Network theory

1.Nodes

2. Connections

3. Quantify elements: a. ascribe a value. b.how connected is any node c. How central is it within the overall network

4. Overall structure : dynamics of the network on the global level feeds back to effect the element on the local level

5. Network density: a. Connnected b. density. c. Distribution —density of the network vary greatly d. patterns of clustering

Q: what happens to these properties if we change some parameter

6. Designed network:

Identify different types of networks

Designed—Most networks aren’t randomly generated—sometimes networks are specifically designed n atop-down fashion

Not designed—Network emerges out of local rules and interactions

7. Network diffusion(传播)

influence factor

Diffusion may b desirable or undesirable

How susceptible is network to failure both from random and strategic attack

8. Network dynamics

How network change over time—lifecycle of the network

9. Main area

A. Graph theory

B. Overall structure

C. types of network

d.network dynamics

Graph theory

1.

Multiple graph: a graph with multiple edges between nodes.—show all the different connections between cities

Simple graph:doesn’t contain loops or multiple edges.—whether there is a connection between cities

2.

Directed graph: have a direction denoted with arrows—have some order to the relations between the nodes.

Undirected graph: unordered pairs of nodes—switch them around

3.

Weighted graph: a number(weight) is assigned to each edge—the weight quantify the degree of interaction between the nodes or the volume of exchange(trading route: ascribe the quantitative value to the amount of trade between different nations)

4.

Multiplex networks: capture how different networks interrelate and overlap to effect each other

Network connection

1. Node connection is a key metric of its significance within the network

2. Degree of connectivity: a. immediate likelihood of catching whatever is flowing through the network b. connectivity is often a positive thing but not always—virus

3.in & out degree :measure for amount of in and out links

Weighted graph: place quantitative values on each edge

4. Adjacent: a edge between a and b

Adjacent matrix: capture all the relations within a network—-1 if adjacent and 0 if not

5. Walks: channel or path

A walk is a sequence of adjacent vertices where repetition is allowed

A path is a walk without revisiting any nodes

E.g traveling sales man —find the shortest path between two nodes on the graph is called geodesic and it represents the fewest number of links we need to traverse

Network centrality

—Node importance

1. Centrality: how influential or significant a node is within the overall network

2. Degree connectivity:

The node position within the overall network

—Degree connectivity : primary metric that define its degree of significance within its local environment

—closeness centrality : how close a node is to any other node in the network that is how quickly or easily can a node reach each other node in the network

Closeness :defined as the reciprocal of farness (farness is defined as the sum its distance to all other nodes) —-the more central a node is, the lower its distance to all other nodes .

closeness—a measurement of node capacity to effect all other elements in the networks—e.g. how long it will take to spread something like information

—betweenness centrality : capture the node role as a connector or bridge between other groups or nodes

Betweenness: vertices that have a high probability of occurring on a randomly chosen shortest path between two vertices

—prestige centrality(eigenvector centrality): decide how significant you are based upon the significance of the nodes you are connected to

Def: it assigns relative scores to all nodes based on the concept that: Highly connected nodes count more than those with a low degree of connectivity.

Application: web search engine —try and rank the relative importance of a website by looking at the importance of a websites that link into

Depend on the context for which one works best

Network theory topology

1. Global matrix

2. Global & local

Network may start random but often develop into some stable overall structure

Overall structure is of central information in network theory

3. Network topology

The way different nodes are placed or interconnected and the overall patterns that emerge out of this

4. Effect topology

Overall topology feed back to affect actions and capabilities of the nodes on the local level

5. Features

(1)connectivity: the density of the connection int he system

—more dense & integrated more defined by the structure and makeup of the network

(2) scale : its size which to be modeled or quantified- the number of nodes

(3)clustering : due to some common set of properties within a set , we get subsystems- cluster forming within networks

Process—The networks are not always but often created by the nodes in the network who create or don’t create connections in response to local level conditions. but once a network can reach a level of maturity, a global structure will emerge to it that feeds back to affect the elements in the system. Then we need to analyze this global structure to the network, what we call its topology in order to understand it.

