EE 232E

Graphs and Network Flows

Homework 2 Report

Pingyuan Yue: 504737715

Ke Xu: 604761427

Yuanyi Ding: 404773978

4/23/2017

Table of Contents

1. Random walk on random networks

Part (a):

Just like what we did in homework 1, we simply created the random network using the following commands:

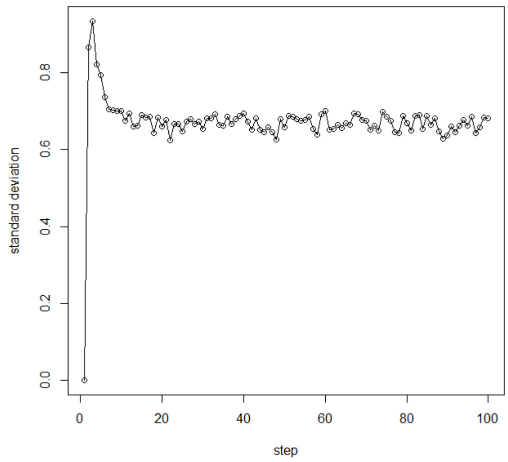
p<-0.01

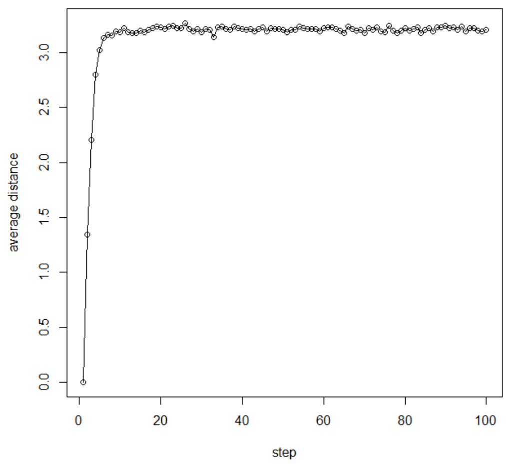
n<-1000

g<-sample_gnp(n, p, directed = FALSE, loops = FALSE)

Part (b):

We utilized the function RandomWalker (netrw) to simulate random walk on the graph with 1000 nodes. The following 2 graphs shows average distance and standard deviation of graph with 1000 nodes:





Figrue1.1: average distance and standard deviation of graph with 1000 nodes

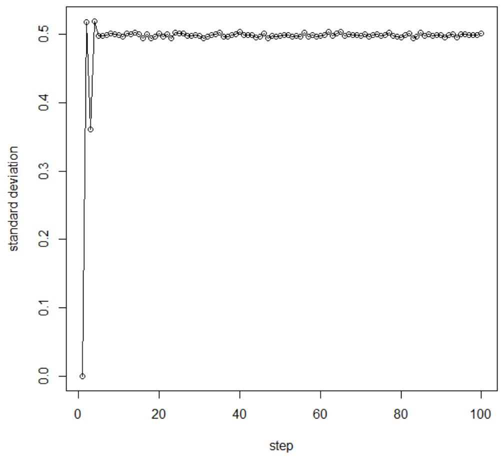
Part (c):

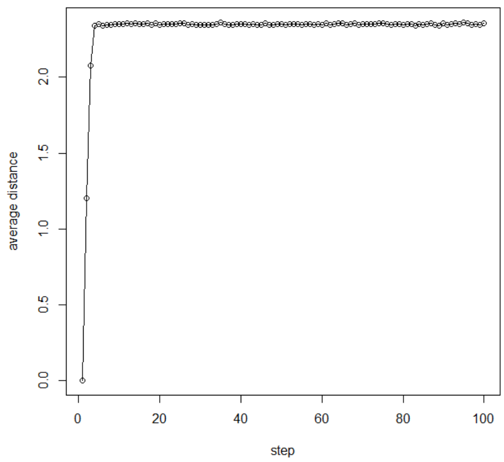
The result in (b) is not the same as the random work in the d-dimensional space. The reason of it is that in the d-dimensional space, the distance could be negative, which means it could have cancellation between positive and negative distance that causes the average distance to be 0. While the random graph we created has no negative distance, so the average distance is around positive 3. As a result, the result is different from a random walker in d dimensional.

