\_\_Author\_\_=’Anand Rawat’

Theory for reference, digital media project report

<https://economics.mit.edu/files/4620>

<http://www.gitta.info/Accessibiliti/en/html/StructPropNetw_learningObject2.html>

<https://bookdown.org/omarlizardo/_main/2-9-density.html>

<https://www.sciencedirect.com/topics/computer-science/network-density>

<https://en.wikipedia.org/wiki/Modularity_(networks)>

<https://www.cs.cmu.edu/~ckingsf/bioinfo-lectures/modularity.pdf>

<https://en.wikipedia.org/wiki/Eigenvector_centrality>

<https://www.sci.unich.it/~francesc/teaching/network/eigenvector.html>

<https://www.cambridge.org/core/journals/combinatorics-probability-and-computing/article/first-passage-percolation-on-the-erdsrenyi-random-graph/67BD7AD45CF42999C3C354ADF26A2C33>

<https://en.wikipedia.org/wiki/Network_theory>

<https://www.pnas.org/content/104/18/7332>

<https://www.facebook.com/notes/facebook-data-science/maintained-relationships-on-facebook/55257228858/>

<https://ndlib.readthedocs.io/en/latest/reference/models/epidemics/Threshold.html>

<https://www.researchgate.net/publication/221611895_Measurement_and_Analysis_of_Online_Social_Networks>

<https://www.ncbi.nlm.nih.gov/pubmed/17456605>

<https://www.danah.org/researchBibs/twitter.php>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029358>

<https://www.usenix.org/legacy/event/wosn10/tech/full_papers/Galuba.pdf>

**PAGE RANK: It can be considered as the important score of a web page or social network node. This important score will always be a non-negative real number and all the scores will add to 1, sometimes it might be presented as a percentage. This score is based on the links made to that page/node from another page/node. The link to a given page per node are called backlinks/ in-degrees for that page/nodes vote for the importance of the other pages by linking to them.**

**It is variant of eigenvector value, but because it uses backlinks/in -degrees it is used in directed networks.**

**Damping factor is d and Epsilon is the stopping criterion value, it determines the stopping point for the algorithm.**

**Cascades:**

An Information **cascade** or informational **cascade** is a phenomenon described in behavioral economics and **network** theory in which several people make the same decision in a sequential fashion. It is similar to, but distinct from herd behavior. An information **cascade** is generally accepted as a two-step process.

The Threshold model was introduced in 1978 by Granovetter.

In this model during an epidemic, a node has two distinct and mutually exclusive behavioral alternatives, e.g., the decision to do or not do something, to participate or not participate in a riot.

Node’s individual decision depends on the percentage of its neighbors have made the same choice, thus imposing a threshold.

The model works as follows: - each node has its own threshold; - during a generic iteration every node is observed: if the percentage of its infected neighbors is grater than its threshold it becomes infected as well.

[**https://www.usenix.org/legacy/event/wosn10/tech/full\_papers/Galuba.pdf**](https://www.usenix.org/legacy/event/wosn10/tech/full_papers/Galuba.pdf)

Microblogging sites are a unique and dynamic Web 2.0 communication medium. Understanding the information flow in these systems can not only provide better insights into the underlying sociology but is also crucial for applications such as content ranking, recommendation and filtering, spam detection and viral marketing. In this paper, we characterize the propagation of URLs in the social network of Twitter, a popular microblogging site. We track 15 million URLs exchanged among 2.7 million users over a 300-hour period. Data analysis uncovers several statistical regularities in the user activity, the social graph, the structure of the URL cascades and the communication dynamics. Based on these results we propose a propagation model that predicts which users are likely to mention which URLs. The model correctly accounts for more than half of the URL mentions in our data set, while maintaining a false positive rate lower than 15%

When @alice, a Twitter user, mentions a URL, all her followers can immediately see her tweet in their feeds. When @bob, a follower of @alice, sees a URL coming from her, he can also tweet that URL to let his followers know about it. This is one of the common mechanisms through which URLs propagate in the Twitter social network.

