# Theory of Random Graphs and Generating Functions

Priyesh Jaipuriar

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur, India
priyeshjaipuriar@gmail.com

## 1 Random Graphs

### 1.1 Introduction

A random graph is obtained by starting with n vertices and adding edges between them at random. In mathematics, a random graph is a graph that is generated by some random process. The theory of random graphs lies at the intersection between graph theory and probability theory, and studies the properties of typical random graphs.

One of the motivations for studying random graphs is the desire to describe a “typical”graph. For example, considering all labelled graphs on n vertices, it is known that the vast majority of the graphs are connected, contain a copy of any fixed graph F, etc. Moreover, the proportion of graphs not having these properties decreases as n grows. Hence, it is justified in saying that “almost all” graphs have these properties. In the terminology of random graphs, it is said that a random graph has these properties with probability tending to 1 as the number of vertices tends to infinity.
Other interesting results can be obtained if restricted to the subclasses of graphs, for example by fixing the number of edges and asking what the typical properties of a graph with n vertices and m edges are. Often random graphs are thought to be states in a process. Beginning at time 0 with an empty graph of n vertices, edges are added to the graph at random, either uniformly or according to some other random procedure.
In graph theory, the Erdos-Renyi model, named for Paul Erdos and Alfred Renyi, is either of two models for generating random graphs, including one that sets an edge between each pair of nodes with equal probability, independently of the other edges. It can be used in the probabilistic method to prove the existence of graphs satisfying various properties, or to provide a rigorous definition of what it means for a property to hold for ”almost all” graphs.

A discovery by Erdos and Renyi was that many graph properties have threshold phenomena: when number of edges in the random graph in significantly smaller than the threshold, it has the property with probability very close to 0, while if the number of edges is significantly greater than the threshold, it has the property with probability very close to 1.

### 1.2 History

Random graphs were first defined by Paul Erdos and Alfred Renyi in their 1959 paper [1], which dealt with evolution and phase transitions of random graphs.
The $G_{n,p}$ model was first introduced by Edgar Gilbert in a 1959 paper [2] which studied the connectivity threshold. According to this model, a graph has n nodes and an edge is added between every pair of nodes with a probability equal to p, with the presence or absence of any two distinct edges in the graph being independent. The $G_{n,M}$ model was introduced by Paul Erdos and Alfred Renyi in their 1959 paper [3]. According to this model, a graph is chosen uniformly at random from the collection of all graphs which have n nodes and M edges. For example, in the $\mathrm{G}_{3,2}$ model, each of the three possible graphs on three vertices and two edges are included with probability $\frac{1}{3}$. As with Gilbert, their first investigation were as to the conectivity of $G_{n,M}$, with the more detailed analysis following in 1960.
Our Concentration will be on the giant components of random graphs. The most influential work in this field has been done since the late 1990's Molloy & Reed [4] and Newman, Strogatz & Watts.

### 1.3 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 Erdos-Renyi Model

Most commonly studied model, usually called the Erdos-Renyi graphs, is written as $G_{n,p}$, in which each of the ${}^{n}C_2$ possible edges occurs independently with probability $p$.
For a node to have degree $k$, it means that it has to select $k$ out of the other $(n-1)$ nodes as its neighbors. The total number of ways in which it can select $k$ out these $(n-1)$ nodes is ${}^{(n-1)}C_k$. Now it will add an edge to each of the $k$ selected nodes with a probability $p$ each and not add any edge to the remaining $(n-1)-k$ nodes with a probability of $(1-p)$ each.

Therefore, the probability of a node having degree $k$ is given by

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

The average degree of a graph is given by

$$\frac{\sum_{1}^{N} degree}{N} = \frac{\sum k.n_k}{N}$$

where $n_k$ is the number of nodes with degree k

$$= \sum k.p_k = \sum k.{}^{n-1}C_k.p^k.(1-p)^{(n-1)-k} = (n-1)p \approx np \; (for\ large\ value\ of\ n) \tag{2}$$

Let $z = np$, therefore, $p = \frac{z}{n}$
Now, when n is large,

$$p_k = \lim_{n\to\infty} {}^{n-1}C_k.p^k(1-p)^{(n-1)-k} = \lim_{n\to\infty} \frac{(n-1)!}{(n-k-1)!\ k!}\left(\frac{z}{n}\right)^k\left(1-\frac{z}{n}\right)^{n-k-1}$$

$$= \lim_{n\to\infty}\left[\frac{(n-1)!}{n^k\ (n-k-1)!}\right].\left(\frac{z^k}{k!}\right)\left(1-\frac{z}{n}\right)^{n-1}\left(1-\frac{z}{n}\right)^{-k} \; (On\ rearrangement)$$

Now,

$$\left(1-\frac{z}{n}\right)^{n-1} \approx e^{-z},\ \left(1-\frac{z}{n}\right)^{-k} \approx 1\ for\ n \to \infty$$

Therefore

$$p_k = \lim_{n\to\infty}\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)....\left(\frac{n-k}{n}\right).\left(\frac{z^k}{k!}\right).e^{-z}$$

Hence,

$$p_k = \frac{e^{-z}.z^k}{k!} \tag{3}$$

, where $z = np$.

### The $\mathbf{G}_{n,M}$ Model

A closely related model, $G_{n,M}$ assigns equal probability to all graphs with exactly M edges. Both models can be viewed as snapshots at a particular time of the random graph process, which is a stochastic process that starts with n vertices and no edges and at each step adds one new edge chosen uniformly from the set of missing edges.

### The Random Dot-Product Model

Another model, which generalizes the Erdos-Renyi graphs, is the random dot-product model. Associated with each vertex $x$ of a random dot-product graph is a real vector $f(x)$. The probability of an edge $uv$ between any vertices $u$ and $v$ is some function of the dot product $f(u).f(v)$ of their respective vectors.

### Configuration Model

In this model of a random graph, we specify a degree distribution $p_k$, such that $p_k$ is the fraction of vertices in the graph with degree $k$. We choose a degree sequence, which is a set of $n$ values of the degrees $k_i$ of the vertices $i = 1, 2, ...n$ from this distribution. We can think of this as giving each vertex $i$ in the graph $k_i$ *spokes* or *stubs* sticking out of it, which are the ends of edges-to-be. Then we can choose pairs of stubs at random from the network and connect them to form edges. This process generates every possible configuration of a graph with a given degree sequence. The *configuration model* is defined as the ensemble of graphs so produced, with each having equal weight [5].

