Structure and Dynamics of Complex Networks (2024/25 Winter)
Homework 2

Mihaly Hanics

This report is about exploring the assortativity and rich cores of a painter network, with various measures. The report is written for the "Structure and Dynamics of Complex Networks" course at the Central European University.

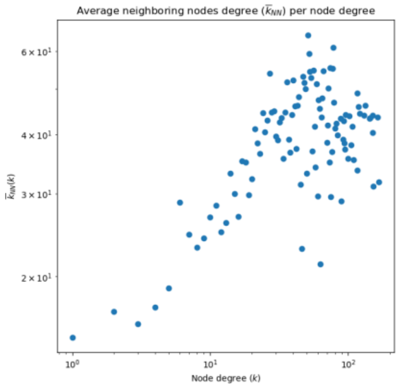
Data

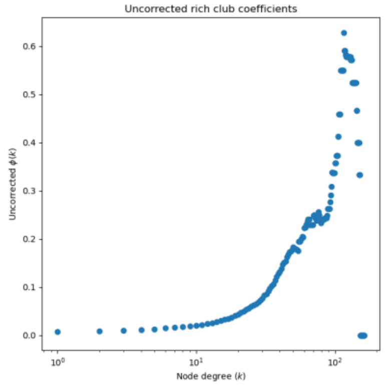
The network explored is a painter network based on locations of painters throughout times, collected from the PainterPalette dataset. Painters are connected if (based on the original dataset) they lived 5 years at the same location; periods are approximated. This is a network with extensive temporal information, with artists dating back to the middle ages all the way to the 20th century. Therefore, the network is longitudinal and connections only span to short range (painters in the same lifetime), making the network's diameter rather large (26) despite its size (2503 nodes, 20267 edges). The network is undirected. Weights equal the strength of the connection based on how long the pair of artists lived at common locations, and based on nationality. We use the connection strengths (weights) here for the second half of the analysis. The network is available here.

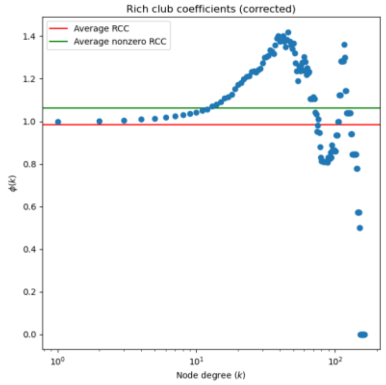
Assortativity and rich club coefficients

Social networks typically have positive assortativity values (they are assortative), therefore in our analysis we expect a positive Newman's r, and also a gradual increase in $< k_{nn,i} >$ for increasing degree i. Values for measures:

Newman's *r*: 0.1385
Average degree of neighbouring nodes (k_{nn}): 28.2570
Average rich club coefficient: 0.9834
Average of nonzero corrected rich club coefficients: 1.0610



