Complex Networks

A2. Community Detection

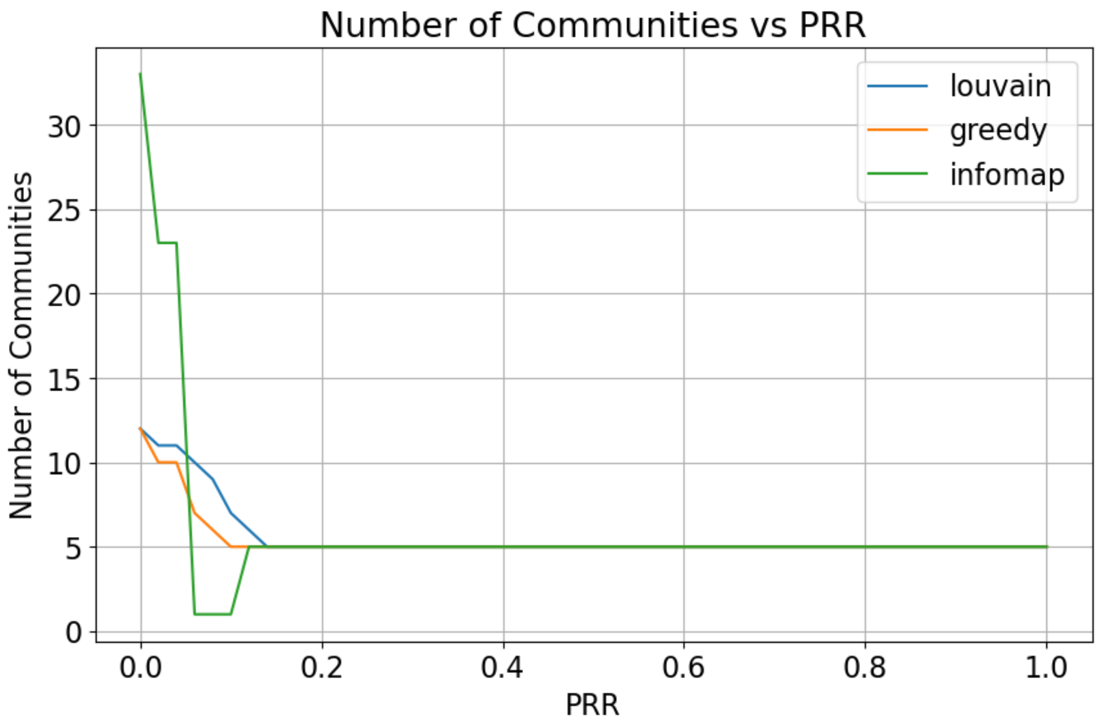
MESIIA 2024-2025
Suiwen Dong
Mark Safrónov

Part 1

We used three algorithms for detecting communities:

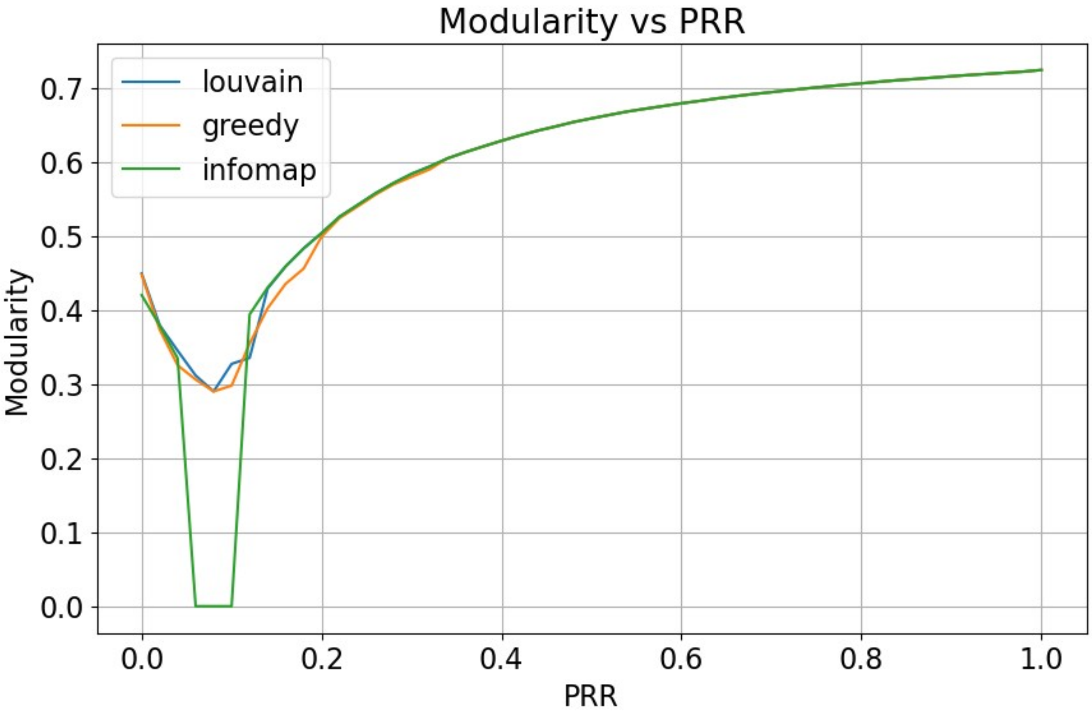
1. Louvain, from the `networkx` package
2. Greedy algorithm, from the `networkx` package
3. Infomap, from `infomap` package