### Microcanonical Ensemble

Still another model of random graphs is with a given arbitrary probability distribution of the degrees of vertices. These graphs are assumed to be entirely random in all respect other than their degree distribution. This means that the degrees of all vertices are independent identically distributed random integers drawn from a specified distribution. The set of random graphs having the degree sequence, which is the given choice of degrees, is called a Microcanonical Ensemble.

In studying the properties of random graphs, graph theorists often concentrate on the limit behavior of random graphs  the 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.

A scale-free network is a noteworthy kind of complex network because many "real-world networks" fall into this category. "Real-world" refers to any of various observable phenomena that exhibit network theoretic characteristics.

In scale-free networks, some nodes act as "highly connected hubs"(high degree), although most nodes are of low degree. 'Scale-free networks' structure and dynamics are independent of the system's size N, the number of nodes the system has. In other words, a network that is scale-free will have the same properties no matter what the number of its nodes is. Their defining characteristic is that their degree distribution follows the Yule-Simon Distribution – A Power law relationship defined by

$P(k) \approx k^{-\gamma}$

### 1.4 Phase Transition

A random graph with $n$ vertices and $0.49n$ edges is very likely to consist of many small components, none of which has more than $O(logn)$ vertices, while a random graph with $n$ vertices and $0.51n$ edges most probably contains a unique large component containing a linear number of vertices[3]. This large component is called giant components(discussed in later section). The particular point in the graph process where the giant component is first formed is generally referred to as the phase transition, because of the similarities to the physical phenomenon of substances turning from one phase to another by a small change in temperature or pressure.

The phenomenon is also related to percolations in statistical mechanics. In percolation theory a typical question is whether, and with which probability, the centre of a porous stone becomes wet if the stone is put into water. The stone can be represented by a graph, with different points being adjacent if they are connected by a hole in the stone.

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. Assume that the number of edges in the system to be denoted by $N(n)$ and $N(n) = c.n$ where c is a positive constant. Erdos and Renyi [3] found out that there is a sharp threshold for the emergence of giant components, which is as follows:

- If $c < 1/2$ then, when $n$ is large, most of the connected components of the graph are small, with the largest having only $O(logn)$ vertices.
- If $c \sim 1/2$ and $n$ is large, the size of the largest component is $O(n^{2/3})$
- And, if $c > 1/2$ then the size of the largest component of the graph is $O(n)$

### 1.5 Giant Components

Giant component is a network theory term referring to a connected subgraph that contains a majority of the entire graph's nodes.

The most studied phenomenon in the field of random graphs is the behavior of the size of the largest component in $G_{n,p}$. A very prominent question is that whether there can exist multiple giant components in a large random graph or not.Multiple giant components cannot exist in the thermodynamic limit(proof later).

#### Multiple Giant Components

One of the major questions that arises in relation to giant components is that whether there can exist multiple giant components in a large random graph or not. Given an ER random graph $G_{n,p}$ of $n$ nodes, what is the probability that there exists two giant components GC$_1$ (size N$_1$) and GC$_2$ (size N$_2$). The $G_{n,p}$ model is used here to find what is the probability that these two components will not get connected by the edges that are randomly *thrown* on the graph. Assume that the graph does consist of two giant components GC$_1$ and GC$_2$. The probability that there is no edge from any vertex of GC$_1$ to any vertex of GC$_2$ is given by

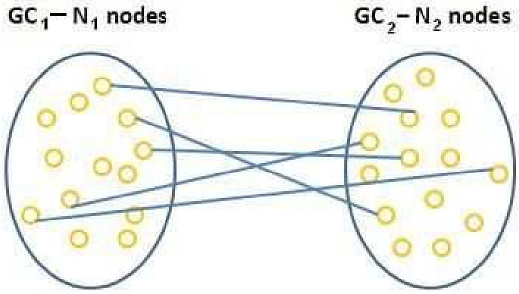

**Fig. 1.** A graph with two giant components.

$P(GC_1\, and\, GC_2\, not\, connected\, by\, the\, edge) = 1 - P(GC_1\, and\, GC_2\, get\, connected\, by\, the\, edge)$

Now, P(GC$_1$ and GC$_2$ get connected by the edge) is equal to the probability that the two endpoints of the edge are not in the same giant component which is $1 - \frac{N_1 * N_2}{^nC_2}$

$$= 1 - \frac{N_1 * N_2}{^nC_2}$$

$$Total\ number\ edges\ in\ G_{n,p} = {^nC_2}.p$$

Therefore,

$$P(none\ of\ those\ edges\ connect\ GC_1\ and\ GC_2) = (1 - \frac{N_1 * N_2}{^nC_2})^{^nC_2.p} \quad (4)$$

The following assumptions were taken:

- $N_1 = O(n)$ and $N_2 = O(n)$ but $n$ >> N$_1$,N$_2$ >> m
- $N_1$ and $N_2$ are so big that addition of a node to $n$ or addition of an edge from $N_1$ to $N_2$ does not make any difference in the probabilities.

Now analyzing the probability at the thermodynamic limit, i.e. for the value where $n \to \infty$.

$$Let\ L = Lim_{n \to \infty}[(1\ -\ \frac{N_1 * N_2}{^nC_2})^{^nC_2 . p}]$$

Now, at $n \to \infty$, $^nC_2 = \frac{n*(n-1)}{2} \approx \frac{n^2}{2}$.
Also as $n \to \infty = \frac{n^2}{2} \to \infty$
Therefore,taking the above assumption in equation and considering that p is independent of n,

$$L = Lim_{\frac{n^2}{2} \to \infty}[(1\ -\ \frac{N_1.N_2}{\frac{n^2}{2}})^{\frac{n^2}{2}}]^p$$

$$L = [Lim_{\frac{n^2}{2} \to \infty}(1\ -\ \frac{N_1.N_2}{\frac{n^2}{2}})^{\frac{n^2}{2}}]^p$$

Now it is known that:

$$Lim_{n \to \infty}(1\ -\ \frac{X}{n})^n = e^{-X}$$

Therefore,

$$L = e^{-N_1.N_2.p}$$

For random graphs $G_{n,p}$, the average degree $z = (n-1).p$, i.e. $z \approx n.p$. Also we know $N_1 = O(n)$, hence $N_1 = \delta_1.n$. Similarly, $N_2 = \delta_2.n$. Substituting these we get,

$$L = e^{-N_1.N_2.p} = e^{-\delta_1.\delta_2.n^2.\frac{z}{n}}$$

$$L = e^{-const.n} \tag{5}$$

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. This means that we cannot have more than 1 giant component in the ER random graph of a very large size.

