Detecting communities of directed networks via a local algorithm

Team: Saishree Godbole, Henish Shah, Tanvi Kolhatkar

ABSTRACT

In our project, we focus on local community detection in directed networks. We have implemented the local community detection algorithm for directed networks by Shigang Liu as published in his paper and applied it to different directed graphs. The fundamental idea of the algorithm is to find an initial community from a node having maximal node strength and expand the community by adding neighbours of this node. We have studied the behaviour of this algorithm on the above-mentioned networks and documented our results in this paper.

KEYWORDS

Directed network, Community detection, Node strength, Zachary Club

INTRODUCTION

Motivation

Detecting communities is a fundamental and relevant problem in the field of network science having multiple applications. This field has garnered a lot of interest in recent years with new algorithms emerging which can be used to detect global community structure. Local communities provide a better measure to understand and visualize the nature of interaction when the global knowledge of networks is unknown.

One application of the local community would be in the field of marketing :

- Nowadays, social network partitioning has become a very important function. One objective for partitioning is to identify interested communities to target for marketing and advertising activities.
- The bottleneck to the detection of these communities is the large scalability of the social network. Previous methods did not effectively address the problem because they considered the overall network.
- Social networks have a strong locality, so designing a local algorithm to find an interested community to address this objective is necessary.

Literature Review

In the past, researchers have proposed many different algorithms for various group structures, in order to divide the communities:

- Eigenvalues spectral bisection algorithm of Laplacian matrix based on graphs[5]
- K-L algorithms based on the greedy algorithm[6].
- In recent years, many algorithms for dividing communities have been invented based on modularity[7].
- In 2004, Newman proposed a fast algorithm for dividing communities[8]

The biggest advantage of the proposed algorithm is that there is no requirement for the whole network structure and the number of communities to be available at the start. Only the local information is enough for this approach. Therefore, this method is easy to use with low complexity and can be used to find the community structure of the whole network.

PROPOSAL

We are replicating the findings of the paper 'Detecting communities of directed networks via a local algorithm' by Shigang Liu. Based on the principle of node strength, the division of the local community structure algorithm is proposed in directed and overlapping networks. We aim to apply this algorithm on the Zachary Club network, a Power Grid network and a Facebook social circles network and observe how it behaves in networks of varying structures and densities.

METHODS

The algorithm is carried out in three major steps:

1. Finding the initial community with its central node having maximal node strength.
2. Expanding the community
3. Finding the nodes that belong to multiple communities

- Finding the initial community
    1. With every node $v \in V(D)$, calculate the in-degree $d_{in}(v)$
    2. Choose $\max\{d_{in}(v), v \in V(D)\}$ as vertex v and its neighbor N(v), these vertices form a community c
    3. Given vertex $u \in c$, if $B(u, c) \leq B^C$ (we choose $B^C = 0.5$), delete vertex u from community c
    4. Repeat step (3) until for every $u \in c$, $B(u, c) > B^C$ is true. And we obtain a new initial community, which we write as c.

Thus, our initial community 'c' is formed by adding the selected node and some of its neighbours. For each of the nodes within our community 'c', we calculate 'B', which is the measure of the compactness of the node within the community. If the value 'B' is less than the empirically derived compactness threshold 'B^C', we remove the node from the community.

- Expanding the community
    1. Find out all the neighbours N(c) of initial community c, for every node $v \in N(c)$, calculate $B(v, c)$
    2. Find out vertices which meet the need of $B(v, c) > B^C$ and $B^L \leq B(v, c) \leq B^C$, write as $N_v = \{ v \mid B(v, c) > B^C \}$ and $N_{lv} = \{ v \mid B^L \leq B(v, c) \leq B^C \}$
    3. If $N_v > 0$, we add all the vertices of N_v into the community c which is larger than before, and we still write as c, back to step (1)
    4. If $N_{lv} > 0$, we add all the vertices of N_{lv} into the community c which is larger than before, and we still write as c, back to step (1)
    5. If $N_v = 0$ and $N_{lv} = 0$, stop expanding the community and finally we obtain a community.

Thus, we start by looking at all the neighbours of the nodes within our community and calculate the compactness B for each of them. For each of the neighbours, if the compactness value 'B' is greater than 'B^C' we directly add them to the community. However, if the value is less than 'B^C', but greater than the lower bound threshold 'B^L', we introduce this node within the community and check the modularity of the network. Only if the modularity increases, the node becomes a part of the community. However, if the modularity were to decrease, we remove the node and move on to the next neighbour.

