Centrality Of One Or More Node In GraphTheory

Bhagat Ram Labana,22m1083 EE2
Guided by:-Prof. Madhu Belur,EE Department, IIT Bombay

***Abstract*—In social networks to share information specially for advertising or marketing there needs to be some influential persons who can diffuse the information within the community. We can define Social networks as graphs and influential persons as the important nodes. Important nodes can be said to critical nodes of the graph. Removal of critical nodes can lead to disrupting the network. Identifying critical nodes can help analyze the topological characteristics of the network, such as vulnerability and robustness. Some groups of critical nodes may be useful for one parameter but may not be useful for some other parameter. From Conventional Centrality degree, closeness, betweenness, eigenvector, Katz and page rank centrality we are finding the important node on the basis of score according to corresponding individual centrality. According to the conventional centrality finding method If any single node is a most influential node in one graph while finding more than one most important node , this single node is also contained in the group of more than one important node. Community Centrality-Based Greedy Approach can be used for Identifying Top-K Influencers in a graph where Top K influencers may not include the The Top K nodes based on its centrality score for any particular centrality. Markov chain algorithm can be used to partition the graph into the communities.And then finding Some Top K nodes from each community and then use Community Centrality based greedy approach to find the imporatant nodes.**

I. INTRODUCTION

CEntrality Of Node of any network plays an important role in social networks, power grids, food delivery systems and cyber attacks. Centrality of Network in graph theory provides the interconnectedness of members of Facebook or any social media platforms. Many Product companies use the Centrality technique in their marketing to detect demand supply chains in the social network. To prevent our network from cyber attacks, centrality plays a very important role. Using centrality we find critical nodes, absence of which can disrupt our network. Finding Centrality can be formulated as a problem to find the most important Top-K nodes in a graph. In Section-1 some prerequisites of graph theory are discussed briefly. In [4] Laplacian matrix, Adjacency matrix, Degree matrix, Algebraic Connectivity and Fiedler Vector is discussed. In Section 2 Some conventional Centralities are discussed. In [2] degree, closeness , betweenness centralities are discussed. In [6] EigenVector And Pagerank Centralities are discussed. In [1] Katz centrality is used to find Top K influential nodes. For A graph Given in figure:01 ranking of importance of all the nodes are found for the conventional centralities discussed in section2. All the centralities discussed in section 2 give the result for Top-K most important node on the ranking based centrality score. For degree, Eigenvector, Katz and Page rank centralities Top 2 important nodes are coming out to be the same but for Different graphs it may vary. For Closeness and Betweenness centrality also the Ranking of the important node is found out for Graph in Figure:01. In Section 3 Community centrality finding approach is Discussed. In[2] Algorithm to find Centrality based on Community partition is given. In [3] Markov Clustering Algorithms to Partition the graph into the communities are given. Community based centrality approach Does not always provide Best-K nodes according to the ranking of their Centrality score.According To the community centrality based greedy workflow approch Nodes which are influential for their own community may be not useful for other community even if its centrality score on the basis of conventional centrality is higher than all nodes in that community. One Example for a graph is taken from [5] and the another example of graph is taken from [3] and found out their centrality from community centrality based approach. In the End of the report some conclusion has been made.

II. PREREQUISITE OF GRAPH THEORY

1) Directed Graph:-We define a Graph by G(V,E). Where V is the set of nodes of graph G while E is the set of Edges of Graph G. We say any graph is directed if and only if
$(v_1, v_2) \neq$ (v_2,v_1) where
v_1, $v_2 \in V$, and
(v_1, v_2) ,(v_2,v_1) $\in E$
2) Undirected Graph:- We say a graph G(V,E) is undirected if and only if
(v_1, v_2) = (v_2,v_1)), where .
v_1, $v_2 \in V$, and
(v_1, v_2) ,(v_2,v_1) $\in E$
3) Walk of Length L:- For any graph Given by G = (V, E), a walk of length l is defined as the set of nodes v_1, v_2, v_3, ... , v_l such that there exists an edge between v_i and v_{i+1}, $\forall 1 \leq i < l$.
4) Degree Matrix:- For Undirected Graph Degree matrix is Defined as D = [D_ij]. Where D_ij is given by
D_ij = {deg(v_i) , if i = j;
=0 otherwise; }
deg(v_i) is degree of node v_i
5) Adjacency Matrix:- For an undirected graph , Adjacency matrix is defined as A = [A_{ij}]. Where A_{ij} is given by
A_{ij} = 1, if v_i and v_j are connected
= 0 otherwise;
6) Laplacian Matrix:- Laplacian matrix for undirected graph G(V,E) is given by L = D - A
7) Algebraic Connectivity:- Algebraic connectivity tells us the connectivity of the graph G(V,E) .Lets say the Laplacian matrix has eigenvalues such that

