Percolation – Theory and Applications (86-889)

Generating Function Method

Shai Carmi

November 2005

Introduction

- Generating function is a powerful tool used to obtain exact solution for complicated combinatorial problems.
- Assume that we have a discrete probability distribution $P_S(s)$.
  The probability that $S = s$ is given by $P_S(s)$, denote it from now on as p_s .
- Define the *generating function* of p_s as follows

$$g(x) \equiv \sum_{s=0}^{\infty} p_s x^s$$

- Note that the generating function is a function of x.

Properties of generating functions

- If p_s is normalized, then $g(x=1)=1$.
- and $g(x) \leq 1$ for every $|x| \leq 1$.
- All p_s's can be obtained from $g(x)$ using derivatives

$$p_s = \left. \frac{1}{s!} \frac{d^s g(x)}{dx^s} \right|_{x=0}$$

- $g(x)$ "generates" the probability distribution p_s.

Properties of generating functions – Moments

- $< S >$ can be generated from $g(x)$ –

$$< S >= \sum_{s=0}^{\infty} s p_s = \left. \frac{dg(x)}{dx} \right|_{x=1}$$

- in the same manner, all higher moments can be generated –

$$< S^n >= \sum_{s=0}^{\infty} s^n p_s = \left[\left(x \frac{d}{dx} \right)^n g(x) \right]_{x=1}$$

Properties of generating functions – Powers

- Let now $S = \sum_{i=1}^{m} S_i$, where the S_i's are iid rvs generated by $g(x)$, then the probability distribution of S is generated by $[g(x)]^m$.
- Example - consider $S = S_1 + S_2$. The probability of the sum being equals s is given by the coefficients of $[g(x)]^2$, as can be seen –

$$
\begin{aligned}
[g(x)^2] \quad &= \quad p_0p_0x^0 + (p_0p_1 + p_1p_0)x^1 \\
&\quad\; +(p_0p_2 + p_1p_1 + p_2p_0)x^2 \\
&\quad\; +(p_0p_3 + p_1p_2 + p_2p_1 + p_3p_0)x^3 + \ldots
\end{aligned}
$$

Calculations on Cayley tree

- Terminology – sites, bonds and clusters.
- Without loss of generality, assume $z = 3$.
- Assume each bond is disconnected with probability p.
- Use the generating function approach to calculate the size of the giant cluster P_∞ and the mean (finite) cluster size S.

Calculations on Cayley tree

- For each site, we define F_s, the probability that this site is connected to a cluster of s bonds via a given bond.
- Because of symmetry, we can assume that this site is a leaf in the cluster.
- F_s is similar to sn_s discussed in class.
- F_s is given by –

$$F_s = \begin{cases} p & \text{If } s = 0 \\ (1-p)\sum_{s_1+s_2=(s-1)} F_{s_1}F_{s_2} & \text{Otherwise} \end{cases}$$

- For a cluster of length 0, the bond should simply be disconnected (with probability p); if it is connected (with probability $(1-p)$), the rest of the $(s-1)$ bonds can originate from either of the $(z-1) = 2$ sites connected to the site reached by the original bond.

Calculations on Cayley tree

- The generating function $g(x)$ of F_s is as explained before –

$$g(x) \equiv \sum_{s=0}^{\infty} F_s x^s$$

- Note that –

$$x[g(x)^2] = F_0F_0x^1 + (F_0F_1 + F_1F_0)x^2 + (F_0F_2 + F_1F_1 + F_2F_0)x^3 + ...$$

$$= \sum_{s_1+s_2=(1-1)} F_{s_1}F_{s_2}x^1 + \sum_{s_1+s_2=(2-1)} F_{s_1}F_{s_2}x^2 +$$

$$\sum_{s_1+s_2=(3-1)} F_{s_1}F_{s_2}x^3 + ... = \sum_{s=1}^{\infty} \sum_{s_1+s_2=(s-1)} F_{s_1}F_{s_2}x^s = \frac{\sum_{s=1}^{\infty} F_s x^s}{1-p}$$

Calculations on Cayley tree

- Using the definition of $g(x)$, we find that –

$$g(x) = F_0 + (1-p)\frac{\sum_{s=1}^{\infty} F_s x^s}{(1-p)} = p + (1-p)x[g(x)^2]$$

- This is a simple quadratic equation, the solution is –

$$g(x) = \frac{1 - \sqrt{1 - 4p(1-p)x}}{2(1-p)x}$$

- We have taken the smaller solution since for $x \leq 1$, $g(x)$ must be smaller than 1 too.

