6010 Assignment 3 Report

Hongyang Wang & Yunhe Song

- ## Part 1: Random Graph Generation

Analysis of the topologies generated by part 1 graphgen program: In this program, linked list is used to generate the scale free network graph and memory it.

In the beginning, initialize the first 3 nodes to construct a fully connected topology such that node 0 has 2 edges: 0-1,0-2; node 1 has 2 edges: 1-0,1-2; and node 2 has 2 edges: 2-0,2-1.

For the forth node: node 3, it has 1/3 probability to connect with node0, also 1/3 with node1, and also 1/3 with node2. Randomly choose one node i with its own probability p(i). After connecting, add a degree on node3 and another degree on the chosen node i, (in total we have 2 more degrees with one adding node).

Repeat the adding process until the graph get total N nodes.

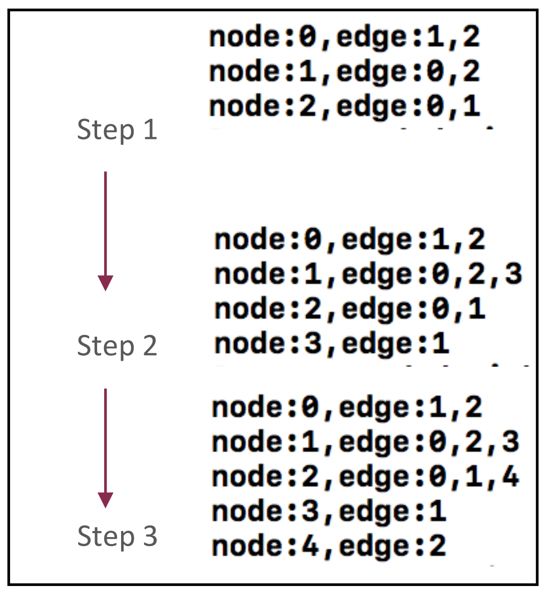
And here is a sample for this generation.

First step: initialize

Second step: add one more node 3

Third step: add one more node 4

Repeat adding step.

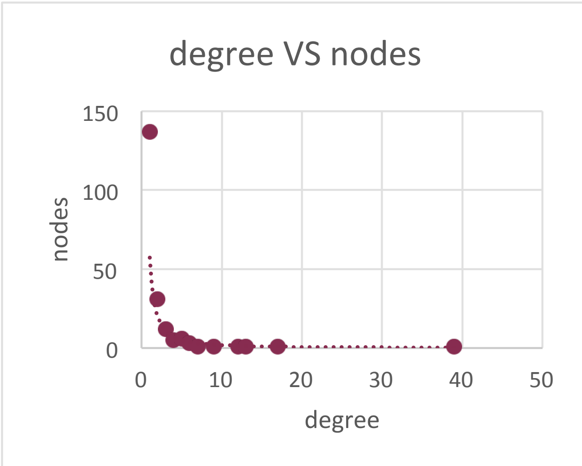


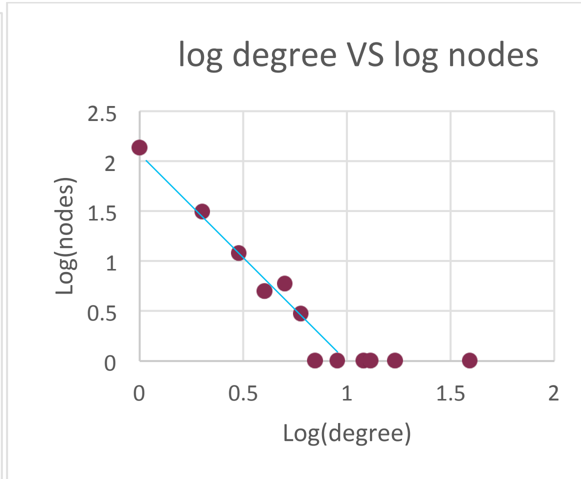
Finally, a topology is generated, and now the task is to show it's a scale free net work graph.