The following is the plot of detected number of communities depending on the PRR, for all three algorithms.



We can see that all three algorithms converge as PRR reaches approximately 0.16.

In the same manner we plot the modularity for the communities detected by each algorithm. It should be noted that despite the number of communities reaching equilibrium at PRR approximately 0.16, modularity continues to grow. More than that, after convergence, it grows strictly following the logarithmic function behavior.



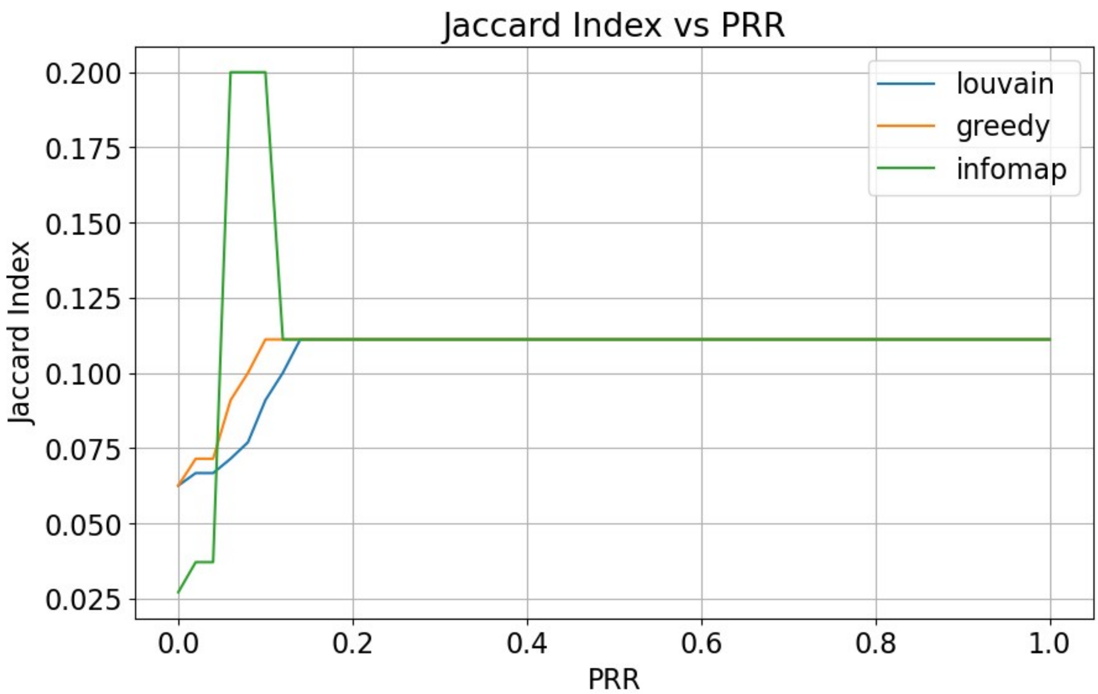
It worth mentioning that just the modularity alone is not enough to draw conclusions on the community structure, especially in our case in the segment with small PRR values.

For example, both PRR ~ 0.02 and PRR ~ 0.16 has modularity 0.4 as can be seen on the plot, for all three algorithms. However, Infomap for 0.02 detects more than 20 communities, and for 0.16 converges at 5 communities. These are wildly varying results for the actual community structure, but having the same modularity values.

