An Overview of Random Graphs

drafted by

Tathagata Das

{03CS3022,tathagata.das1565@gmail.com}

I. Introduction

A random graph is a collection of points, with edges, connecting pairs of them at random. The theory of random graphs began in the late 1950s in several papers by Paul Erdos and Albert Renyi, which is why these graphs are also known as Erdos-Renyi graphs.

II. Literature Survey

The field of random graphs was started by the publications from Paul Erdos and Albert Renyi(1959-1960s), which dealt with evolution and phase transitions of random graphs. After them, there has been an enormous number of publications in this field. In this report we will be concentrating on the giant components of random graphs and the most influential work in this field has been done since the late 1990's Molloy & Reed and Newman , Strogatz & Watts.

III. Random Graph Models

A random graph is obtained by starting with a set of n vertices and adding edges between them at random. Different random graph models produce different probability distributions on graphs. The most commonly studied model, usually called the Erdos-Renyi graphs, is written as $G_{n,p}$, where n is the number of nodes in the graph and p is the probability of any edge existing between any pair of nodes. This probability for one edge is independent of the existence of any other edge in the graph. Based on these assumptions we can find out that the average degree of $G_{n,p}$ is $z = \frac{p}{n-1} \approx \frac{p}{n}$ as n is large (n>>1). Also, the probability of a node having degree k is given by $p_k = ({}^N C_k) p^k.(1-p)^{N-k} = \frac{e^{-z}.z^k}{k!}$.
A closely related model, $G_{n,m}$ defines the set of graphs having n vertices and m randomly selected edges. Still another model of random graphs is a random graph with a given arbitrary probability distribution of the degrees of their vertices. In all respects other than their degree distribution, these graphs are assumed to be entirely random. This means that the degrees of all vertices are independent identically distributed random integers drawn from a specified distribution. For a given choice of these degrees, also called the "degree sequence", the set of random graphs having the degree sequence is called a Microcanonical Ensemble.

IV. Microcanonical Ensemble

In studying the properties of random graphs, graph theorists often concentrate on the limit behavior of random graphsthe values that various probabilities converge to as n grows very large. In such cases, a Microcanonical Ensemble is a set of all large graphs having the same degree sequence that matches as closely as possible to the desired degree probability distribution. Properties of such graphs are calculated by averaging over the whole ensemble of graphs of the given degree sequence.

V. Phase Transition

One of the most interesting aspects of this addition of the edges to form the random graph is the Phase Transition. There are two distinct phases in the formation of random graphs. Initially, the graph is disconnected and later, after addition of a certain number of edges, the graph becomes largely connected. Largely connected need not mean fully connected, it only means a large majority of the nodes is connected. Here comes the concept of Giant Components. Giant components are large connected components of a random graph, whose size is proportional to the size of the whole graph, i.e. O(n). So it increases linearly as the size of the graph increases. The emergence of GC in a evolving random graph marks the transition of the graph to the connected phase. Erdos and Renyi found out that there is a sharp threshold for the emergence of giant components, which is as follow: [7]

- If $p = c/n$ and $c < 1$ then, when n is large, most of the connected components of the graph are small, with the largest having only $O(logn)$ vertices.
- In contrast if $c > 1$ there is a constant $\Theta(c) > 0$ so that for large n the largest component has $\sim \Theta(c)n$ vertices and the second largest component is $O(logn)$.

VI. Giant Components

Giant Components is perhaps the most studied phenomenon in the field of random graphs is the behavior of the size of the largest component in $G_{n,p}$. The major question on which we will be concentration in this discussion is that whether there can exist multiple giant components in a large random graph or not. For that purpose let us first understand the definitions of the terms to be used, then we prove that in the thermodynamic limit multiple giant components cannot exist.

A. Multiple Giant Components

One of the major question that arises in relation to giant components is that whether there can exist multiple giant components in a large random graph or not. So let us try to find out whether two giant components can exist is a random graph. That is given a ER random graph $G_{n,p}$ of n nodes, what is the probability that there exist two giant components GC1 (size n1) and GC2 (size N2). We are using the $G_{n,p}$ model, so we want to find out that what is the probability that these two components will not get connected by the edges that are randomly *thrown* on the graph.

$$P(GC1\ and\ GC2\ not\ connected\ by\ the\ edge) = 1\ -\ P(GC1\ and\ GC2\ gets\ connected\ by\ the\ edge)$$

$$=\ 1 - \frac{N1 * N1}{^{n}C_2} \qquad [1]$$

$$Total\ number\ edges\ in\ G_{n,p}\ =\ ^{n}C_2.p \qquad [2]$$

Therefore,

$$P(none\ of\ those\ edges\ connect\ GC1\ and\ GC2) = (1\ -\ \frac{N1 * N1}{^{n}C_2})^{^{n}C_2.p} \qquad [3]$$

Here we have taken the following assumptions.

- N1 = O(n) and N2 = O(n) but n >> N1,N2 >> m
- N1 and N2 are so big that addition of a node to n or addition of an edge from N1 to N2 does not make any difference in the probabilities.