### Existence of a node in Giant Component of a ER Graph

At the point of formation of the single giant component, the size of the component is $O(n^{\frac{2}{3}})$ [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 not in the giant component, then it implies all its neighbors are also not in the giant component.
Prob of one node not in GC $(u)$ = Prob of having k neighbors × Prob of all k neighbors not in GC. Now $p_k$ is the probability of having degree $k$ and the

individual probabilities of the neighbors not being in GC is $u$. We can sum this for all possible values of degree $k$ to obtain the expression for $u$. Therefore the probability $u$ can be expressed as

$$u = \sum_{k=0}^{\infty} p_k.u^k$$

where $p_k$ = probability of a node having k neighbors $= ({}^N C_k)p^k.(1-p)^{N-k} = \frac{e^{-z}.z^k}{k!}$ from Eqn. 3. Therefore,

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

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

$$u = e^{-z(1-u)}$$

Therefore s = probability that a node is in the GC $= 1 - u$

$$s = 1 - e^{-z.s} \tag{6}$$

The first non-zero solution of this equation is the required probability. We can see from this equation that if $z < 1$, then $s < 0$. This means that if the average degree of the nodes is less than 1 then a giant component does not exist. This is obvious because an average degree of less than 1 implies that the graph is disconnected.

$s$ is also known as the 'order parameter' of the network.

Fig. 2 shows the mean component sizes and the size of the giant component. We can see that the giant component emerges at $z = 1$. This is also the point at which the mean component size in the absence of giant component diverges.

### Existence of Giant Component in a generalized random graph of given degree sequence

In the paper by *Molloy & Reed* [4], 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 ($n \rightarrow \infty$)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 the network like a graph by maintaining a list of unexplored nodes, then for a

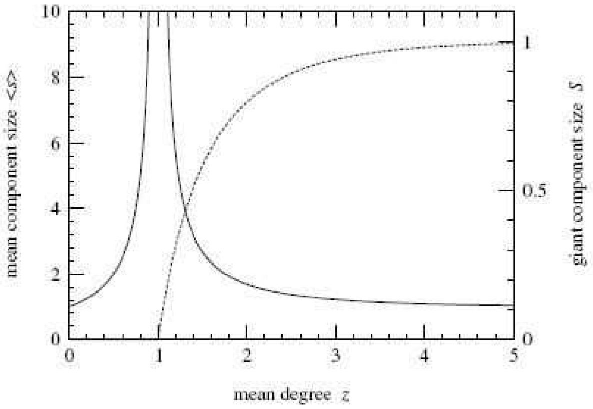

**Fig. 2.** The solid line shows the mean component size excluding the giant component if there is one, the dotted line shows the giant component size, for the Poisson random graph.

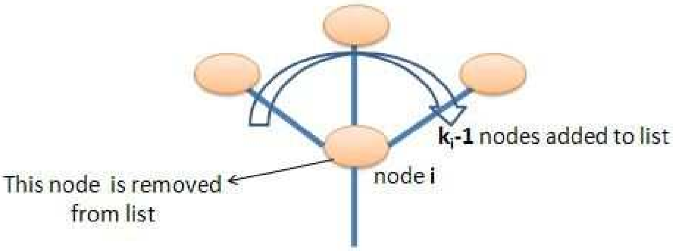

**Fig. 3.** Change in the list of unexplored nodes : The *ith* node is removed from the list and $k_i - 1$ are added

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 $k$ (referred to as $k_i$), we now have $k_i - 1$ new nodes to traverse, which we have got at the cost of traversing the node $i$. These new nodes are added to the unexplored list and the vertex $i$ is removed from the list. So 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,

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

Therefore, the increase in size of the unexplored list for a node of degree $k$ will be $p_k.(k-2)$ (where $p_k$ is the probability of arriving at the node with degree $k$). Therefore, for the total increase in the size of the list, this has to be summed over all $k$. 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 $S$.

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

$$\sum k_i(k_i - 2) > 0 \tag{7}$$

This is the result we have.

## 2 Generating Functions

### 2.1 Introduction

In mathematics, a generating function is a formal power series whose coefficients encode information about a sequence $a_n$ that is indexed by the natural numbers.
There are various types of generating functions, including ordinary generating functions, exponential generating functions, Lambert series, Bell series and Dirichlet series. Every sequence has a generating function of each type. The particular generating function that is most useful in a given context will depend upon the nature of the sequence and the details of the problem being addressed.
Generating functions are often expressed in closed form as functions of a formal argument $X$. Sometimes a generating function is evaluated at a specific value of $X$. However, it must be remembered that generating functions are formal power series, and they will not necessarily converge for all values of $X$. It is very important to note that generating functions are not functions in the formal sense of a mapping from a domain to a codomain; the name merely stems from the historical study of the structures.
A general approach to random graphs with given degree distribution was developed by Newman, Strogatz, and Watts (2001) using a generating function formalism (Wilf, 1990). It turns out that many properties of the network model are exactly solvable in the limit of large network size. The crucial trick for finding the solution is that instead of working directly with the degree distribution $p_k$, we work with generating functions of the sequence of the degree distribution $p_k$, which is defined as

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k. \tag{8}$$

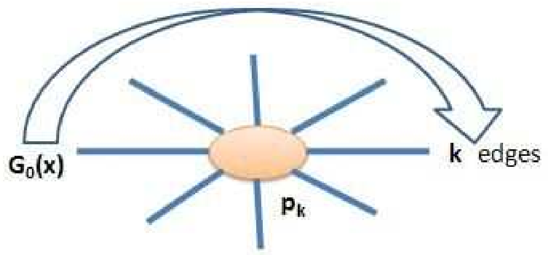

**Fig. 4.** The first neighbors of a node are given by G$_0$(x).

Formalism for calculating various local and global quantities on large undirected graphs with arbitrary probability distribution of the degrees of their vertices is presented here.

