SOCIAL MEDIA ANALYSIS

Course: SC205

Course Instructor: Dr. Manish K. Gupta

Discrete Mathematics

DA-IICT

GANDHINAGAR

Made by: Aditya Zala

ID: 201901202

Contents

June 7, 2020

1 INTRODUCTION

Network theory is one of the most exciting and dynamic areas of science today with new Breakthroughs coming every Few years as we're piecing together a whole new way of looking at the world a true paradigm shift that is all about Connectivity. From the metabolic networks that fuel the cells in our body to the social networks that shape our lives Networks are everywhere. We see them in the rise of the internet the flow of global air traffic and in the spread of financial crises.Learning to model and design these networks is central to 21st century science and engineering.

To begin with, we need to have some idea about Network. There are several ways of formally defining a network, depending on the branch of mathematics used. The most usual and flexible definition is derived from graph theory, a social network is conceptualized as a graph, that is, a set of vertices representing social entities or objects and a set of lines representing one or more social relations among them.

A graph is made up of nodes; just like that a social media is a kind of a social network, where each person or organization represents a node. These nodes in a social media are interdependent on each other via common interests, relations, mutual friends, knowledge, common dislikes, beliefs etc. The overall graphical structure of a social media can be very complex with millions of nodes and thousands of interconnections amongst them based upon various grounds.

Thus here i have made my project where i have analyzed social media which is indeed a very useful tool for extracting knowledge from unstructured data using graph theory.The knowledge obtained from this field provides a vivid knowledge of various kinds interactions and relations amongst various individuals on social media.

2 NETWORKING OVERVIEW

Formally, a network N can be defined as N = (U, L, FU, FL) containing a graph G = (U, L), which is an ordered pair of a unit or vertex set U and a line set L, extended with a function FU specifying a vector of properties of the units (f: U X) and a function FL specifying a vector of properties of the lines (f: L Y). The set of lines L may be regarded as the union of a set of undirected edges E and a set of directed arcs A (L E = A). Each element e of E (each edge) is an unordered pair of units u and v (vertices) from U, that is, e(u: v), and each element a of A (each arc) is an ordered pair of units u and v (vertices) from U , that is, a(u: v).

Based on the contents of the nodes Network can be divided into two major types:

1. Social and Economic Network – It consists of a group of people connected with some sort of interactions or pattern of communication.
    - e.g. - Facebook, Twitter, business relation between companies and clients, interrelationship between families involved in a marriage etc.
2. Information Network -The connection between information objects.
    - e.g. – Semantic (links between various words and symbols), World Wide Web (link between various web pages; new page connecting to another through hyperlinks.

Here, we are going to concentrate only on Social Network and the working principle of Graph Theory on it.

2.1 BRIEF IDEA ON SOCIAL MEDIA

When we need to represent any form of relations in the society in the form of links, it can be termed as Social Network. The pattern of interdependency between each individual (node) can be based on different aspects, viz. - friendship, interconnection between families, common interest, financial exchange, dislike, or relationships of beliefs, knowledge or prestige.

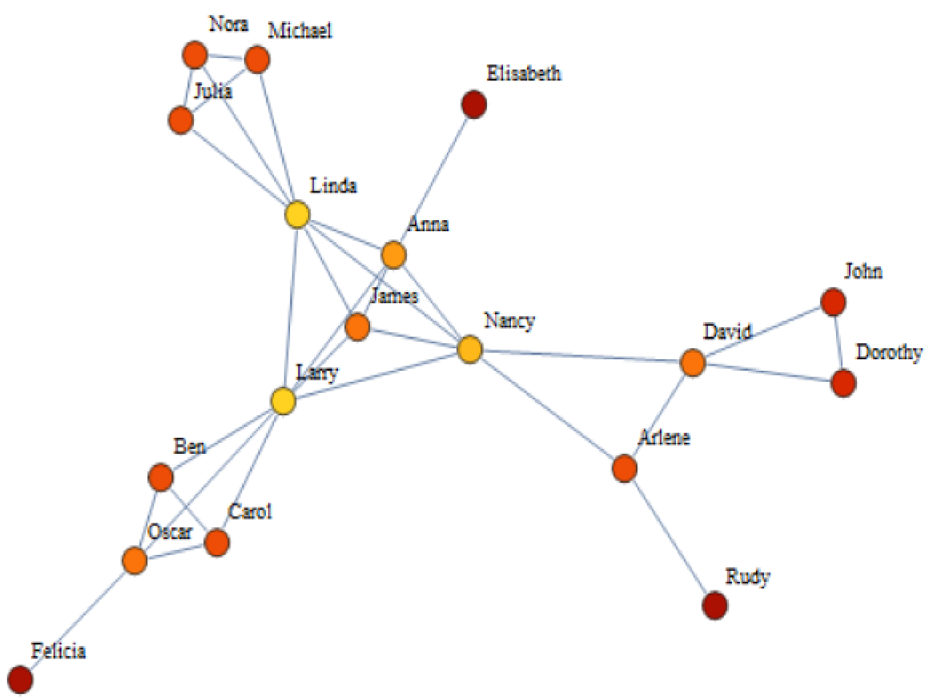


Figure 1: Network of friends

Figure 1 shows the relationship of a particular group of school friends. Each vertex represents a key person in the friendship, and the edges denote a connection by relationship and interaction of any kind that has been denoted by four different colors.

2.2 SOCIAL NETWORK ANALYSIS

Just like with classical graph theory, there are many points of view from which one can enter the study of social networks. One might come to social networks as a sociologist, or anthropologist, or linguist, or economist, or computer scientist, or engineer, or biologist, or businessperson, or investigator.