$$\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq .... \leq \lambda_{max}.$$

Eigenvalue λ_2 is called algebraic connectivity. If λ_2= 0, graph is connected otherwise it is disconnected.

8) Fiedler Vector:- Eigenvector corresponding to algebraic connectivity is called a fiedler vector.
9) Perron Frobenius Theorem:- For $A \in R^{(n*n)}$. Where A is an adjacency matrix for strongly connected graphs. Then there exists an unique eigenvalue λ_{max} such that eigenvector U corresponding to λ_{max} has all positive entries.
10) Geodesics:- From any path from vertex V_i to V_j is called geodesic if it is the minimum length between V_i and V_j. And this path length is called the Geodesic length.
11) Stochastic Matrix:- Stochastic matrix is formed by transition probability for each node.While forming a stochastic matrix we assume a self loop exists and then form the matrix corresponding to the group.

III. Centralities Of The Graph

In this section Centrality of a graph is discussed. In graph theory we assign every node some ranking of importance by assigning some score to each node. But In real Life There are many Networks where no. of nodes are in very large in number. Lets say in social network on facebook or any social media website have billions of the users. But to propagate some information to maximum no. of users through network it will take a lot of time if we try to propagate to individual node. To solve this problem we need to find some influential person or user through we can propagate news to whole network. To solve the above problem we use graph theory where the social network is considered to be a graph and all the users as the Nodes of the graph. Similarly Influential users are considered to be the Important nodes of the graph. Our aim is to find the Important nodes of the graph. Here we will describe some of the conventional centralities used in many applications. With the help of these centralities we can find the Important nodes of the Graph. Here Unweighted and undirected graph is considered.

A. Degree Centrality

Degree Centrality is defined by the parameter degree of nodes of the graph. d(v_i) is defined as the number of nodes in graph G connected to the node v_i directly. For the figure:01 Given below are given the degree centrality score.
d(v_1) = 4, d(v_2)= 3, d(v_3)= 3 , d(v_4)= 3,
d(v_5)= 2, d(v_6)= 3 , d(v_7)= 1, d(v_2)= 1.
Hence With Degree centrality we get the most influential node in terms of degree score. The Node(Vertex) with the highest degree is the most influential node. And if we want to find K most influential nodes then rank nodes on the basis of score. Select top K nodes(Vertices) with higher degree score. Suppose we want to select any top two influential nodes. Then Top 1 most influential node is Obvious. Node 1 has highest degree of 4 so we considers the node 1 as the most important node acoording to the degree centrality . Second most important can be considered as any one of node from the the set $G^{'}$ = {2, 3, 4, 6,} , Becuase all these nodes have same degree centrality score. Then for top-2 most important nodes we can select any of the sets out of {1,2}, {1,3}, {1,4} and {1,6}.

