Social Networks and Online Markets
Homework 1

Flavia Monti, 1632488

The Erdös-Rènyi G_{np} random graph model

This type of model is a random graph model, it has two parameters: n and p; n is the number of nodes of the graph and p is the probability that an edge exists.
By some analysis done during the lectures, the degree distribution followed by this model is the Binomial one. The clustering coefficient of one node is define as $cc_v = \frac{\binom{n}{2} \cdot p}{\binom{n}{2}} = p$, so the clustering coefficient of the graph is $cc = p$. For what regard the diameter there is not a precise value, it is of the order of $\frac{lnn}{lnnp}$ but for the hand-waving done during the computation of that value it is said that the diameter is double that value.
A particular aspect of this model is the Threshold Phenomenon, this phenomenon holds for monotone probability. It says that a property will not hold with high probability for small values of p until some value p* (that depends on the property). Then, as soon as p > p* the property will hold with high probability. So there is a threshold at the value p*.
The property of connectivity is affected by this phenomenon. When p is less than 1/n the graph contains lot of small connected components (size O(lnn)), when it is almost 1/n there is a phase transition and a large connected component (of size $n^{2/3}$) appears. When p reaches $\frac{lnn}{n}$ there is another phase transition where there is a giant connected component and a constant number of isolated nodes, when p is greater than that value the graph becomes connected.

In the following table there are some results from the experiments done with different values of, in particular with 15k, 10k and 5k nodes, and for different values of p (chosen looking at the Threshold phenomenon).

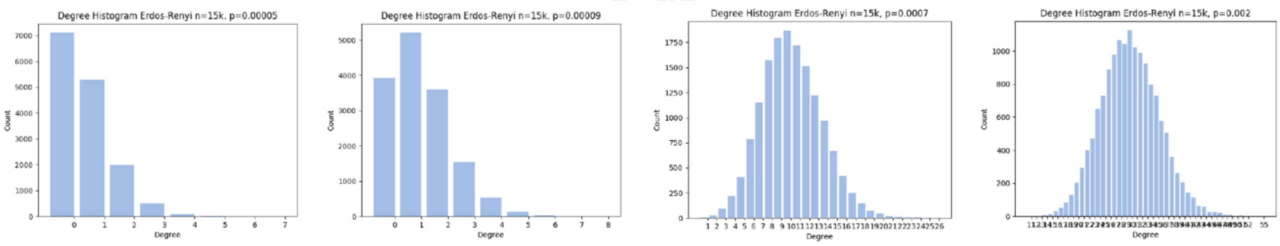
	p	Clustering coefficient	Diameter	Size bigger connected component	Connected
n=15000	0.00005	0.0	30	47	No
	0.00009	0.0	59	7296	No
	0.0007	0.0006534	7	78414	Yes
	0.002	0.0020057	4	225416	Yes
n=10000	0.00008	0.0	22	107	No
	0.0004	0.000408	14	19756	No
	0.0009	0.000901	7	44797	No
	0.002	0.001990	5	100260	Yes
	0.005	0.004985	4	250365	Yes
n=5000	0.0002	0.0003	33	194	No
	0.0008	0.000711	14	9811	No
	0.002	0.001625	7	25169	Yes
	0.005	0.005087	4	62472	Yes

Taking in consideration the observations done before, we can see that the values of the different parameters for different values of n and p are respected:

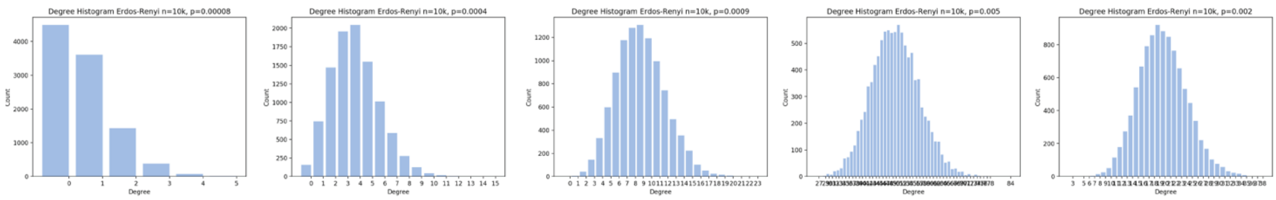
- The clustering coefficient is quite equal to p. When the clustering coefficient is 0.0 it means that the graph does not contains any 'triangle'.
  The fact that the clustering coefficient is quite equal to p is a problem because p is usually small and in reality the clustering coefficient is higher.
- The diameter is the double of $\frac{lnn}{lnnp}$. In some cases the graph is not connected and the diameter is computed to be the maximum between the diameters of the different connected components of the graph. In these cases the diameter looks strange this is because the connected component can have many different shapes.
- The study of the connectivity follow the Threshold Phenomenon.
  For different values of the probability it is possible to see the trend of the size of the biggest connected component going from small size to a size much bigger. Also the connectivity of the graph switch to "connected" only when the probability is greater than $\frac{lnn}{n}$.