The notion of characterizing social networks by key properties is important when it comes to the question of modeling social networks. By modeling social networks, we mean finding an algorithmic way to generate graphs whose key properties are essentially the same as those of social networks found in nature. These methods typically depend upon a random process- maybe at each stage you add/remove a node or edge with some given probability. People have used the idea of "social network" loosely for over a century to connote complex sets of relationships between members of social systems at all scales, from interpersonal to international. Overall and Local Network Structure – These are the two perspective to analyze social networks. Overall Network structure concentrates on ties and interaction between persons or other social objects. This approach to social networks is known as the socio-centered approach. The other approach to social networks focuses on the individual element and its immediate network neighborhood and is analyzed through Local Network structure. This is known as the ego-centered approach.

In order to explain and analyze networks, we need to focus on the following topics:

1. Interpersonal Ties and bridges
2. Triadic Closure
3. Structural holes
4. Diameter and average path length
5. Connectedness
6. Degree Distributions
7. Density
8. Centrality

Now in the following section we will formulate and solve mathematics of each of the above mentioned topics for analyzing networks.

3 FORMULATING MATHEMATICS

In this section we would go to each of the 8 topics mentioned above and formulate and solve the mathematics involved.

3.1 Interpersonal Ties and bridges

In mathematical sociology, interpersonal ties define the type of connection between two or more people in a relationship. These ties are important and relevant in social network interactions and can be classified into 3 different type based on the strength of interaction: strong ties, weak ties and absent ties.

1. Strong ties: the stronger links, corresponding to friends, dependable sources of social or emotional support.
2. Weak ties: the weaker links, corresponding to acquaintances.
3. Absent ties: the one for which we have no information regarding its nature

Tie Strength in Large-Scale Data- Tie strength refers to a general sense of closeness with another person. Refer to the Figure 3, we extend the strong and weak ties as in a continuous quantity to measure the Neighborhood Overlap (NO) of an edge (x,y).

$$
\begin{aligned}
NO(x,y) &= \frac{|\text{common neighbors of } x \text{ and } y|}{|\text{neighbors of at least one of } x \text{ or } y|} \\
&= \frac{(N(x) \cap N(y)|}{|(N(x) - \{y\}) \cup (N(y) - \{x\})|}
\end{aligned}
$$

Figure 2:Neighbourhood overlap formula

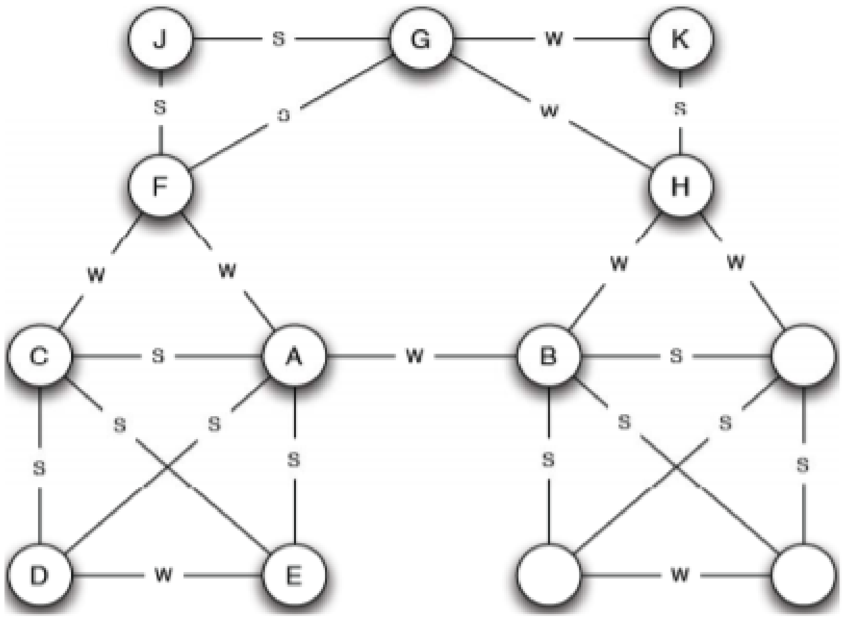


FIGURE 3: NO(A,B) = 0 and NO(A,F) = 1/6 in the above figure

Weak ties play a crucial role in binding a group of strong ties together. They help in strengthening the relationship and forming new bonds between existing relations. So, weak ties can operate as bridges between two groups. To fully understand what this means, one first have to understand the term bridge. A link between two nodes A and B is a bridge if deleting this link separates the nodes into two components. The bridge is the only way information can move from A to B. There are two types of bridges- local bridge and a regular bridge.

A local bridge is a link between two nodes which when broken increases the distance between those nodes to more than two. It's the shortest route through which information can flow from a group of nodes to another group. Based on the definition of *strong triadic closure, a local bridge is necessarily a weak tie.

The occurrence of a bridge in large social networks is rare due to the fact that in most cases there will be a different path that connects the nodes A and B together. Local bridges however are more often seen and can connect us with parts of the network that would otherwise been neglected.

***The Strong Triadic Closure Property–It is the property among three nodes A, B and C, such that if a strong tie exits between A-B and A-C, there is a weak or strong tie between B-C.**

Local bridges are especially those with large span.Local bridges are the edges of neighborhood overlap 0 -and hence we can think of edges with very small neighborhood overlap as being almost local bridges.

For example in figure 3 AB is the only local bridge

3.2 TRIADIC CLOSURE

The strength of a tie is a combination of services, the amount of time, the intimacy, and the emotional intensity between the subjects. There is always a possibility of forming new acquaintances when two persons have a strong connection to a third person, i.e.**triadic closure.**

Triadic Closure is a principle that implies that if two friends in social network have a friend in common, then there is increased likelihood that they will become friends too.This principle can explain the evolving of network over times in many situations.