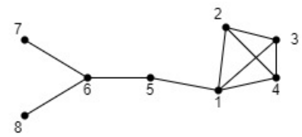

Fig. 1. Graph G

B. Closeness Centrality

Closeness of a node u is defined as the reciprocal of the average of the shortest path length from node u to all other nodes of a connected network. If a node has strong closeness centrality , then it can spread information quickly. Closeness is the measure of how fast a node can spread the information to all other nodes.Let d(u, v) be the distance from a node u to a node v within a Graph G. Then, the closeness of u is defined as:

$$C_c(u) = (\frac{\Sigma_{\forall v \in G;} d(u,v)}{|G|-1})^{-1} \quad (1)$$

Here in the given equation above $u \neq v$. Where $|G|$ denotes the cardinality of the graph G. For the graph given in figure:01 Centrality Value for each node is calculated as follows
$C_B(1)$ = $\frac{7}{12}$, $C_B(2)$ = $\frac{7}{16}$, $C_B(3)$ = $\frac{7}{16}$, $C_B(4)$ = $\frac{7}{16}$, $C_B(5)$ = $\frac{7}{12}$, $C_B(6)$ = $\frac{7}{15}$, $C_B(7)$ = $\frac{7}{20}$, $C_B(8)$ = $\frac{7}{20}$,
Node1 and Node5 is the most important node corresponding to the Closeness centrality for graph G.

C. Betweenness Centrality

Betweenness of a node is defined as the no. of the individual pair passing through a node in order to reach one another in minimum no. of steps. In any Graph G let st be total no. of the shortest length(hops) to go from node s to t and let st(u) be the total no. of the shortest length(hops) to go from node s to t via node u. Betweenness of node u is given by

$$C_B(u) = \Sigma_{\forall s,t \in G;}(\frac{\sigma_{s_t}(u)}{\sigma_{s_t}}) \quad (2)$$

For the equation(2) ($s \neq t \neq u$). For the graph in figure:01 the between centrality for all the nodes is given by, $C_B(v_1)$= 12 ,$C_B(v_2)$= 0,$C_B(v_3)$= 0 ,$C_B(v_4)$= 0,$C_B(v_5)$= 12,$C_B(v_6)$= 10,$C_B(v_7)$=0 ,$C_B(v_8)$=0 , Hence node1 and node5 are the most important nodes for the graph G.

D. Eigenvector Centrality

Eigenvector Centrality is an extension to degree centrality. Degree centrality does not depend on the score of the neighbor node. But Eigenvector Centrality depends on the

Neighboring Node score also. Nodes connected to highly influential nodes also get some relative importance depending on the score of neighboring nodes score. Eigenvalue and Eigenvector is defined as follows

$$AX = \lambda X \tag{3}$$

Where A any n*n matrix and λ is eigenvalue corresponding to eigenvector X. In case of eigenvector centrality A is an adjacency matrix of Graph G and eigenvector Centrality is given by principal eigenVector Values. Where principal Eigenvector is the eigenvector corresponding to principal eigenvalue. Principal eigenvalue is the largest eigenvalue λ_{max} as defined in above section() forAdjacency matrix A.

1) PROCEDURE TO COMPUTE EIGENVECTOR CENTRALITY:

1) Compute eigenvalues of A
2) Select the largest eigenvalue i.e., λ_{max}
3) Let's say eigenvector of λ_{max} is Ce
4) Based on the Perron Frobenius theorem , all the components of Ce will be positive .
5) Components of the Ce are the eigenvector Centralities for the graph. Hence Eigenvector Centrality is given by

$$C_e(v_i) = \frac{1}{\lambda}\Sigma_{j=1}^{n} d_i a_{j,i} C_e(v_j) \tag{4}$$

Where $a_{j,i}$ is the Element of adjacency matrix and d_i is the degree of node v_i. For the graph given in the figure:01 Adjacency matrix A is given by

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

.

λ_{max} corresponding to adjacency matrix A is 3.099. And Eigenvector Ce corresponding to λ_{max} is given by Ce = [0.5237 , 0.4762, 0.4762, 0.4762, 0.1945, 0.0793, 0.0256, 0.0256] Hence the ranking of importance of nodes of graph G given in figure:01 based on eigenvector centrality is given by $Node1 > Node2 = Node3 = Node4 > Node5 > Node6 > Node7 = Node8$

