EE 232E Project 1
Random Graph and Random Walks

Tianyi Liu, Sheng Chang, Wandi Cui, and Hengjie Yang

I. Generating Random Networks

A. Create random networks using Erdös-Rényi model

(a) By creating the undirected networks with number of nodes $N = 1000$ and $p = 0.003$, 0.004, 0.01, 0.05, and 0.1, we observe the binomial distribution.

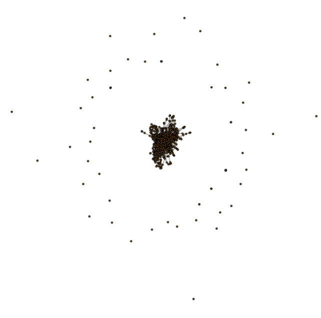

Fig. 1: Random network G_1 with $N = 1000$ and $p = 0.003$.

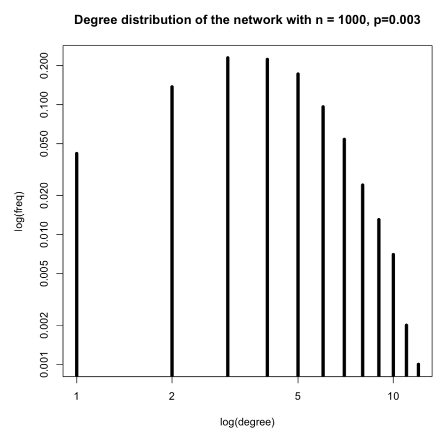

Fig. 2: Degree distribution of G_1

Fig. 3: Random network G_2 with $N = 1000$ and $p = 0.004$.

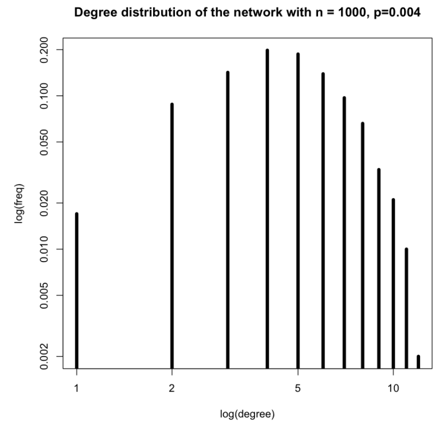

Fig. 4: Degree distribution of G_2

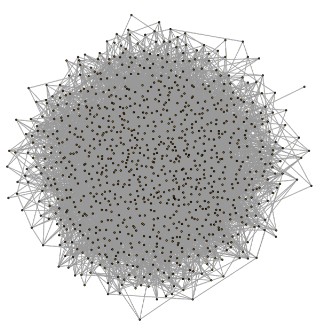

Fig. 5: Random network G_3 with $N = 1000$ and $p = 0.01$.

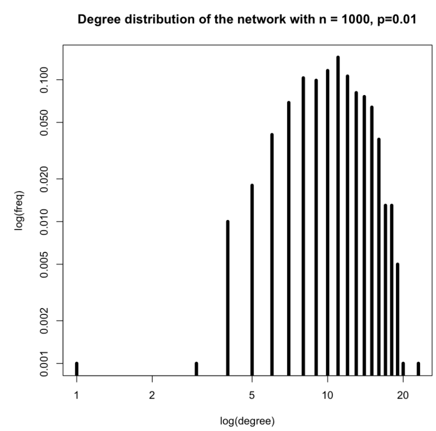

Fig. 6: Degree distribution of G_3

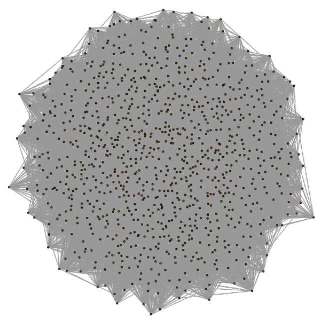

Fig. 7: Random network G_4 with $N = 1000$ and $p = 0.05$.

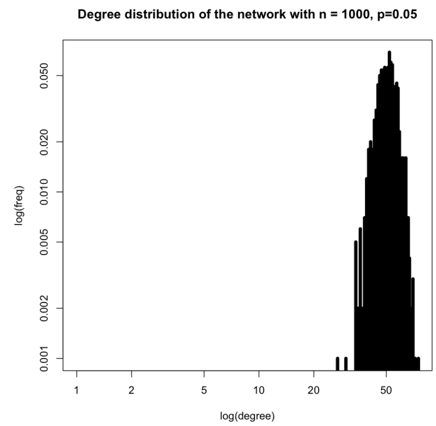

Fig. 8: Degree distribution of G_4

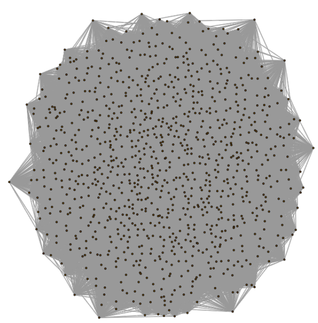

Fig. 9: Random network G_5 with $N = 1000$ and $p = 0.05$.

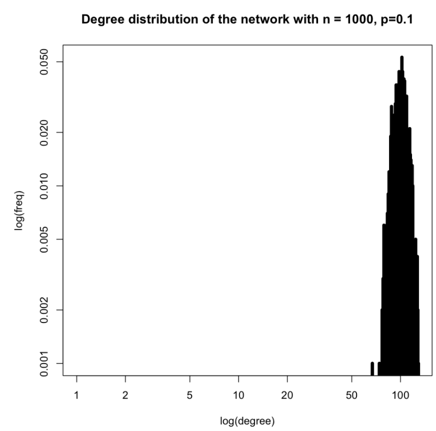

Fig. 10: Degree distribution of G_5

Below is the comparison between the empirical and theoretical values of mean and variance for the above 5 random networks.