Network connectivity

1. Density = the number of edges/ the number of possible edges

When we increasing the coupling parameter, We are increasing the density of the network and the average degree of connectivity

2. Ease of connection: the easier it is for elements to create a connection, the more connections and the longer these connections can be

Network diameter &Scale

1. Network diameter: the longest of all the geodesics (geodesics is the shortest path between two nodes )

2. Average path length : the shortest path between all the pairs of nodes, adding them and dividing by total number of pairs— show the number of steps it takes on average to get from one member of the network to another

3. Facebook :721 million users with a average path length of just 4.74

The small-world phenomena

Size as a product of structure

Clustering & connectedness

1. Clique : a group of people who interact with each other more regular and intensely than others in the same settings

2. Homophily 同质性

birds of a feather flock together(famous saying)

— interaction between components with similar attributes requires less resources.

—clustering may be due to physical constraints of the resource expenditure required to maintain them over a greater distance

3. Clustering conditions—to better integrate them

4. Clustering coefficient of a node : measure the degree of a local cluster— capture the ratio of existing links connecting a nodes neighbors to each other relative to the maximum possible number of link that could exist between them.

Clustering coefficient : the average of the clustering coefficients of all the nodes.

1. Network transitivity:

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

2. degree distribution?

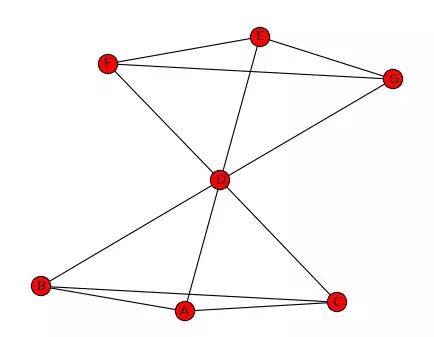
3. node centrality (example and definition):

<https://blog.csdn.net/yyl424525/article/details/103108506>

A) Degree centrality

In undirected networks, we can measure centrality by the degree of a node (equivalent to the number of your Facebook friends). E.G. with more than 90 million followers on Facebook , this guy's degree centrality is high.

The principle behind it is that important nodes are nodes with many connections. The more connections you have, the more influence you have.



In the network above, the connection of node D is 6 as it connects every node in this network. The connections of other nodes are all 3. Thus, node D has the highest degree centrality.

There are 7 nodes in this network, which means each node can have 6 connection at most. So, the degree centrality of D is 6/6 = 1 and that of others are 3/6=0.5

B) Betweenness centrality

Interactions between two non-adjacent members of a network depend on other members, especially those in the path between them. They control and constrain the interaction between two non-adjacent members. Freeman (1979) believed that intermediate members have a "greater influence on interpersonal relationship" on members at both ends of the path.

Therefore, the principle of centrality is that a member is a core member and has greater betweenness centrality if it lies on multiple shortest paths for other members. Compute all shortest paths for any two nodes in the network. If many of these shortest paths pass through a node, the node is considered to have high betweenness centrality.

Going back to the bow network above, suppose we want to calculate the mediation centrality of node D.

First, we calculate the number of shortest paths between all node pairs other than node D, which is 15 in this case (the number of node pairs to choose two out of six).

And then let's see how many of these shortest paths go through node D, for example, node A has to go through node D in order to find node E. There are nine shortest paths through node D.

Finally, we divide the shortest path through node D by the total number of shortest paths for all node pairs, and this ratio is the betweenness centrality of node D. The betweenness centrality of node D is 9/15=0.6.

When many members of a network depend on me for access or low-cost access, I have control over others. I can use this relationship to control the flow of information and reap huge benefits.

C) Closeness centrality

Degree centrality only makes use of the local characteristics of the network, that is, the number of connections of nodes, but the number of connections of a person does not mean that he/she is at the core of the network. Like betweenness centrality, closeness centrality makes use of the characteristics of the entire network, that is, the location of a node in the entire structure. If the shortest distance from a node to any other node in the graph is small, then its closeness centrality is high. Closeness centrality is closer to geometric centrality than to betweenness centrality.( 相比中介中心性，接近中心性更接近几何上的中心位置。)

Suppose we want to calculate the closeness centrality of node D, we first calculate the shortest distance from node D to all the other nodes. It can be judged from the figure that the distance from node D to all other nodes is 1, and the sum of the distances is 6.Therefore, the closeness centrality of node D is (7-1)/6=1.The numerator is the total number of nodes in the network minus 1. In other words, if a person has direct contact with everyone else in the network, then his or her proximity centrality is 1. For other nodes, such as node A, the proximity centrality is (7-1)/9=0.667.