Degree distribution:

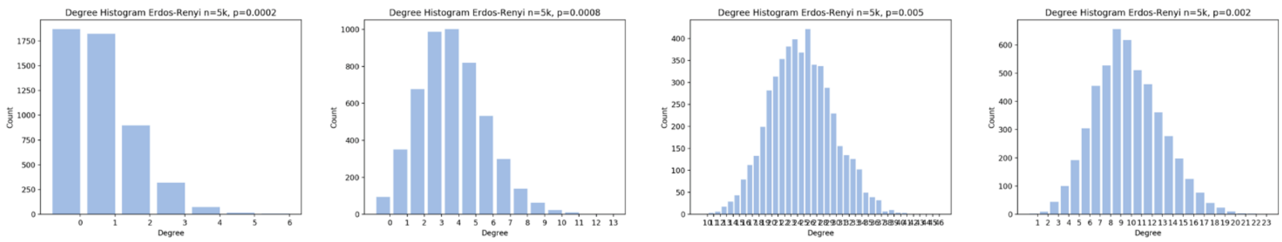
n = 15k



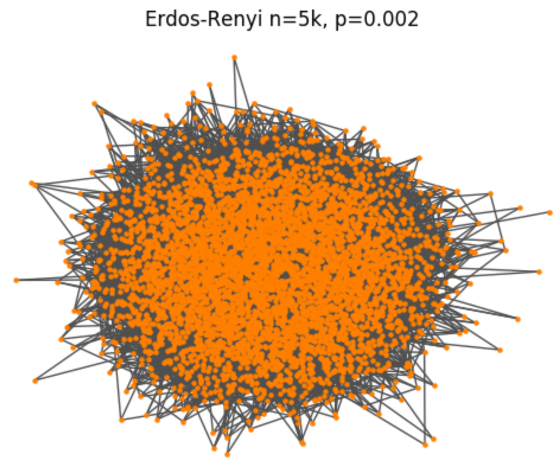
n = 10k



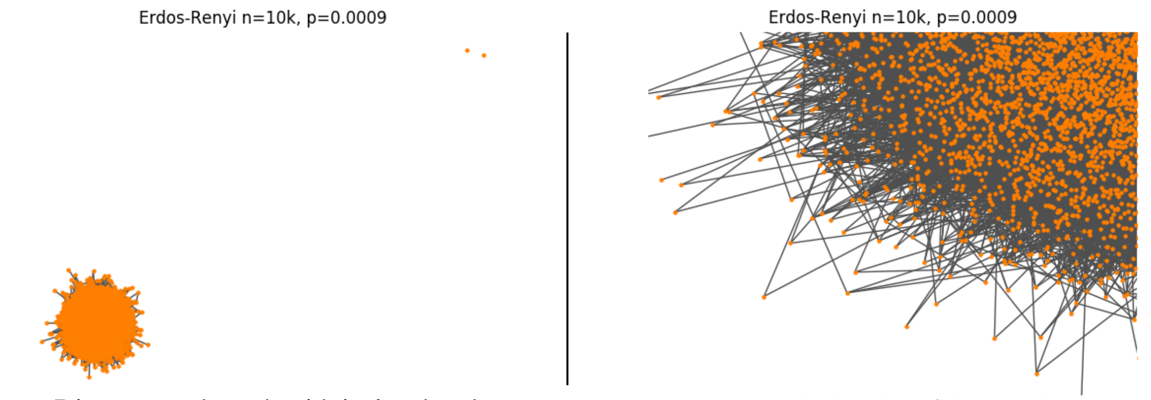
n = 5k



The degree distribution of the Erdos-Renyi random graph model follows the binomial distribution. When p assume small value there is a particular distribution, it seems to follow the exponential distribution, maybe due to the fact that there are lot of isolated nodes and a connected component with not to many nodes. When the probability grows, it is possible to see how the distribution transforms into a binomial one. There are lot of nodes with $d_v = \frac{\max(d)}{2}$ and less nodes with high degree.