Part (d):

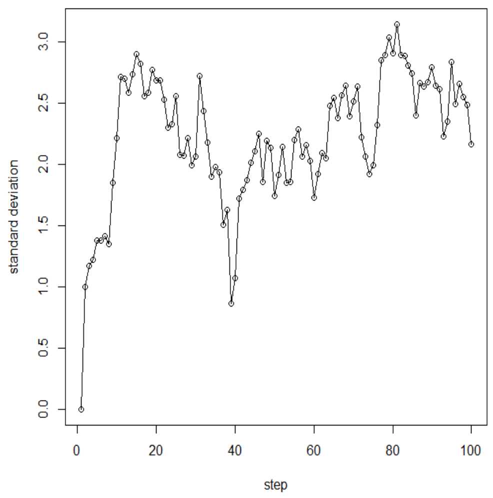
We just repeated the process for random networks with 100 and 10000 nodes.

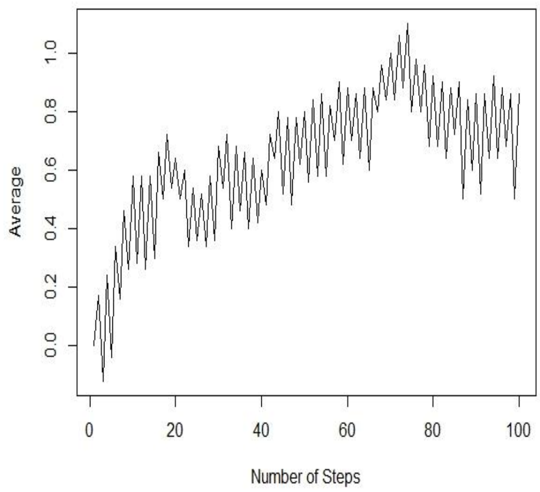
The following 4 graphs shows average distance and standard deviation of graph with 100 and 10000 nodes:





Figrue1.2: average distance and standard deviation of graph with 10000 nodes





Figrue1.3: average distance and standard deviation of graph with 100 nodes

Also the diameters of graphs with 100, 1000 and 10000 nodes are 10,6,3 respectively. It can be obviously seen from figures above that for the graph with 100 nodes the distribution converge slowly or even cannot converge if we do not set the start node fixed. The smaller the diameter is, the ave and std converge more quickly and turn out to have smaller fluctuation. Also since the average distance of graphs with 100, 1000 and 10000 nodes are 0.6, 2.8, 3.2 respectively, we can say that generally, the graph with larger diameter will have larger average distance.

Part (e):

Firstly, the number of total steps of the random walk is set to 100.

Then the following figure shows the comparison between the degree distribution of the graph and the degree distribution of the nodes reached at the end of the random walk on the graph.

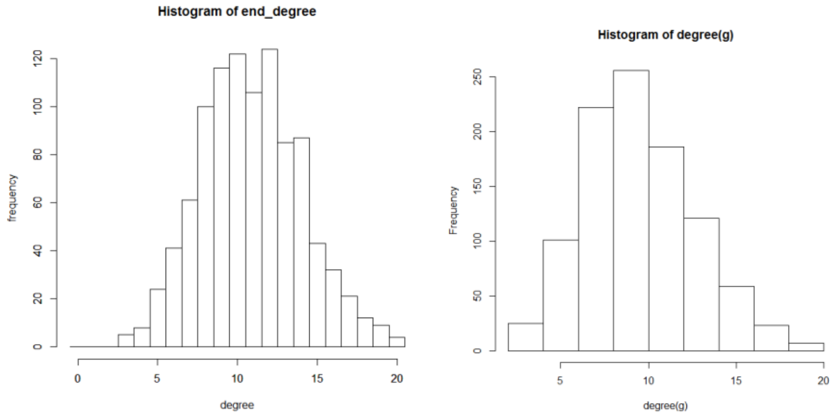


Figure 1.4 Comparison between two degree distributions

Running several times to get several figures, we could see that after a certain large number of steps of random walk, the two degree distributions are similar to each other.

2. Create a Network with Fat-tailed Distribution

Part (a):

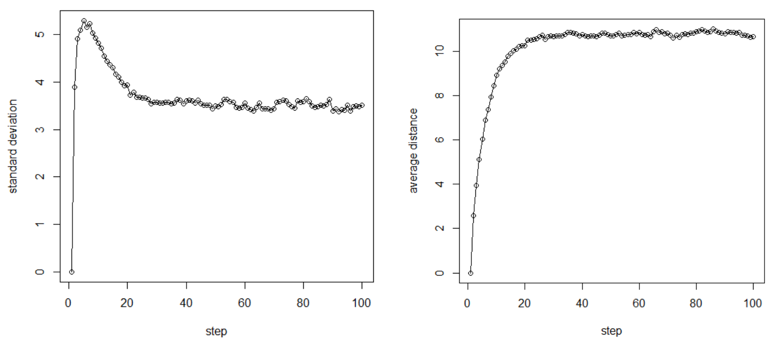
Just like what we did in problem 1 part a, we simply created the fat-tailed network using the following commands:

n<-1000

g <- barabasi.game(n, power = -3, directed = FALSE)