Calculations on Cayley tree – finding P_∞

- Since $g(1) = \sum_{s=0}^{\infty} F_s$ is the probability of a site to belong to some *finite* cluster, then P_∞, which is the probability of a site to belong to the *infinite* cluster is given by – $P_\infty = 1 - g(1)$.
- At $x = 1$ the positive value of the root is equal to $|2p - 1|$.
- If $p > 1/2$ then $|2p - 1| = (2p - 1)$, $\Rightarrow g(1) = 1$, and $P_\infty = 0$.
- If $p < 1/2$ then $|2p - 1| = (1 - 2p)$, $\Rightarrow g(1) < 1$, and $P_\infty > 0$.
- Hence, $p_c = 1/2$.
- Above the threshold, $P_\infty = 1 - g(1) = \frac{1-2p}{1-p} = \frac{2(1/2-p)}{1-p} = \frac{2(p_c-p)}{1-p} \sim (p_c - p)^\beta$, with $\beta = 1$.

Calculations on Cayley tree – finding S

- The mean cluster size $S = < s > = \sum_{s=0}^{\infty} sF_s = \left.\frac{dg(x)}{dx}\right|_{x=1}$.
- The derivatives can be performed easily.
- For $p < p_c$ (or $p < 1/2$)–

$$S = \frac{p^2}{2(1/2 - p)(1-p)} = \frac{p^2}{2(p_c - p)(1-p)} \sim |p - p_c|^{-1}$$

- For $p > p_c$ (or $p > 1/2$)–

$$S = \frac{1-p}{2(p - 1/2)} = \frac{1-p}{2(p - p_c)(1-p)} \sim |p - p_c|^{-1}$$

- Hence in both cases $S \sim |p - p_c|^{-\gamma}$ with $\gamma = 1$.

Calculations on networks

- Terminology – nodes, edges (or links) and components.
- For networks, we will define the generating function $g_0(x)$ through the degree distribution p_k which is the probability of a node to have a degree k.

$$g_0(x) \equiv \sum_{k=0}^{\infty} p_k x^k$$

- Using some combinatorics, we will be able to write few more generating functions, and calculate the mean (finite) component size S, and the giant component size P_∞.

Demonstration on Erdős-Rényi networks

- Let us demonstrate the calculation of $g(x)$ for Erdős-Rényi random networks.
- The number of nodes is N, the average degree $< k > \equiv z$ and the probability for a link to exist is $p = z/N$.
- The degree distribution is binomial. We will use generating function to show that it becomes Poisson in the limit of large N and small p –

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

Demonstration on Erdős-Rényi networks

- Substituting p_k into the definition of $g_0(x)$, we get –

$$
\begin{aligned}
g_0(x) &= \sum_{k=0}^{\infty} \binom{N}{k} p^k (1-p)^{N-k} x^k \\
&= \sum_{k=0}^{N} \binom{N}{k} (px)^k (1-p)^{N-k} \\
&= [px + (1-p)]^N = [(x-1)p + 1]^N = [\frac{z(x-1)}{N} + 1]^N
\end{aligned}
$$

- In the limit of $N \to \infty$, $g_0(x) = e^{z(x-1)}$.

Demonstration on Erdős-Rényi networks

- Can derive now – $g_0'(x) = ze^{z(x-1)}$. Hence $g_0'(1) = z$, as expected (z is the average degree).
- Can generate the probability distribution –

$$\begin{aligned} p_k &= \frac{1}{k!}\frac{d^k g_0(x)}{dx^k}\bigg|_{x=0} \\ &= \frac{1}{k!}\frac{d^k}{dx^k}e^{z(x-1)}\bigg|_{x=0} \\ &= \frac{1}{k!}z^k e^{z(x-1)}\bigg|_{x=0} = \frac{z^k e^{-z}}{k!} \end{aligned}$$

- We got a Poisson distribution, as expected.