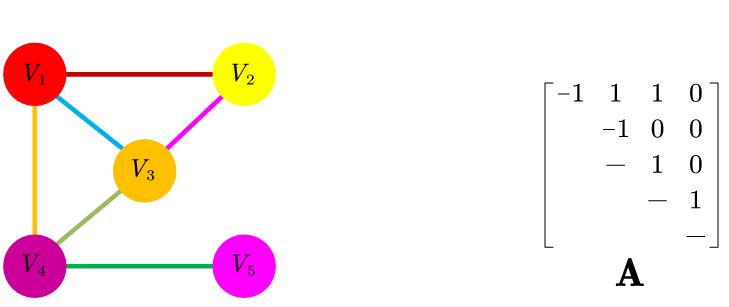
The nodes with high closeness centrality generally play the role of "gossiper". They are not necessarily celebrities, but they are willing to send messages between different groups of people.

D) Eigenvector centrality

The basic idea of eigenvector centrality is that the centrality of a node is a function of the centrality of adjacent nodes. In other words, the more important the people you connect with, the more important you become.

Eigenvector centrality is different from degree centrality, a high degree centrality means that nodes with many connections are not necessarily high eigenvector centrality, because all connectors may have low eigenvector centrality. Similarly, a high eigen vector centrality does not mean that it has a high degree centrality, as one has very few but very important connectors can also have a high eigenvector centrality.

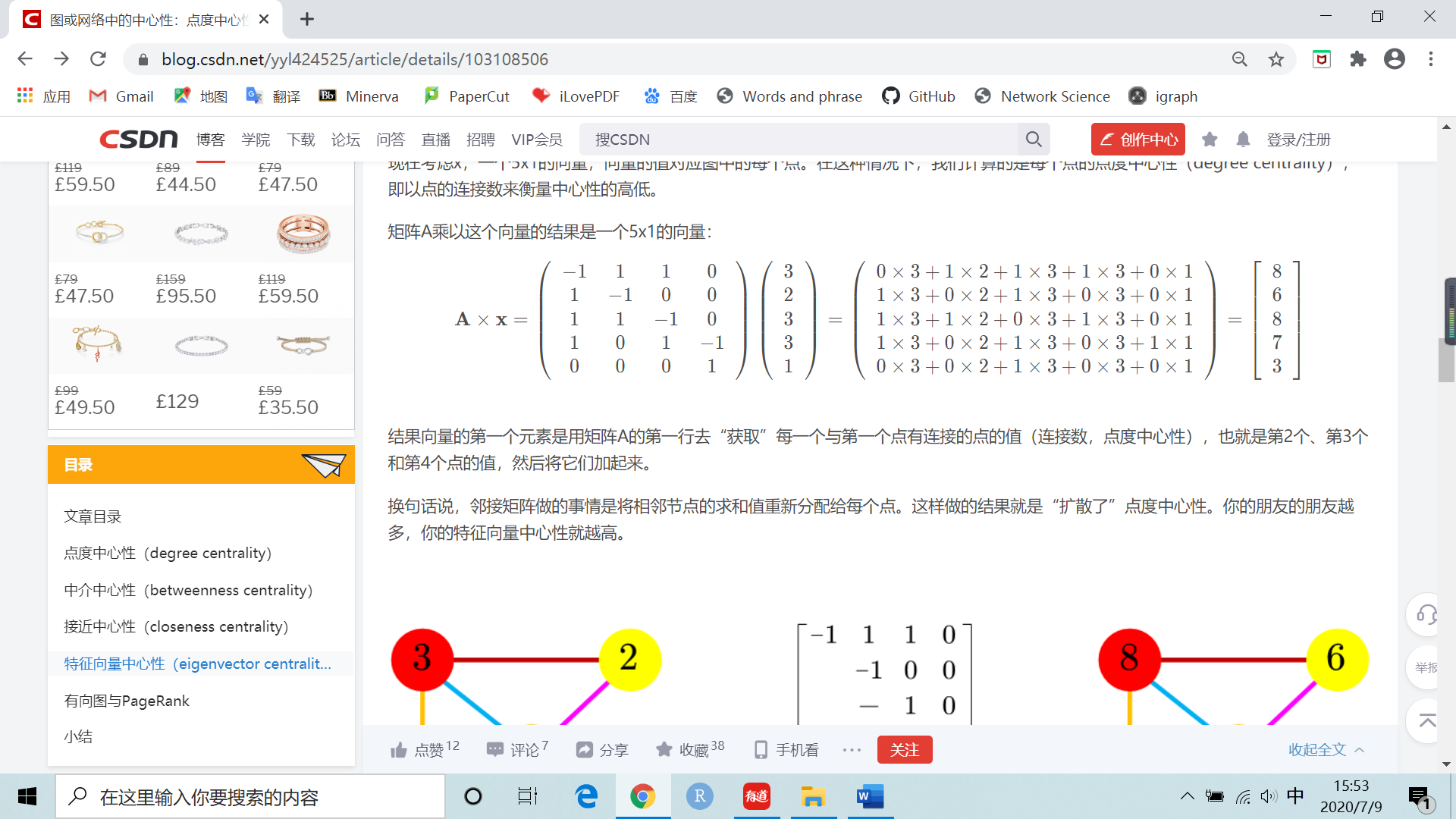
Consider the figure below, and the corresponding 5x5 Adjacency Matrix, A.