Connected graph



Disconnected graph with isolated nodes

Zoom on the border of the graph

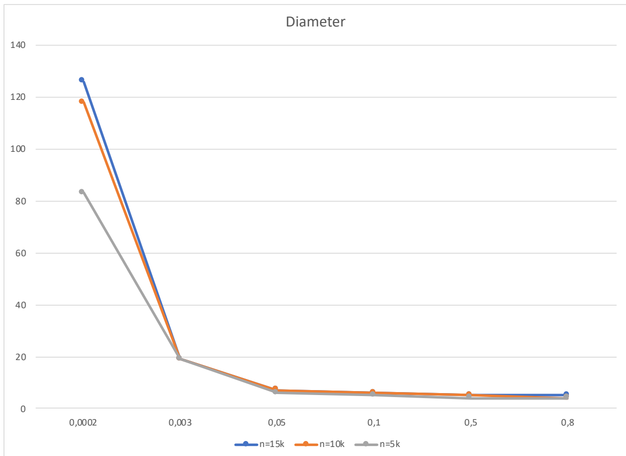
The Watts-Strogatz small-world model

This model is a particular model that follows the small-world property. This type of graph is obtained from a graph structured as a ring where each edge has the same number of neighbors, then with some small probability edges are replaced with new random ones.
This model has two parameters: k and beta, where k is the number of neighbors of a single node (k/2 on the left and k/2 on the right) while beta is the probability of rewiring one edge with a new random one.
An important aspect of this model is the relationship between the diameter and the clustering coefficient. From the calculation done during the lecture, the clustering coefficient is approximately k/2 and the diameter is n/k.

In the following table there are the results of the experiments:

	beta	Clustering coefficient	Diameter
n=15000 k=25	0.0002	0.716981	126
	0.003	0.710860	19
	0.05	0.616189	7
	0.1	0.525272	6
	0.5	0.091445	5
	0.8	0.006936	5
n=10000 k=25	0.0002	0.717027	118
	0.003	0.711542	19
	0.05	0.613727	7
	0.1	0.524386	6
	0.5	0.092164	5
	0.8	0.007991	4
n=5000 k=25	0.0002	0.716939	83
	0.003	0.711457	19
	0.05	0.615069	6
	0.1	0.522429	5
	0.5	0.091925	4
	0.8	0.010042	4

Observing the values taken by the clustering coefficient and by the diameter, it is possible to notice how they variate depending on the value taken by the probability beta.

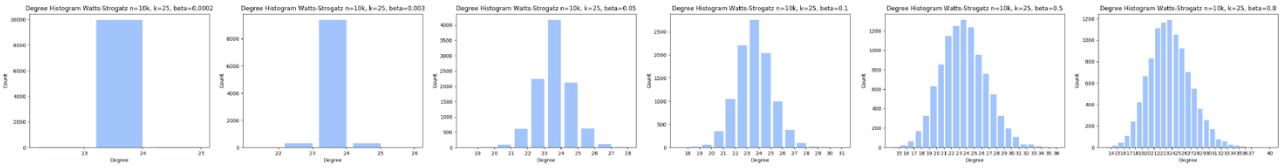


There are six different values taken by p going from 0.0002 to 0.8, it is possible to see that on one side the clustering coefficient stays high and at the end drops down, while on the other side the diameter drops down immediately.

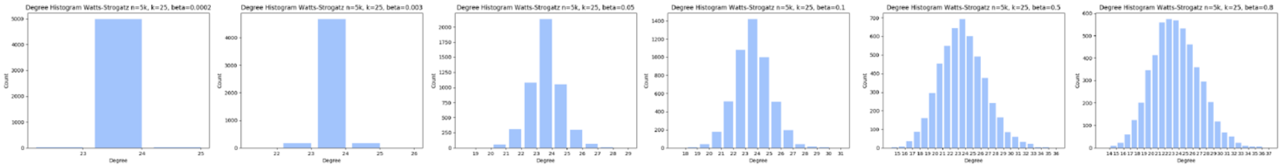
When p is small, the clustering coefficient is large and does not change too much because a small number of edges are rewired and edges between neighbors exist and this small number of random edges between neighbors reduce a lot the value of the diameter. When this happens we can say that we are in a situation like the reality.