### 2.2 Degree Distribution of a randomly picked node

Considering an undirected graph of $N$ vertices, with $N$ large, the generating function $G_0(x)$ for the probability distribution of vertex degrees $k$ is defined as follows:

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \tag{9}$$

where $p_k$ is the probability that a randomly chosen vertex on the graph has degree $k$. The distribution $p_k$ is normalized, so that

$$G_0(1) = 1 \; as \; G_0(1) = \sum_k p_k = 1$$

The function $G_0(x)$ encapsulates all the information of the degree distribution $p_k$. Therefore the probability $p_k$ is given by the $k^{th}$ derivative of $G_0$,

$$p_k = \frac{1}{k!} \frac{d_k G_0}{dx^k} |_{x=0}.$$

Hence, the function $G_0(x)$ is said to ”generate” the probability distribution $p_k$.

The average degree $z$ of a vertex is nothing but the average over the probability distribution generated by the generating function $G_0(x)$. It is given by

$$z = \langle k \rangle = \sum_k k p_k \tag{10}$$

Now we know that

$$G_0(x) = \sum_k p_k x^k$$

$$\Rightarrow G_0'(x) = \sum_k k.p_k x^{k-1}$$

$$Therefore\ G_0'(1) = \sum_k kp_k = \langle z \rangle$$

Higher moments of the distribution can be calculated from higher derivatives of $G_0(x)$. In general we have

$$\langle k_n \rangle = \sum_k k^n p_k = [(x\frac{d}{dx})^n G_0(x)]_{x=1} \tag{11}$$

If a distribution of a property $k$ of an object is generated by a given generating function, then the distribution of the total of $k$ summed over $m$ independent realizations of the object is generated by the $m^{th}$ power of that generating function. This is also referred to as the *Powers* property of the generating functions. Therefore the distribution of the sum the degrees of $m$ randomly chosen vertices is generated by $[G_0(x)]^m$. For example, the degree distribution of two nodes taken together (if they are independent) is given by $[G_0(x)]^2$ Now,

$$[G_0(x)]^2 = p_0^2 + (p_1.p_0 + p_0.p_1)x + (p_2.p_0 + p_1^2 + p_0.p_2)x^2 + .....\ to\ \infty$$

Consider the coefficient of $x^2$ (which is the probability that the sum of the degrees of the two nodes is 2) - $p_2.p_0 + p_1^2 + p_0.p_2$. It is the sum of three probabilities namely - node$_1$ has degree 2 and node$_2$ 0, both the nodes have degree 1, node$_1$ has degree 0 and node$_2$ 2.

### 2.3 Degree Distribution of a node arrived at from a random edge

Now considering the degree distribution of the vertices that are arrived at by following a random edge. The probability to arrive at a node by a random edge is proportional to the degree of the vertex. Therefore the probability distribution of degree of a vertex is proportional to $kp_k$. After arriving at a node, the number of residual(all outgoing connections apart form the one which was used to arrive at the node) connections are $k-1$. Hence the exponent of $x$ is $k-1$. The correctly normalized distribution known as $G_1(x)$ is given by

$$G_1(x) = \frac{\sum_k kp_k x^{k-1}}{\sum_k kp_k} = \frac{G_0^{'}(x)}{G_0^{'}(1)}. \tag{12}$$

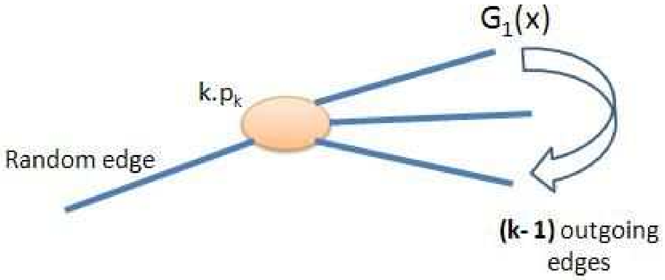

**Fig. 5.** The degree distribution of a node reached from a random edge, given by $G_1(x)$.

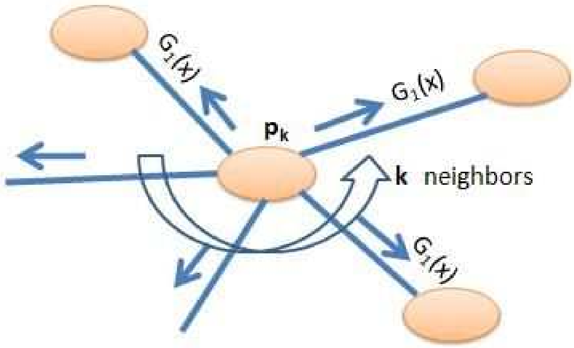

**Fig. 6.** The distribution of second neighbors of a node, given by $G_0(G_1(x))$.

### 2.4 Distribution of Second Neighbors

To calculate the generating function of the second neighbors of a node, we can pick a $k$ degree node with probability $p_k$ and then move along all $k$ of its edges with degree distribution of its neighbors given by $G_1(x)$ Thus, the generating function for the probability distribution of the number of *second* neighbors of the node can be written as

$$\sum_k p_k[G_1(x)]^k \ which\ is\ same\ as\ G_0(G_1(x)).$$

The average number of second neighbors, $z_2$ is given by

$$z_2 = [\frac{d}{dx}G_0(G_1(x))]_{x=1} = G_0^{'}(1)G_1^{'}(1) = G_0^{''}(1) \tag{13}$$

### 2.5 Distribution of Third Neighbors

To get the distribution of three hop neighbors we have to consider three steps -
(a) Choose a $k$ degree node at random with probability $p_k$
(b) Now, traverse to its neighbors through its edges and obtain their outgoing degree(which is its second neighbors) distribution as $G_1(x)$
(c) Now repeat this process for all its first neighbors to get the 3 hop degree distribution of the original node.

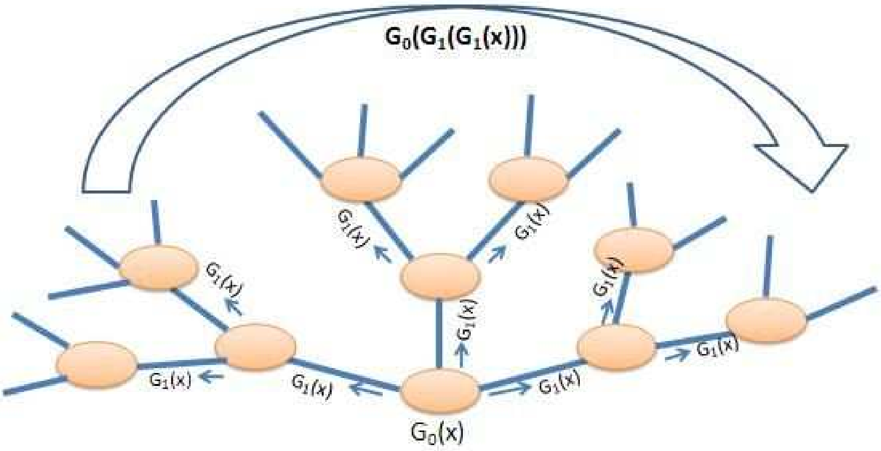

**Fig. 7.** The distribution of third neighbors of a node, given by $G_0(G_1(G_1(x)))$.

Hence the third neighbor distribution can be written as $G_0(G_1(G_1(x)))$. Hence, the average number of third neighbors $z_3$ is given by