Now let us try to analyze what happens to the probability at the thermodynamic limit, i.e. $n \to \infty$.

$$Let\ L = Lim_{n\to\infty} = [(1\ -\ \frac{N1 * N1}{^{n}C_2})^{^{n}C_2.p}] \qquad [4]$$

Now, at $n \to \infty$, $^{n}C_2 \approx \frac{n^2}{2}$. Therefore,

$$L = Lim_{n\to\infty}[(1\ -\ \frac{N1.N1}{\frac{n^2}{2}})^{\frac{n^2}{2}}]^p \qquad [5]$$

$$L = [Lim_{n\to\infty}(1\ -\ \frac{N1.N1}{\frac{n^2}{2}})^{\frac{n^2}{2}}]^p \qquad [6]$$

$$L = e^{-N1.N2.p} \qquad [7]$$

For random graphs $G_{n,p}$, the average degree $z = (n-1).p$, i.e. $z \approx n.p$. Also we know $N1 = O(n)$, hence $N1 = \delta_1.n$. Similarly, $N1 = \delta_2.n$. Substituting these we get,

$$L = e^{-N1.N2.p} = e^{-\delta_1.\delta_2.n^2.\frac{z}{n}} \qquad [8]$$

$$L = e^{-const.n} \qquad [9]$$

At thermodynamic limit, $Lim_{n\to\infty}L = 0$. Therefore we can conclude that as the size of the ER random graph increase to infinity, the probability of having 2 giant components tends to zero.

B. Existence of a node in GC

At the point of formation of the single giant component, the size of the component is $O(n^{\frac{2}{3}})$.
Let us try to analyze the probability of a node being in the giant component. Let u be the probability that the node is not in a giant component. Probability that all its k neighbors are not in giant component is u^k.
Now, if a node is in a giant component, then it implies all its neighbors are not in the giant component.
Prob of one node = Prob of having k neighbors X Prob of all k neighbors not in GC

$$u = \sum_{k=0}^{\infty} p_k.u^k \qquad [10]$$

where p_k = probability of a node having k neighbors $= ({}^N C_k)p^k.(1-p)^{N-k} = \frac{e^{-z}.z^k}{k!}$

$$u = \sum_{k=0}^{\infty} \frac{e^{-z}.z^k}{k!}.u^k \qquad [11]$$

$$u = e^{-z}.\sum_{k=0}^{\infty} \frac{(z.u)^k}{k!} = e^{-z}.e^{z.u} \qquad [12]$$

$$u = e^{-z(1-u)} \qquad [13]$$

Therefore s = probability of node not being in GC $= 1 - u$

$$s = 1 - e^{-z.s} \qquad [14]$$

The first non-zero solution of this equation is the required probability.

C. Existence of GC in a generalized random graph of given degree sequence

In the paper *The size of the giant component of a random graph with a given degree sequence* by *Molloy & Reed*, they have suggested the following:

Given a sequence of nonnegative real numbers $\lambda_0, \lambda_1, \lambda_2, \ldots$ which sum to 1, a random graph having approximately $\lambda_i n$ vertices of degree i will have a giant component at the thermodynamic limit if $\sum i(i-2)\lambda_i > 0$.

This essentially means that given a degree sequence $k_0, k_1, k_2, \ldots$, a large random graph $(n \to \infty)$having that degree sequence will have a giant component if $\sum k_i(k_i - 2) > 0$.

This can be understood in an intuitive manner. If we are trying to traverse a the network like a graph by maintaining a list of unexplored nodes, then for a giant component to exist we must ensure that the list does not become empty i.e. the connected component can go on expanding. Now when we come to a node i having degree k_i, we now have k_i new nodes to traverse, which we have got at the cost of traversing

the node i. So in the list of unexplored nodes increases by $(k_i - 1) - 1 = k_i - 2$.
Now we can argue that probability of reaching a node of degree k_i is k_i times the probability of reaching a node of degree 1 (because it can be reached by k different edges). Therefore,

$$P(reaching\ node\ of\ deg\ k_i) = k_i.P(reaching\ node\ of\ deg\ 1) = k_i.const$$

Hence we can say that $\sum p_k.(k-2) > 0$ ensures that the list of unexplored nodes will never be empty. Let the sum be S.

$$S = \sum p_k.(k-2) \approx \sum (k_i.const)(k_i - 2) = const. \sum k_i(k_i - 2) > 0 \qquad [15]$$

$$\sum k_i(k_i - 2) > 0 \qquad [16]$$

This is the result we have.

VII. Conclusion

From the above discussions we can conclude that in the asymptotic case, ER graph of the form $G_{n,p}$ cannot have more than one giant components. Also the probability of a node being in the giant component is $s = 1 - e^{-z.s}$. And given a degree sequence $k_0, k_1, k_2, \ldots$, a large random graph at thermodynamic limit having that degree sequence will have a giant component if $\sum k_i(k_i - 2) > 0$.

VIII. References

1. "On Random Graphs" - Erdos P., and A. Renyi, *Publicationes Mathematicae 6, pp.290-297 (1959)*
2. "On the Evolution of Random Graphs" - Erdos P., and A. Renyi, *Hungarian Academy of Sciences 5, pp. 17-61 (1960)*
3. "Random graphs with arbitrary degree distribution and their applications" - M. E. J. Newman, S. H. Strogatz and D. J. Watts, *Reviews of Modern Physics, 2002*
4. "A critical point for random graphs with a given degree sequence" - M. Molloy & B. Reed, *Random Structures and Algorithms 6, 161-179 (1995)*
5. "The size of the giant component of a random graph with a given degree sequence" - M. Molloy & B. Reed, *Combinatorics, Probability and Computing 7, 295-305(1998).*
6. Random Graphs on Wikipedia - $http://en.wikipedia.org/wiki/Random_graph$
7. Random Graph Dynamics - Book by Rick Durrett, Cornell U., Published by Cambridge U. Press, October 2006 *http://www.math.cornell.edu/ durrett/RGD/fch1.pdf*