CC(A) = P(B$\in N(C)|B, C\in N(A)$)

CC(A) = P(Two randomly selected friends of A are friends)

CC(A) = P(fraction of pairs of A's friends linked to each other)

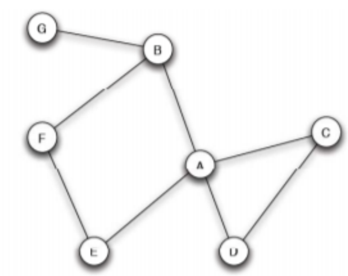


FIGURE 4(a):Before B-C edge forms

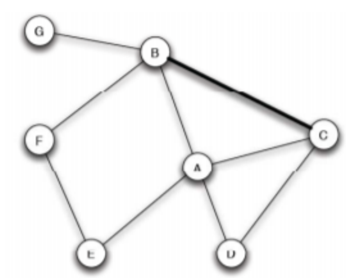


FIGURE 4(b):After B-C edge forms

The formation of the edge between B and C illustrates the effects of triadic closure,since they have a common neighbor A.For example, in Figure 4(a) CC(A) = 1/(4*3/2)=1/6

3.2.1 REASONS FOR TRIADIC CLOSURE

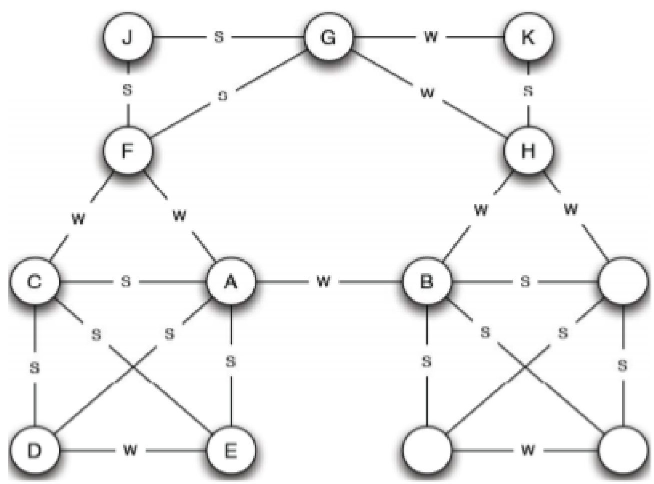
1. **Opportunity**: if A spends time with both B and C, then there is an increased chance that B and C will end up knowing each other and potentially becoming friends

2. **Trusting**: the fact that each of B and C is friends with A (provided they are mutually aware of this) gives them a basis for trusting each other that an arbitrary pair of unconnected people might lack.

3. **Incentive**: if A is friends with B and C, then it becomes a source of latent stress in these relationships if B and C are not friends with each other.

3.2.2 THE STRONG TRIADE CLOSURE PROPERTY

1. A node A violates the Strong Triadic Closure Property if it has strong ties to two non-linked nodes B and C.



2. No node in the above figure violates the Strong Triadic Closure Property

3. If we change AF and AB to strong ties, then it violates the Strong Triadic Closure Property due to the absence of link BF

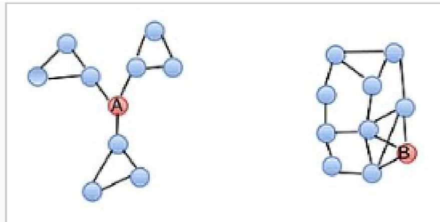
3.3 STRUCTURE HOLES

A structural hole refers to an "empty space" between contacts in a person's network. It means that these contacts do not interact closely (though they may be aware of one another). Actors on either side of the structural hole have access to different flows of information.The theory of structural holes was developed to explain how to benefit from competition in social networks and their intersecting relationships.

3.3.1 CONCEPT

Most social structures tend to be characterized by dense clusters of strong connections, also known as network closure. The theory relies on a fundamental idea that the homogeneity of information, new ideas, and behavior is generally higher within any group of people as compared to that in between two groups of people.[1] An individual who acts as a mediator between two or more closely connected groups of people could gain important comparative advantages. In particular, the position of a bridge between distinct groups allows him or her to transfer or gatekeep valuable information from one group to another. In addition, the individual can combine all the ideas he or she receives from different sources and come up with the most innovative idea among all.At the same time, a broker also occupies a precarious position, as ties with disparate groups can be fragile and time consuming to maintain.

If we compare two nodes, node A is more likely to get novel information than node B, even though they have the same number of links. This is so because nodes connected to B are also highly connected between each other. Therefore, any information that any of them could get from B, it could easily get from other nodes as well. Furthermore, the information, which B gets from different connections, is likely to be overlapping, so connections involving node B are said to be redundant. On contrary, the position of node A makes it serve as a bridge or a 'broker' between three different clusters. Thus, node A is likely to receive some non-redundant information from its contacts. The term 'structural holes' is used for the separation between non-redundant contacts. As a result of the hole between two contacts, they provide network benefits to the third party (to node A).



Two types of network structure

3.4 DIAMETER AND AVERAGE PATH LENGTH

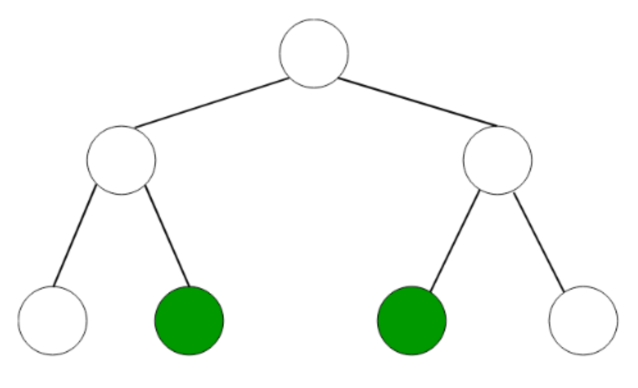
DIAMETER:

Diameter is an index measuring the topological length or extent of a graph by counting the number of edges in the shortest path between the most distant vertices. It is: maxs(i,j)where s(i, j) is the number of edges in the shortest path from vertex i to vertex j. With this formula, first, all the shortest paths between all the vertices are searched; then, the longest path is chosen. This measure therefore describes the longest shortest path between two random vertices of a graph.