**ROBUSTNESS:**

The structural robustness of networks is studied using [percolation theory](https://en.wikipedia.org/wiki/Percolation_theory) When a critical fraction of nodes (or links) is removed randomly (random failures), the network becomes fragmented into small disconnected clusters. This phenomenon is called percolation, and it represents an order-disorder type of [phase transition](https://en.wikipedia.org/wiki/Phase_transition) with [critical exponents](https://en.wikipedia.org/wiki/Critical_exponents). Percolation theory can predict the size of the largest component (called giant component), the critical percolation threshold and the critical exponents. The failures discussed above are random, as usually assumed in percolation theory. However, when generalizing percolation also to non-random but targeted attacks, e.g., on highest degree nodes, the results, such as p\_c, change significantly. Recently, a new type of failures in networks has been developed, called localized attacks. In this case one chose randomly a node and remove its neighbors and next nearest neighbors until a fraction of 1-p nodes are removed. One such realistic example of random percolation is the use of percolation theory to predict the fragmentation of biological virus shells (capsids), with the percolation threshold of Hepatitis B Virus capsid predicted and detected experimentally: a molecular, randomly played game of Jenga on a rhombically tiled sphere.

Electronic databases, from phone to e-mails logs, currently provide detailed records of human communication patterns, offering novel avenues to map and explore the structure of social and communication networks. Here we examine the communication patterns of millions of mobile phone users, allowing us to simultaneously study the local and the global structure of a society-wide communication network. We observe a coupling between interaction strengths and the network's local structure, with the counterintuitive consequence that social networks are robust to the removal of the strong ties but fall apart after a phase transition if the weak ties are removed. We show that this coupling significantly slows the diffusion process, resulting in dynamic trapping of information in communities and find that, when it comes to information diffusion, weak and strong ties are both simultaneously ineffective.

Uncovering the structure and function of communication networks has always been constrained by the practical difficulty of mapping out interactions among many individuals. Indeed, most of our current understanding of communication and social networks is based on questionnaire data, reaching typically a few dozen individuals and relying on the individual's opinion to reveal the nature and the strength of the ties. The fact that currently an increasing fraction of human interactions are recorded, from e-mail to phone records , offers unprecedented opportunities to uncover and explore the large-scale characteristics of communication and social networks. Here we take a first step in this direction by exploiting the widespread use of mobile phones to construct a map of a society-wide communication network, capturing the mobile interaction patterns of millions of individuals. The data set allows us to explore the relationship between the topology of the network and the tie strengths between individuals, information that was inaccessible at the societal level before. We demonstrate a local coupling between tie strengths and network topology and show that this coupling has important consequences for the network's global stability if ties are removed, as well as for the spread of news and ideas within the network.

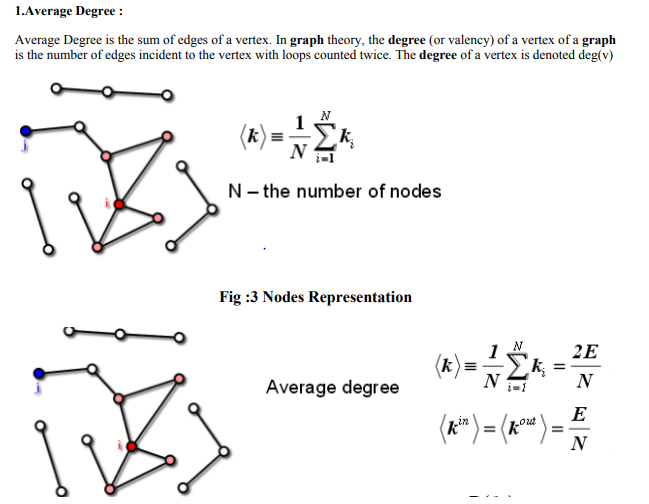
The social networks we study in this paper exist in the databases of online social networking sites. However, other online social networks are implemented as overlay networks. For instance, the graph formed by people who exchange email, or the graph formed by Skype users who include each other in their contact lists can be viewed as another social network on top of the Internet. If future distributed online social networks are popular and bandwidth-intensive, they can have a significant impact on Internet traffic, just as current peer-to-peer content distribution networks do. Understanding the structure of online social networks is not only critical to understanding the robustness and security of distributed online social networks, but also to understanding their impact on the future Internet.

**AVERAGE DEGREE:** The **average degree** of an undirected **graph** is used to measure the number of edges compared to the number of nodes. To do this we simply divide the summation of all nodes' **degree** by the total number of nodes.

Average degree is simply the average number of edges per node in the graph. It is relatively straightforward to calculate.

Total Edges/Total Nodes=Average Degree

It doesn’t really make sense to talk about the average degree in a directed network. This is because the direction of the ties is likely to be meaningful. Instead, what is likely of theoretical interest is the in-degree and out-degree. Additionally, because for every tie in the network there is a sender and a receiver, any attempt to calculate the average in- or out-degree will result in the same answer as the average degree calculation.