The Result sample is different nodes number varies from 200 to 80,000.

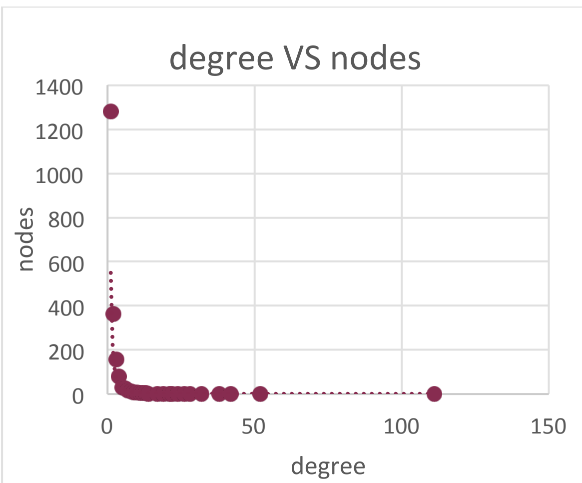
The degree and node relationship will be represented by the plots below.

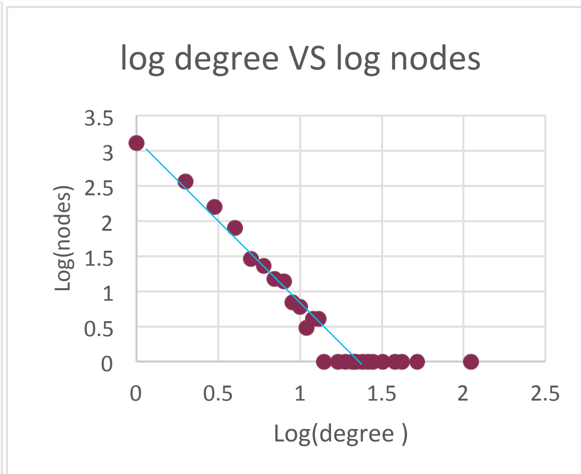
Nodes number: 200



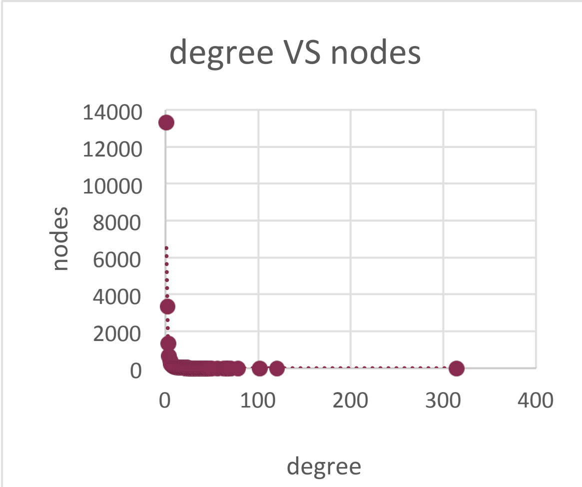


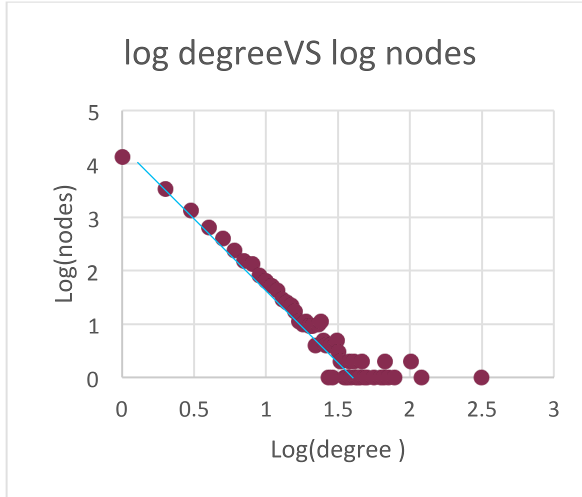
Nodes number: 2000



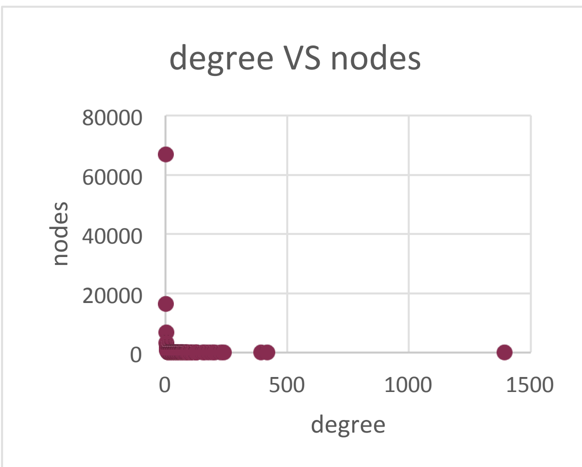


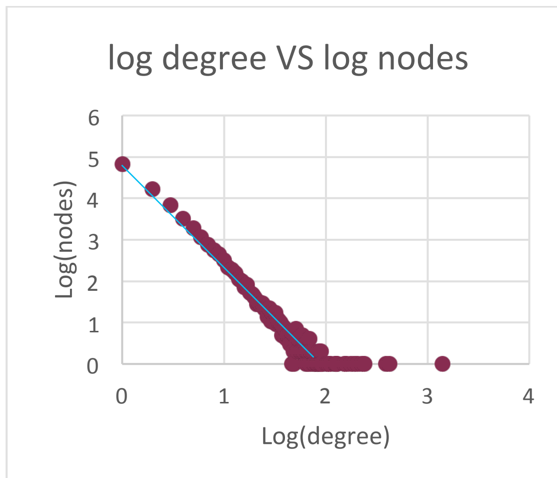
Nodes number: 20000





Nodes number: 80000





In order to show that the graph the program made is scale free net work graph, the definition or characteristics for scale free graph is needed.

An important characteristic of scale-free networks is the clustering coefficient distribution, which decreases as the node degree increases. This distribution also follows a power law.[1]

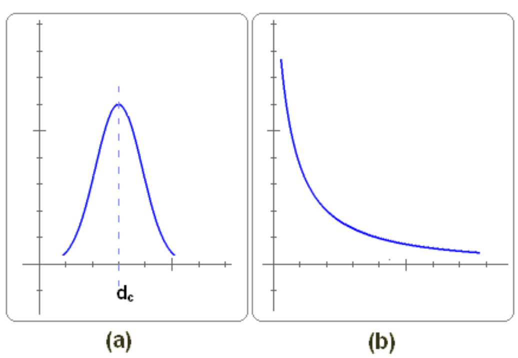


Figure.1 network degree distribution of random & scale free

(b) in figure.1 shows the power law and that represents scale free; meanwhile (a) represents network degree distribution of random.

This character is easily confirmed by the plots in the result part above.

We can recognize that a degree distribution has a power-law form by plotting it on a log-log scale. As shown in the above scatter plot, the points will tend to fall along a line. The line gets pretty messy, though, for large degree, as there are few points to average out the noise.[2][3][4]