The detected community structure has been compared with the true community structure using the following three metrics:

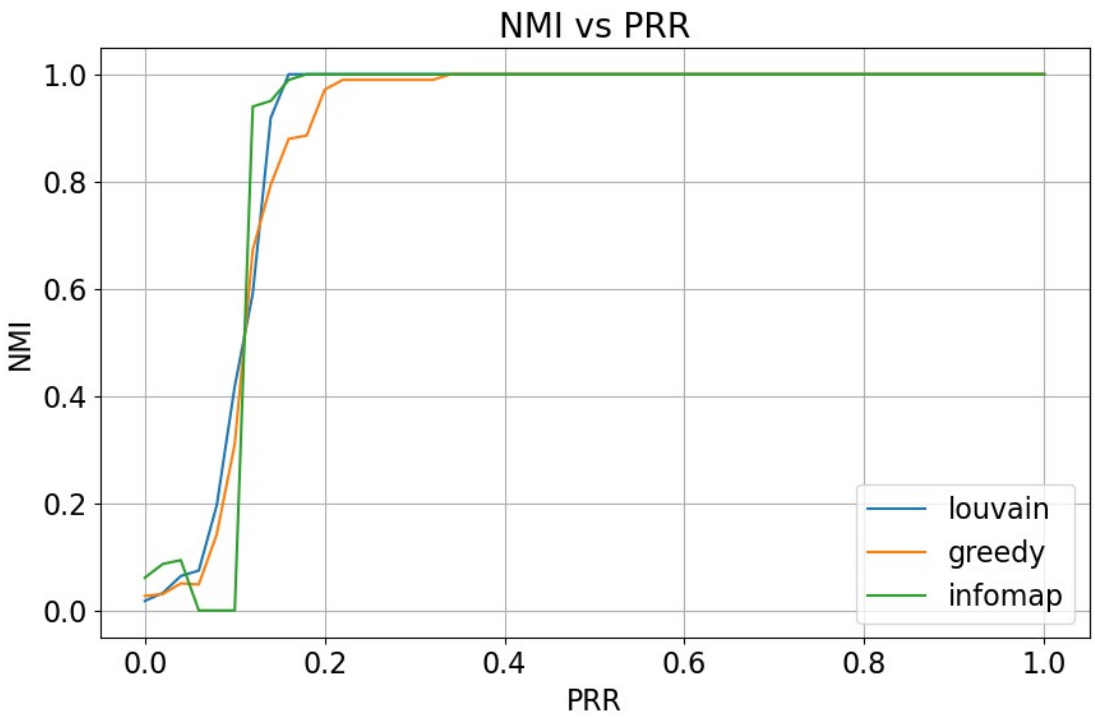
1. Jaccard Index, manual implementation
2. Normalized Mutual Information, from the scikit-learn project
3. Normalized Variation of Information, manual implementation

For all three community detection algorithms, we get the following Jaccard Index behavior.

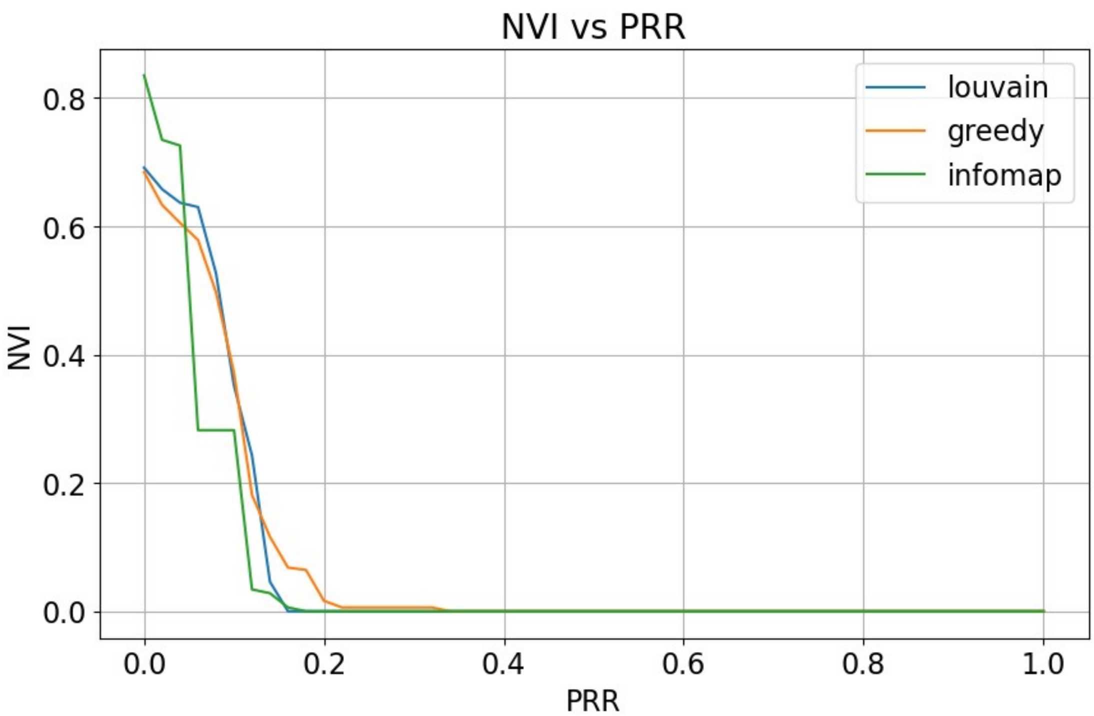


It again, converges for all three algorithms at around PRR 0.16 despite all three algorithms behaving very differently before that.