E. Katz Centrality

Katz Centrality is an extension of the EigenVector Centrality. As we have discussed in above eigenvector centrality that eigenvector centrality considers the effect of the immediate neighbor but in katz centrality it considers the effect of immediate as well as non immediate neighbor. Katz Centrality is given by

$$C_{katz}(v_i) = \alpha\Sigma_{j=1}^{n} A_{i,j} C_{katz}(v_j) + \beta \tag{5}$$

Where α is damping factor where condition on α is given by

$$0 < \alpha < \frac{1}{\lambda_{max}} \tag{6}$$

And β is bias constant and is known as an exogenous vector. This exogenous vector is used to eliminate zero centrality values. And if

$$\alpha > \frac{1}{\lambda_{max}} \tag{7}$$

then the centrality tends to diverge.
In the case of degree centrality it does not consider the effect of the neighbors . And In case of Eigenvector Centrality it considers only the effect of influence of immediate neighbors. But in the case of Katz centrality It considers the effect of influence of immediate as well as non immediate neighbors. Degree centrality considers the local influence and EigenVector Centrality considers the global influence but the Katz centrality takes in picture both the local influence and global influence.In Katz centrality, walks of all the length from any single node to all the other nodes are considered. But as walk length increases it attenuates the effect of influence by attenuation factor . For immediate neighbor path length l =1 , so influence gets attenuated by α . But for non immediate neighbor of length l = 2, Influence gets attenuated by α^2 . Similarly for any nodes at walk length of l = K the the influence gets attenuated by α^K. Effect of the attenuation factor on the the graph G can be shown by the given following equation,

$$(I-\alpha A)^{-1} = I+\alpha A+(\alpha A)^2+.......+(\alpha A)^k = \Sigma_{k=0}^{\infty}(\alpha A)^k \tag{8}$$

Katz Centrality for Graph can be generalized as follows,

$$C_{katz} = \beta(I - \alpha A)^{-1} \tag{9}$$

Where 1 is the column vector of all ones in it. For above graph given in figure 1 , for different values of α and β centrality ranking varies. For graph in figure 1 we have calculated the centrality for two sets of values of α and β .
For α = 0.3, β = 0.5, The centrality ranking comes out to be $Node1 > Node2 = Node3 = Node4 > Node5 > Node6 > Node7 = Node8$
For α = 0.1, β = 0.5, The centrality ranking comes out to be $Node1 > Node2 = Node3 = Node4 > Node6 > Node5 > Node7 = Node8$
Effect of and on the centrality:-
case1).If $\alpha \to 0^+$, The Katz centrality converges to the degree centrality.
case2).If $\alpha \to \frac{1}{\lambda}$, The Katz centrality converges to the Eigenvector centrality.

F. Page Rank Centrality

Page Rank was founded by the google founder to rank the importance of the webpages. It works well for directed graph but it can also used to find the centrality of Undirected graph nodes. Page Rank centrality for undirected graph is given by

$$C_p(v_i) = \alpha \Sigma_{j=1}^{n} A_{j,i} \frac{C_p(v_j)}{{d_j}^{out}} + \beta \tag{10}$$

Where d_j^{out} is the outdegree of any node of the graph G. D is the Diagonal matrix defined by

D = diag(${d_1}^{out}$, ${d_2}^{out}$,......,${d_n}^{out}$)

In case of d_j^{out} = 0 , above equation is coming out to be not defined. So in that case set d_j^{out} = 1. Hence D_{ii}= max (${j}^{out}$,1)

Page Rank centrality for graph G in vector form is given by

$$C_p = \alpha A^T D^{-1} C_p + \beta .1 \tag{11}$$

$$C_p = \beta(I - \alpha A^T D^{-1})^{-1} .1 \tag{12}$$

Where 1 is a column vector of size of length n . And α and β is same as the defined above for the katz centrality. Condition on the α is that α should be less than reciprocal of the maximum eigenvalue of the matrix $A^T D^{-1}$.

$\alpha < \frac{1}{\lambda_{max}}$

λ_{max} is the maximum eigenvalue of $A^T D^{-1}$.