3.4.1 CODE FOR CALCULATION OF DIAMETER

Now lets device a code for calculation of diameter in a tree which is one of the examples of a graph:

The diameter of a tree (sometimes called the width) is the number of nodes on the longest path between two leaves in the tree.



Diameter, 5 nodes through root

After observing above tree we can see that the longest path will always occur between two leaf nodes. We start DFS from a random node and then see which node is farthest from it. Let the node farthest be X. It is clear that X will always be a leaf node and a corner of DFS. Now if we start DFS from X and check the farthest node from it, we will get the diameter of the tree.

CODE:

The C++ implementation uses adjacency list representation of graphs. STL's list container is used to store lists of adjacent nodes.

```
// C++ program to find diameter of a binary tree
// using DFS.
#include <iostream>
#include <limits.h>
#include <list>
using namespace std;

// Used to track farthest node.
int x;

// Sets maxCount as maximum distance from node.
void dfsUtil(int node, int count, bool visited[],
                int& maxCount, list<int>* adj)
{
    visited[node] = true;
    count++;
    for (auto i = adj[node].begin(); i != adj[node].end(); ++i) {
        if (!visited[*i]) {
            if (count >= maxCount) {
                maxCount = count;
                x = *i;
            }
            dfsUtil(*i, count, visited, maxCount, adj);
        }
    }
}

// The function to do DFS traversal. It uses recursive
// dfsUtil()
void dfs(int node, int n, list<int>* adj, int& maxCount)
{
    bool visited[n + 1];
    int count = 0;
```

```
    // Mark all the vertices as not visited
    for (int i = 1; i <= n; ++i)
        visited[i] = false;

    // Increment count by 1 for visited node
    dfsUtil(node, count + 1, visited, maxCount, adj);
}

// Returns diameter of binary tree represented
// as adjacency list.
int diameter(list<int>* adj, int n)
{
    int maxCount = INT_MIN;

    /* DFS from a random node and then see
    farthest node X from it*/
    dfs(1, n, adj, maxCount);

    /* DFS from X and check the farthest node
    from it */
    dfs(x, n, adj, maxCount);

    return maxCount;
}
```

```
/* Driver program to test above functions*/
int main()
{
    int n = 5;

    /* Constructed tree is
        1
        / \
        2 3
    / \
    4 5 */
    list<int>* adj = new list<int>[n + 1];

    /*create undirected edges */
    adj[1].push_back(2);
    adj[2].push_back(1);
    adj[1].push_back(3);
    adj[3].push_back(1);
    adj[2].push_back(4);
    adj[4].push_back(2);
    adj[2].push_back(5);
    adj[5].push_back(2);


    /* maxCount will have diameter of tree */
    cout << "Diameter of the given tree is "
        << diameter(adj, n) << endl;
    return 0;
}
```

The Output Of Above Code:

Diameter of the given tree is 4

AVERAGE PATH LENGTH:

The average path length is the average distance between any two nodes in the network.

- Average path length is bounded from above by the diameter; in some cases, it can be much shorter than the diameter.
- If the network is not connected, one often checks the diameter and the average path length in the largest component

THE FORMULA FOR CALCULATING AVERAGE PATH LENGTH IS:-

Diameter= max I (i, j)

The average path length is the average distance between any two nodes in the network:

$$\textbf{Average path length} = \frac{\sum i \geq j \; I\,(i\,,j\,)}{\frac{n(n-1)}{2}}$$

3.5 CONNECTEDNESS

Whether it is possible to traverse a graph from one vertex to another is determined by how a graph is connected. Connectivity is a basic concept in Graph Theory. Connectivity defines whether a graph is connected or disconnected.Connectedness has a great importance while we study networks and is used in measuring centrality and clusterdness.

A graph is said to be connected if there is a path between every pair of vertex. From every vertex to any other vertex, there should be some path to traverse. That is called the connectivity of a graph. A graph with multiple disconnected vertices and edges is said to be disconnected.

3.5.1 CODE FOR CHECKING CONNECTIVITY OF GRAPH

To check connectivity of a graph, we will try to traverse all nodes using any traversal algorithm. After completing the traversal, if there is any node, which is not visited, then the graph is not connected.

```
#include<iostream>
#define NODE 5
using namespace std;
int graph[NODE][NODE] = {{0, 1, 1, 0, 0},
{1, 0, 1, 1, 0},
{1, 1, 0, 1, 1},
{0, 1, 1, 0, 1},
{0, 0, 1, 1, 0}};
void traverse(int u, bool visited[]) {
   visited[u] = true; //mark v as visited
   for(int v = 0; v<NODE; v++) {
      if(graph[u][v]) {
         if(!visited[v])
            traverse(v, visited);
      }
   }
}
bool isConnected() {
   bool *vis = new bool[NODE];
   //for all vertex u as start point, check whether all nodes are visible or not
   for(int u; u < NODE; u++) {
      for(int i = 0; i<NODE; i++)
         vis[i] = false; //initialize as no node is visited
         traverse(u, vis);
      for(int i = 0; i<NODE; i++) {
         if(!vis[i]) //if there is a node, not visited by traversal, graph is not connected
            return false;
      }
   }
   return true;
}
```

```
int main() {
    if(isConnected())
        cout << "The Graph is connected.";
    else
        cout << "The Graph is not connected.";
}
```

OUTPUT:

The Graph is connected.