$$z_3 = \frac{d}{dx}\left(G_0(G_1(G_1(x)))\right)|_{x=1} = G_1'(1).G_1'(1).G_0'(1)$$

$$z_3 = [G_1'(1)]^2.G_0'(1) \tag{14}$$

This can be generalized for m hop neighbors as well. so, the distribution for $mth$ neighbor would be

$$G^{(m)}(x) = G_0(G_1(...G_1(x)...\textit{(m-1) times})) \tag{15}$$

and the average number of m-hop neighbors is given by

$$z_m = [G_1'(1)]^{m-1}.G_0'(1) \tag{16}$$

## 2.6 Examples

### Poisson distributed graphs

In this distribution model the probability $p = z/N$ of the existence of an edge between any two vertices is the same for all vertices. The $G_0(x)$ is given by

$$G_0(x) = \sum_{k=0}^{N} {}^{N}C_k p^k (1-p)^{N-k} x^k$$

$$= (1 - p + px)^N = e^{z(x-1)},$$

where the last equality applies in the limit $N \longrightarrow \infty$. The average degree of a vertex is $G_0^{`}(1) = z$. In this case

$$G_1(x) = \frac{G_0^{'}(x)}{G_0^{'}(1)} = \frac{ze^{z(x-1)}}{z} = G_0(x).$$

Therefore the distribution of out-going edges at a vertex is the same, regardless of whether we arrived there by choosing a vertex at random, or by following a randomly chosen edge. This property makes the theory of the random graphs simple.
The average number of first neighbors of a poisson graph is given by

$$z \approx Np \tag{17}$$

The average number of second neighbors of a node is given by $z_2 = \mathrm{G}_0''(1)$.

$$G_0(x) = e^{z(x-1)}$$

$$\Rightarrow G_0'(x) = z.e^{z(x-1)}$$

$$\Rightarrow G_0''(x) = z^2.e^{z(x-1)}$$

So,

$$z_2 = G_0''(1) = z^2 \tag{18}$$

### Exponentially distributed graphs

The exponential distribution of vertex degrees is

$$p_k = (1 - e^{-1/\kappa})e^{-k/\kappa},$$

where $\kappa$ is a constant. The generating function for this distribution can obtained as follows:

$$G_0(x) = \sum_k (1 - e^{-1/\kappa})e^{-k/\kappa} x^k$$

$$G_0(x) = (1 - e^{-1/\kappa}) \sum_{k}^{\infty} e^{-k/\kappa} x^k = \frac{1 - e^{-1/\kappa}}{1 - xe^{-1/\kappa}},$$

$$G_0'(x) = \frac{1 - e^{-1/\kappa}}{(1 - xe^{-1/\kappa})^2} . e^{-1/\kappa}$$

$$G_0''(x) = \frac{1 - e^{-1/\kappa}}{(1 - xe^{-1/\kappa})^3} . 2e^{-2/\kappa}$$

The average degree $z$ is

$$\langle z \rangle = G_0^{'}(1) = \frac{e^{-1/\kappa}}{1 - e^{-1/\kappa}}$$

and

$$G_1(x) = [\frac{1 - e^{-1/\kappa}}{1 - xe^{-1/\kappa}}]^2 = [G_0(x)]^2$$

The average number of second neighbors is

$$z_2 = G_0''(1) = \frac{2e^{-2/\kappa}}{(1 - e^{-1/\kappa})^2}$$

**Power-Law distributed graphs**

The distribution is given by $p_k = Ck^{-\tau} e^{-k/\kappa}$ for $k \geq 1$ where $C, \tau and \kappa$ are constants. The exponential cutoff is included as many real-world graphs shoe this cutoff and also it makes the distribution normalizable for all $\tau$. The value of $C$ is fixed as $C = [Li_\tau(e^{-1/\kappa})]^{-1}$ as per the requirement of normalization, where $Li_n(x)$ is the $n^{th}$ polylogarithm of $x$. $G_0(x)$ can be calculated as

$$G_0(x) = \sum_{k=0}^{\infty} Ck^{-\tau} e^{-k/\kappa} x^k$$

$$= CLi_\tau(xe^{-1/\kappa})$$

$$= \frac{Li_\tau(xe^{-1/\kappa})}{Li_\tau(e^{-1/\kappa})}.$$

The first The function $G_1(x)$ is given by

$$G_1(x) = \frac{Li_{\tau-1}(xe^{-1/\kappa})}{xLi_{\tau-1}(e^{-1/\kappa})}.$$

The average number of neighbors of a randomly chosen vertex is

$$z = G_0^{'}(1) = \frac{Li_{\tau-1}(e^{-1/\kappa})}{Li_{\tau}(e^{-1/\kappa})},$$

and the average number of second neighbors is

$$z_2 = G_0^{''}(1) = \frac{Li_{\tau-2}(e^{-1/\kappa}) - Li_{\tau-1}(e^{-1/\kappa})}{Li_{\tau}(e^{-1/\kappa})}.$$

Explanation of Li$_n$(x) -
The function Li$_n$(x) or the n$^{th}$ polylogarithm of x is a special function defined by the sum

$$Li_n(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^n}$$

The polylogarithm function is a generalization of the logarithm : the generalized logarithm. It has also been given the name of Jonquiere's functions. Some special cases are -
(i) n = 1 : the function equivalent to the logarithm
(ii)n = 2 : the dilogarithm. Li$_2$(1-x) is also known as the Spence's integral.