- Finding nodes that belong to multiple communities

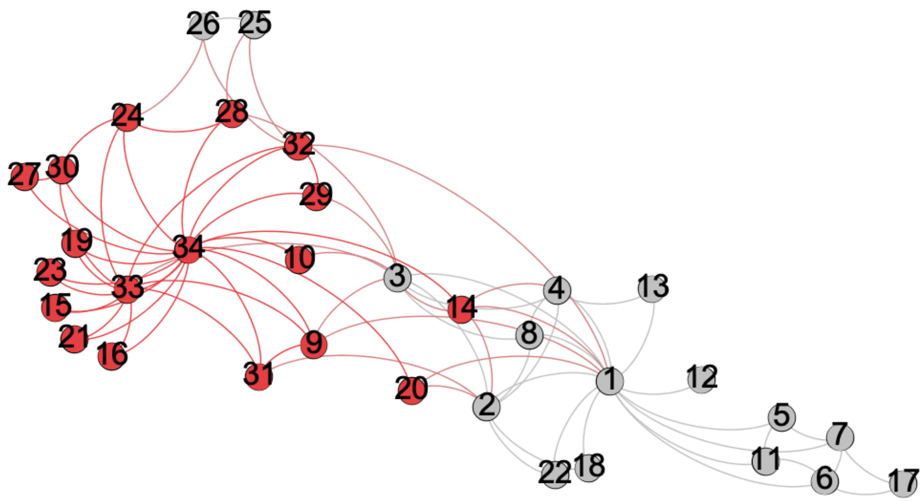
For the given network, if we encounter a node which has an equal number of edges going into more than one community, we mark it as belonging to multiple communities.

APPLICATIONS AND RESULTS

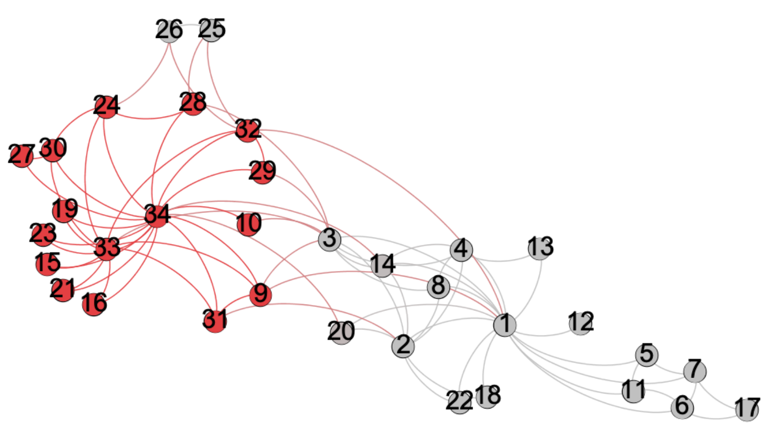
- Zachary Club Network

  The following steps are repeated till all communities are formed.

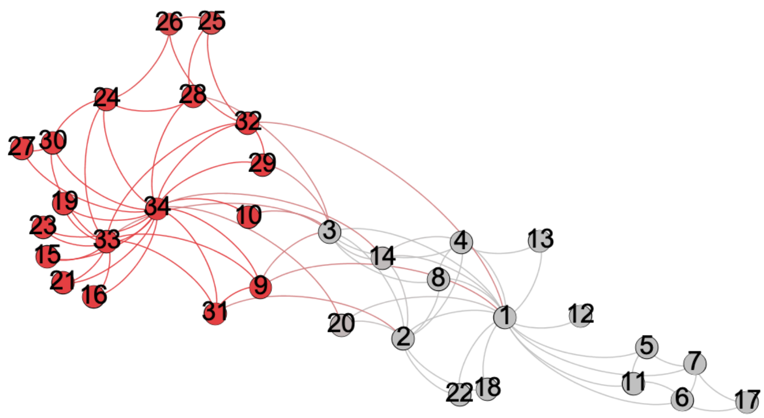
  - Step 1: We begin by creating a dictionary of the nodes with their node strengths. We select node 34 since it has the highest node strength in the network. We begin by creating a temporary community of the neighbors of 34.



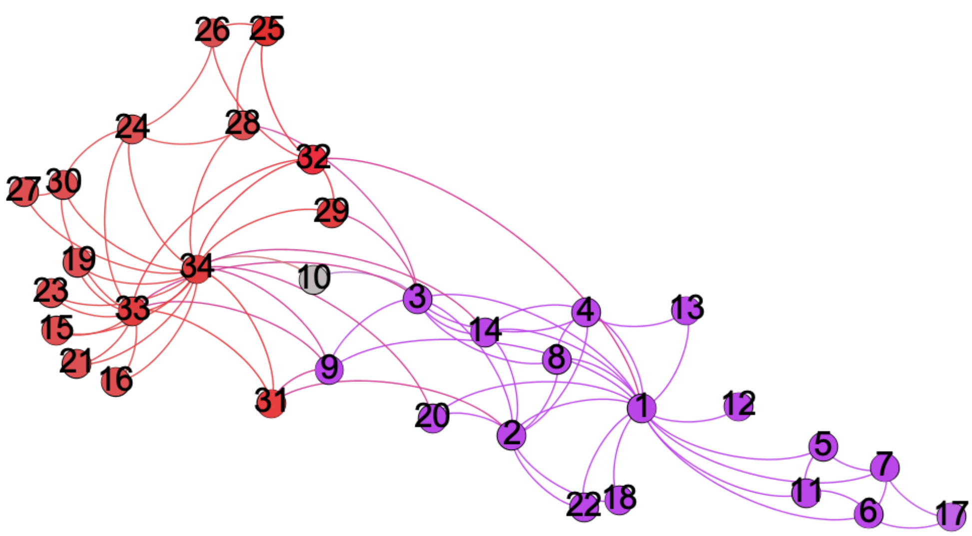
  - Step 2: We remove the nodes 14 and 20 since their compactness values fall below the threshold of 0.5 which is the compactness of the temporary community