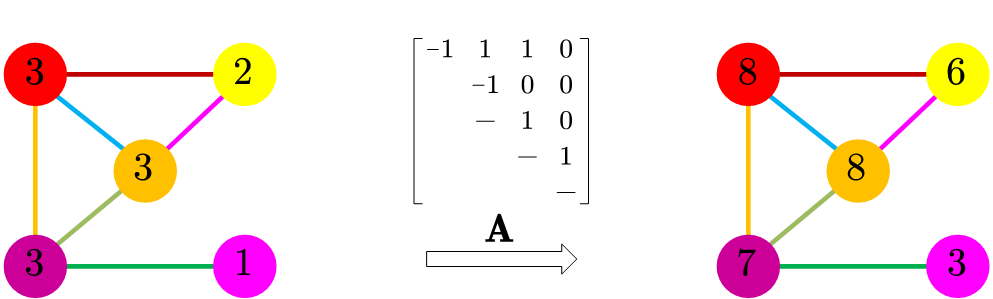
The adjacency matrix means that if two nodes are not directly connected, it's 0, otherwise it's 1.

Now let's think about x, a vector of 5x1, and the value of that vector corresponds to every point in the graph. In this case, we calculate degree centrality of each point, that is, the degree centrality is measured by the number of connections of points.

The matrix A times this vector is going to be A 5x1 vector:

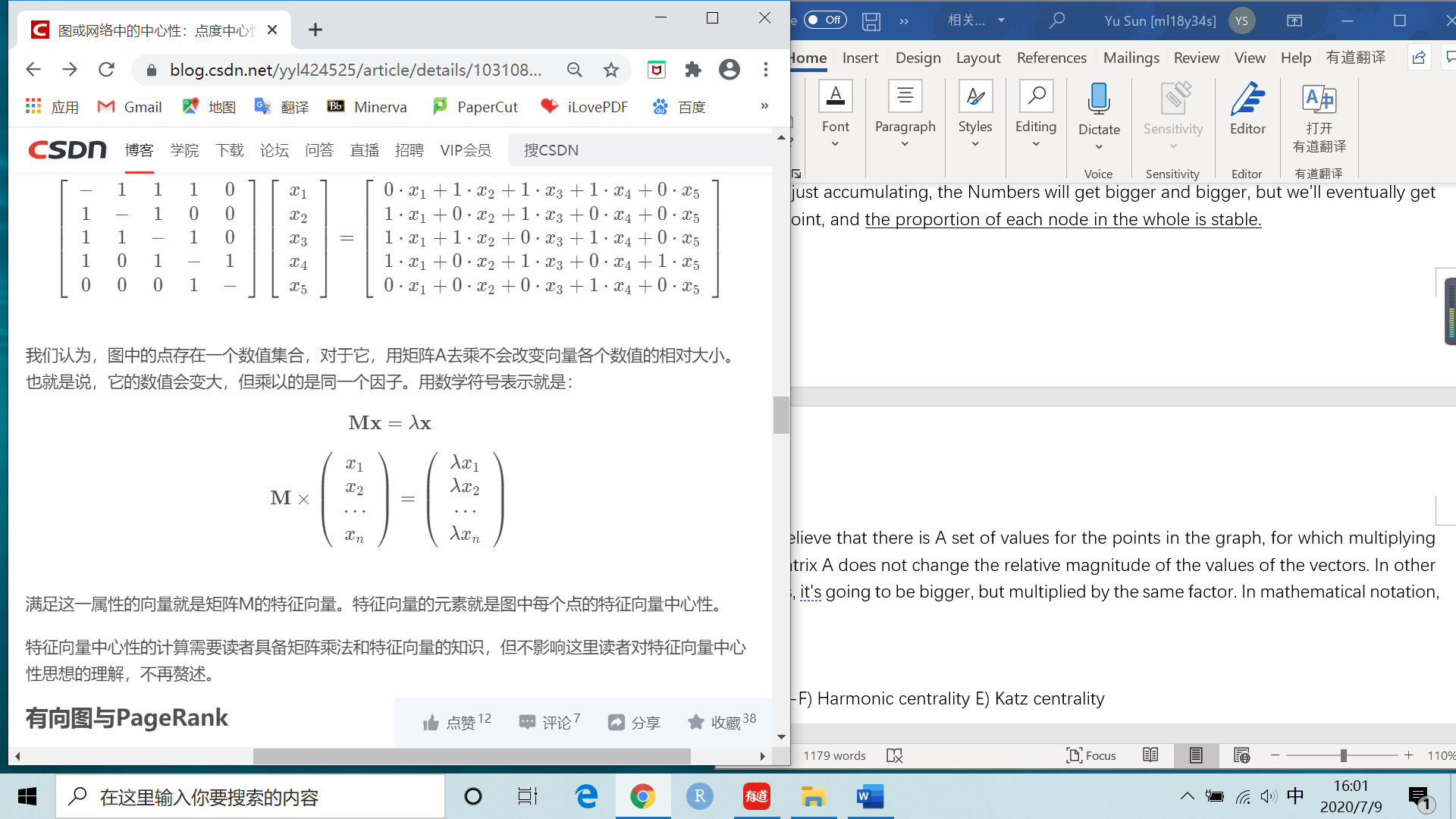


In other words, what the adjacency matrix does is it reassigns the sum of adjacent nodes to each point. The result is "diffused" point degree centrality. The more friends of your friends, the more centrality of your eigenvectors(扩散了点的中心性)



In fact, we allow this one of the central values to spread again along the boundary of the map. We will observe the diffusion in both directions (points both give and reap the adjacent nodes). We guess that this process will eventually reach a balance, and the number of certain points will be balanced by the number of adjacent nodes. Now that we're just accumulating, the Numbers will get bigger and bigger, but we'll eventually get to a point, and the proportion of each node in the whole is stable.

We believe that there is A set of values for the points in the graph, for which multiplying by matrix A does not change the relative magnitude of the values of the vectors. In other words, it's going to be bigger, but multiplied by the same factor. In mathematical notation, it is:



The vector that satisfies this property is the eigenvector of the matrix M. The elements of an eigenvector are the eigenvector centrality at each point in the graph.

--F) Harmonic centrality E) Katz centrality

Random walk

<https://stanford.edu/~rezab/classes/cme305/W16/Notes/6.pdf>

Given a graph and a starting vertex, select a neighbor of it uniformly at random, and move to this neighbor; then select a neighbor of this point at random, and move to it etc. The random sequence of vertices selected this way is a random walk on the graph.