IV. Community Centrality-Based Greedy Approach for Identifying Top-K Important node in a graph

As we have discussed in the above section, degree, closeness, betweenness , Eigenvector, Katz and Page rank centrality briefly. Finding Top k important nodes in a graph can give different results for different graphs. All Centrality finding methods that we have discussed in the above section rank the importance of the node on the basis of the centrality score for each Centrality Method. But sometimes in any social network finding the top k influential person or Finding top K important nodes in Any graph on the basis of the above conventional centrality method cannot always give us the best results. Sometimes Sets of Nodes with the not having highest centrality score can give us best results. In [2] there has been discussion of the approach to find Top Most K influencers in social networks which gives better results as compared to the naive approaches discussed in the above section. Below are the four processes which we should implement to find the top K most influential node in the following flow

A. Community Detection

While finding the most influential node in any graph we first partition the graph in different communities. Then we pick some of the most influential nodes from each community.

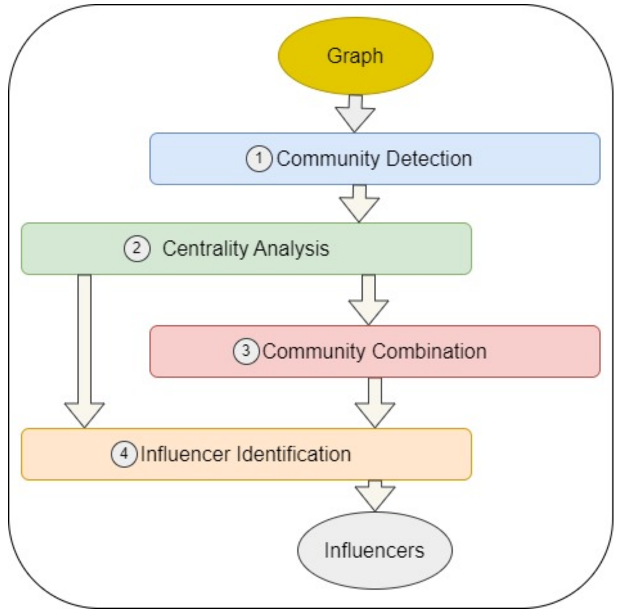

Fig. 2. Algorithm Flow

Motivation behind forming communities is that nodes in the same community interact more with each other so if we select influential nodes from each community then that will give us better results in spreading any news among communities. Hence it is more likely that any node within any particular community will influence nodes of the same community better than some other communities' most influential nodes.

In paper[2] described the method to partition the community from Markov Clustering Algorithm. From the given graph according to its topological structure Graph G can be partitioned in communities G_1, G_2, ... , G_k. G = $\{G1$,G_2,. . . . . . . ,G_k $\}$. Markov clustering assumes the random walk approach to find the community. If someone walks randomly in a graph from any node within the graph then if some set of nodes is frequently countered then that set of nodes will form a community. In MCL we need not to give the input parameter for no. of communities. Algorithm automatically forms suitable communities.

Here Markov Clustering algorithm (MCL) includes two steps which are given as follows. Markov Clustering takes a stochastic matrix as the input. The after that Expansion and Inflation operators are applied to the stochastic matrix of the graph.

Expansion:- Expansion is done by usual power of stochastic matrix.We denote the power t to matrix A by $Exp_t A = A^t$. Matrix Expansion is done to get higher step transition probabilities.

Inflation:- inflation (denoted r) means taking the Hadamard power with coefficient r of a stochastic matrix and subsequently scaling its columns to have sum 1 again. Matrix Inflation is done to promote high transition probability and demotion of the low transition probability. For the figure 01 we can implement MCL algorithm to find the communities.

Stochastic matrix for the given graph in figure 01 is given by

$$\begin{bmatrix} 0.2 & 0.25 & 0.25 & 0.25 & 0.3333 & 0 & 0 & 0 \\ 0.2 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0.2 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0.2 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0.3333 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3333 & 0.25 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0.5 \end{bmatrix}$$

After Applying Expansion and The inflation in the given stochastic matrix above seven times the matrix converges to the following matrix.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

From Above matrix structure we can infer that there are two communities formed from the graph given in the figure 01. Graph G is partitioned into 2 communities G_1 and G_2 given in fig.3 and fig.4 . G= { 1,2,3,4,5,6,7,8 }. And G_1 = { 1,2,3,4 } and G_2 = { 5,6,7,8 }.