For G_1

$$\tilde{\mathbb{E}}[\deg(G_1)] = 4.122, \quad \tilde{\mathrm{var}}(\deg(G_1)) = 3.305116$$

$$\mathbb{E}[\deg(G_1)] = 2.997, \quad \mathrm{var}(\deg(G_1)) = 2.988009$$

For G_2

$$\tilde{\mathbb{E}}[\deg(G_2)] = 5.028, \quad \tilde{\text{var}}(\deg(G_2)) = 4.461216$$

$$\mathbb{E}[\deg(G_2)] = 3.996, \quad \text{var}(\deg(G_2)) = 3.980016$$

For G_3

$$\tilde{\mathbb{E}}[\deg(G_3)] = 10.796, \quad \tilde{\text{var}}(\deg(G_3)) = 9.142384$$

$$\mathbb{E}[\deg(G_3)] = 9.99, \quad \text{var}(\deg(G_3)) = 9.8901$$

For G_4

$$\tilde{\mathbb{E}}[\deg(G_4)] = 51.222, \quad \tilde{\text{var}}(\deg(G_4)) = 45.54472$$

$$\mathbb{E}[\deg(G_4)] = 49.95, \quad \text{var}(\deg(G_4)) = 47.4525$$

For G_5

$$\tilde{\mathbb{E}}[\deg(G_5)] = 100.772, \quad \tilde{\text{var}}(\deg(G_5)) = 97.11202$$

$$\mathbb{E}[\deg(G_5)] = 99.9, \quad \text{var}(\deg(G_5)) = 89.91$$

(b) The network G_1, G_2 and G_3 are disconnected whereas the network G_4 and G_5 are connected. Taking G_1 as an instance, the size of GCC is 941 and the diameter is 13.

(c) The normalized GCC size versus the probability p in the range of $(0, \frac{\ln n}{n}]$ is given in Fig 11, where n is the number of nodes in the network. The result matches the theoretical values derived in class.

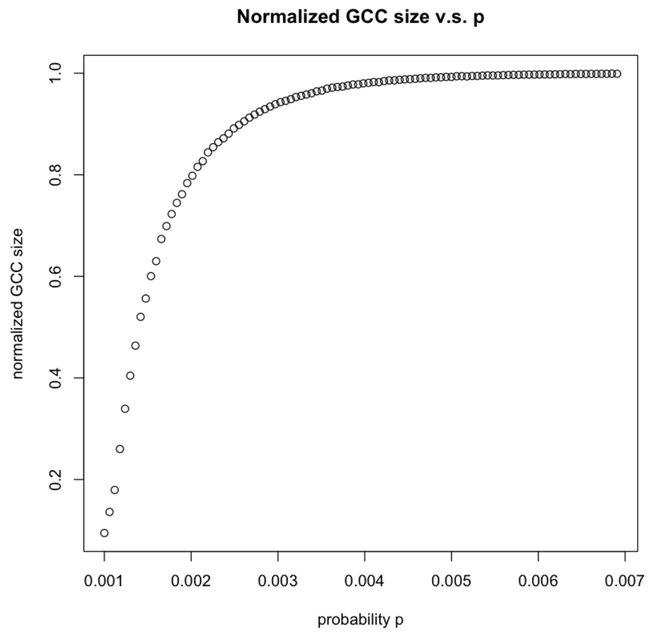

Fig. 11: The normalized GCC size versus the probability p in the range of $(0, \frac{\ln n}{n}]$

(d) The expected size of the GCC v.s. the number of nodes n where $c = np = 0.5$.

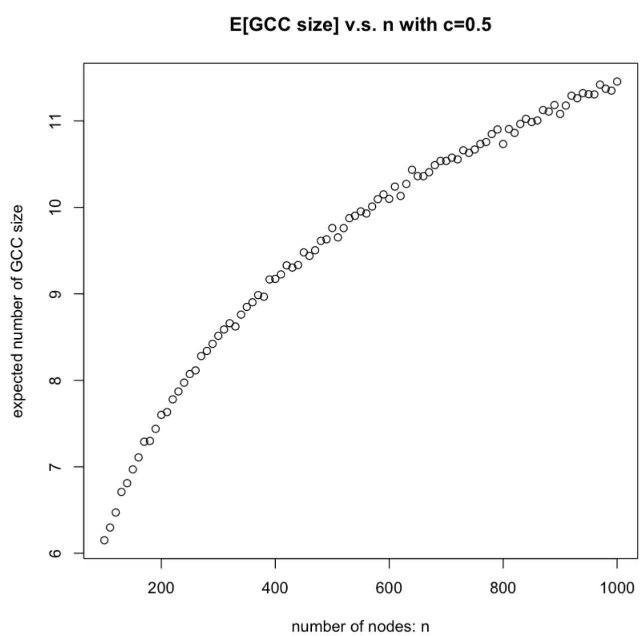

Fig. 12: The expected size of the GCC v.s. the number of nodes n when $c = np = 0.5$

The result for $c = np = 1$ is shown in Fig 13.

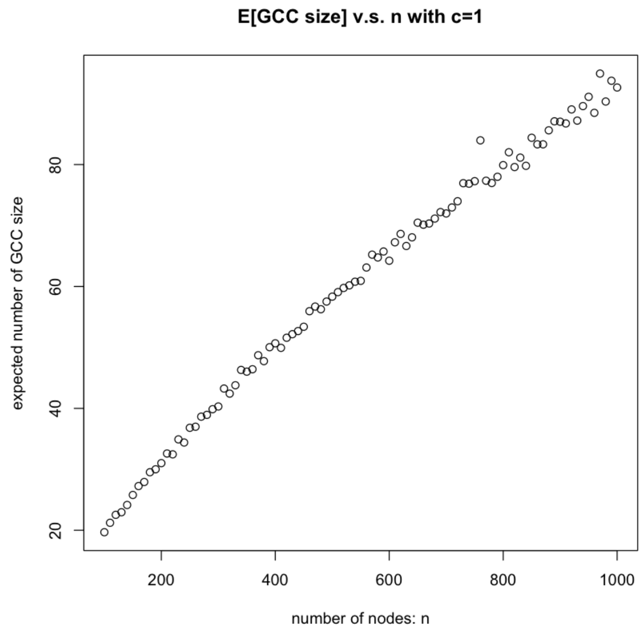

Fig. 13: The expected size of the GCC v.s. the number of nodes n when $c = np = 1$

The result for $c = 1.1$, 1.2 and 1.3 is shown in Fig 14.