- Step 3: We add nodes 25 and 26 to the community since their compactness values lie above 0.4 which is the compactness value of the local community. Adding these nodes to the community also increases the modularity of the network.

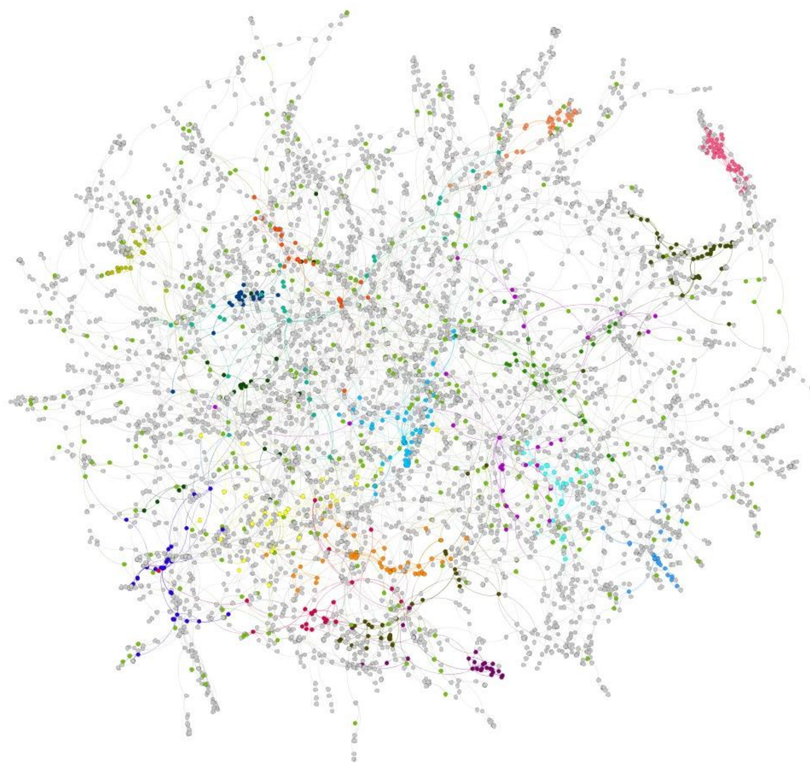


- Step 4: We remove node 10 which has an incoming edge from community 1 and an outgoing edge to community 2. The incoming and outgoing edges of node 10 make it difficult to assign it to either community and hence is labeled as belonging to both.