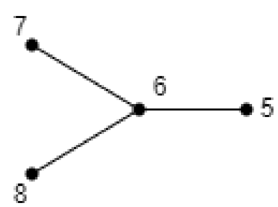

Fig. 3. G_1

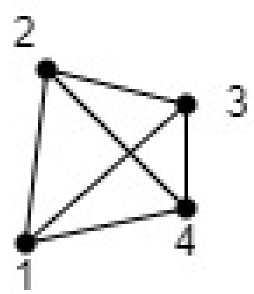

Fig. 4. G_2

B. Centrality Analysis

After communities are formed we find the most influential nodes from each community. The centrality can be found using any of the above methods. We can find some K most important nodes in each community. For the graph G given in Figure01 the top two important nodes are Node 1 and Node 2. Let's set of most important is given by ω = {1,2}. By using degree centrality we find that the most important nodes in graph G1 and G2 are {1 or 2 or 3 or 4} and {6} respectively. For Katz Centrality also the most important node for graph G1 and G2 is given by {1 or 2 or 3 or 4} and {6} repectively.

C. Community Combination

MCL Algorithm sometimes produces Very small and Dispersed communities . For that reason sometimes we combine the communities if any top-K influential node present in each communities are connected then we merge those communities and again find the top K influential node from merged communities. Again Repeat the procedure if the Influential nodes from new communities are connected to each other. Stop ,when no influential nodes from any community are connected within each other. For Fig.3 and Fig.4 We can see Communities G_1 and G_2 are not merging beacause top most node 6 of community G1 is is not connected to any of the node from community G2 . So Both communities are considered to be final communities for graph G given in the Figure:01.

D. Influencer Identification

Now from final communities extract Top K nodes from each community and merge together to get final Top most important nodes for a graph.
The above algorithm is implemented on the the given graph in Fig.5. given below. Stochastic Matrix for the given above graph is given in the Fig.6.

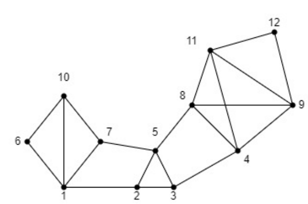

Fig. 5. Graph C

Applying Markov Chain Algorithm Given in [3], applying Expansion and inflation to the stochastic matrix given in Fig.6, The Matrix Converges to the the matrix given in the Fig.7.

From the Convergent Matrix in Fig.7 , we can see that there three communities are forming For the graph C given in the fig.5. Graph C is partitioned into C1, C2 and C3 communities. C1 community contains nodes 1, 6, 7,and 10. Hence C1 can be given as C1 = { 1,6,7,10}. Similarly C2 and C3 communities are given by C2 = {2,3,5} and C3 = {4,8,9,11,12}. Lets say we need to find Top two important nodes from graph C. To find Top 2 important node we are Applying Degree Centralities

$$\begin{bmatrix}
0.2 & 0.25 & 0 & 0 & 0 & 0.33 & 0.25 & 0 & 0 & 0.25 & 0 & 0 \\
0.2 & 0.25 & 0.25 & 0 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.25 & 0.25 & 0.2 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.25 & 0.2 & 0 & 0 & 0 & 0.2 & 0.2 & 0 & 0.2 & 0 \\
0 & 0.25 & 0.25 & 0 & 0.2 & 0 & 0.25 & 0.2 & 0 & 0 & 0 & 0 \\
0.2 & 0 & 0 & 0 & 0 & 0.33 & 0 & 0 & 0 & 0.25 & 0 & 0 \\
0.2 & 0 & 0 & 0 & 0.2 & 0 & 0.25 & 0 & 0 & 0.25 & 0 & 0 \\
0 & 0 & 0 & 0.2 & 0.2 & 0 & 0 & 0.2 & 0.2 & 0 & 0.2 & 0 \\
0 & 0 & 0 & 0.2 & 0 & 0 & 0 & 0.2 & 0.2 & 0 & 0.2 & 0.33 \\
0.2 & 0 & 0 & 0 & 0 & 0.33 & 0.25 & 0 & 0 & 0.25 & 0 & 0 \\
0 & 0 & 0 & 0.2 & 0 & 0 & 0 & 0.2 & 0.2 & 0 & 0.2 & 0.33 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0.2 & 0.33
\end{bmatrix}$$