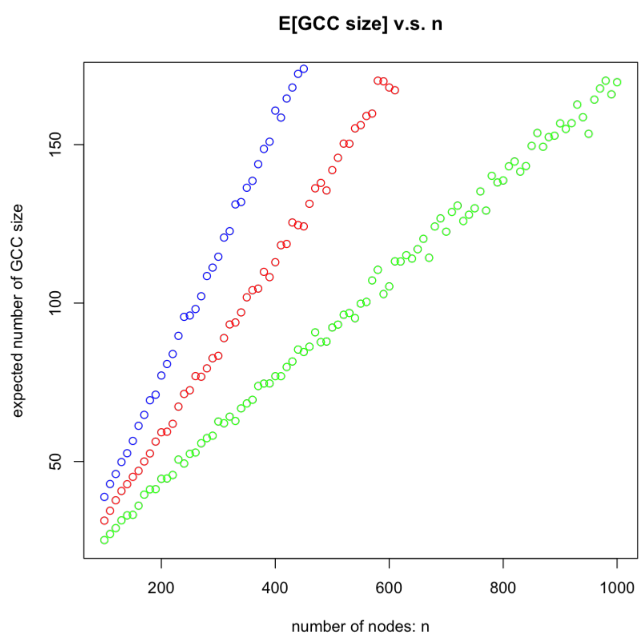

Fig. 14: The expected size of the GCC v.s. the number of nodes n when $c = 1.1$, 1.2 and 1.3

B. *Create networks using preferential attachment model*

(a) We created an undirected network under preferential attachment model with $n = 1000$ and $m = 1$. We can see that the network we generated is always connected from both practical result and theory analysis.

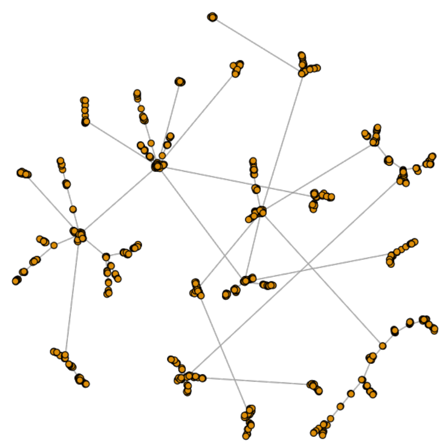

Fig. 15: undirected graph with 1000 nodes based on preferential attachment Model

(b) We used fast greedy method to find the community structure. Measure modularity.

II. Random Walk on Networks

A. *Random walk on Erdös-Rényi networks*

(a) Same as what we have done in the previous part, we generated an undirected random network containing 1000 nodes with the probability of 0.01 to add an edge between any pair of nodes. The graph we got is shown below:

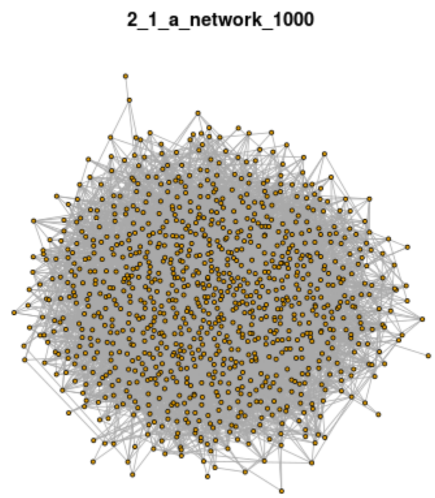

Fig. 16: undirected graph with 1000 nodes based on Erdös-Rényi Model

(b) After simulating random walk, we got following results:

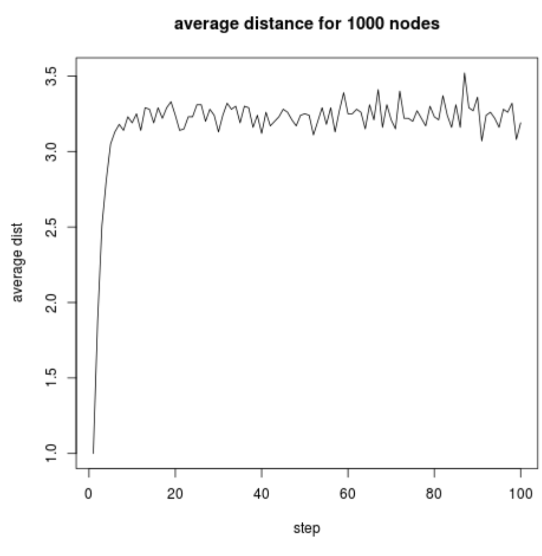

Fig. 17: s(t) v.s. t with $N = 1000$

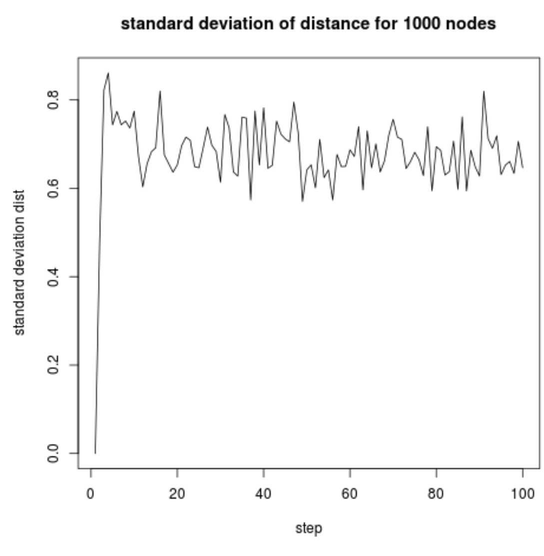

Fig. 18: 2(t) v.s. t with $N = 1000$

(c) The degree distribution of the last reached nodes in N times of the random walk and the degree distribution of the graph are shown below.

