AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

19CSE337 Social Networking Security

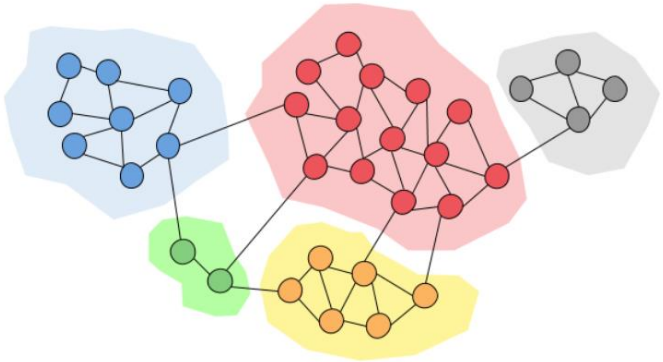
Lecture 12

Topics to Discuss

- Community
- Clustering
- Girvan-Newman Algorithm

Community

- A community, with respect to graphs, can be defined as a subset of nodes that are densely connected to each other and loosely connected to the nodes in the other communities in the same graph.

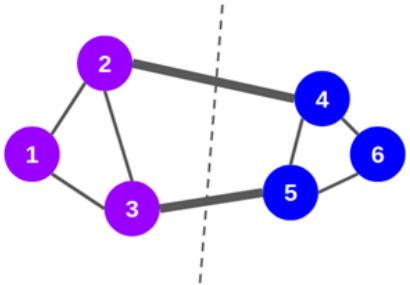


Community

- Think about social media platforms such as Facebook, Instagram, or Twitter, where we try to connect with other people.
- Eventually, after a while, we end up being connected with people belonging to different social circles.
- These social circles can be a group of relatives, school mates, colleagues, etc.
- These social circles are nothing but communities!

Community Detection

- Community detection in a social network is identifying sets of nodes in such a way that the connections of nodes within a set are more than their connection to other network nodes.
- Community detection techniques are useful for social media algorithms to discover people with common interests.

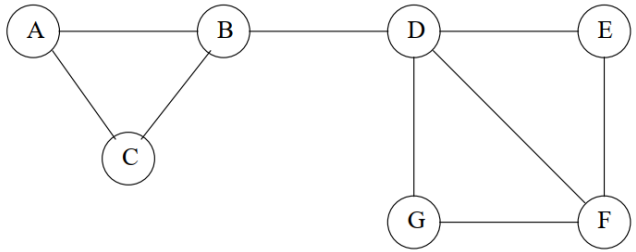


Clustering

- Often clustering and community detection are used interchangeably.
- Both focus on grouping of nodes.
- In short, clustering group nodes based on its features whereas community detection group nodes based on its connectivity.
- Community not partition nodes into sets rather overlap.

Clustering Algorithms

- There are many popular clustering algorithms like hierarchical clustering, K-means clustering etc. These algorithms can be used to detect communities.
- In the given graph, we can identify two communities {A,B,C} and {D,E,F,G}. Also, two sub communities {D,E,F} and {D,F,G} of {D,E,F,G}.
- Unfortunately, a pure clustering algorithm sometimes fails to detect overlapping communities, sometimes fails to place nodes in appropriate groups.
- Say for example in the given graph, node D may be placed as part of community I or II.



Clustering

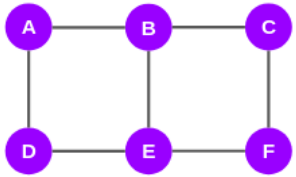
- Since there are problems with standard clustering algorithms, we need more specialised algorithms.
- One such algorithm is Girvan-Newman algorithm.
- It works based on finding the edges that are least likely to be inside a community.

Girvan-Newman Algorithm

- The Girvan-Newman algorithm finds the communities in a graph by iteratively removing the edges of the graph, based on the edge betweenness centrality value.
- The edge with the highest edge betweenness is removed first.
- The edge betweenness centrality (EBC) can be defined as the number of shortest paths that pass through an edge in a network.
- Each and every edge is given an EBC score based on the shortest paths among all the nodes in the graph.
- The betweenness of an edge (a,b) is the number of pairs of nodes (x,y) such that the edge (a,b) lies on the shortest path between x and y.

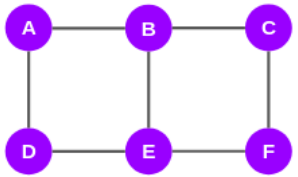
Girvan-Newman Algorithm

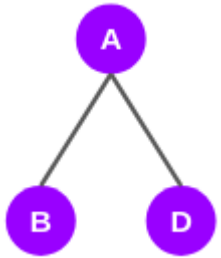
- Let’s take an example to find how EBC scores are calculated.
- Consider the given graph. It has 6 nodes and 7 edges.
- Steps:
  - Take one node at a time and plot the shortest paths to the other nodes from the selected node
  - Based on the shortest paths, compute the EBC scores for all the edges
  - Repeat this process for every node in the graph.
  - After all iterations, every edge will get EBC scores. These scores will be added edge-wise
  - Finally, the total score of each edge will be divided by 2 to get the EBC score since it is an undirected graph.