- Power Grid Network

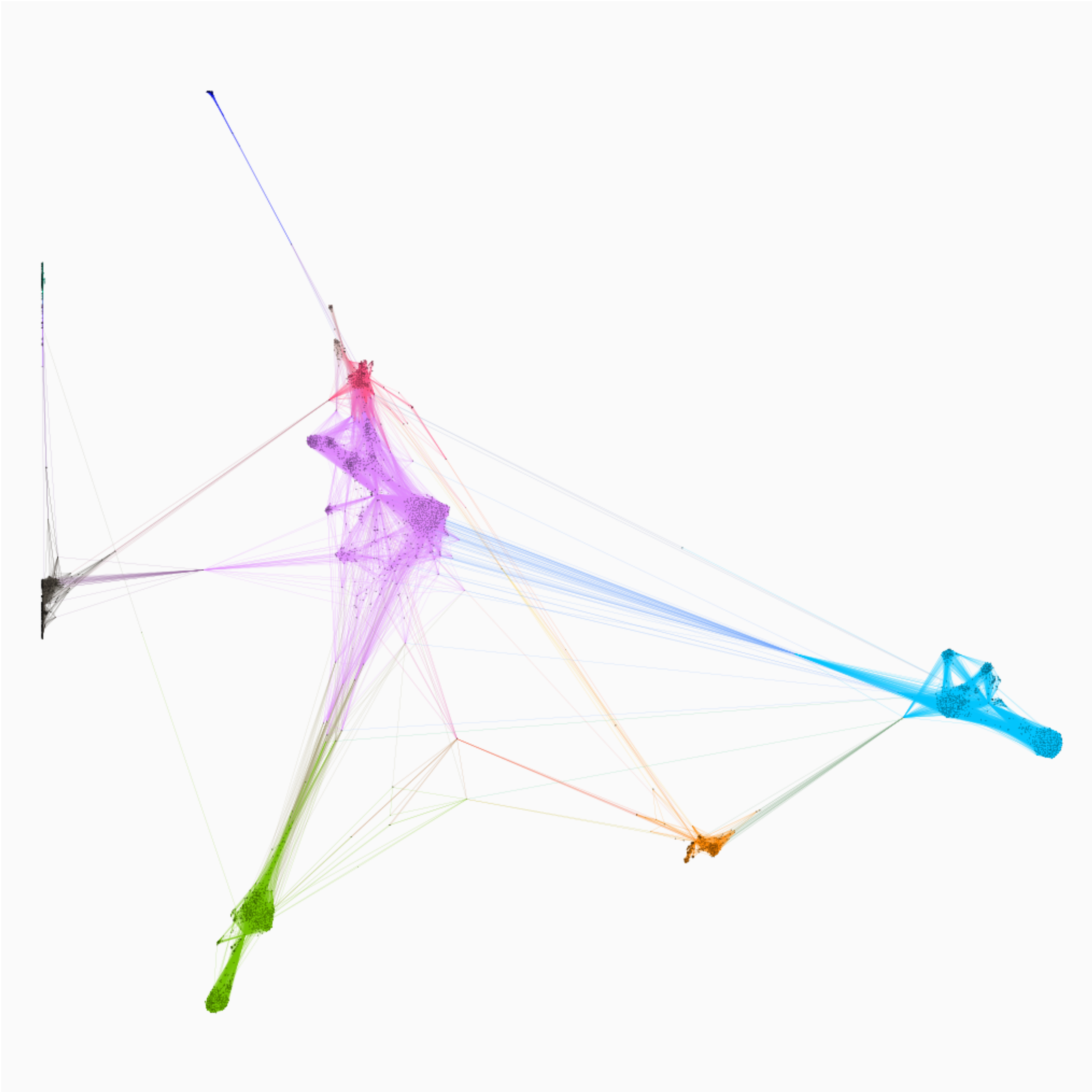
  The algorithm performs well when we are concerned with local communities and fails for networks which have large sparsely connected communities.



- Ego-Alter Facebook Social Network

  To evaluate the performance of this algorithm on a social network, we have used a Facebook friends circle network. This is an ego-alter network consisting of the alters of 11 egos.
  As with the application on the Zachary club network, we will treat the network as a two-way network, that is to say, the in-degree and out-degree of every vertex are equal to its vertex degree. We set the 'B^C' value as 0.6 and 'B^L' values as 0.5 to get 13 initial communities. We eliminate any communities having a single node and are left with 10 communities.

  Since we do not have the ground truth of the actual number of communities in this network, we calculate the modularity to gauge the algorithm's ability to detect local communities within the alter. The modularity of the network with the final communities is 0.73.



DISCUSSION

Conclusion

- In conclusion, we have studied an algorithm that detects local communities in a directed network.
- We have applied this algorithm to networks having different network structures and densities to identify local communities.
- We discussed our observations of the performance of this algorithm on the different networks.
- We have outlined the limitations of the algorithm.

Limitations

- One of the major issues we faced was that there were no detailed explanations of the parameters involved in the algorithm and its application.
- The algorithm proceeds with assumptions and changes needed to accommodate the Zachary Club network.

Further Scope

- For large scale networks, one could make modifications in the second step ‘Expand the community’ to accommodate an additional set of neighbours, essentially finding communities with three-layers.

REFERENCES

[1] Shigang Liu *‘Detecting communities of directed networks via a local algorithm’* by Shigang Liu https://www.researchgate.net/publication/316273095_Detecting_communities_of_directed_networks_via_a_local_algorithm

[2] Zachary Karate Club dataset - http://www-personal.umich.edu/~mejn/netdata/

[3] Power Grid dataset - http://www-personal.umich.edu/~mejn/netdata/

[4] Facebook social circles - https://snap.stanford.edu/data/ego-Facebook.html

[5] M.Fiedler, Algebraic connectivity of graphs,. Czech Math. J, 23(98) (1973) 298-305.

[6] B.W.Kernighan and S.Lin, An efficient heuristic procedure for partition graphs, Bell Sys Tech J., 49 (1970) 291-307.

[7] M.E.J.Newman and M.Girvan, Finding and evaluating community structure in networks, Phys. Rev E, 69 (2004), p.026113.

[8] M.E.J.Newman, Fast algorithm for detecting community structure in networks, Phys Rev E, 69(6) (2004) 1-5.