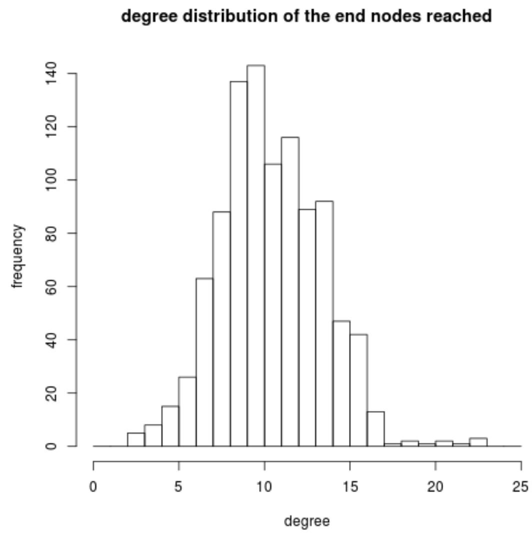

Fig. 19: degree distribution of the ending node with nodes = 1000 in network

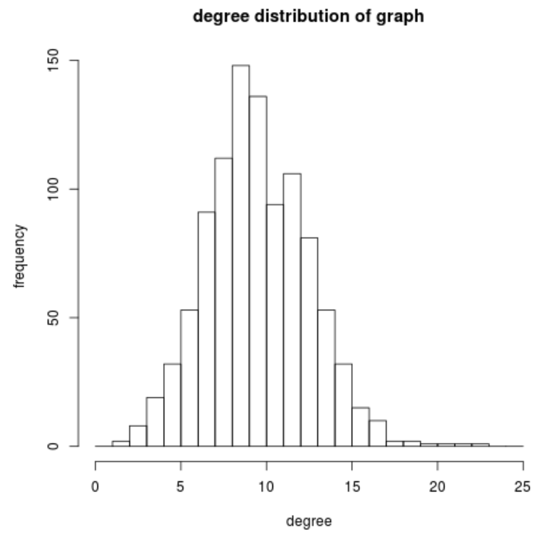

Fig. 20: degree distribution of the graph with nodes = 1000 in network

The results illustrate that the real graph degree distribution has similar trends with the degree distribution of the ending node by random walk. Both of them follows a Gaussian distribution. This could be proved by doing some simple calculation under the Erdös-Rényi Model. Every time a new node joins the network has the same probability (p) to add an edge to each existing nodes. That is to say, for all the nodes, the degree of a node follows binomial distribution.

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{1}$$

When the number of nodes N is large, Gaussian distribution could be seen the approximation of binomial distribution.

(d) To figure out the difference of the shortest path length with different number of nodes in the network, we repeated what we have done in problem(b) under 100 nodes and 10000 nodes respectively in the network. Figure 6 and 7 show the average distance and standard deviation of distance under 100 nodes respectively, while figure 8 and 9 show the results under 10000 nodes. The diameter of the nodes $N = 100$, $N = 1000$ and $N = 10000$ is about 2.0, 3.5 and 2.5 respectively. This seems not what we have expected, so we ran program several times to generate different networks with certain number of nodes. Not surprisingly, we found that graph with more nodes tend to have smaller diameter and smaller average distance. In addition, we can get another consequence from figures bellow that for the graph with 100 nodes, the average distance converge slowly or even cannot converge if we do not set the start node fixed. On the contrary, the results show less fluctuation with 10000 nodes. In conclusion, the smaller the diameter is, the average distance and standard deviation of the distance converge more quickly and turn out to have smaller fluctuation.