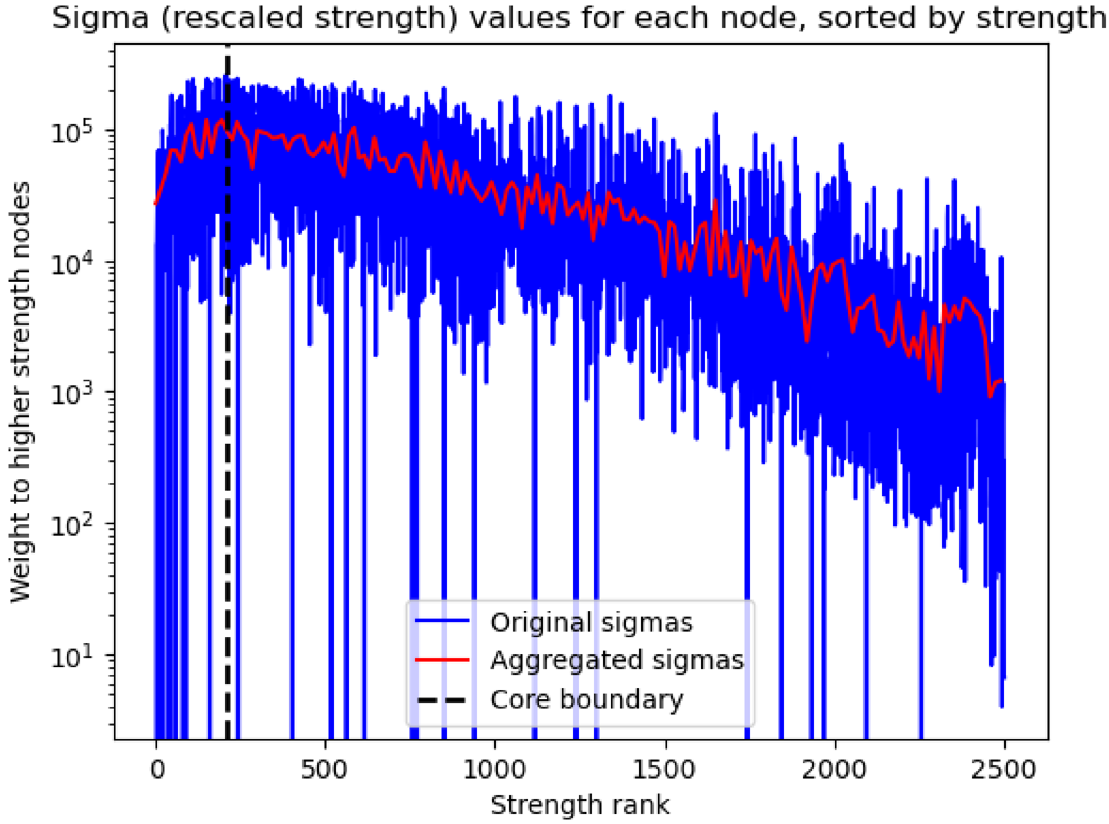




From both the k_{nn} and the corrected rich club coefficients plots we can see that there is assortativity and rich-club behavior in the network by the increase in values (till some point), but at the highest values there is a sharp dropoff for the highest values. Alongside the fact that for the highest values, this A possible explanation behind highest degree nodes being unconnected (disassortative) is the historical nature of the network – the included artists span 11 centuries, and likely the most connected painters are the most connected in their time, but across centuries these painters cannot be connected. I imagine the network as 11 power-law distributed networks cascaded after eachother, each representing a respective century, and the few hubs in each network are not interconnected because they lived in different centuries (mostly). Nonetheless, there is assortativity and our hypotheses seem to hold.

Rich core (Ma, Mondragon 2014 PLoS One)

This algorithm finds the rich core by sorting nodes by strength (weighted by the edge weights, or degree) and calculates (rescaled) weights to nodes with higher strength (typically, with higher degree, but as I work with a weighted network, I use node strengths instead). The algorithm selects the nodes with strength above a boundary value as part of the core.



The ρ coreness (as defined by Borgatti and Everett) for this partition is 2095 with 209 nodes, in comparison to a partition configuration model (we randomly shuffle which nodes are in the core and which ones are in the periphery, keeping sizes) the score is 14.5 times higher than to a randomized configuration model.

Core-periphery by the Kojaku-Masuda configuration algorithm

This algorithm on our “time-longitudinal” network selected 1696 nodes as core, twice as many nodes as core than as periphery. Likely the longitudinality enables more nodes to be core nodes, as there is no clear “shell” of the network (the network is far from being “radial”). The choice behind the algorithm is that it takes edge weights into account, unlike e.g. the Borgatti-Everett algorithm.

The ρ coreness is 13491, which is 1.49 times higher than a configuration with the same core and periphery sizes.

Metrics on the two core-periphery partitions: Rand index and Jaccard index

The Rand index for two partitions (X,Y) is calculated as:

$$\text{Rand(X,Y)} = \frac{a_{00} + a_{11}}{a_{00} + a_{11} + a_{01} + a_{10}}$$

where:

- a_{00}: number of pairs of nodes where the two nodes are in **different communities** (core/periphery in our case) **in both** X and Y

- a_{11}: number of pairs of nodes where the two nodes are in the **same community in both** X and Y

- a_{10}: number of pairs of nodes where the two nodes are in the **same community in X but in different communities in Y**

- a_{01}: number of pairs of nodes where the two nodes are in **different communities in X but in the same community in Y**

The Jaccard index for two partitions (X,Y) is modified to account for many pair of nodes being in different communities in any partition, for the reason that a_{00} can be too large in certain settings:

$$\text{Jaccard(X,Y)} = \frac{a_{11}}{a_{11} + a_{01} + a_{10}}$$

Since we only have two communities, the core and the partition, and in both partitions the difference in core-periphery sizes is so large that pairs of nodes in different communities are relatively rare (e.g. in the first partition, there are 209·2294 pairs of nodes in different communities, but there are 2294·2293/2 pairs of nodes in periphery), a_{00} must be relatively low, and the difference between the two indices are not as large as with many communities.

Rand index for rich core and KM-config partitions: 0.522

Jaccard index for rich core and KM-config partitions: 0.494

We can see that there is not much difference in the values for the two indices, which indicates that a_{00} is indeed not large. As the two algorithms interpret "being core" differently (as seen by the large difference in core sizes), the index values are not as high as possibly expected. In comparison, while the "divisive" algorithm (also defined by Kojaku and Masuda) has lower "corrected" coreness score, it yields a 78.9% Jaccard similarity score with the rich core, meaning a more similar partition.

Conclusion

The network shows signs of assortativity, as expected by a social network, but for high degrees there seems to be no high inter-connectedness, which may be partially because of the longitudinal nature of the network. Testing different core-periphery partitions, different models typically yield quite different partitions, as shown by the Rand and Jaccard measures, and in these cases the two measures do not differ by much.