Below are represented the various degree distribution for the different values taken by the probabilities:

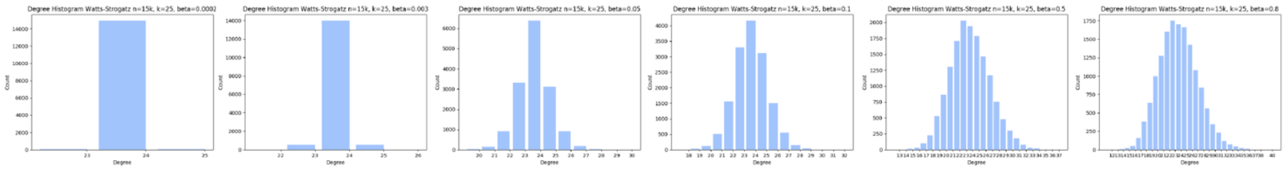
n = 15k



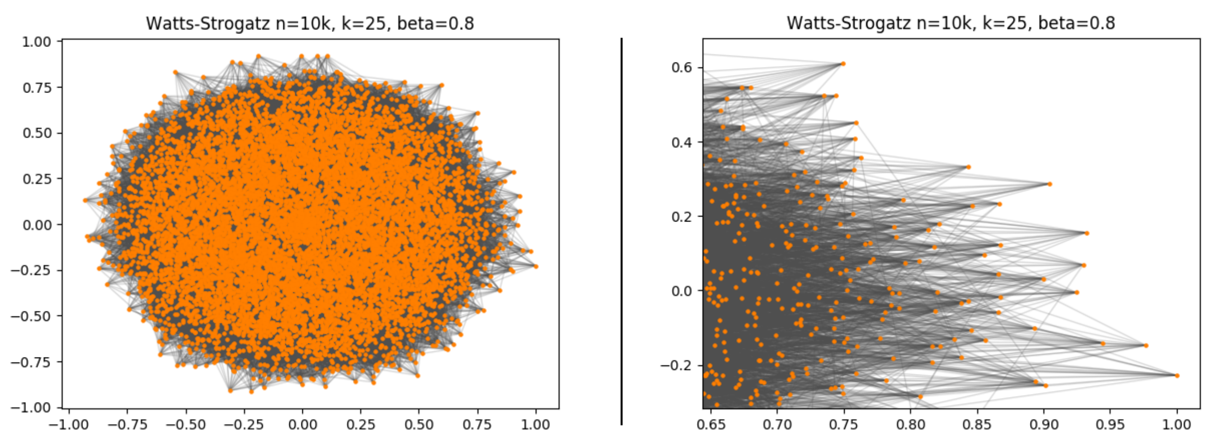
n = 10k



n = 5k



In all above images it is possible to see the binomial distribution, there is a small number of nodes having low and high degree and there are many nodes that have degree with intermediate values.



Zoom on the border of the graph

The Barabási-Albert preferential attachment model

This model is a random graph model that have the characteristic to follow the power-law distribution for the degree.

It contains initially a small number of nodes and as time goes on new nodes are added to the graph. This model has two parameter n, which is the number of nodes, and l, that is the number of edges that a new node has. A particular aspect of this model is the preferential attachment characteristic consisting on the following: when a new node u comes, the probability that a node v is its neighbor is proportional to the degree of node v. Higher is the degree of v, higher is the chance that u will be attached to v. The probability that a node is v is attached to the new node u at time t-1 is $\frac{d_v}{2(t-1)l}$, the denominator is the sum of the degree of all the nodes (except for node u) at time t-1.

Studying the flow of the degree as t increase it is possible to come up with the power law distribution:

- number of nodes with small degree (k < l): 0
- number of nodes with degree equal to l (k = l): $\frac{2}{l+2}$
- number of nodes with large degree (k > l): $\frac{2l(l+1)}{k(k+1)(k+2)}$. And for large k the denumerator is approximately k^3. So the degree distribution follow approximately a power-law distribution with exponent equal to 3.