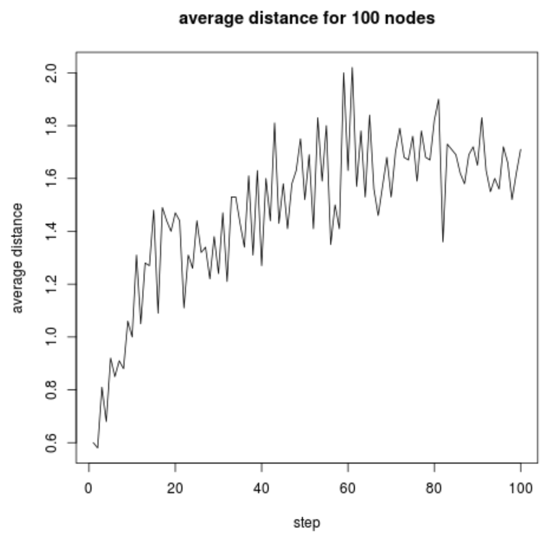

Fig. 21: s(t) v.s. t with nodes = 100

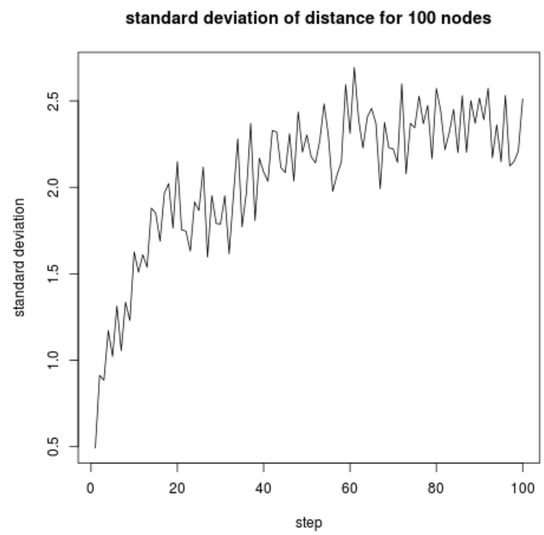

Fig. 22: 2(t) v.s. t with nodes = 100

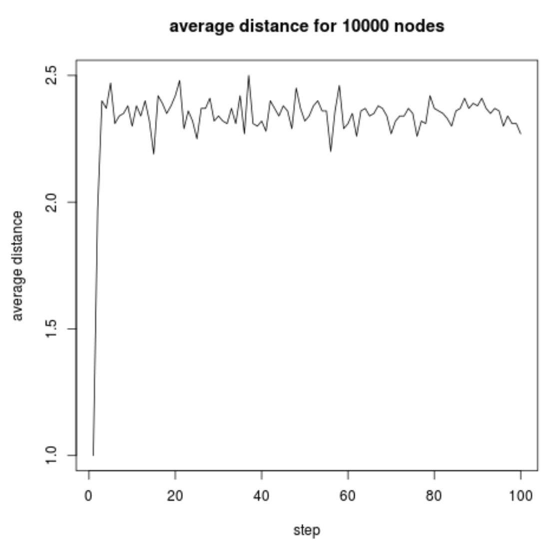

Fig. 23: s(t) v.s. t with nodes = 10000

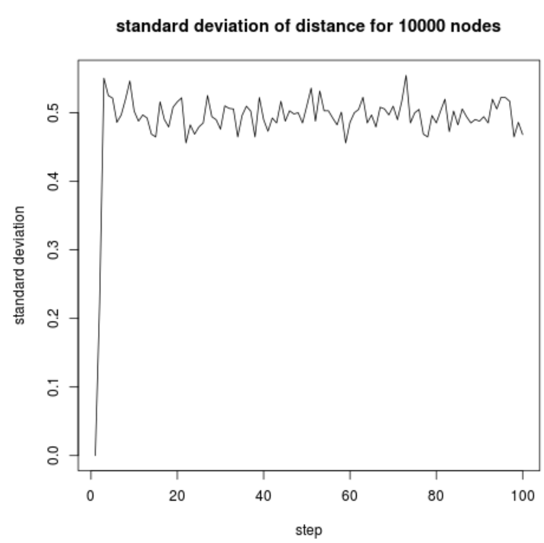

Fig. 24: 2(t) v.s. t with nodes = 10000

B. Random walk on networks with fat-tailed degree distribution

(a) We used function *sample_pa()* to generate a fat-tailed network according to the Barabasi-Albert Model. The graph we got is shown bellow.

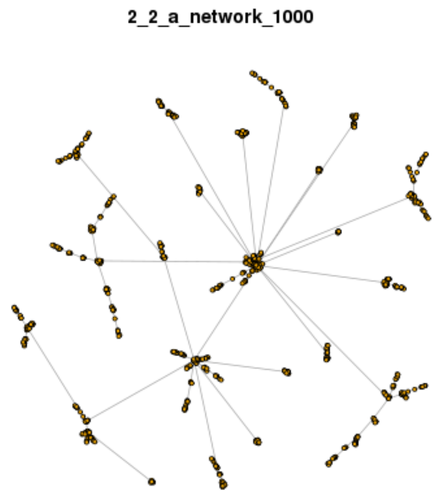

Fig. 25: undirected graph with 1000 nodes based on Barabasi-Albert Model

(b) As what we have dealt with Erds-Rnyi networks, we also simulated random walk on preferential attachment network and Figure 11 and 12 demonstrate the plot s(t) v.s. t and 2(t) v.s. t respectively. (c) After running several times to get several figures(shown bellow), we could see that after a certain large number of steps of random walk, the results of two kinds of degree distribution are similar to each other.

(d) Repeated the process in (b), we got the diameters of graphs with 100 and 10000 nodes are 4.5 and 5 respectively. Though we got this result, according to the theory, we knew that this seemed incorrect. Therefore, we generate network several times and calculate the average of the diameter. Then, we got the conclusion that in general, the graph with more nodes has smaller diameter, which means it also has smaller average distance.

The results also illustrate some differences between two different graph generation models. Considering the graph with 1000 nodes, in previous model(Erds-Rnyi Model), we can see that the average distance increases dramatically within not more than 10 steps, and then the value