Define more generating functions

- We are interested in the degree distribution of nodes we arrive at by following a randomly chosen edge.
- Such an edge arrives at a node with probability proportional to the degree of that node, thus the node itself has a probability distribution of degree proportional to kp_k.
- The correctly normalized distributed is generated by –

$$\frac{\sum_k kp_k x^k}{\sum_k kp_k} = x\frac{g_0'(x)}{g_0'(1)}$$

Define more generating functions

- Assume now that we start with some random node, and pick one of its links. We follow the link and arrive to some other node. We look now for the distribution of the number of remaining links that the second node have.
- This distribution is generated by the function from the last slide, except for less one power of x, since we are generating the distribution of the number of the *remaining* links.
- Denote this generating function as $g_1(x)$ –

$$g_1(x) = \frac{g_0'(x)}{g_0'(1)} = \frac{1}{z} g_0'(x)$$

Define more generating functions

- The probability that the edges outgoing from the second node connects to the original node (or any other first or second neighbors of it) goes like N^{-1} and hence can be neglected in the limit of large N. In other words, we neglect any loops.
- We look now for the distribution of the number of *second neighbors* (= nodes 2 hops away) of the original node.
- To obtain this distribution, let's start by assuming the original node has degree k.
- The probability of the original node to have n second neighbors, is the probability that the sum of the remaining links of each of its k neighbors sums into n.
- Since the distribution of the number of remaining links is generated by $g_1(x)$, and using the 'power' property of generating functions discussed above, the probability distribution of second neighbors is generated by $[g_1(x)]^k$.

Define more generating functions

- What is left now is only to multiply this generating function by p_k and sum over all possible k's.

$$\sum_{k=0}^{\infty} p_k [g_1(x)]^k = g_0(g_1(x))$$

  where the equality follows from the definition of $g_0(x)$.

- The average number z_2 of second neighbors is –

$$\begin{aligned} z_2 &= \left[\frac{d}{dx}[g_0(g_1(x))]\right]_{x=1} = [g_0'(g_1(x))g_1'(x)]_{x=1} \\ &= g_0'(g_1(1))g_1'(1) = g_0'(1)g_1'(1) = g_0''(1) \end{aligned}$$

- Similarly, the distribution of third-nearest neighbors is generated by $g_0(g_1(g_1(x)))$ and so on.

Demonstrate on Cayley tree

- For Cayley tree, $p_k = \delta_{k,z}$. Hence, $g_0(x) = x^z$.
- $g_1(x) = \frac{1}{z} g_0'(x) = x^{z-1}$. Hence, the average number of remaining links is $g_1'(1) = \left[(z-1)x^{z-2}\right|_{x=1} = z-1$ as expected.
- The average number of the second neighbors, is given by $z_2 = g_0''(1) = \left[z(z-1)x^{z-2}\right]_{x=1} = z(z-1)$ as we know from the course.

Calculating component sizes

- Define $h_0(x)$ – The generating function of the distribution of the size of the component reached when choosing a random node.
- Define $h_1(x)$ – The generating function of the distribution of the size of the component reached when choosing a random edge.
- The giant component is not taken into account when calculating h_0 and h_1. Also, we ignore loops.
- We will find a self-consistency condition for h_1.

Calculating component sizes

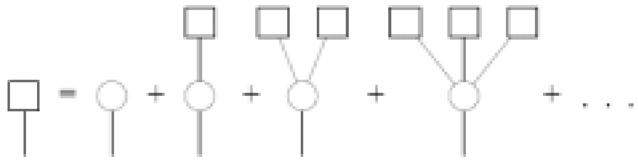

Figure 1: Schematic representation of the sum rule for the connected components

- We look at the component that emerges from a given edge. On the left hand side, is the whole component. On the right hand side, is the sum over all possibilities of the degree of the node reached through the original edge.
- We can look at the components reached when travelling the edges emanating from the initial node, and write a self-consistency equation.

Calculating component sizes

- If we denote by q_k the probability that the initial node has k edges going out of it (other than the edge we came in along), then using again the 'power' property, we can write an equation for h_1 –

$$\begin{aligned} h_1(x) &= xq_0 + xq_1 h_1(x) + xq_2[h_1(x)]^2 + ... \\ &= x \sum_{k=0}^{\infty} q_k [h_1(x)]^k \end{aligned}$$

The extra x is because we must add the initial node to the component size.

- Recalling that q_k is just the coefficient of x^k in $g_1(x)$, we can write –

$$h_1(x) = xg_1(h_1(x))$$

Calculating component sizes

- If we start from a random node (as opposed to random edge before), only thing we need to change is to count all the edges coming out of the initial node, i.e. replace q_k by p_k or $g_1(x)$ by $g_0(x)$ –

$$h_0(x) = xg_0(h_1(x))$$

- We have arrived at a transcendental equations for $h_1(x)$ and $h_0(x)$.
- The probability of a randomly chosen node to belong to a component of size s is given by –

$$P_s = \frac{1}{s!} \frac{d^s h_0(x)}{dx^s}\bigg|_{x=0}$$

- Usually equations are difficult to solve analytically.