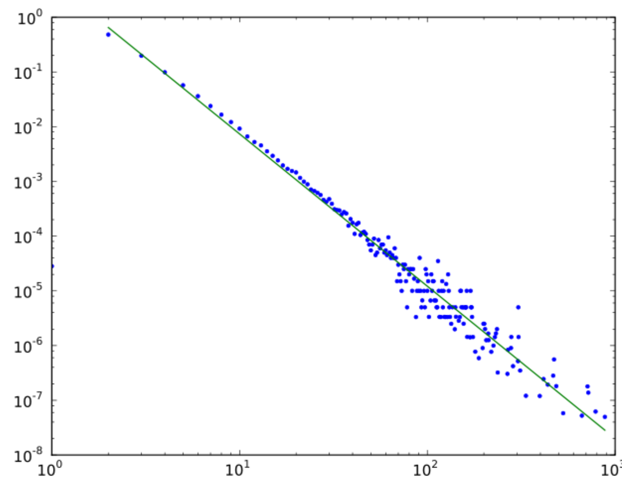


Figure.2 log-log plots for scale free distribution

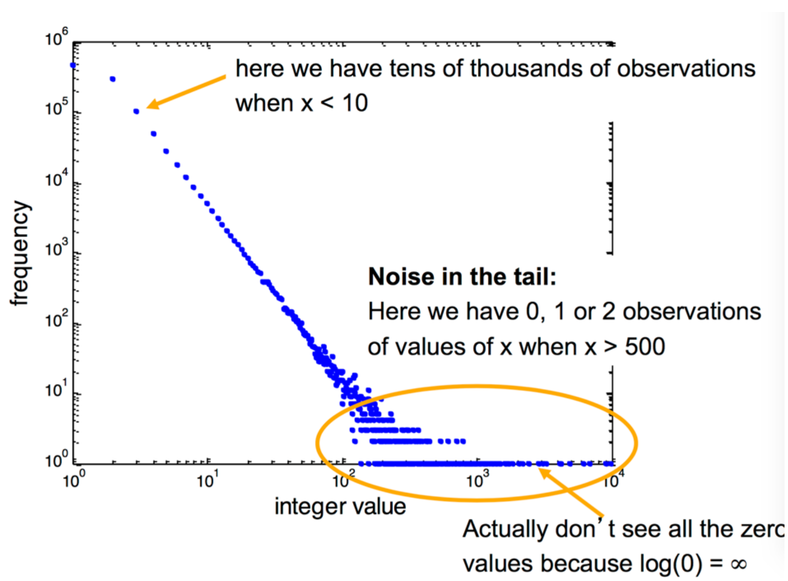


Figure.3 Analysis

Compared with the plots obtained from the results, in any case (though it's not clear when nodes number is small such as 200), the scatter plot of a power-law degree distribution on log-log scale, points lie approximately along a line. (The line is marked in blue on the plots.) The results fit the description in the assignment as well as what we got from the literature research.

Reference:

[1] Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. SIAM review, 51(4), 661-703.

[2] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001), The degree sequence of a scale-free random graph process. Random Struct. Alg., 18: 279–290. doi:10.1002/rsa.1009

[3]Barab AL and Albert R. Statistical mechanics of complex networks. Rev. Modern Physics, 7:47-97, 2002.

[4] Barabási, A. L. (2009). Scale-free networks: a decade and beyond. science, 325(5939), 412-413.

- File format

The format for the graph file is as follows:

```
10
1       0
2       0
2       1
3       1
4       1
5       2
6       1
7       4
8       1
9       6
```

Sample graph.txt

The first line is number of nodes (e.g. here we have 10 nodes).

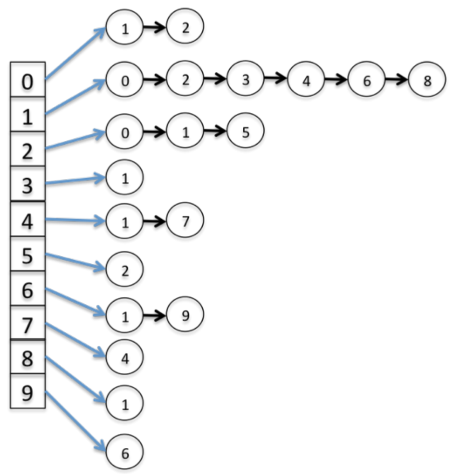
Then the following line: each line represents an edge, and those two numbers represents 2 vertices as the edge ends (e.g. First edge is an edge between node 0 and node 1).

- Part 2

This section is mainly deal with the analysis of the network size and contains 3 parts: load and store of the graph, breadth-first search and save the output.