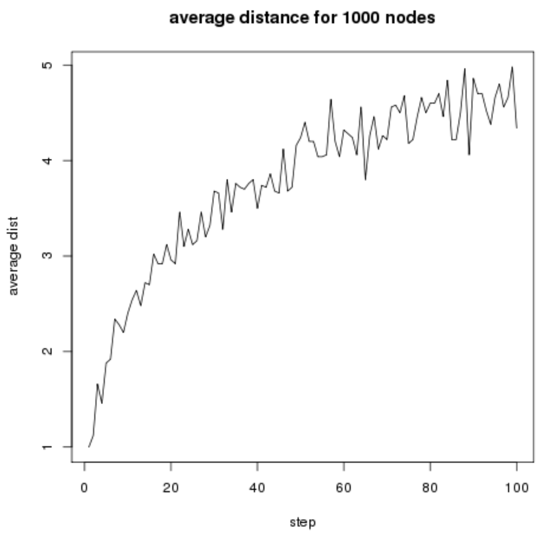

Fig. 26: s(t) v.s. t with nodes = 1000

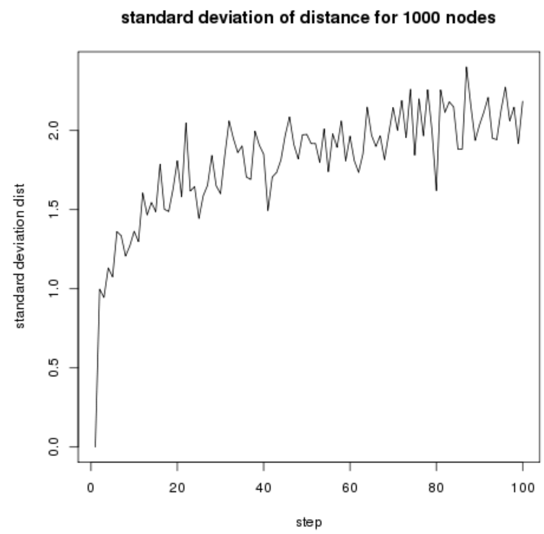

Fig. 27: 2(t) v.s. t with nodes = 1000

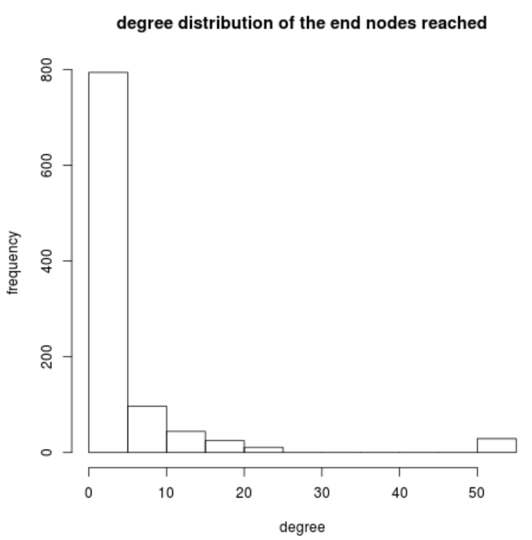

Fig. 28: degree distribution of the ending node with nodes = 1000 in network

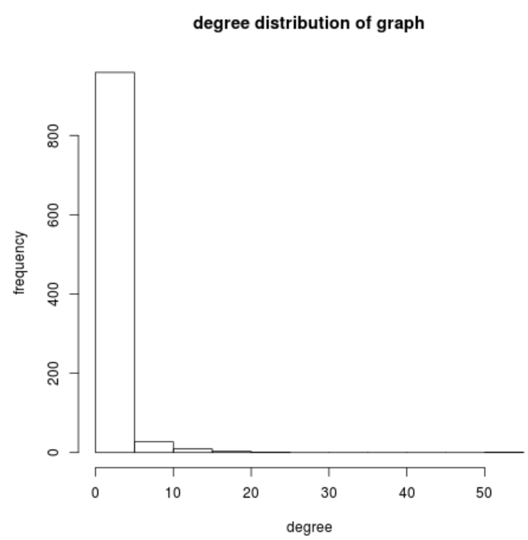

Fig. 29: degree distribution of the graph with nodes = 1000 in network

keeps fluctuating along the step t and gradually converges. However in Barabasi-Albert Model, the average distance shows a slight increase along the step t.

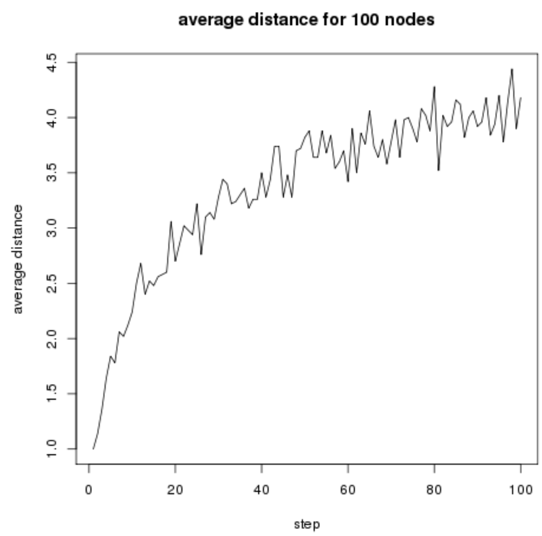

Fig. 30: s(t) v.s. t with nodes = 100

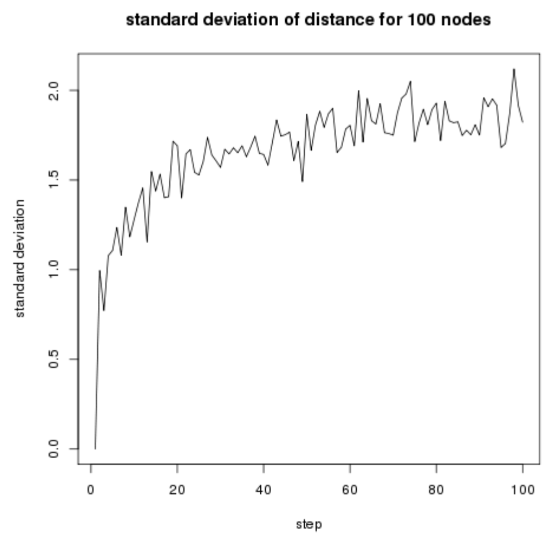

Fig. 31: 2(t) v.s. t with nodes = 100

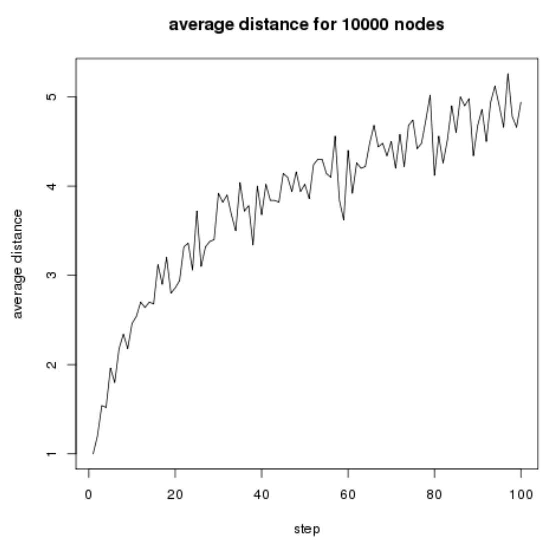

Fig. 32: s(t) v.s. t with nodes = 10000

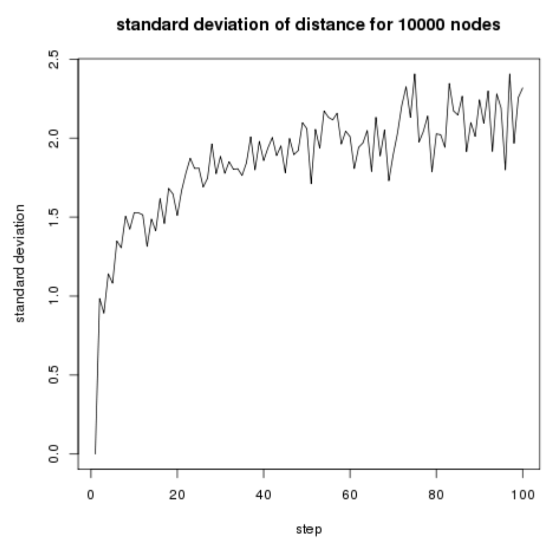

Fig. 33: 2(t) v.s. t with nodes = 10000

C. PageRank

PageRank is a famous algorithm designed by Larry Page, one of the founder of Google. Although there are a few limitations of this algorithm, many other ranking algorithms are invented under its inspiration. In this section, we applied random walk based on PageRank and got some interesting results which illustrate the principle of PageRank.