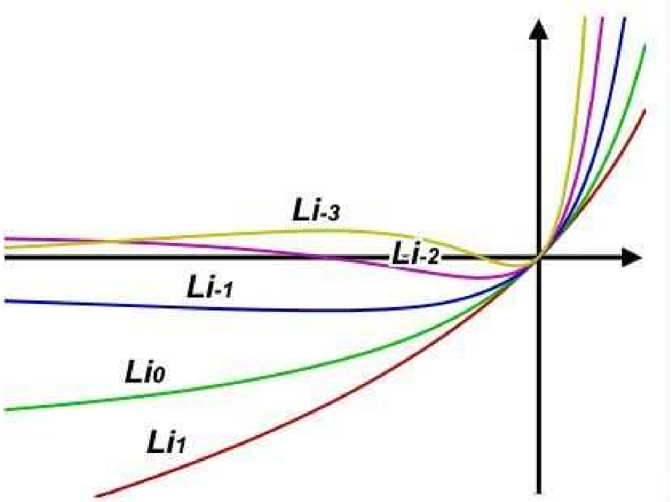

**Fig. 8.** The polylogarithmic function at certain values of n.

### 2.7 Component Sizes

Interesting properties like the distribution of the size of the connected components in the graph can be calculated using the generating function approach as discussed here. Let $H_1(x)$ be the generating function for the distribution of the sizes of components which are reached by choosing a random edge and following it to one of its ends. The giant component is excluded from $H_1(x)$. The chances of a component containing a closed loop of edges is $N^{-1}$, which is negligible in the limit of large N. The distribution of components generated by $H_1(x)$ can be represented as shown in Fig.1; each component is tree-like in structure, consisting of the single site we reach by following our initial edge, plus any number of other tree-like clusters, with the same size distribution, joined to it by single edges. If $q_k$ is the probability that the initial site has $k$ edges coming out of it other than the edge it is arrived through, then $H_1(x)$ can be written as

$$H_1(x) = xq_0 + xq_1H_1(x) + xq_2[H_1(x)]^2 + ...$$

Here the first term $xq_0$ is the case where we reach a node via a random edge and it has no other component connected to it. Here $q_0$ is the probability that the number of neighbors of a node reached from a random edge is 0 (same as the constant term in G$_1$(x)). The second term $xq_1H_1(x)$ is for the case when we reach a node from a random edge and there is one another component attached to it. Here q$_1$ is the probability that the number of neighbors of the node reached from a random edge is 1 (same as the coefficient of x in G$_1$(x)) and since the neighbor is a component itself, the term is multiplied by H$_1$(x). Similarly we have the rest of the terms of the expression. Here the coefficients q$_i$'s are the same as that in G$_1$(x).

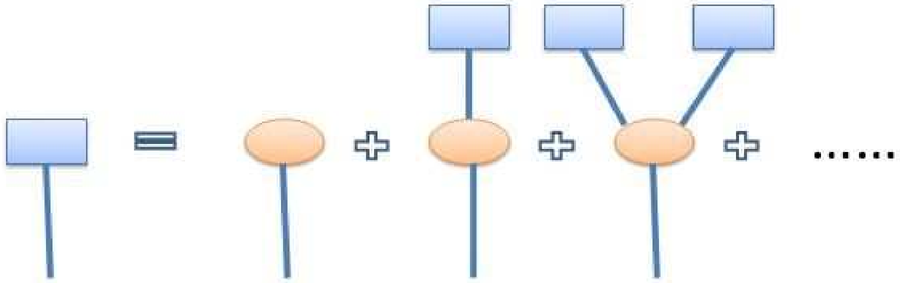

**Fig. 9.** Schematic representation of the sum rule for the connected component of vertices reached by following a randomly chosen edge.

Since, $q_k$ is the coefficient of $x^k$ in the generating function $G_1(x)$, $H_1(x)$ can be written as

$$H_1(x) = xG_1(H_1(x)). \tag{19}$$

Similarly, the distribution of component size when a node is randomly can be written as

$$H_0(x) = xp_0 + xp_1 H_1(x) + xp_2[H_1(x)]^2 + ...$$

$$H_0(x) = xG_0(H_1(x)). \tag{20}$$

Here the coefficients p$_i$'s are the same as those in G$_0$(x).

The average size of the component a randomly chosen vertex belongs, where there is no giant component, is given by

$$\langle s \rangle = H_0^{'}(1) = 1 + G_0^{'}(1)H_1^{'}(1).$$

We have

$$H_1^{'}(1) = 1 + G_1^{'}(1)H_1^{'}(1),$$

therefore, we have

$$\langle s \rangle = 1 + \frac{G_0^{'}(1)}{1 - G_1^{'}(1)} = 1 + \frac{z_1^2}{z_1 - z_2}, \tag{21}$$

where $z_1 = z$ is the average number if neighbors of a vertex and $z_2$ is the average number of second neighbors.

This equation diverges when $G_1^{'}(1) = 1$. This is the point at which a giant component first appears. Since,

$$G_1^{'}(x) = \frac{\sum_k k(k-1)p_k x^{k-2}}{\sum_k kp_k},$$

we can also write the condition for phase transition as

$$\sum_k k(k-2)p_k = 0.$$

This sum is a monotonically increasing with the addition of edges to the graph. Therefore, it can be said that the giant component exists if and only if the this sum is positive. From the equation of $\langle s \rangle$ we also obtain the condition $z_2 > z_1$ for the existence of a giant component.

Till now we have dealt with the case where the giant component hasn't emerged. But the generating function formalism still works when the giant component is present, but according to the definition of $H_0(x)$ which excludes the giant component, $H_0(1)$ will not be unity. $H_0(1)$ in this case would be equal to $1 - S$, where $S$ is the fraction of the graph occupied by the giant component.

We can write

$$S = 1 - G_0(u)$$

, since $H_0(x) = G_0(H_1(x))$, where $u \equiv H_1(1)$ is the smallest non-negative real solution of

$$u = G_1(u).$$

The general expression for the average component size excluding the giant component is given as follows,

$$\langle s \rangle = \frac{H_0^{'}(1)}{H_0(1)}$$

$$= \frac{1}{H_0(1)}[G_0(H_1(1)) = \frac{G_0^{'}(H_1(1))G_1(H_1(1))}{1 - G_1^{'}(H_1(1))}]$$

$$= 1 + \frac{zu^2}{[1-S][1-G_1'(u)]}. \tag{22}$$

This is the same as the equation derived for $< s >$ in absence of giant components in which case $S = 0$, and $u = 1$.