1. Graph Representation

We use the linked list approach to store the graph, which is very space-efficient especially for sparse graphs. Concretely, for the sample graph.txt, an array of 10 pointers will be created, each of which is pointing the head of a linked list. The following figure shows the data structure:



2. Breadth-first Search

We use an iterative way to do breadth-first search of the graph. To achieve this, we use the queue from assignment 2 to store the nodes that we are going to process in the next round, and we also use a Boolean array to record whether the a node has been visited. Following is the pseudo code:

```
q = Queue()
visited = [False] * N
q.push(startNode)  # BFS starts with this startNode
dist = -1
while q is not empty:
    dist++
    siz = q.size()  # record total number of nodes at this level
    nodesAtThisLevel = []  # store nodes at this level and will be returned if last level
    for i in range(siz):  # process every node at this level
        curNode = q.pop()
        visited[curNode] = True
        nodesAtThisLevel.add(curNode)
        for node in curNode.adjacentList and not visited[node]:
            q.push(node)  # push node of next level to the queue
    if q is empty:  # done with BFS for the startNode
        return dist, nodesAtThisLevel  # return node list with max dist to the startNode
```

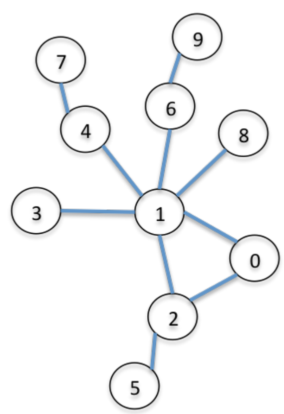
We need to do BFS for N times, each time starting at one of the nodes in the graph. For each BFS the time complexity is O(N), where N is the total number of nodes as well as the total number of edges in the graph. So the total time complexity for this step will be $O(N^2)$.

3. Output Format

The following shows the analysis result for the sample graph.txt. Concretely, the first number "4" denotes the diameter of the graph. Then the following 10 lines show the max distance and the list of nodes with this max distance for each of the 10 nodes.

```
4
0: 3, [9, 7]
1: 2, [9, 7, 5]
2: 3, [9, 7]
3: 3, [9, 7, 5]
4: 3, [9, 5]
5: 4, [9, 7]
6: 3, [7, 5]
7: 4, [9, 5]
8: 3, [9, 7, 5]
9: 4, [7, 5]
```

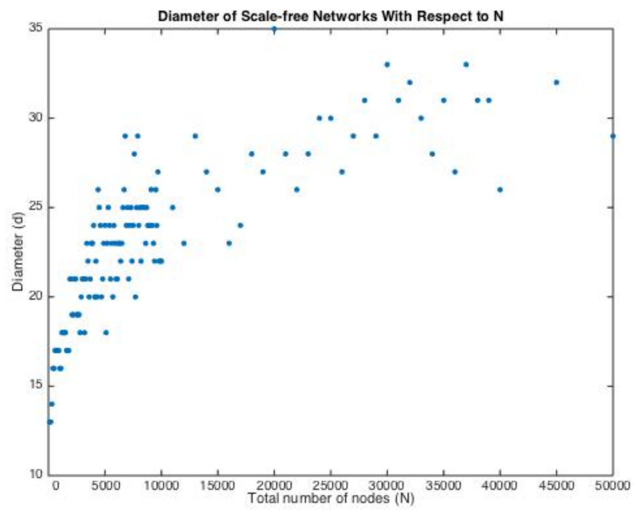
To verify the correctness of the algorithm, the following is a visualization of the graph topology:



Say for node 1, as we can see from the graph above, node 7, 9 and 5 reach the max distance from node 1, which matches our analysis result. And for another example, node 5, both node 9 and 7 have the max distance, which also matches the analysis result.

There is an interesting phenomenon in the result: most of the nodes actually have the same list of max-distant nodes, for example most of the nodes in the result above have node 9 and 7 as their max-distant nodes. This phenomenon can be more obvious with larger graphs.