(a) The graph we generated based on preferential attachment is shown bellow.

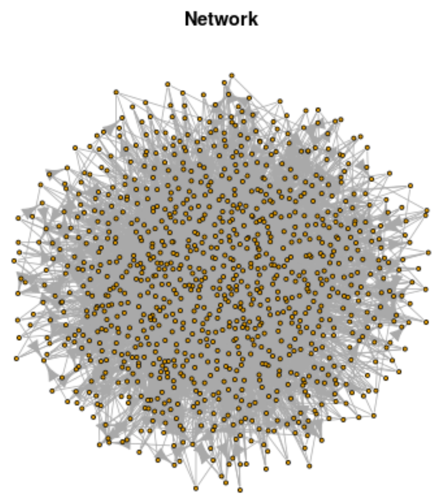

Fig. 34: directed graph with 1000 nodes using preferential attachment

From the results shown bellow, 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. (b) Considering the teleportation probability of $\alpha = 0.15$, we get slightly different results. It demonstrates that the difference of visiting probability along degree of the node is getting larger.

D. Personalized PageRank

(a) In Personalized PageRank, we set the teleportation probability proportionally to PageRank of each node. The results are shown bellow. Comparing to 3(a), the results show that Personalized PageRank can make degree of the node influence more to the visiting probability, which is in fact similar to what we get in 3(b). The higher degree the node is, the more possible it will be

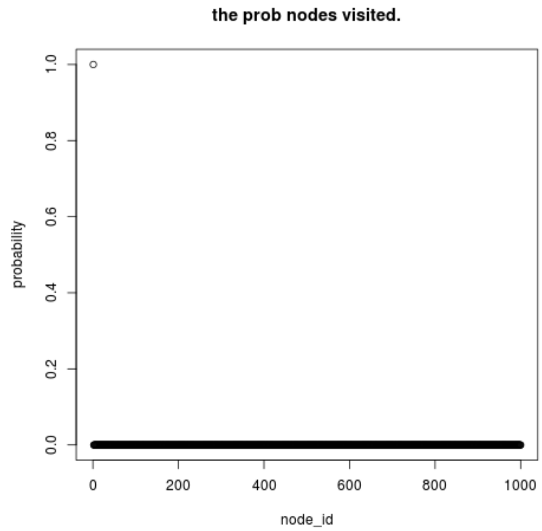

Fig. 35: probability that the walker visits each node

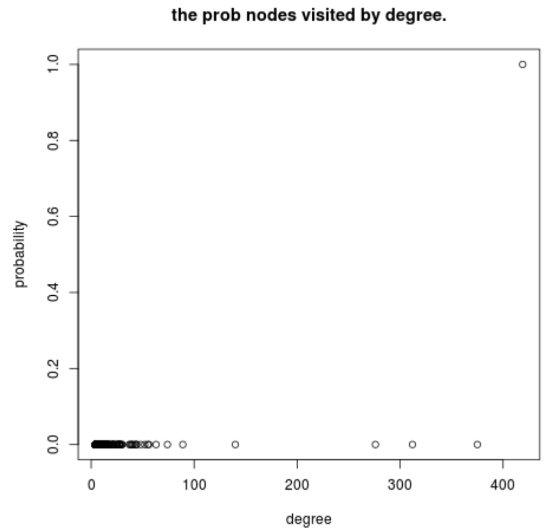

Fig. 36: visit probability v.s. degree

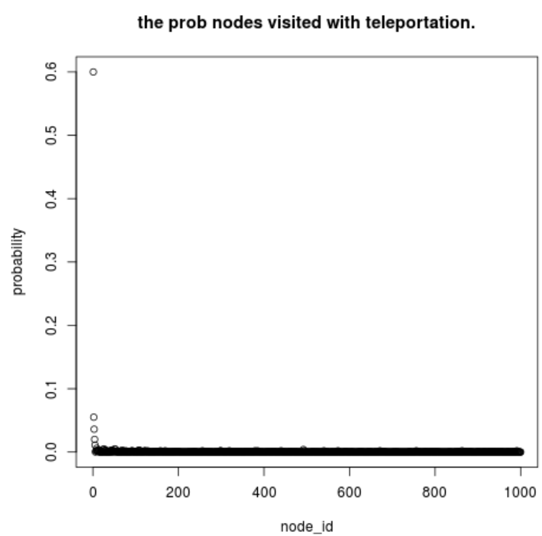

Fig. 37: probability that the walker visits each node with teleportation probability $\alpha = 0.15$

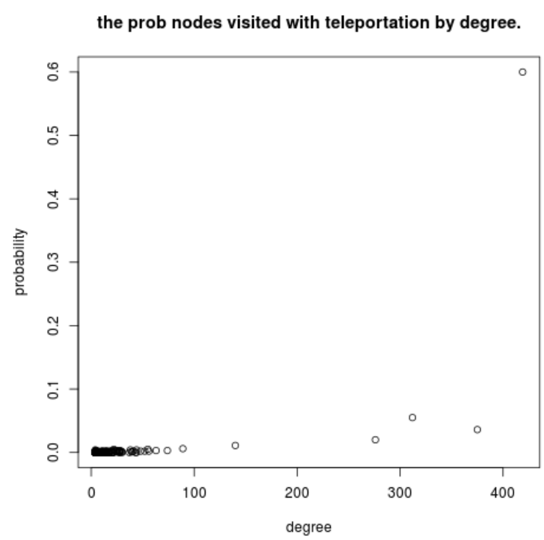

Fig. 38: visit probability v.s. degree with teleportation probability $\alpha = 0.15$