**AVERAGE WEIGHTED DEGREE:**

**Average** sum of weights of the edges of nodes. The **graph** is designed in such a way that, **weight** of an edges represents, how many times that edges is traversed between a pair of nodes. If **weight** of node is higher, it means it has been visited many times than any other low **weight degree** node.

Average sum of weights of the edges of nodes. The graph is designed in such a way that, weight of an edges represents, how many times that edges is traversed between a pair of nodes. If weight of node is higher, it means it has been visited many times than any other low weight degree node. The weighted degree of a node is like the degree. It’s based on the number of edges for a node but pondered by the weight of each edge. It’s doing the sum of the weight of the edges. For example, a node with 4 edges that weight 1 (1+1+1+1=4) is equivalent to: • a node with 2 edges that weight 2 (2+2=4) or • a node with 2 edges that weight 1 and 1 edge that weight 2 (1+1+2=4) or • a node with 1 edge that weight 4 etc.

**Betweenness centrality :**

To visualize the concept, you take all the shortest path from all nodes to all nodes on your graph. In each path, if one node is traveled, you add him « one point ». When all the paths are done, you have a ranking where some nodes are traveled a lot and some a few (almost never). This describes the Betweenness centrality; if a node has a high number, it has a high betweenness centrality.

If you are navigating onto the graph, you will most probably be traveled nodes that have a high Betweenness. And if you remove theses nodes first, there is a high probability to cut your graph into multiple unconnected components.

**NETWORK DIAMETER:**

The diameter of a graph is the maximum eccentricity of any vertex in the graph. That is, it is the greatest distance between any pair of vertices. To find the diameter of a graph, first find the shortest path between each pair of vertices. The greatest length of any of these paths is the diameter of the graph.

Diameter, D, of a network having N nodes is defined as the maximum shortest paths between any two nodes in the network

Diameter, D, of a network having N nodes is defined as the longest path, p, of the shortest paths between any two nodes D ¼ max (minp[pij length( p)). In this equation, pij is the length of the path between nodes i and j and length (p) is a procedure that returns the length of the path, p. For example, the diameter of a 4 4 Mesh D ¼ 6.

As another means of measuring network graphs, we can define the diameter of a network as the longest of all the calculated shortest paths in a network. It is the shortest distance between the two most distant nodes in the network. In other words, once the shortest path length from every node to all other nodes is calculated, the diameter is the longest of all the calculated path lengths. The diameter is representative of the linear size of a network. If node A-B-C-D are connected, going from A->D this would be the diameter of 3 (3-hops, 3-links)

# Diameter of a graph

Another measure for the structure of a graph is its diameter. Diameter δ 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:  
  
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.

## **GRAPH DENSITY:**

## The density of a graph is a measure of how many ties between actors exist compared to how many ties between actors are possible. As such, the density of an undirected graph is quite simply calculated as,

UndirectedNetworkDensity=TotalEdgesTotalPossibleEdges=CardinalitySize=mn(n−1)/2(1)(1)UndirectedNetworkDensity=TotalEdgesTotalPossibleEdges=CardinalitySize=mn(n−1)/2where n is the number of nodes in the network.

### **Graph density:**

Twitter networks vary in terms of their interconnectedness. Some networks are more tightly interconnected, by mentioning and replying to one another. In other networks, users are only sparsely connected, rarely mentioning or replying to others. 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. The calculation is a slightly different for directed and [undirected networks](https://www.sciencedirect.com/topics/computer-science/undirected-network), as directed networks have twice as many possible edge (i.e., from vertex A to vertex B, and from vertex B to vertex A).

The extent to which a network is densely interconnected impacts the rate of information flow within it. Interaction between individuals leads to shared knowledge, and shared knowledge leads to even more interaction. Granovetter states that tightly interconnected individuals are typically connected by strong and redundant ties. Burt notes that networks in which people are very highly interconnected are better at transmitting information. Others also demonstrate an important outcome of strongly embedded close relationships as an increase in trust between individuals, which can lead to increased information transfer. On Twitter the rate at which information is spread through a network was found to depend on its density; the greater the density, the faster information spreads.

What is the density of your network?

Your network size often affects whether the density value is high or low. If you have a large network, and chances are you do, the density value will be rather low. As a network grows, its density is likely to shrink. Like the discussion about reciprocity earlier, density should be evaluated in comparison with other networks. Within a network, it is hard to determine whether a density value should be considered low or high. Often, analysts collect several datasets of the same topic-networks over time, which allows them to evaluate a change in density. Later in this chapter, networks clusters will be discussed. As a network may contain several clusters, this will give us an opportunity to compare clusters, in terms of their density.