Part (b):

We utilized the function RandomWalker (netrw) to simulate random walk on the fat-tailed graph with 1000 nodes. The following 2 graphs shows average distance and standard deviation of graph with 1000 nodes:



Figrue2.1: average distance and standard deviation of graph with 1000 nodes

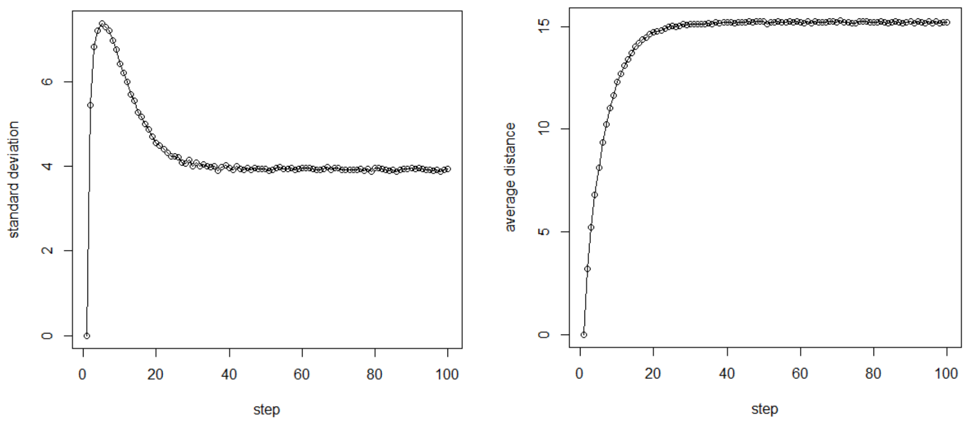
Part (c):

The result in (b) is not the same as the random work in the d-dimensional space. The reason of it is that in the d-dimensional space, the distance could be negative, which means it could have cancellation between positive and negative distance that causes the average distance to be 0. While the random graph we created has no negative distance, so the average distance is around positive 11. As a result, the result is different from a random walker in d dimensional.

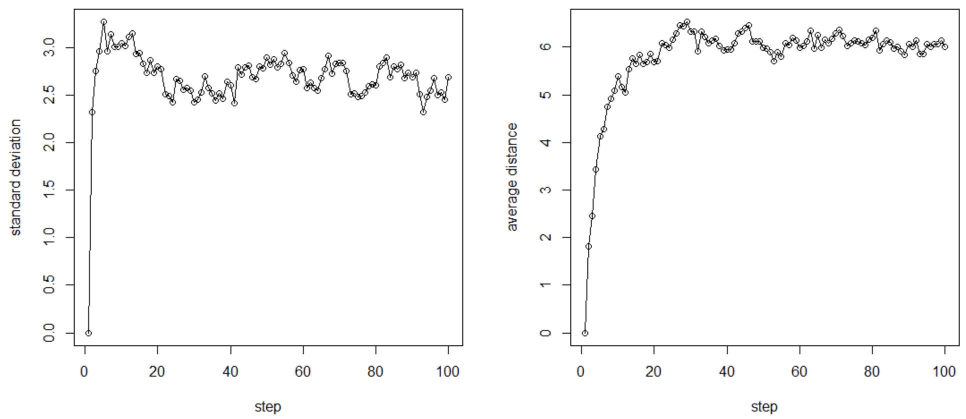
Part (d):

We just repeated the process for fat-tailed networks with 100 and 10000 nodes.

The following 4 graphs shows average distance and standard deviation of graph with 100 and 10000 nodes:



Figrue2.2: average distance and standard deviation of graph with 10000 nodes



Figrue2.3: average distance and standard deviation of graph with 100 nodes

Also the diameters of graphs with 100, 1000 and 10000 nodes are 15,24,33 respectively. Since the average distance of graphs with 100, 1000 and 10000 nodes are 6, 11, 15 respectively, we can say that generally, the graph with larger diameter will have larger average distance.

Part (e):

Firstly, the number of total steps of the random walk is set to 100.

Then the following figure shows the comparison between the degree distribution of the graph and the degree distribution of the nodes reached at the end of the random walk on the graph.