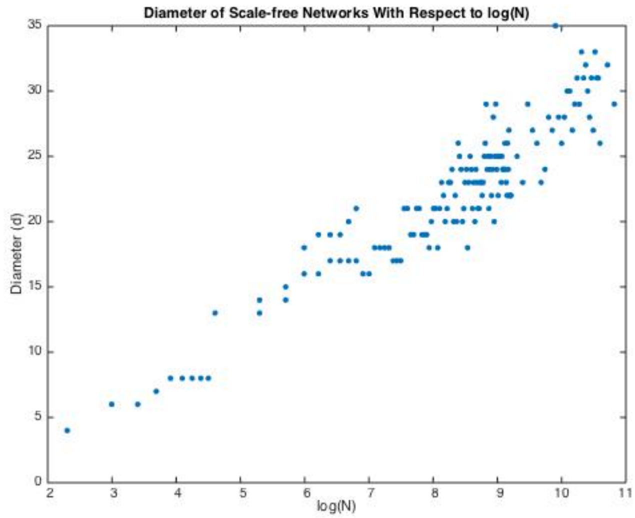
4. Result



This plot shows how the diameter changes as the number of nodes N grows. Overall, the diameter increases as N becomes larger, especially when N is relatively small, say less than 10000. However, when N gets large, this increasing trend becomes subtle. And one thing we can notice is that the diameter is actually very small (approximately 30 or so) even though the graph size can be really large.

Previous works have been done regarding to the diameter of a scale-free network. Bela and Oliver[1] shows that the diameter of a scale-free random graph is proportional to log(n) when m = 1. Here, they consider a random graph process in which vertices are added to the graph one at a time and joined to a fixed number m of earlier vertices. They also show that for m >= 2 the diameter is asymptotically log(n) / log(log(n)).

For our result, we change the x-axis to be log(N), and the plot is as below:



As we can see, the diameter has an approximately linear relationship with log(N). In our graph, when a new node is to be added, it will only choose one from all the existing nodes and create a connection, which means m = 1 as describe in the paper. Therefore, our result is a good match to the previous works.

Reference

[1] Bela Bollobas, Oliver Riordan. The Diameter of a Scale-Free Random Graph. Combinatorica Bolyai society, 24(1), (2004) 5-34.

- Literature Report-1

Scale free networks have several different models that arise in different applications.

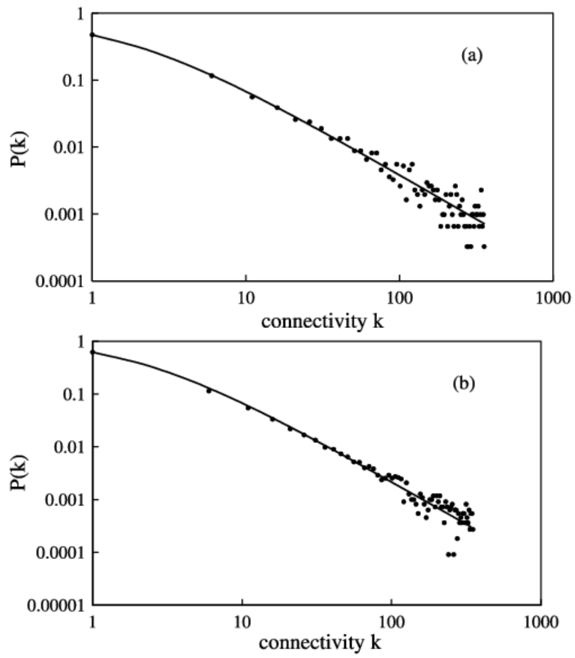
Applications where scale-free networks are believed to arise.