Fig. 6. Stochastic matrix for graph C

$$\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0.5 & 0.5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0.5 & 0.5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Fig. 7. Convergent Matrix Formed from Stochastic matrix after applying MCL

to the all Communities C1 , C2 and C3. For community C1 and C2 Among top two important nodes , node1 and node2 is connecting by an edge. So we combine the community C1 and community C2 and Lets call that new community C^* . where C^* can be given by C^*= {1,2,3,5,6,7,10}. Now Top one Node for C^* is {1} and Top One node for community C3 is {9} or {11}. Now For Identifying top two node for graph C we select top one node from each community C^* and C3 . After That we can combine them to form top two nodes for the graph C. Hence the possible top two nodes for graph C are {1,9} or {1,11}. But from degree centrality on graph C topmost node centrality score is equal for five nodes and forms the set D = {1,4,5,8,9,11}. Any combination of two nodes from set D will give us the result for the best two nodes from conventional degree centrality, Which is different from the the top two nodes we find from community cebtrality based approach.

V. CONCLUSION

In this seminar problem of finding the centrality of a top-K node for a given graph is discussed. First Some basics of graph theory has been discussed. After that some of the

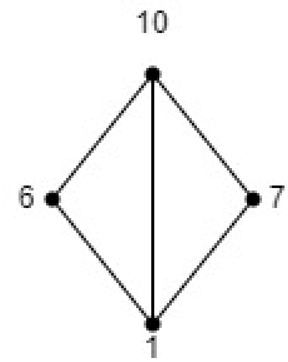

Fig. 8. Community C1

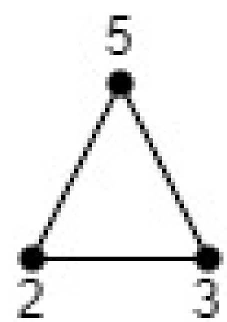

Fig. 9. Community C2

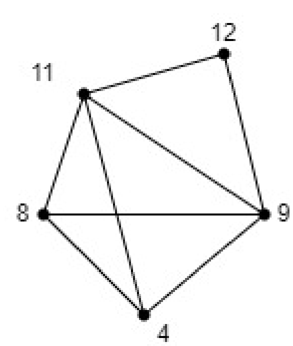

Fig. 10. Community C3

conventional centralities e.g., degree, closeness, betweeness, eigenvector , katz and pagerank centralities are discussed. Applying all these centralities to a graph, it was seen that nodes ranking according to centrality score may be different for every centrality depending upon the given graph. After that Community centrality based greedy approach has been discussed to find top-K important nodes. In the result It was found that centrality result from community centrality based greedy approach (CCGA) does not exactly match with conventional centralities. CCGA Does not give the most important nodes according to ranking of nodes based on centrality score of any conventional centralities.

REFERENCES

1)Identification of top-K nodes in large networks using Katz centrality Justin Zhan* , Sweta Gurung† and Sai Phani Krishna Parsa†

2)Community Centrality-Based Greedy Approach for Identifying Top-K Influencers in Social Networks Bundit Manaskasemsak(B) , Nattawut Dejkajonwuth, and Arnon Rungsawang

3)GRAPH CLUSTERING VIA A DISCRETE UNCOUPLING PROCESS STIJN VAN DONGEN†

4)THE THIRD SMALLEST EIGENVALUE OF THE LAPLACIAN MATRIX SUKANTA PATI

5)Manufacturing Consent Vivek S. Borkar, Fellow, IEEE, Aditya Karnik, Member, IEEE, Jayakrishnan Nair, and Sanketh Nalli, Student Member, IEEE

6) Basic Of Graph Theory