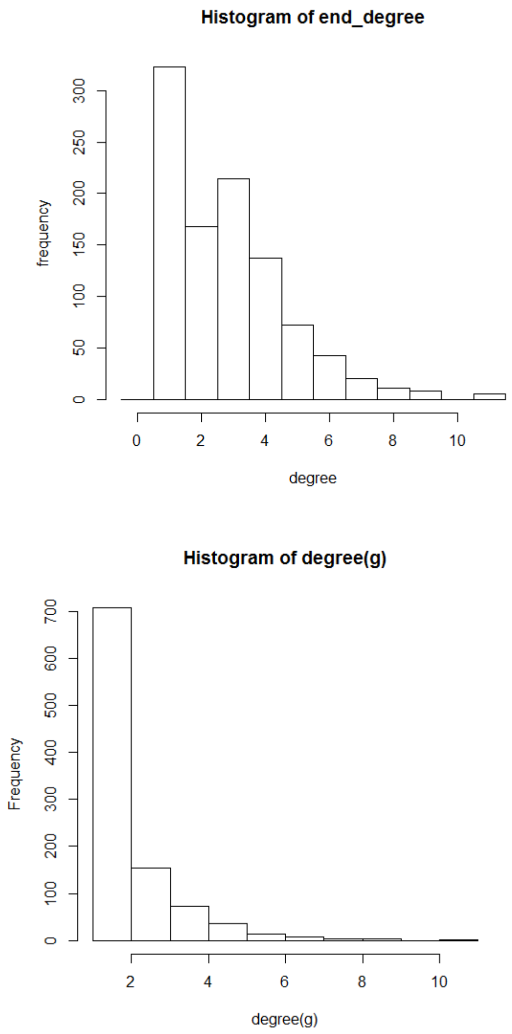


Figure 2.4 Comparison between two degree distributions

Running several times to get several figures, we could see that after a certain large number of steps of random walk, the two degree distributions are similar to each other.

3. PageRank

PageRank (PR) is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after Larry Page, one of the founders of Google. PageRank is a way of measuring the importance of website pages. Here in this problem, we use random walk to simulate page rank.

Part (a):

In this part, we simulate random walk on an undirected network with 1000 nodes and sampling factor of 1. The probability that the walker visits each node is measured by calculating the degree of nodes and plotting the relationship between the visit probability and the degree of each node. To get a better understanding of the relationship, the correlation between the probability and degree is also computed. The results are given below:

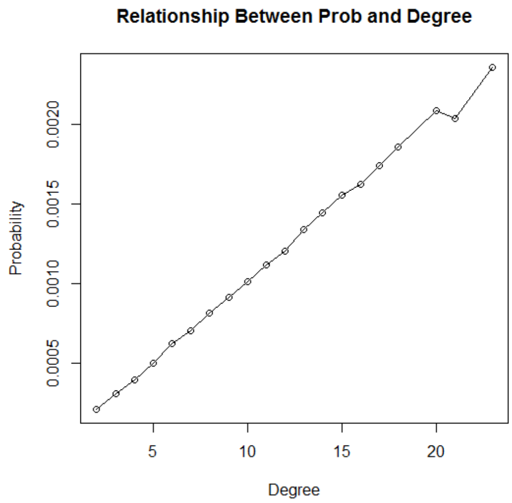


Fig3.1 The relation between probability and degree (n = 1000, damping factor = 1)

The correlation between degree and probability is : 0.9989891. From the results above, it could be concluded that the visit probability is proportional to the node degree. The higher the node degree is, the more possible the walker would visit the node.

Part (b):

In this part, we simulate random walk on a directed network with 1000 nodes and sampling factor of 1. The probability that the walker visits each node is measured by calculating the in degree of each node and plotting the relationship between the visit probability and the degree of each node. To get a better understanding of the relationship, the correlation between the probability and degree is also computed. The results are given below:

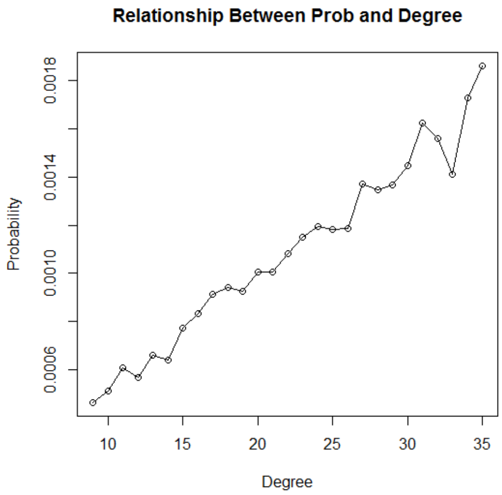


Fig3.2 The relation between probability and degree (directed)

The correlation between degree and probability is: 0.9470152. From the results above, it could be concluded that the relationship between visit probability and degree in a directed graph is also proportional. However, the probability is not as strongly related as it is in undirected graph.