3.6 DEGREE DISTRIBUTIONS

The degree of a node in a network (sometimes referred to incorrectly as the connectivity) is the number of connections or edges the node has to other nodes. If a network is directed, meaning that edges point in one direction from one node to another node, then nodes have two different degrees, the in-degree, which is the number of incoming edges, and the out-degree, which is the number of outgoing edges.

The degree distribution P(k) of a network is then defined to be the fraction of nodes in the network with degree k. Thus if there are n nodes in total in a network and n$_k ofthemhavedegreek, wehaveP(k) = n_k/n$.

The same information is also sometimes presented in the form of a cumulative degree distribution, the fraction of nodes with degree smaller than k, or even the complementary cumulative degree distribution, the fraction of nodes with degree greater than or equal to k (1 - C) if one considers C as the cumulative degree distribution; i.e. the complement of C.

3.6.1 OBSERVED DEGREE DISTRIBUTIONS

The degree distribution is very important in studying both real networks, such as the Internet and social networks, and theoretical networks. The simplest network model, for example, the (Erdős–Rényi model) random graph, in which each of n nodes is independently connected (or not) with probability p (or 1 p), has a binomial distribution of degrees k:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k},$$

(or Poisson in the limit of large *n*, if the average degree $\langle k \rangle = p(n-1)$ is held fixed). Most networks in the real world, however, have degree distributions very different from this. Most are highly right-skewed, meaning that a large majority of nodes have low degree but a small number, known as "hubs", have high degree. Some networks, notably the Internet, the world wide web, and some social networks were argued to have degree distributions that approximately follow a power law: $P(k) \sim k^{-\gamma}$, where γ is a constant. Such networks are called scale-free networks and have attracted particular attention for their structural and dynamical properties[1][2][3][4]. However, recently, there have been some researches based on real-world data sets claiming despite the fact that most of the observed networks have fat-tailed degree distributions, they deviate from being scale-free.[5]

3.7 DENSITY

A dense graph is a graph in which the number of edges is close to the maximal number of edges. The opposite, a graph with only a few edges, is a sparse graph. Thus in network system density is of great importance as it gives us an idea of which area has more weightage over others.For example in case of a disease outbreak by looking at a graph we can see that which areas are more affected than others just by looking how dense a specific area of a graph is.

3.7.1 MATHEMATICAL FORMULA

Network density is measured as the number of possible or potential connections (i.e., edges), over the number of actual connections. Density values range between zero and one, and can be thought of as the percent of all possible edges that are realized.

(density) = (the number of edges) / (the number of possible edges)

For undirected simple graphs, the graph density is:-

$$D = \frac{|E|}{\binom{|V|}{2}} = \frac{2|E|}{|V|(|V|-1)}$$

For directed simple graphs, the maximum possible edges is twice that of undirected graphs to account for the directedness, so the density is:

$$D = \frac{|E|}{2\binom{|V|}{2}} = \frac{|E|}{|V|(|V|-1)}$$

3.8 CENTRALITY

Social network theory is becoming more and more significant in social science, and the centrality measure is underlying this burgeoning theory. In perspective of social network, individuals, organizations, companies etc. are like nodes in the network, and centrality is used to measure these nodes' power, activity, communication convenience and so on. Meanwhile, degree centrality, betweenness centrality and closeness centrality are the popular detailed measurements.

Centrality is such an important index because it indicates which node takes up critical position in one whole network. Central positions always get equated with remarkable leadership, good popularity or excellent reputation in the network. As soon as the social actor gets a higher centrality, it mean he/she gets closer to the center of network, that higher power, influence, convenience from the network he/she may acquire

Now we will look at each of the 3 centrality mentioned above in detail:-

3.8.1 DEGREE CENTRALITY

Degree is a simple centrality measure that counts how many neighbors a node has. If the network is directed, we have two versions of the measure: in-degree is the number of in-coming links, or the number of predecessor nodes; out-degree is the number of out-going links, or the number of successor nodes. Typically, we are interested in in-degree, since in-links are given by other nodes in the network, while out-links are determined by the node itself.

Degree centrality thesis reads as follows:

Node is important if it has many neighbors, or, in the directed case, if there are many other nodes that link to it, or if it links to many other nodes

MATH OF DEGREE CENTRALITY:

Let $A = (a_{i,j})$ **be the adjacency matrix of a directed graph. The in-degree centrality** x_i **of node** i **is given by:**

$$x_i = \sum_k a_{k,i}$$

or in matrix form (1 is a vector with all components equal to unity):

$$x = 1A$$

The out-degree centrality y_i of node i is given by:

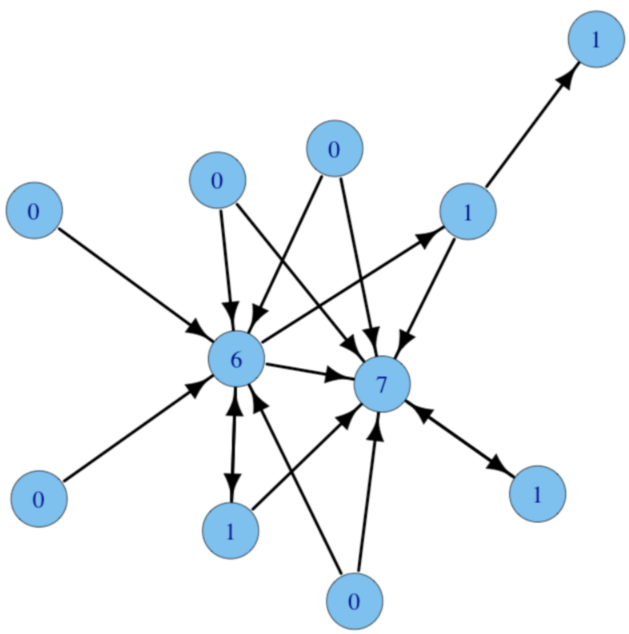
$$y_i = \sum_k a_{i,k}$$

or in matrix form:

