INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, THIRUVANANTHAPURAM

Mathematical Foundations to Data Science (DSC 212)

Submitted by: - Anay Mishra, IMS24274, BS - MS 2024

Instructor: - Dr. Saptarshi Bej

Refer to following files here *https://github.com/astroanay/DSC-212-Assignment---Karate-Club-Analysis-by-Anay-Mishra-IMS24274.git* :

1) Assignment Question: DSC212__Assignment.pdf
2) Code and it's working: DSC212_Assignemnt_GraphTheory_AnayMishra_IMS24274.ipynb

Overview of this document:

It provides a short explanation on which nodes consistently remain central across splits, and how community structure influences these metrics.

Which nodes consistently remain central across splits?

- Node 33 remained strongly central across multiple iterations. It has the highest mean degree and a very high betweenness score. In the context to Zachary's Karate Club ground truth, node 33 is a hub for different factions. Node 33 acts as a central connector in the cluster on the right of the layout. Its high degree makes it persistently central.
- Node 0 is second - highest mean degree and top betweenness. Node 0 sits near the centre of the other faction and connects many nodes which makes it both a hub (degree) and a broker (betweenness).
- Node 32 is also a consistently important node (high degree and betweenness). It sits in the dense right-hand community, adjacent to node 33, and participates in many shortest paths.
- Nodes 2 and 1 have moderate-to-high degree and betweenness that's why they are in the denser or bridge region between the two sides and remain important across the split.

Why does these nodes remain central throughout each iteration?

Degree centrality is local which means nodes with many direct friendships (means high degree) remain high in degree regardless of how the network is split, since the underlying graph does not change.

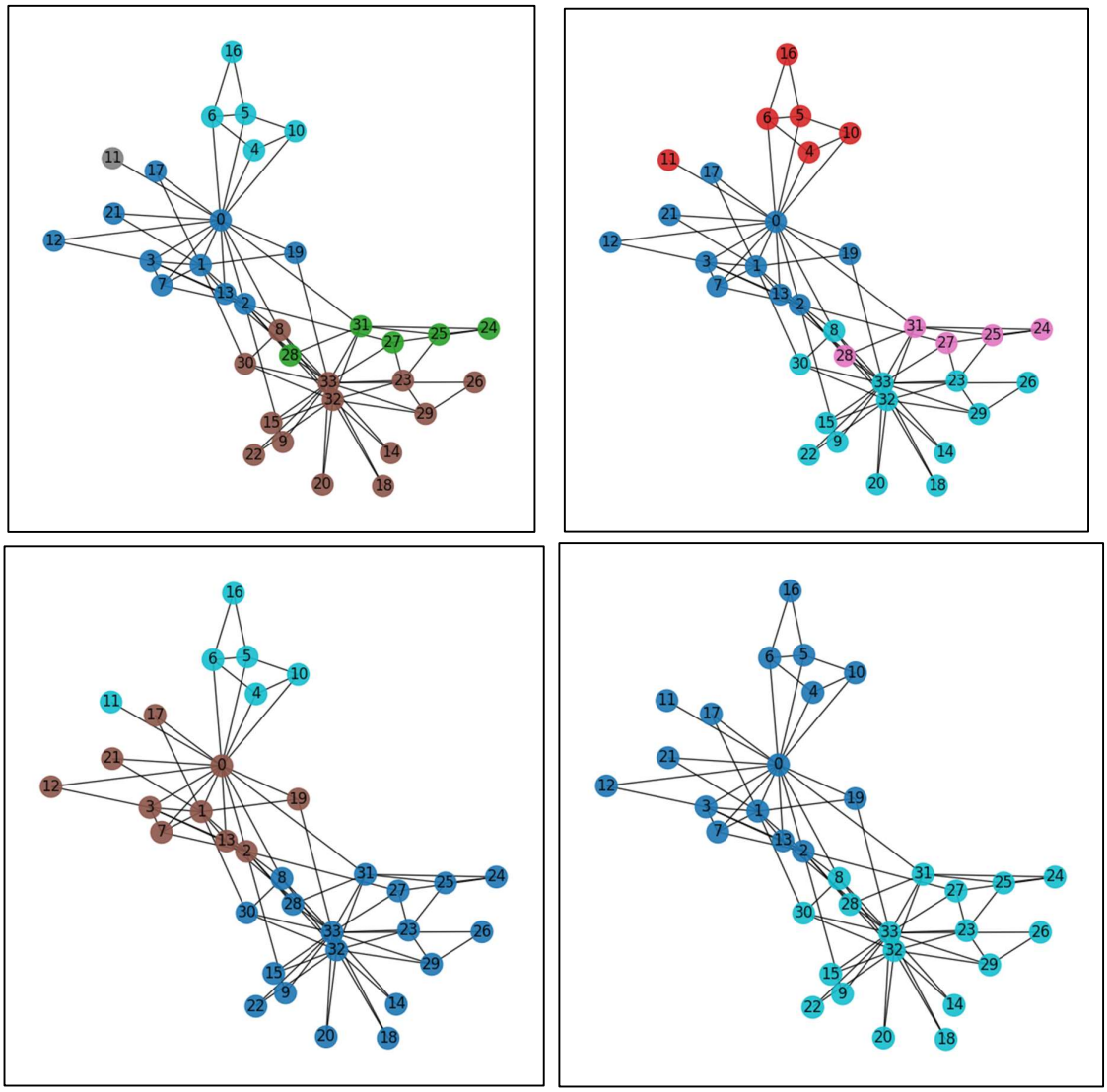
Nodes that sit between communities or connect tightly knitted groups see high betweenness. If the split separates the network into two cohesive blocks, the nodes that connect those blocks keep their high betweenness (or even increase relatively). Closeness centrality favours nodes that are overall near others in path-length terms. In a small graph like the Karate Club, the hubs (0, 33) have high closeness as well because shortest-path distances from them to others are low. Clustering coefficient helps reflecting local closeness (means how many of a node's neighbours connect to each other). High-degree hubs often have lower clustering (because their many neighbours are not all connected), while nodes inside tight groups show higher clustering.

How community structure influences these metrics?

- After the first split, nodes partition into two communities roughly matching the known factions. This structural change does not change node degrees (edges are unchanged), so degree centrality is stable across iterations. That's why the degree-evolution plot is flat, the numbers are identical across iterations.
- Betweenness can change if a node's position relative to community boundaries changes. If a node becomes internal to a community (means that the split assigns many of its neighbours to the same side), shortest paths that previously traversed it to reach other parts of the network may shift, but in the Karate Club split we ran, the main bridges (0 and 33 and neighbours) remain in bridging positions so betweenness remains high for them.
- Closeness is relatively stable in this small example because path-lengths across the full graph don't change. However, local rearrangements in community structure can alter the average distance to nodes in other communities, slightly affecting closeness.
- Nodes inside dense pockets show high clustering. When a split isolates such a pocket as a community, its clustering remains high. Hubs bridging communities typically have low clustering because their neighbours are in distinct groups and are less likely to link with each other.

Conclusion

Degree centrality is the most stable metric across splits (because edges do not change), while betweenness/closeness/clustering can be more sensitive to how splits isolate or group nodes. In this dataset, nodes 0 and 33 (and their close neighbours such as 32) are clear stable leaders/hubs, they are consistently top-ranked by degree and betweenness which is an expected outcome given their roles in the real club.



Figure(above): Visualizing graph splits across 4 iterations (from below right to above left)

Figure(below) : Heatmaps showing variation in different parameter used while analysing the splits