### 2.8 Number of Neighbors and Average Path Length

The number of neighbors which are at a distance of m steps from a randomly chosen vertex can be expressed as follows:

$$G^{(m)}(x) = G_0(G_1(...G_1(x)...))$$

Therefore $G^{(m)}(x)$ can be defined as the generating function for the $m^{th}$ neighbor.

$$G^{(m)}(x) = \begin{cases} G_0(x), & \text{for m = 1;} \\ G^{m-1}(G_1(x)), & \text{for } m \geq 2. \end{cases}$$

The average number of $m^{th}$ neighbors, $z_m$ is

$$z_m = \left.\frac{dG^m}{dx}\right|_{x=1} = G_1'(1)G^{(m-1)'}(1) = G_1'(1)z_{m-1}$$

Note that here $G^{(m-1)'}(x)$ represents the differentiation of $G^{m-1}(x)$ and not the $mth$ derivative of $G(x)$

Along with the initial condition $z_1 = z = G_0'(1)$. Therefore,

$$z_m = [G_1'(1)]^{m-1}G_0'(1) = [\frac{z_2}{z_1}]^{m-1}z_1. \tag{23}$$

Now we can make an estimate the length of the shortest path between two randomly chosen vertices on the graph, $l$. This length is reached approximately when the total number of neighbors of a vertex out to that distance is equal to the number of vertices on the graph, i.e., when

$$1 + \sum_{m=1}^{l} z_m = N.$$

Therefore,

$$z_1 + z_1.\frac{z_2}{z_1} + z_1.\left(\frac{z_2}{z_1}\right)^2 + ... + z_1.\left(\frac{z_2}{z_1}\right)^{l-1} = N - 1$$

$$z_1.\frac{1-(\frac{z_2}{z_1})^l}{1-\frac{z_2}{z_1}} = N - 1$$

On rearranging,

$$\left(\frac{z_2}{z_1}\right)^l = 1 - \frac{N-1}{z_1^2}.(z_1 - z_2)$$

On taking log both sides and rearranging, we have

$$l = \frac{\log[(N-1)(z_2 - z_1) + z_1^2] - \log z_1^2}{\log \frac{z_2}{z_1}} \tag{24}$$

In the common case where $N \gg z_1$ and $z_2 \gg z_1$, this results

$$l = \frac{log(N/z_1)}{log(z_2/z_1)} + 1 \tag{25}$$

This method assumes that all vertices are reachable from a randomly chosen starting vertex, which is not true in general unless there is a giant component which fills the entire graph. And also the conditions used to derive are only an approximation, hence this result is only an approximation. Even with such shortcomings it has a number of remarkable features: i)average vertex-vertex distance for all random graphs scale logarithmically with the size of N, regardless of the degree distribution, ii) the average distance, which is a global property, can be calculated using only the knowledge of the average number of first- and second-nearest neighbors, which are local properties, and iii) two random graphs with completely different distribution of vertex degrees, but the same values of $z_1$ and $z_2$, will have the same average distances.

### 2.9 Bipartite Networks

This section illustrates the use of generating functions in the case of bipartite networks. The example presented is a movie actor network.

Here $\mathrm{f}_0(\mathrm{x})$ is the generating function for the degree (number of movies acted in) distribution of the actors. $\mathrm{g}_0(\mathrm{x})$ is the generating function for the degree (number of actors in the movie) distribution of the movies.

Let us assume that

$$Average/Actor = \mu \quad movies \tag{26}$$

and

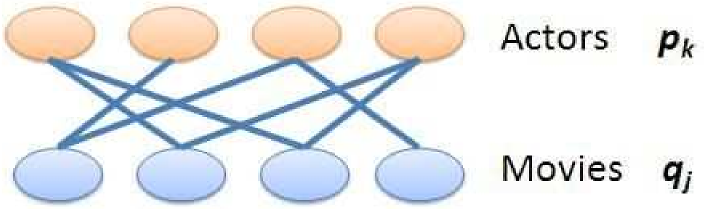

**Fig. 10.** An Actor Movie bipartite network.

$$Average/Movie = \nu \quad actors \tag{27}$$

We have

$$f_0(x) = \sum_{j=0}^{\infty} p_j x^j \quad , \quad f_1(x) = \frac{1}{\mu} f_0'(x) \tag{28}$$

and

$$g_0(x) = \sum_{k=0}^{\infty} q_k x^k \quad , \quad g_1(x) = \frac{1}{\nu} g_0'(x) \tag{29}$$

Consider G$_0$(x) to be the degree distribution of the co-actor network. It can be obtained from the original bipartite network in the following manner - pick a random actor node with a probability p$_k$, then go along an edge to a movie node, the outgoing degree distribution of the movie node gives the distribution of co-actors for the starting actor node. The generating function would thus be -

$$G_0(x) = p_0[g_1(x)]^0 + p_1[g_1(x)]^1 + p_2[g_1(x)]^2 + ..... + p_k[g_1(x)]^k + .........to\infty$$

$$\Rightarrow G_0(x) = f_0(g_1(x)) \tag{30}$$

Similarly we can also calculate the outgoing degree distribution G$_1$(x) of a node reached form a random edge of the co-actor network in the following manner - pick a random edge in the original network to reach an actor node with a probability p$_k$, then go along an edge to a movie node, the outgoing degree distribution of the movie node gives the required distribution. The generating function would thus be -

$$G_1(x) = 0.p_0[g_1(x)]^0 + 1.p_1[g_1(x)]^1 + ...... + k.p_k[g_1(x)]^k + ....to\infty$$

$$\Rightarrow G_1(x) = f_1(g_1(x)) \tag{31}$$

The average number of neighbors in the co-actor network will be given by

$$z_1 = G_0'(1) = g_1'(1).f_0'(g_1(1))$$

$$= g_1'(1).f_0'(1) \tag{32}$$

The average number of second neighbors will be given by

$$z_2 = G_0'(1).G_1'(1) = g_1'(1).f_0'(1).f_1'(g_1(1)).g_1'(1)$$
$$= f_0'(1).f_1'(1).[g_1'(1)]^2 \tag{33}$$

We can also study the existence of a giant component in the co-actor network. We know the expression for the average component size of a graph to be

$$\langle s \rangle = 1 + \frac{G_0'(1)}{1 - G_1'(1)}$$

from Eqn. 21 So, for the existence of a giant component ⟨s⟩ → ∞

$$\Rightarrow G_1'(1) = 1 \quad \Rightarrow \quad f_1'(1).g_1'(1) = 1 \tag{34}$$

$$\Rightarrow \frac{1}{\mu} f_0''(1).\frac{1}{\nu} g_0''(1) = 1 \quad \Rightarrow \quad f_0''(1).g_0''(1) = f_0'(1).g_0'(1)$$

$$We \quad have, f_0(x) = \sum p_j x^j \quad and \quad g_0(x) = q_k x^k$$

$$\Rightarrow f_0'(1) = \sum j.p_j \quad and \quad f_0''(1) = \sum j(j-1)p_j$$

So, our condition for the existence of giant component reduces to

$$\sum_{jk} [j(j-1)p_j.k(k-1)q_k - jp_j.kq_k] = 0$$

$$\Rightarrow \sum_{jk} p_j q_k [jk[jk - j - k]] = 0 \tag{35}$$

The equation obtained is symmetric in j and k. This implies that if we consider a co-movie network, it will have the same percolating condition as the co-actor network.