For Normalized Mutual Information metric we can see that greedy algorithm converges later than others, at around PRR = 0.35



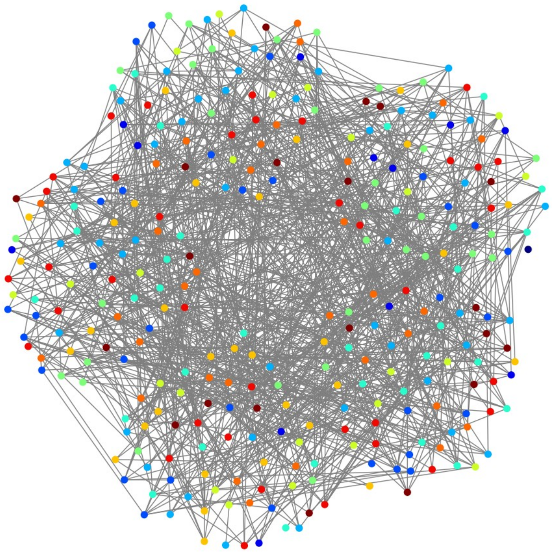
And the same behavior can be seen on the plot of the Normalized Variation of Information: Louvain and Infomap converge at around PRR=0.16, but greedy algorithm gets to zero at around 0.35.



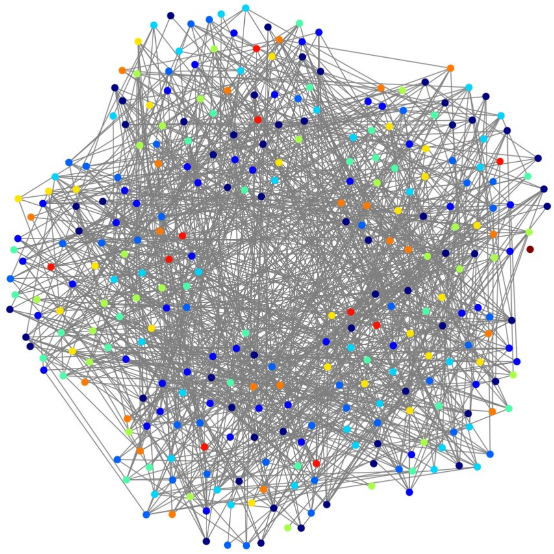
To render the network, first the positions of the nodes were obtained using the network with PRR = 1.00 and the Kamada-Kawai algorithm. Then, the color coding was derived from the detected communities for each algorithm in question.

For PRR = 0.02 the assignment looks chaotic for all three algorithms.