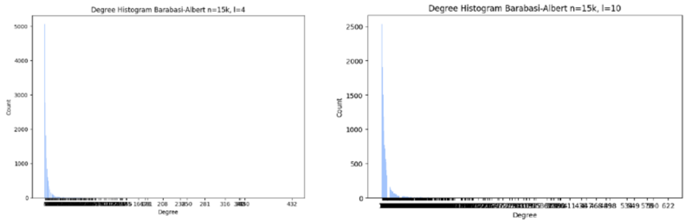
Many experiments are done with different values of n and l.
Below there is a table containing results about the clustering coefficient and diameter.

	l	Clustering coefficient	Diameter
n=15000	4	0.004494	7
	10	0.008423	5
n=10000	4	0.006852	6
	10	0.011199	5
n=5000	4	0.011140	6
	10	0.019523	4

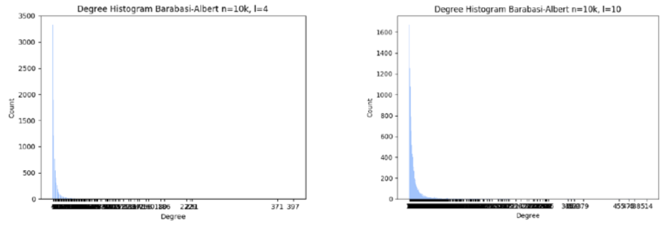
Clustering coefficient is quite small in all the cases and the diameter does not take big value.
A clustering coefficient with these small value is not like the reality where instead networks are more clustered. The reason of this small clustering is the preferential attachment property and the effect of "rich get richer" on the graph. When a new node arrives, the neighbors are chosen with some probability that depends on their degree, this does not create a graph like the previous model (Watts-Strogatz) where the probability of rewiring an edge is not a depending on the degree of each node.

Down here there are some plots of the degree distribution of the graphs.

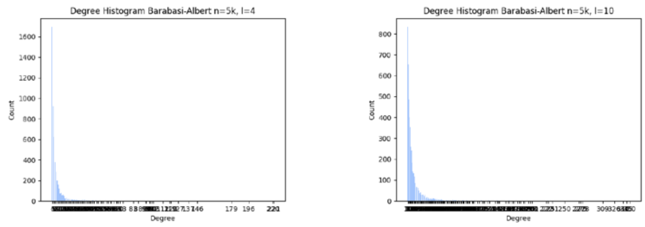
n = 15k



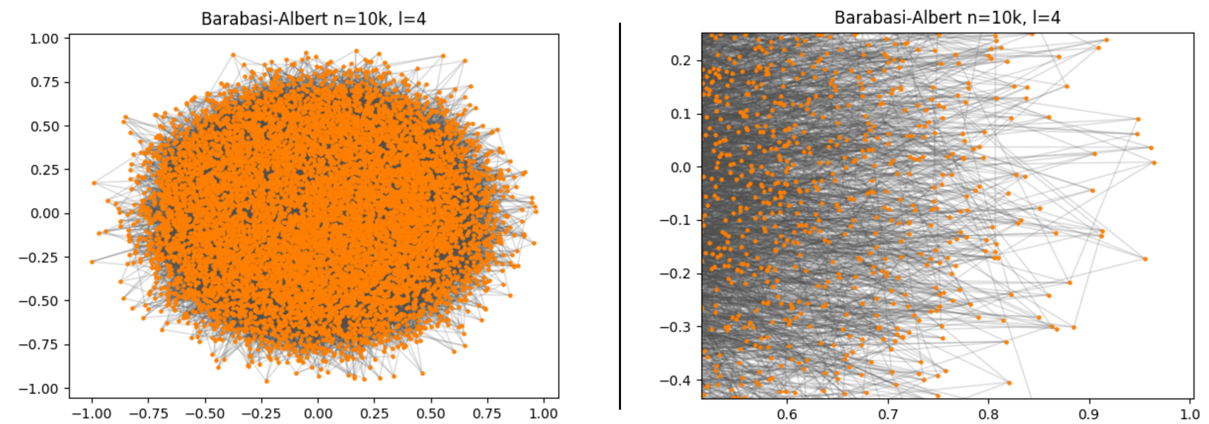
n = 10k



n = 5k



From these plots regarding it is possible to see the flow of the power-law distribution. There are a lot of nodes with low degree and some single nodes having degree much higher. This is because of the preferential attachment, nodes with higher degree have more probability to be the endpoint of new nodes that comes. This phenomenon is also called “rich get richer”.



Zoom on the border of the graph

References:

- Anagnostopoulos A., Social Networks and Online Markets – Notes
- Easley D., Kleinberg J., Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Cambridge University Press, 2010
- https://networkx.github.io/documentation/stable/