### 2.10 Directed Graphs

We will now see the use of generating functions in the case of directed graphs. An example of directed graphs can be the world-wide web, since a hyperlink between two pages on the web goes in only one direction.

Directed graphs introduce a subtlety that is not present in undirected ones, in a directed graph it is not possible to talk about a "component" - i.e. a group of connected vertices - because even if vertex $A$ can be reached by vertex $B$ following a directed edge, it does not guarantee that vertex $B$ can be reached by vertex $A$. This leads to two generalizations to the idea of component of a directed graph : the set of vertices that are reachable from a given vertex, and the set form which a given vertex can be reached referred as "out-components" and ”in-components” respectively. The generating functions for a directed graph are discussed below -

In a directed graph, each vertex has separate in-degree and out-degree for edges running into and out of that vertex. Let us define $p_{j,k}$ to be the probability that a randomly chosen vertex has in-degree $j$ and out-degree $k$. The distribution $p_{j,k}$ is a joint probability distribution on $j$ and $k$. So, the generating function for the joint probability distribution of in-degrees and out-degrees, which is necessarily a function of two independent variables $x$ and $y$,

$$G(x,y) = \sum_{jk} p_{jk} x^j y^k \tag{36}$$

The average in-degree is given by

$$\left.\frac{\partial G}{\partial x}\right|_{x,y=1} = \sum_{jk} j.p_{jk}$$

. The average out-degree is given by

$$\left.\frac{\partial G}{\partial y}\right|_{x,y=1} = \sum_{jk} k.p_{jk}$$

Now, since every edge on a directed graph must leave some vertex and enter another, the net average number of vertices entering a vertex is zero

$$Average\, number\, of\, edges\, entering \;=\; Average\, number\, of\, edges\, leaving \;=\; z \tag{37}$$

$$\Rightarrow \sum_{jk} j.p_{jk} = \sum_{jk} k.p_{jk}$$

$$\Rightarrow \sum_{jk} (j-k)p_{jk} = 0 \tag{38}$$

Let $z$ be the average degree(both in-degree and out-degree) of the vertices in the graph. Using the function $G(x,y)$ we can define the functions $G_0$ and $G_1$ for the number of outgoing edges leaving a randomly chosen vertex, and the number of edges leaving the vertex reached by following a randomly chosen edge.

Similarly we can define $F_0$ and $F_1$ for the number of edges arriving at such an edge. These functions are given by

$$F_0(x) = G(x,1), \quad F_1(x) = \frac{1}{z}\left.\frac{\partial G}{\partial y}\right|_{y=1} \tag{39}$$

$$G_0(y) = G(1,y), \quad G_1(y) = \frac{1}{z}\left.\frac{\partial G}{\partial x}\right|_{x=1} \tag{40}$$

The equations are similar to Eqn. 9 and 12 derived for a undirected graph. These equations can be used to calculate further results, for e.g. the average number of second neighbors reachable from a node(out-degree) can be calculated similarly as in Eqn. 13 as

$$z_2 = G'_0(1)G'_1(1) = z.\frac{1}{z}\left(\frac{\partial G}{\partial y}\left(\left.\frac{\partial G}{\partial x}\right|_{x=1}\right)\right)\bigg|_{y=1} = \left.\frac{\partial^2 G}{\partial y\, \partial x}\right|_{x,y=1} \qquad (41)$$

The average number of first neighbors $z$ reachable from a node is given by Eqn. 37.

Now, since the equations for $z$ and $z_2$ are symmetric in x and y, these are also the average number of first and second neighbors from which a random vertex can be reached (in-degree).

The probability distribution for the number of vertices that can be reached from a randomly chosen vertex in a directed graph i.e. the size of the out-components is generated by the function $H_0(y) = yG_0(H_1(y))$, where $H_1(y)$ is the size of the out-components reachable from a vertex and is given by $H_1(y) = yG_1(H_1(y))$. These equations can be derived just as those for undirected graphs like Eqn. 20, 19. The size of the in-components can be calculated similarly as $J_0(y) = yF_0(J_1(y))$ and $J_1(y) = yF_1(J_1(y))$

## 3 Conclusion

In this chapter we have studied various intriguing topics related to random graphs. We've looked at various models which are used to generate random graphs, the phases in the formation of random graphs. We've seen some important properties related to the existence of giant components in random graphs especially the ER graph.

We have also seen how generating functions can be used to model random graphs and for calculation of interesting properties of the graph mathematically like average number of neighbors, average number of second neighbors, different moments. They can also be used to determine component sizes of a graph and various properties of bipartite graphs and their one-mode projections.

Hence we have seen that there are a number of interesting features which are related to the class of random graphs and generating functions offer a very powerful mathematical tool to investigate these features.

## References

1. Erdos, P.; Rnyi, A. "On Random Graphs. I.". Publicationes Mathematicae 6: 290297 (1959).
2. Gilbert, E.N. "Random Graphs". Annals of Mathematical Statistics 30: 11411144 (1959).

3. Erdos, P.; Rnyi, A. "The Evolution of Random Graphs". Magyar Tud. Akad. Mat. Kutat Int. Kzl. 5: 1761 (1960)
4. Molloy, M. ;Reed B. "The Size of the Giant Component of a Random Graph with a Given Degree Sequence" in Combinatorics, Probability and Computing Vol.7: 295 - 305 (1998).
5. Newman, M.E.J. "The Structure and Function of Complex Networks" in Siam Review Vol. 45, 167-256(2003).
6. Newman, M.E.J ;Strogatz, S.H. ;Watts, D.J. "Random Graphs with Arbitrary Degree Distributions and their Applications" in Physical Review E, Vol. 64, 026118 (2001).