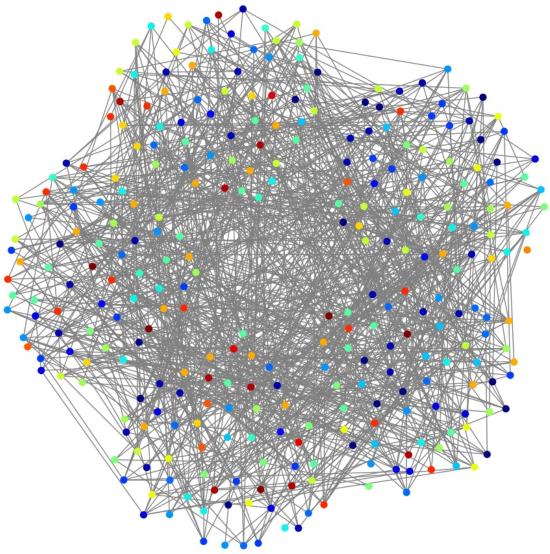
Louvain, prr=0.02



Greedy, prr=0.02

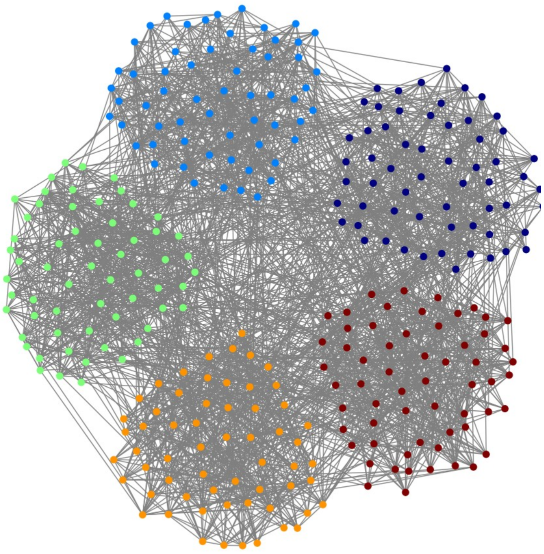


Infomap, prr=0.02

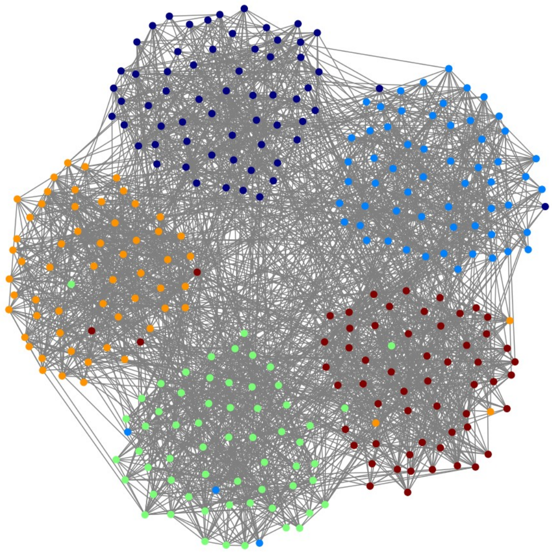


For PRR = 0.16 the assignment is almost correct, with small errors in some clusters. Greedy algorithm has more errors than the other two algorithms. Infomap has just a single misplaced node. Louvain is perfectly correct.

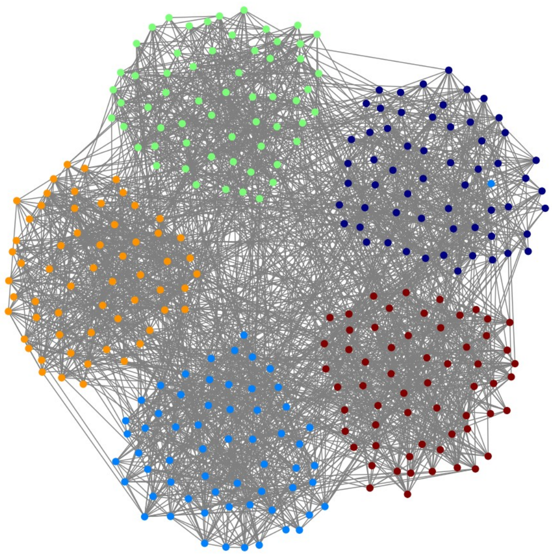
Louvain, prr=0.16



Greedy, prr=0.16

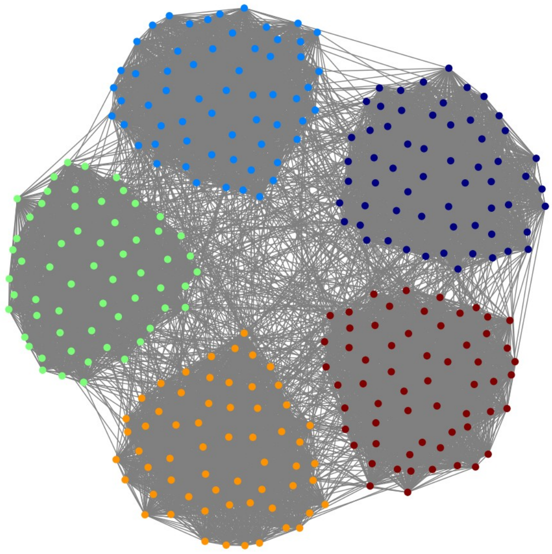


Infomap, prr=0.16

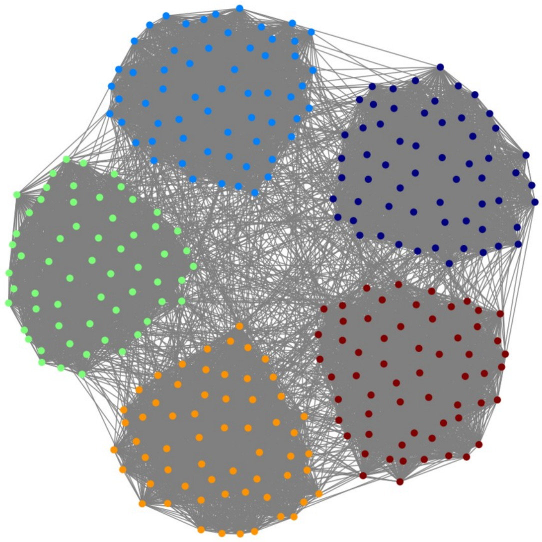


And with PRR = 1.00 the assignment is perfectly correct for all three algorithms.