Part (c):

In this part, we simulate random walk on an undirected network with 1000 nodes and sampling factor of 0.85. The probability that the walker visits each node is measured by calculating the degree of nodes and plotting the relationship between the visit probability and the degree of each node. To get a better understanding of the relationship, the correlation between the probability and degree is also computed. The results are given below:

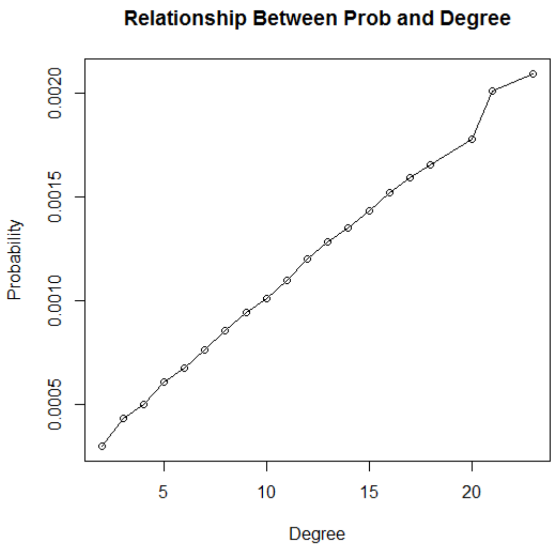


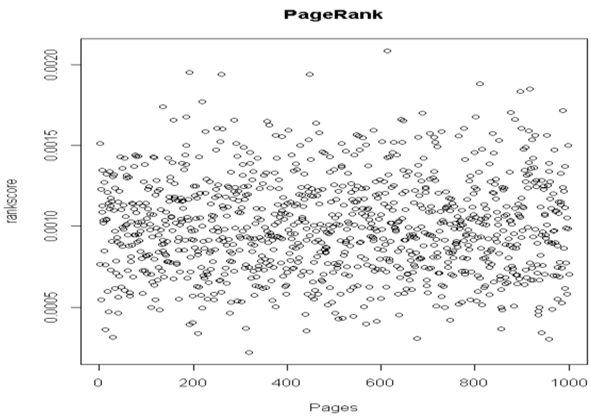
Fig3.3 The relation between probability and degree (n = 1000, damping factor = 0.85)

The correlation between degree and probability is: 0.9864352. From the results above, it could be concluded that in the undirected graph with sampling factor 0.85, the visit probability is proportional to the node degree. The higher the node degree is, the more possible the walker would visit the node.

4. Personalized PageRank

Part (a):

In this problem, we create a directed random network with 1000 nodes, where the probability p for drawing an edge between any pair of nodes is 0.01. We use the page.rank function in igraph package and random walk function with damping parameter 0.85 to successfully simulate the PageRank of the nodes. Below are PageRank of the nodes.



Figrue4.1: page rank

Part (b):

In this part, we are asked to personalize our own PageRank, assuming our interest to a node is proportional to the node's PageRank. We modified our random walk function and use the personalized.pagerank function in igraph package. The result PageRank graph is shown below:

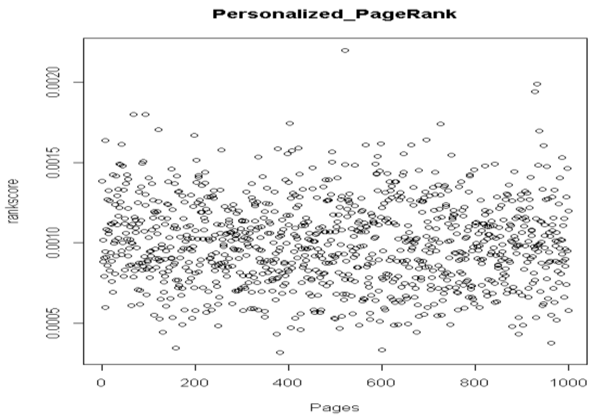


Figure4.2: Personalized PageRank

Part (c):

In real world, people can have different interest to different nodes(website). However, this is against the original assumption of the normal PageRank, where we assume that people have the same interest in all nodes (website). So we change the teleportation probability from 1/N to N.