Girvan-Newman Algorithm

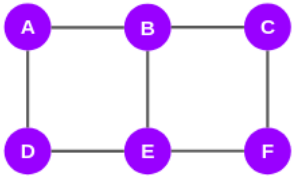
- Let’s start with node A. The directly connected nodes to node A are nodes B and D. So, the shortest paths to B and D from A are AB and AD respectively.

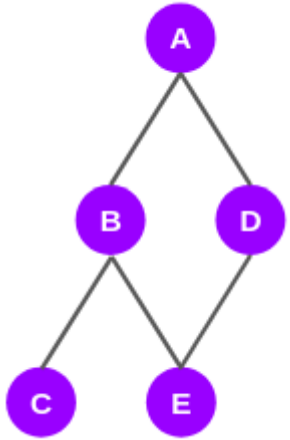




Girvan-Newman Algorithm

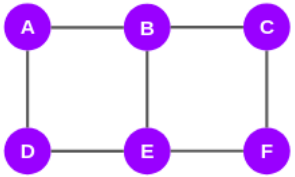
- It turns out that the shortest paths to nodes C and E from A go through B and D.

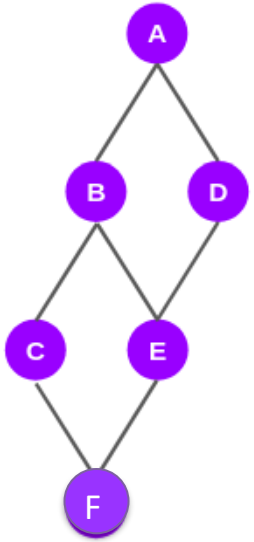




Girvan-Newman Algorithm

- The shortest paths to the last node F from node A, pass through nodes B, D, C, and E.





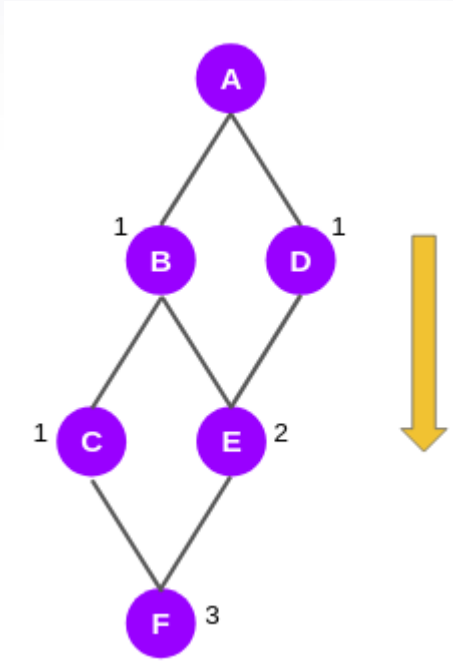
Girvan-Newman Algorithm

- The graph above depicts only the shortest paths from node A to all the other nodes.
- Now edge scores need to be assigned.
- Before giving scores to the edges, assign a score to the nodes in the shortest-path-graph.
- To assign these scores, traverse the graph from the root node. (i.e., node A to the last node F).

Girvan-Newman Algorithm

Assigning scores to nodes

- Nodes B and D have been given a score of 1 each. This is because the shortest path to either node from node A is only one. For the very same reason, node C has been given a score of 1 as there is only one shortest path from node A to node C.
- Moving on to node E. It is connected to node A through two shortest paths, ABE and ADE. Hence, it gets a score of 2.
- The last node F is connected to A through three shortest paths — ABCF, ABEF, and ADEF. So, it gets a score of 3.