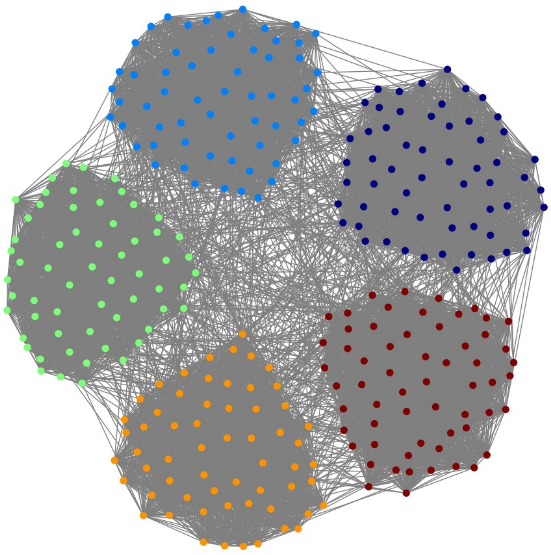
Louvain, prr=1.0



Greedy, prr=1.0



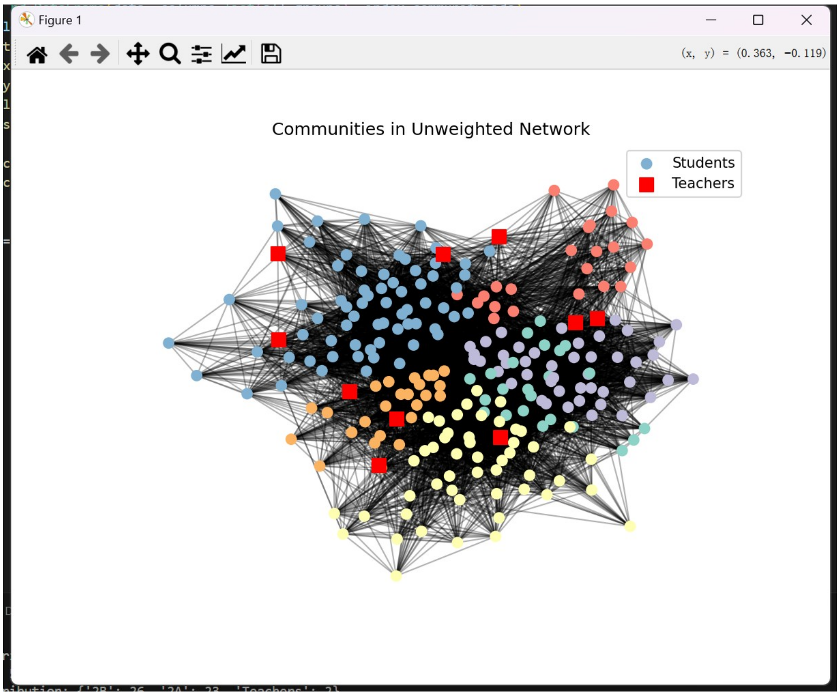
Infomap, prr=1.0



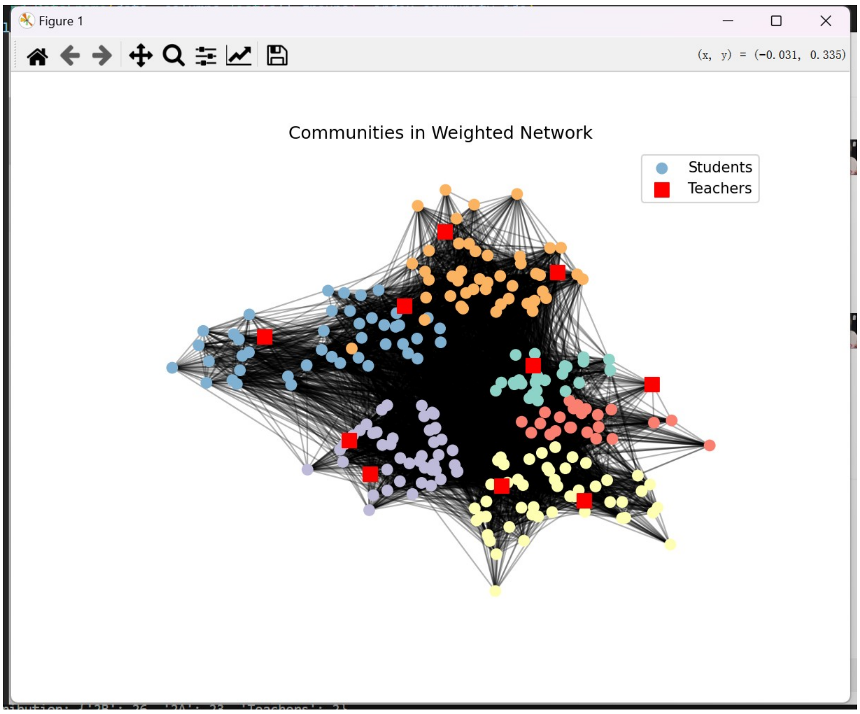
Part 2

The output image of the community structure of the given two networks, the dots are the students, the squares are the teachers.