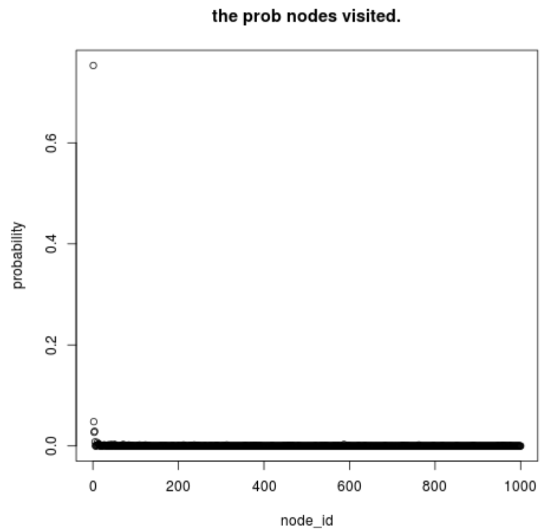

Fig. 39: probability that the walker visits each node under Personalized PageRank

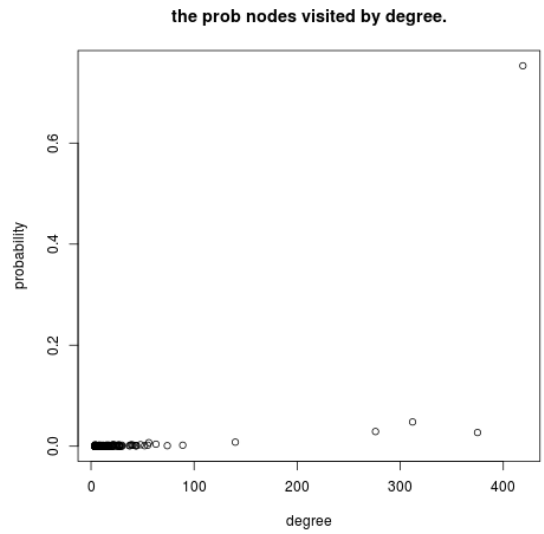

Fig. 40: visiting probability v.s. degree under Personalized PageRank

visited. In other words, since it is a preferential attachment model, the longer time a node stays in the network, the more possible it will be visited.

(b) The results are shown bellow. Comparing to all the results we get from PageRank or

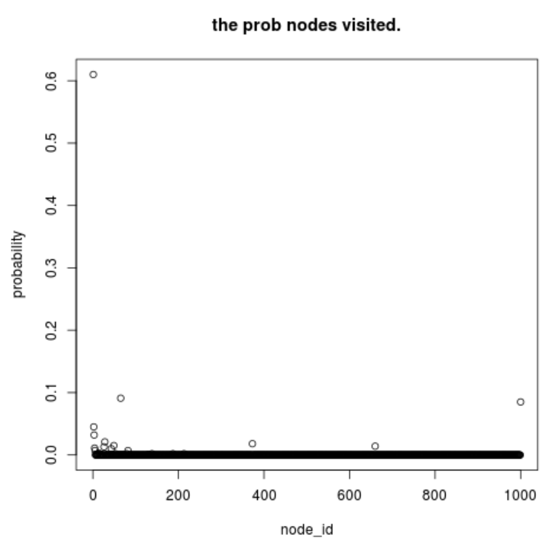

Fig. 41: probability that the walker visits each node under fixed teleportation

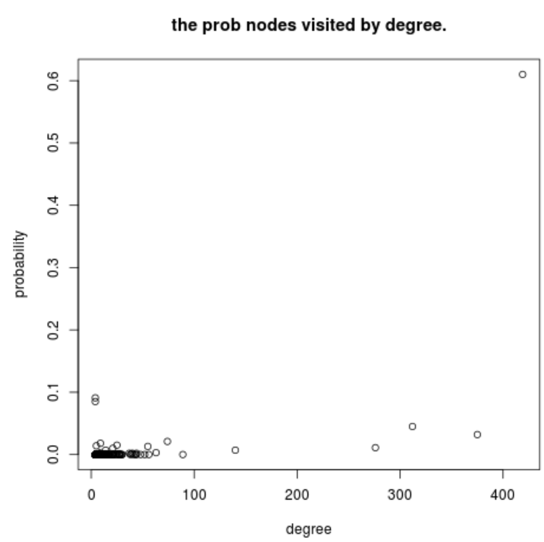

Fig. 42: visiting probability v.s. degree under fixed teleportation

Personalized PageRank, under this circumstance, the visiting probability varies more significantly along the degree of a node. Specifically, some nodes with less than 100 degree have high probability to be visited than some other nodes with degree around less than 50. Nevertheless, this rarely happens in previous model, where nodes with degree of 300 or higher may be more possible to be visited than other nodes, while nearly most of nodes within the degree of 100 have similar visiting probability of around 0. Though it appears some irregular results within the degree of 100 (the visiting probability does not follow the rule exactly), the general rule that nodes with higher degree have more chance to be visited does not change. (c) Since at this time, nodes can only teleport to trusted nodes, which is similar to the previous problem that only two nodes are allowed to teleport to. Here, only the nodes belong to trusted node set are allowed to teleport to. Thus, the equation would be like this:

$$r = \alpha \cdot T \cdot r + (1 - \alpha) \cdot d$$

where T is the transition matrix and vector d can be used to assign a non-zero score to the set of trusted pages only. In this way, trusted pages will have higher probability to be visited, and what these trusted pages point to can also be regarded as trustworthy pages, and their visiting probability will increase since we make restriction to teleport to trusted pages only.