1. World-wide web[1]

   The world-wide web forms a large directed graph, whose vertices are documents and edges are links pointing from one document to another. Despite its apparent random character, the topology of this graph has a number of universal scale-free characteristics. In Albert's discussion, he introduced a model that leads to a scale-free network, capturing in a minimal fashion the self-organization processes governing the world-wide web. To show that www is scale-free network topology, he built a model to represent www by scale free network. A major assumption in the model was the use of a linear relationship between $\Pi(k_i)$ and k_i, given by $\Pi(k_i) = k_i / \sum_j k_j$. However, at this point there is nothing to guarantee that $\Pi(k)$ is linear, i.e., in general we could assume that $\Pi(k)$□k^{α}, where $\alpha \neq 1$. The precise form of $\Pi(k)$ could be determined numerically by comparing the topology of real networks at not too distant times. In the absence of such data, the linear relationship seems to be the most efficient way to go. In principle, if nonlinearities are present (i.e., $\alpha \neq 1$), that could affect the nature of the power-law scaling.

2. earthquakes [2]

   The district of Southern California and Japan are divided into small cubic cells, each of which is regarded as a vertex of a graph if earthquakes occur therein. Two successive earthquakes define an edge or a loop, which may replace the complex fault-fault interaction. In this way, the seismic data are mapped to a random graph. It is discovered that an evolving random graph associated with earthquakes behaves as a scale-free network of the Barabási-Albert type. The distributions of connectivities in the graphs thus constructed are found to decay as a power law, showing a novel feature of earthquake as a complex critical phenomenon. This result can be interpreted in view of the facts that the frequency of earthquakes with large values of moment also decays as a power law (the Gutenberg-Richter law) and aftershocks associated with a mainshock tend to return to the locus of the mainshock, contributing to the large degree of connectivity of the vertex of the mainshock. Thus, a mainshock plays the role of a "hub". It is also found that the exponent of the distribution of connectivities is characteristic for the plate under investigation.



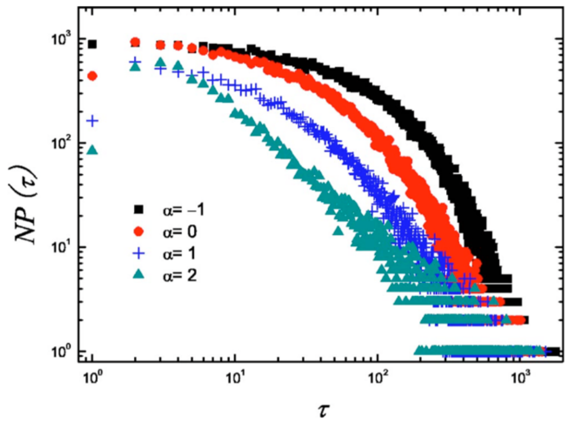
The scale-free nature of the earthquake network may accept the following natural interpretation. The frequency of earthquakes with large values of moment decays as a power law due to the Gutenberg-Richter law, on the one hand, and the analysis of the seismic data, on the other hand, shows that aftershocks associated with a mainshock tend to return to the locus of the mainshock geographically and therefore contribute to the large degree of connectivity of the vertex of the mainshock. In this way, the principle of preferential attachment is satisfied and the scale-free nature of the earthquake network is realized.

3. Traffic dynamics based on local routing protocol [3]

Wang proposed a packet routing strategy with a tunable parameter based on the local structural information of a scale-free network. As free traffic flow on the communication networks is key to their normal and efficient functioning, he focused on the network capacity that can be measured by the critical point of phase transition from free flow to congestion. Simulations showed that the maximal capacity corresponds to =−1 in the case of identical nodes' delivering ability. To explain this, Wang investigate the number of packets of each node depending on its degree in the free flow state and observe the power law behavior. The results indicate that some fundamental relationships exist between the dynamics of synchronization and traffic on the scale-free networks.

The traffic model is described as follows: at each time step, there are R packets generated in the system, with randomly chosen sources and destinations, and all nodes can deliver at most C packets towards their destinations. To navigate packets, each node performs a local search among its neighbors. If the packet's destination is found within the searched area, it is delivered directly to its target. Otherwise, it is forwarded to a nodes i, one of the neighbors of the searching node, according to the preferential probability: $\Pi_i = \frac{k_i^{\alpha}}{\sum k_j^{\alpha}}$

where the sum runs over the neighbors searched area of the searching node, ki is the degree of node i and is an adjustable parameter.