**Clustering coefficient:**

The clustering coefficient is a measure of an "all-my-friends-know-each-other" property. This is sometimes described as the friends of my friends are my friends. More precisely, the clustering coefficient of a node is the ratio of existing links connecting a node's neighbors to each other to the maximum possible number of such links. The clustering coefficient for the entire network is the average of the clustering coefficients of all the nodes. A high clustering coefficient for a network is another indication of a [small world](https://en.wikipedia.org/wiki/Small-world_experiment).

**MODULARITY:**

**Modularity** is one measure of the structure of [networks](https://en.wikipedia.org/wiki/Complex_network) or [graphs](https://en.wikipedia.org/wiki/Graph_(discrete_mathematics)). It was designed to measure the strength of division of a network into modules (also called groups, clusters or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. Modularity is often used in optimization methods for detecting [community structure](https://en.wikipedia.org/wiki/Community_structure) in networks. However, it has been shown that modularity suffers a resolution limit and, therefore, it is unable to detect small communities. Biological networks, including animal brains, exhibit a high degree of modularity.

Modularity is the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random. The value of the modularity for unweighted and undirected graphs lies in the range [ -1/2,1]{\displaystyle [-1/2,1]}[]. It is positive if the number of edges within groups exceeds the number expected based on chance. For a given division of the network's vertices into some modules, modularity reflects the concentration of edges within modules compared with random distribution of links between all nodes regardless of modules.

High modularity ⇒ more edges within the module that you expect by chance

**PATH LENGTH:**

In a graph, a **path** is a sequence of nodes in which each node is connected by an edge to the next. The path length corresponds to the number of edges in the path.

**EIGENVECTOR CENTRALITY:**

In [graph theory](https://en.wikipedia.org/wiki/Graph_theory), **eigenvector centrality** (also called **eigen centrality** or **prestige score**[[1]](https://en.wikipedia.org/wiki/Eigenvector_centrality#cite_note-:0-1)) is a measure of the influence of a [node](https://en.wikipedia.org/wiki/Node_(networking)) in a [network](https://en.wikipedia.org/wiki/Network_(mathematics)). Relative scores are assigned to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. A high eigenvector score means that a node is connected to many nodes who themselves have high scores.

A natural extension of degree centrality is **eigenvector centrality**. In-degree centrality awards one centrality point for every link a node receives. But not all vertices are equivalent: some are more relevant than others, and, reasonably, endorsements from important nodes count more. The eigenvector centrality thesis reads:

*A node is important if it is linked to by other important nodes.*

Eigenvector centrality differs from in-degree centrality: a node receiving many links does not necessarily have a high eigenvector centrality (it might be that all linkers have low or null eigenvector centrality). Moreover, a node with high eigenvector centrality is not necessarily highly linked (the node might have few but important linkers).

[**Erdős–Rényi model**](https://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93R%C3%A9nyi_model)**:**

In the mathematical field of [graph theory](https://en.wikipedia.org/wiki/Graph_theory), the **Erdős–Rényi model** is either of two closely related models for generating [random graphs](https://en.wikipedia.org/wiki/Random_graph). They are named after mathematicians [Paul Erdős](https://en.wikipedia.org/wiki/Paul_Erd%C5%91s) and [Alfréd Rényi](https://en.wikipedia.org/wiki/Alfr%C3%A9d_R%C3%A9nyi), who first introduced one of the models in 1959,[[1]](https://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93R%C3%A9nyi_model#cite_note-er59-1)[[2]](https://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93R%C3%A9nyi_model#cite_note-b01-2) while [Edgar Gilbert](https://en.wikipedia.org/wiki/Edgar_Gilbert) introduced the other model contemporaneously and independently of Erdős and Rényi.[[3]](https://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93R%C3%A9nyi_model#cite_note-g59-3) In the model of Erdős and Rényi, all graphs on a fixed vertex set with a fixed number of edges are equally likely; in the model introduced by Gilbert, each edge has a fixed probability of being present or absent, [independently](https://en.wikipedia.org/wiki/Statistical_independence) of the other edges. These models can be used in the [probabilistic method](https://en.wikipedia.org/wiki/Probabilistic_method) to prove the existence of graphs satisfying various properties, or to provide a rigorous definition of what it means for a property to hold for [almost all](https://en.wikipedia.org/wiki/Almost_all) graphs.