U.net



W.net



Analysis of the Distribution of Unweighted Network Communities

Differences in Community Sizes: There are certain differences in the sizes of various communities in the unweighted network. The smallest communities (such as Community 0 and Community 3) have a size of 24, and the largest community (such as Community 4) has a size of 70. This indicates that in the communities divided based on simple connection relationships, the distribution of the number of nodes is uneven.

Characteristics of the Composition of School Groups

- Communities Dominated by a Single Class: For example, Community 0 is mainly composed of 23 students from Class 1A and 1 teacher, Community 3 is mainly composed of 23 students from Class 4B and 1 teacher, and Community 5 is mainly composed of 25 students from Class 1B and 1 teacher. This shows that under the unweighted network, some communities present a composition pattern with students from a single class as the main body, accompanied by a small number of teachers.

- Multi-class Mixed Communities: Community 1 includes students from Class 2B and Class 2A as well as 2 teachers; Community 2 includes students from Class 3A and Class 3B as well as 2 teachers; Community 4 includes students from Class 5B, Class 5A, and Class 4A as well as 3 teachers. This reflects that some communities are composed of students from multiple classes, together with a certain number of teachers.

Analysis of the Distribution of Weighted Network Communities

Differences in Community Sizes: There are also differences in the sizes of communities in the weighted network. The smallest community (such as Community 3) has a size of 24, and the largest community (such as Community 5) has a size of 49. Compared with the unweighted network, although the size range is similar, the size of each specific community has changed. For example, Community 4, which had the largest size (70) in the unweighted network, has a size of 45 in the weighted network.

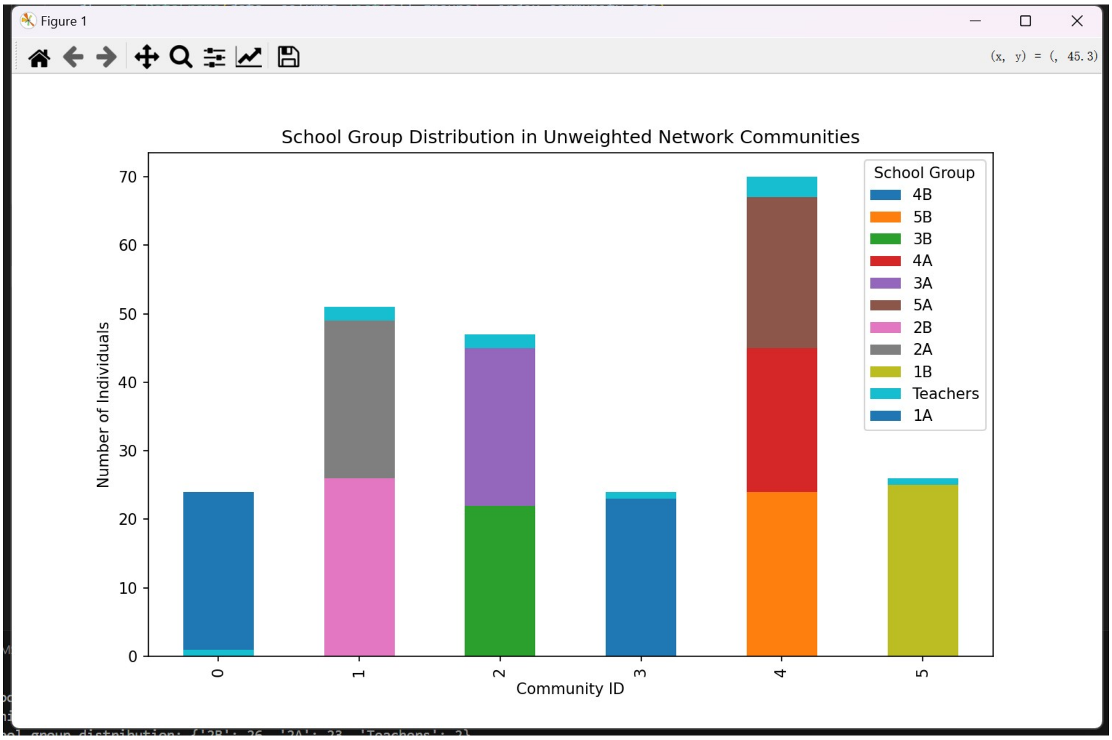
Characteristics of the Composition of School Groups

- Changes in Class Composition: Compared with the unweighted network, the class composition of some communities has changed. For example, Community 0 was mainly composed of Class 1A in the unweighted network, but it has become mainly composed of Class 1B in the weighted network; Community 3 was mainly composed of Class 4B in the unweighted network, but it has become mainly composed of Class 1A in the weighted network. This reflects the influence of the weight factor on community division, causing nodes to be re-divided into different communities due to considerations of connection strength, and thus changing the class composition within the community.

