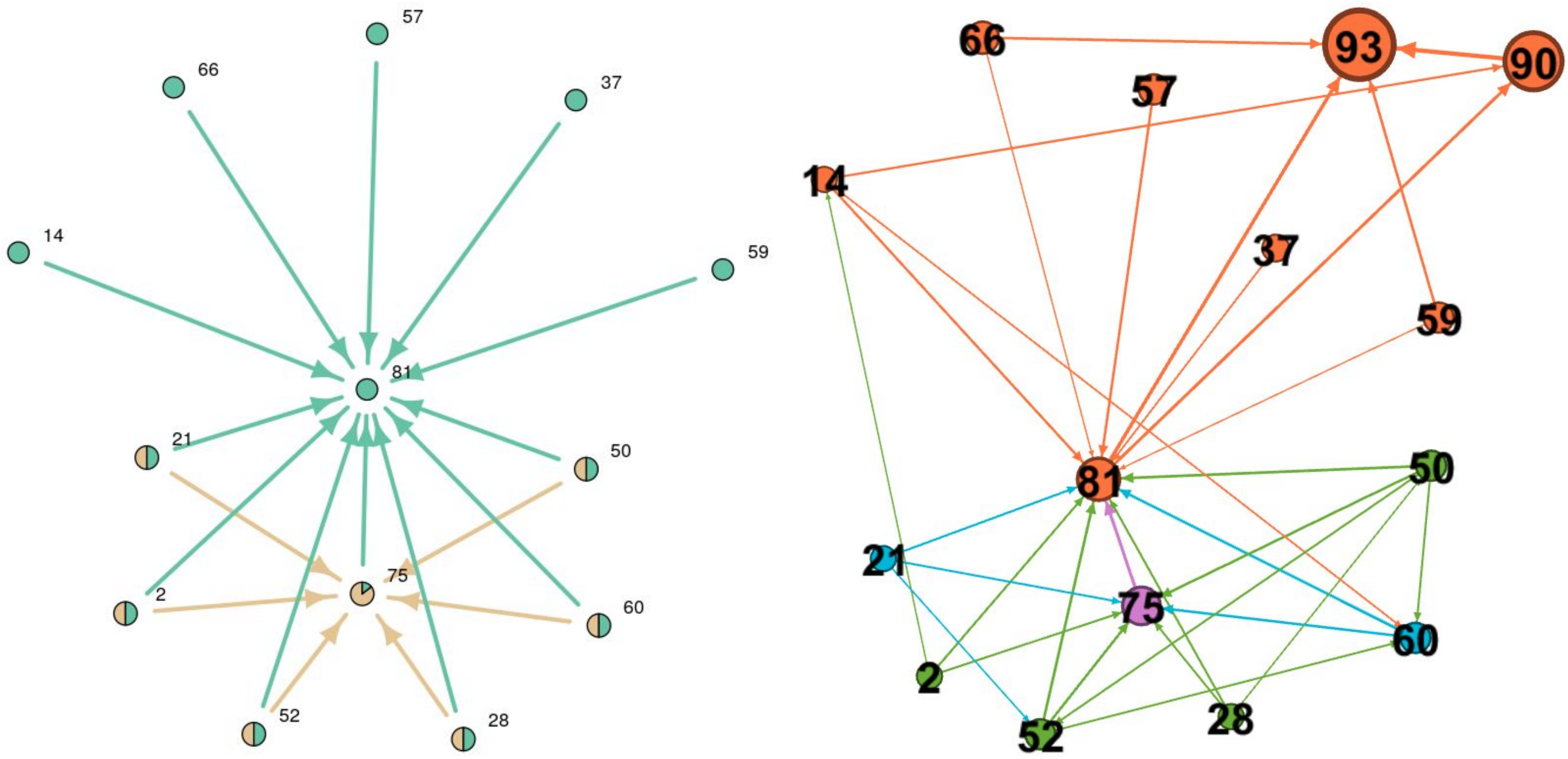
Results



Walktrap Algorithm heatmaps for each merged graph



Linkcomm detected nested community and ego network of node 71 on test75 graph



Linkcomm detected nested community and ego network of node 81 on test100 graph

Conclusion

Disjoint community detection algorithms are only able to recover one of the underlying structures. Spinglass algorithm partly discovers information from both structures partly.

Infomap algorithm has the highest matching ratios for all graphs except the graph with 1000 nodes. Walktrap method resulted the best on the last merged graph test1000.

Information from both structures at the same time can only be obtained by overlapping community detection methods. Clique percolation method mismatches some communities due to low connectivity. Linkcomm method is able to detect communities from both original graphs with high ratios, however this method suffers on graphs with high average degree or high connectivity.

Community detection algorithms; especially overlapping and nested community detection methods are expected to recover valuable information from graphs with multiple layered community structures.