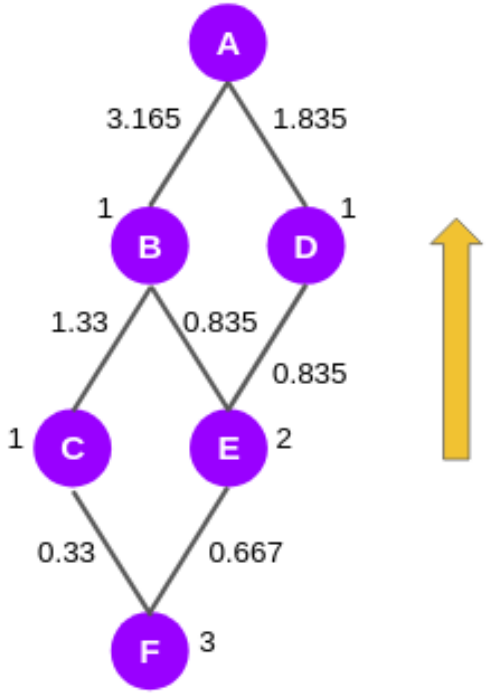


Girvan-Newman Algorithm

Computing Scores for Edges

- Next, proceed with computing scores for the edges. Here computation begins in the backward direction from node F to node A.
- Firstly, compute the score for the edges FC and FE. The edge score for edge FC is the ratio of the node scores of C and F, i.e. 1/3 or 0.33. Similarly, for FE the edge score is 2/3.
- Now, calculate the edge score for the edges CB, EB, and ED. According to the Girvan-Newman algorithm, from this level onwards, every node will have a default value of 1 and the edge scores computed in the previous step will be added to this value.
- So, the edge score of CB is (1 + 0.33)/1. Similarly, edge score EB or ED is (1 + 0.667)/2. Then we move to the next level to calculate the edge scores for BA and DA.

Girvan-Newman Algorithm



FC = ⅓ = 0.33
FE = ⅔ = 0.667

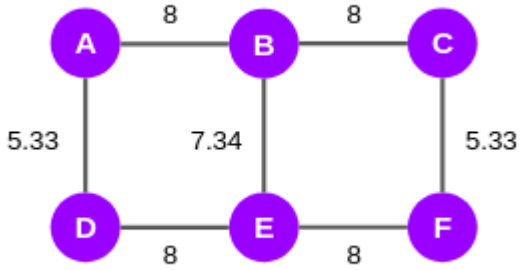
CB = 1 + 0.33 = 1.33
EB = (1 + 0.667)/2 = 0.835
ED = (1 + 0.667)/2 = 0.835

BA = (1 + 1.33 + 0.835)/1 = 3.165
DA = (1 + 0.835)/1 = 1.835

Girvan-Newman Algorithm

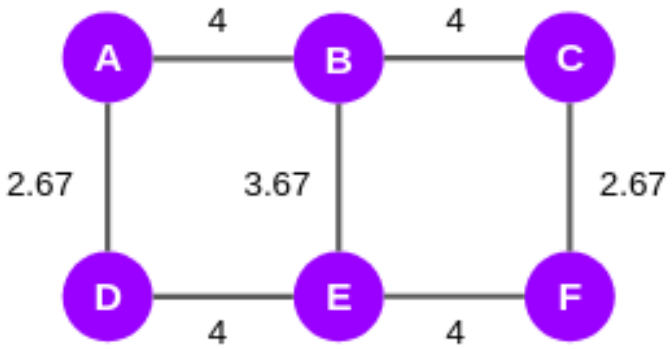
Computing Final Scores for Edges

- So far, we have computed the edge scores of the shortest paths with respect to node A. We need to repeat the same steps again from the other remaining five nodes.
- In the end, we will get a set of six scores for all the edges in the network. Then add these scores and assign them to the original graph as shown in diagram.



Girvan-Newman Algorithm

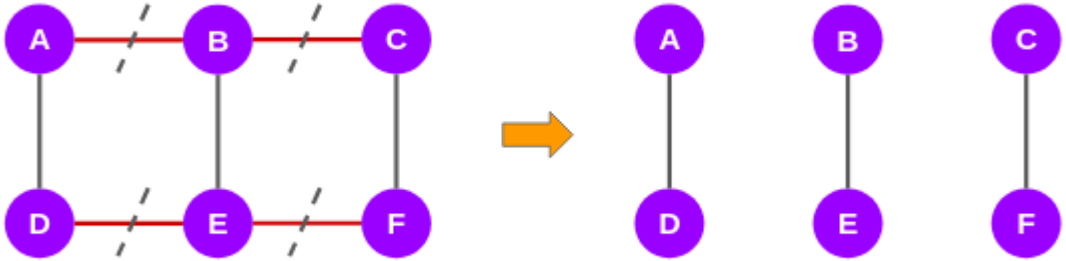
- Since it is an undirected graph, divide these scores by two and finally, we will get the EBC scores.



Girvan-Newman Algorithm

- According to the Girvan-Newman algorithm, after computing the EBC scores, the edges with the highest scores will be taken off till the point the graph splits into two.
- So, in the graph above, we can see that the edges AB, BC, DE, and EF have the highest score, i.e., 4.
- We will strike off these edges and it gives us 3 subgraphs that we can call communities.

Girvan-Newman Algorithm



Thanks……