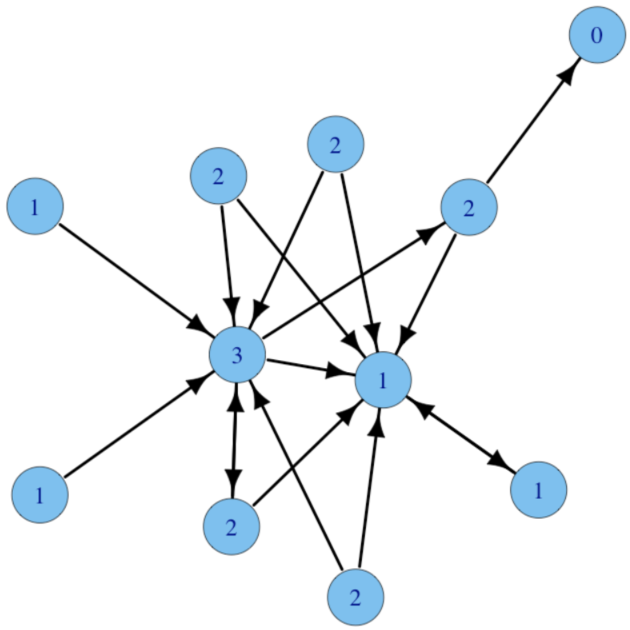
$$y = A1$$

<u>CODE FOR DEGREE CENTRALITY:</u>

```
# Degree centrality
# INPUT
# g = graph
# mode = degree mode ("in" for in-degree, "out" for out-degree, "all" for total degree)
degree.centrality = function(g, mode) {
  A = get.adjacency(g);
  o = rep(1, n);
  if (mode == "in") c = o %*% A else
    if (mode == "out") c = A %*% o else
      c = o %*% (A + t(A));
  return(as.vector(c));
}
```



A graph with nodes labelled with their in-degree centrality



the same graph with nodes labelled with their out-degree centrality:

3.8.2 CLOSENESS CENTRALITY

Closeness centrality is a measure of how long will it take to spread something such as information form the node of interest to all the other nodes sequentially.It is the measurement of node's capacity to effect all other elements in the network.

Node with highest closeness centrality is closest to all other nodes or rather closeness centrality is the average length of the shortest path between the specific nodes to all other nodes.

Closeness centrality measures the mean distance from a vertex to other vertices. Recall that a geodesic path is a shortest path through a network between two vertices. Suppose $d_{i,j}$ is the length of a geodesic path from i to j, meaning the number of edges along the path. Then the mean geodesic distance for vertex i is:

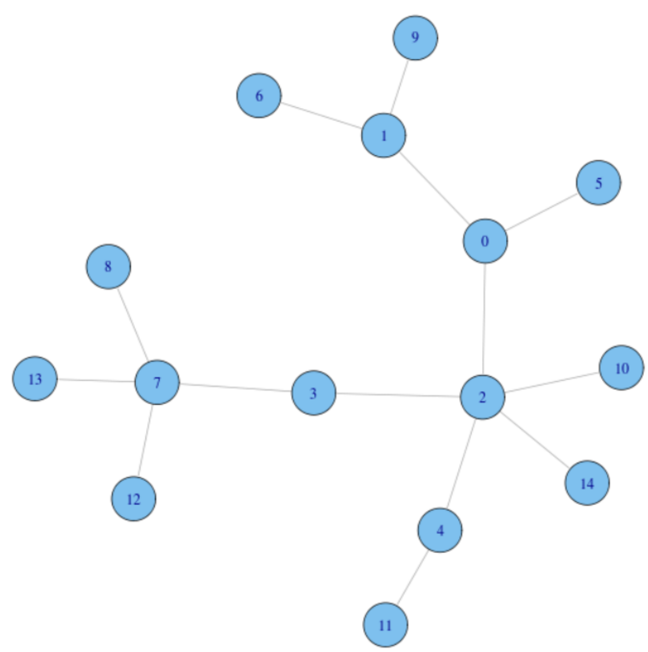
$$l_i = \frac{1}{n}\sum_j d_{i,j}$$

The mean distance l_i is not a centrality measure in the sense of the previous ones, since it gives low values to more central nodes and high values to less central ones, which is the opposite of other centrality measures. In the social network literature, therefore, researchers commonly calculate its inverse, called closeness centrality:]This quantity takes low values for vertices that are separated from others by only a short geodesic distance on average. Such vertices might have better access to information at other vertices or more direct influence on other vertices. In a social network, for instance, a person with lower mean distance to others might find that their opinions reach others in the community more quickly than the opinion of someone with higher mean distance. Some authors exclude from the sum the term $j = i$ for which $d_{i,i} = 0$ and hence divide the sum for $n - 1$ instead of n.