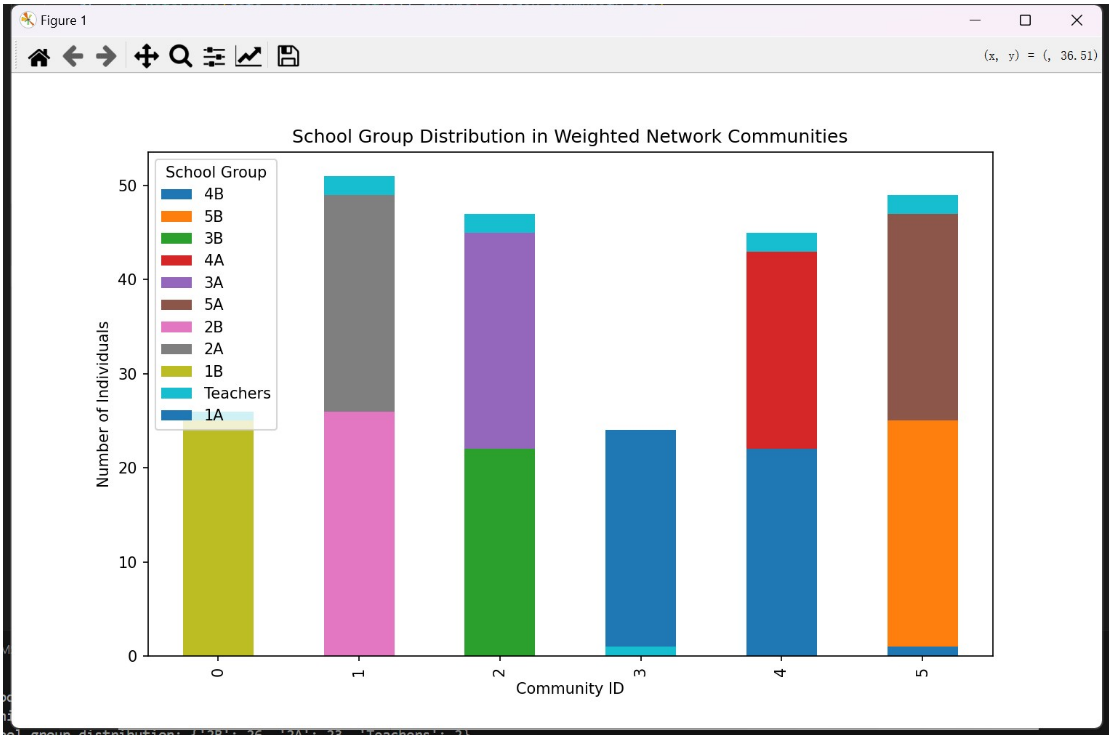
- Adjustments of Mixed Communities: In multi-class mixed communities, the weighted network has also adjusted the proportion of the number of students in each class. For example, in Community 5 of the weighted network, in addition to the students and teachers from Class 5B and Class 5A, a student from Class 4B has been newly added. This further illustrates that the weight affects the composition structure of different school groups within the community.

The school group distribution for the given networks looks as follows.

U.net



W.net



Summary of the Comparison between Unweighted and Weighted Networks

- Stability of Community Structure: For some communities (such as Community 1, 2, and 3), the sizes and the compositions of school groups are relatively similar in both the unweighted and weighted networks. This indicates that these communities are relatively stable under the two network analysis methods, which may mean that the connection relationships between nodes within these communities, whether they are simple connections or connections considering weights, present a relatively consistent aggregation pattern.

- Influence of Weights: From the differences of multiple communities under the two networks, it can be seen that the weight factor has a significant influence on community division and the composition of school groups. It changes the attribution of nodes, and thus changes the overall structure and internal composition of the community. This shows that when studying such networks, considering weights can uncover information about the community structure based on the actual connection strength, complementing the analysis results of the unweighted network that only considers the presence or absence of connections, and providing a basis for a more comprehensive understanding of the relationships between individuals in the network.

A brief discussion on the differences among the communities detected in the weighted and unweighted network. Why weights are relevant?

In an unweighted network, each edge is regarded as the same, while a weighted network takes into account the strength of the connections. Weights can highlight the actual frequency and closeness of interactions, making the division of communities more accurate. They can filter out weak and noisy connections, thus more truthfully reflecting the relationships within groups.

All the code which has been used for generating the figures is located in the GitHub repository at https://github.com/Suiwen-D/CN-A2