Reference:

[1] Barabási, A. L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. Physica A: Statistical Mechanics and its Applications, 281(1), 69-77.

[2] Abe, S., & Suzuki, N. (2004). Scale-free network of earthquakes. EPL (Europhysics Letters), 65(4), 581.

[3] Wang, W. X., Wang, B. H., Yin, C. Y., Xie, Y. B., & Zhou, T. (2006). Traffic dynamics based on local routing protocol on a scale-free network. Physical Review E, 73(2), 026111.

- Literature Report-2

We consider a random graph process in which vertices are added to the graph one at a time and joined to a fixed number m of earlier vertices, where each earlier vertex is chosen with probability proportional to its degree.[1] This process was introduced by Barab́asi and Albert [3], as a simple model of the growth of real-world graphs such as the world-wide web. Computer experiments presented by Barab́asi, Albert and Jeong [2,4] and heuristic arguments given by Newman, Strogatz and Watts [5] suggest that after n steps the resulting graph should have diameter approximately logn. We show that while this holds for m=1, for m≥2 the diameter is asymptotically logn/loglogn.

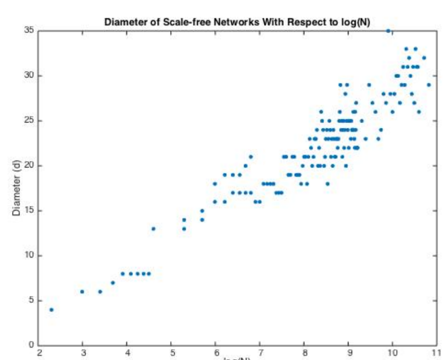
In several papers random graphs modeling complex real-world networks are studied, in particular with respect to their diameter or average diameter, the average distance between two vertices. One model studied is the process G_m^n considered here. Another model is given by taking a function P(d), d≥1, to represent the probability that a vertex has degree d, and choosing a graph on n vertices at random from all graphs with this distribution of degrees.

In [2,4,5] results from computer experiments have been published suggesting that for the second model the diameter of the random graph will vary as A+Blogn when n is large, for a wide variety of functions P(d). In [5] a heuristic argument is given, based on the standard neighbourhood expansion method; roughly speaking, if one follows a random edge to one of its end vertices v, the probability P’ (d) that v will have degree d should be proportional to dP(d). Thus, writing N (v) for the set of vertices within distance k of a given vertex v, one would expect $|N_{k+1}(v)|$ to be about

$$f = \sum dP'(d) = \left(\sum d^2 P(d)\right) / \left(\sum dP(d)\right)$$

times larger than $|N_k(v)|$. One then expects the diameter of the graph to be about logn/logf.

The plot for our program(in Log scale)：



the diameter has an approximately linear relationship with log(N). In our graph, when a new node is to be added, it will only choose one from all the existing nodes and create a connection, which means m = 1 as describe in the paper. Therefore, our result is a good match to the previous works.

Reference:

[1] Bela Bollobas, Oliver Riordan. The Diameter of a Scale-Free Random Graph. Combinatorica Bolyai society, 24(1), (2004) 5-34.

[2] R.Albert,H.JeongandA.-L.Baraba si:Diameteroftheworld-wideweb,Nature 401 (1999), 130–131.

[3] A.-L. Baraba si and R. Albert: Emergence of scaling in random networks, Science 286 (1999), 509–512.

[4] A.-L. Baraba si, R. Albert and H. Jeong: Scale-free characteristics of random networks: the topology of the world-wide web, Physica A 281 (2000), 69–77.

[5] M. E. J. Newman, S. H. Strogatz and D. J. Watts: Random graphs with arbi- trary degree distribution and their applications, Physical Review E 64 (2001), 026118.