The mean distance l_i is not a centrality measure in the sense of the previous ones, since it gives low values to more central nodes and high values to less central ones, which is the opposite of other centrality measures. In the social network literature, therefore, researchers commonly calculate its inverse, called closeness centrality:

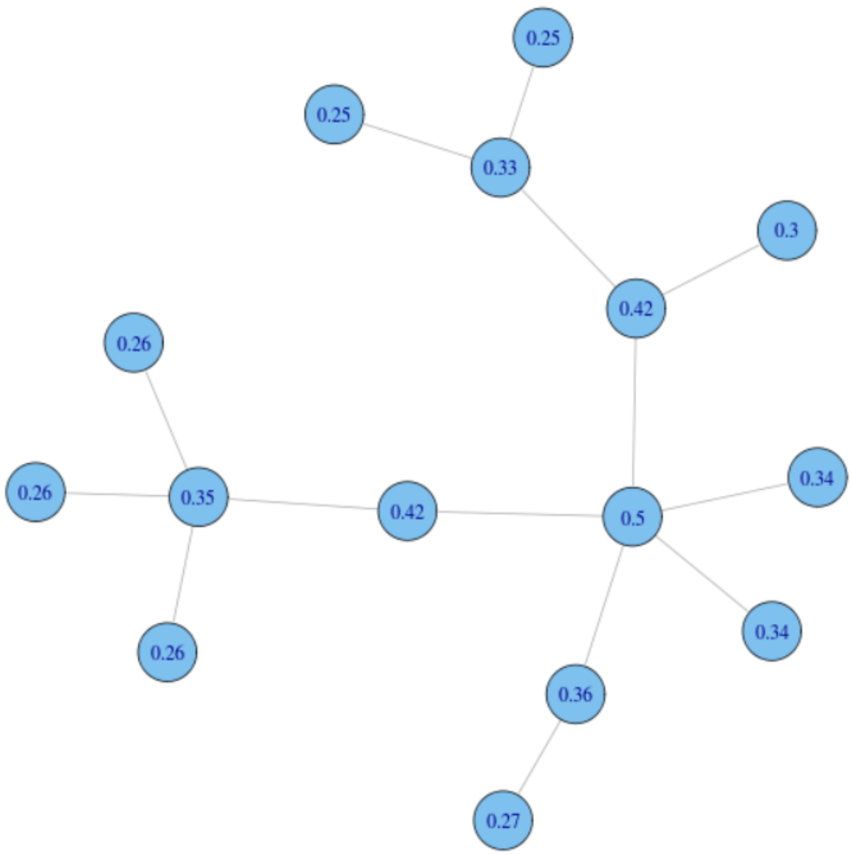
$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{i,j}}$$

Consider the following simple network:



Function closeness (R, C) computes closeness centrality (this function uses $n-1$ as the numerator of the formula and assigns a distance equal

to the number of nodes of the graph to pairs of nodes that are not reachable). This is the same network with nodes labelled with their closeness centrality (the scores are rounded to the second decimal digit):



3.8.3 BETWEENNESS CENTRALITY

Betweenness centrlaity is the measure of how often a node acts as a bridge between other nodes.Vertices that have a high probability of occuring on a randomly chosen shortest path between two vertices can be said to have a higher degree of betweenness centrality.

BS CENTRALITY OF NODE X=

$$\frac{FRACTIONOFSHORTESTPATHTHATGOESTHROUGHNODEX}{ALLSHORTESTPATHBETWEENEVERYNODE}$$

CALCULATING BETWENNESS CENTRALITY:-

Betweenness centrality $C_B(v)$ for a vertex v is defined as

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

where σ_{st} is the number of shortest paths with vertices and as their end vertices, while σ_{st}(v)is the number of those shortest paths that include vertex . High centrality scores indicate that a vertex lies on a considerable fraction of shortest paths connecting pairs of vertices.

- Every pair of vertices in a connected graph provides a value lying in (0,1) to the betweenness centrality of all other vertices.
- If there is only one geodesic joining a particular pair of vertices, then that pair provides a betweenness centrality 1 to each of its intermediate vertices and zero to all other vertices. For example, in a path graph, a pair of vertices provides a betweenness centrality 1 to each of its interior vertices and zero to the exterior vertices. A pair of adjacent vertices always provides zero to all others.
- If there are geodesics of length 2 joining a pair of vertices, then that pair of vertices provides a betweenness centrality to each of the intermediate vertices.

RELATIVE BETWEENNESS CENTRALITY:

The betweenness centrality increases with the number of vertices in the network, so a normalized version is often considered with the centrality values scaled to between 0 and 1. Betweenness centrality can be normalized by dividing C$_B(v) byitsmaximumvalue.$

$$C'_B(v) = \frac{C_B(v)}{\operatorname{Max} C_B(v)} = \frac{2C_B(v)}{(n-1)(n-2)} \quad 0 \leq C'_B(v) \leq 1.$$

BETWEENNESS CENTRALITY OF A GRAPH:

The betweenness centrality of a graph measures the tendency of a single vertex to be more central than all other vertices in the graph. It is based on differences between the centrality of the most central vertex and that of all others. The betweenness centrality of a graph is defined as the average difference between the measures of centrality of the most central vertex and that of all other vertices.

The betweenness centrality of a graph G is defined as

$$C_B(G) = \frac{\sum_{i=1}^{n}[C_B(v^*) - C_B(v_i)]}{\text{Max}\sum_{i=1}^{n}[C_B(v^*) - C_B(v_i)]}, \tag{3}$$

where $C_B(v^*)$ is the largest value of $C_B(v_i)$ for any vertex v_i in the given graph G and Max $\sum_{i=1}^{n}[C_B(v^*) - C_B(v_i)]$ is the maximum possible sum of differences in centrality for any graph of n vertices which occur in star with the value $n-1$ times $C_B(v)$ of the central vertex, that is, $(n-1)\binom{n-1}{2}$.

4 CASE STUDY:NETWORKS ON FACEBOOK

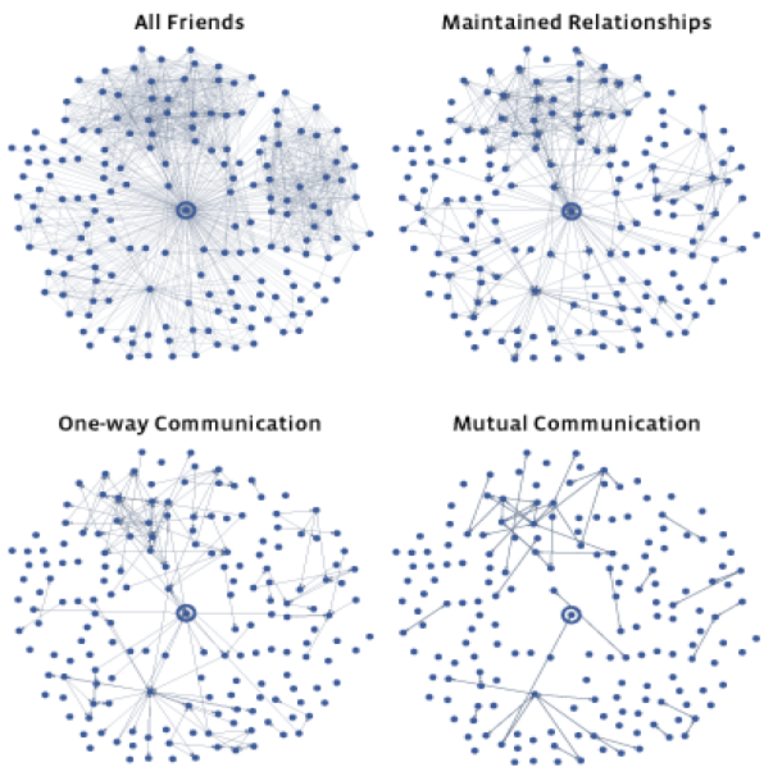
This case study depicts the size of Facebook networks. Few random users have been studies for the course of a month and the relationship is classified into four different network patterns:

1. **All Friends:** This network represents the list of all friends a user has, hence this is the largest among all the representations.

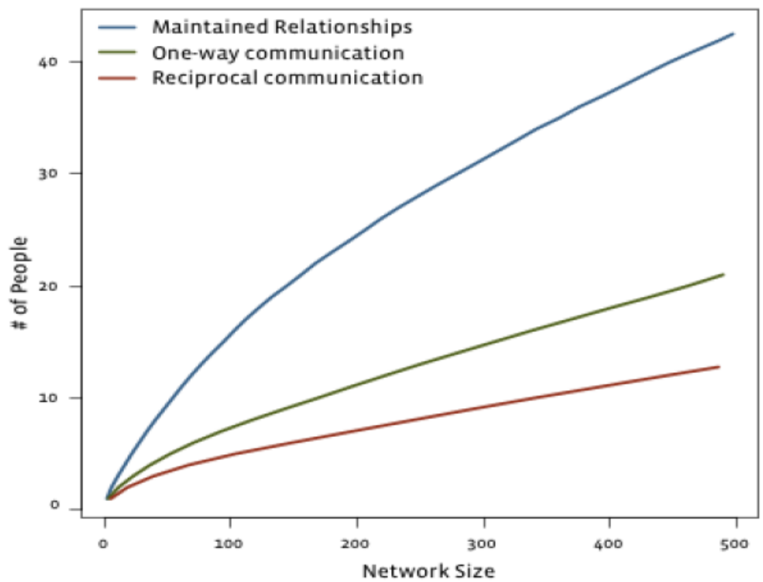
2. **Reciprocal Communication:**This representation shows the mutual communication between two parties, this type of network forms when there is mutual exchange of information between two parties.

3. **One-way Communication:** It consists of people with whom a user has communicated.

4. **Maintained Relationships:**This pattern of relationship consists of people whose profile has been checked by the user more than once to maintain engagement.



NETWORK PATTERNS OF A USER ON A FACEBOOK



In the diagram, the red line shows the number of reciprocal relationships, the green line shows the one-way relationships, and the blue line shows the passive relationships is a function of your network size.

5 FURTHER APPLICATIONS

Well on the basis of graph theory there can be various applications that can be implemented other than social media:

1. **Google maps** uses graphs for building transportation systems, where intersection of two(or more) roads are considered to be a vertex and the road connecting two vertices is considered to be an edge, thus their navigation system is based on the algorithm to calculate the shortest path between two vertices.

2. In **Facebook**, users are considered to be the vertices and if they are friends then there is an edge running between them. Facebook's Friend suggestion algorithm uses graph theory. Facebook is an example of undirected graph.

3. In **Operating System**, we come across the Resource Allocation Graph where each process and resources are considered to be vertices. Edges are drawn from resources to the allocated process, or from requesting process to the requested resource. If this leads to any formation of a cycle then a deadlock will occur.

6 REFERENCES

The References used were:

1:NETWORK THEORY COURSE:

https://www.youtube.com/watch?v=7HkXkAZye1Ylist=PLsJWgOB5mIMAuH3cHa-MXukX6-RPpDXglindex=1

2:SOCIAL NETWORKS COURSE NPTEL:

https://www.youtube.com/watch?v=v9GQyenwwzwlist=PLyqSpQzTE6M8CLBcLnq-f3vHRH-klC39L

3:ARTICLE ON SOCIAL MEDIA:

https://towardsdatascience.com/how-to-visualize-social-network-with-graph-theory-4b2dc0c8a99f

4:FACEBOOK API:

https://medium.com/tow-center/the-graph-api-key-points-in-the-facebook-and-cambridge-analytica-debacle-b69fe692d747