Trick for calculating probabilities

- Solve the self-consistency equation by numerical iterations.
- To find P_s, calculate the derivatives numerically.
- This is simple, but numerical derivatives are prone to machine-precision errors.
- Use Cauchy formula –

$$P_s = \frac{1}{s!} \frac{d^s h_0(x)}{dx^s}\bigg|_{x=0} = \frac{1}{2\pi i} \oint \frac{h_0(z)}{z^{s+1}} dz$$

- Calculate the integral numerically with the contour being for unit circle $|z| = 1$.

Back to component sizes

- Assume all components are finite.
- The average size of the component to which a randomly chosen node belongs is given by –

$$
\begin{aligned}
S = \; < s > \; &= h'_0(1) = [g_0(h_1(x)) + x g'_0(h_1(x)) h'_1(x)]_{x=1} \\
&= 1 + g'_0(1) h'_1(1)
\end{aligned}
$$

  Where the equalities follow from the normalization property of the generating functions.

- In the same manner, taking derivatives of the equation for $h_1(x)$ yields –

$$h'_1(1) = 1 + g'_1(1) h'_1(1)$$

Component sizes

- After some algebra, we find that –

$$S = \ < s > \ = 1 + \frac{g_0'(1)}{1 - g_1'(1)} = 1 + \frac{z_1^2}{z_1 - z_2}$$

  Where $z_1 \equiv z$, the average degree or the average number of first neighbors.

- The average component size diverges when $g_1'(1) = 1$.

- More intuitively, giant component appears first when $(z_2 > z_1)$, i.e. a node has more second neighbors than first neighbors.

The giant component

- If there is a giant component, then $h_0(x)$ is no longer normalized.
- Denote by P_∞ the fraction of network occupied by the giant component. Then $h_0(1) = 1 - P_\infty$.
- To find P_∞, first we substitute $x = 1$ in the equation - $h_1(x) = xg_1(h_1(x))$, and define $u \equiv h_1(1)$ –

$$u = g_1(u)$$

- After solving for u, recall that $h_0(x) = xg_0(h_1(x))$, thus $h_0(1) = g_0(h_1(1)) = g_0(u)$.
- P_∞ is given by $P_\infty = 1 - g_0(u)$.

Application to a one-dimensional chain

- One dimensional chain can be considered as a Cayley tree with $z = 2$ ($p_k = \delta_{k,2}$).
- Each edge is removed with probability $q \equiv (1 - p)$.
- Degree distribution is now $p_0 = (1-p)^2, p_1 = 2p(1-p), p_2 = p^2; p_{k>2} = 0$ (Using binomial distribution with $N = 2$).
- $g_0(x) = (1-p)^2 + 2p(1-p)x + p^2x^2$.
- $g_0'(x) = 2p(1-p) + 2p^2x$. Hence $z = g_0'(1) = 2p(1-p) + 2p^2 = 2p$.
- $g_1(x) = \frac{g_0'(x)}{z} = \frac{2p(1-p)+2p^2x}{2p} = (1-p) + px$. Hence, $g_1'(x) = p = g_1'(1)$.
- $z_2 = g_0'(1)g_1'(1) = 2p^2$.

Application to a one-dimensional chain

- Combining all the above information, we can extract some useful quantities.
- The percolation threshold is given by the condition $g'_{1_{p_c}}(1) = 1 \Rightarrow p_c = 1$.
- The average size of the finite components is given by

$$S = 1 + \frac{g'_0(1)}{1 - g'_1(1)} = 1 + \frac{2p}{1-p} = \frac{1+p}{1-p} = \frac{1+p}{p_c - p} \sim (p_c - p)^{-1}$$

Application to a one-dimensional chain

- To find the size of the giant size, let us first solve the equation $u = g_1(u) \;\Rightarrow\; u = (1-p) + pu \;\Rightarrow\; u = 1.$
- $P_\infty = 1 - g_0(u) = 1 - g_0(1) = 1 - 1 = 0.$
- Explanation - since p is always smaller than p_c, there is never a giant component.
- All the above results, in agreement with previous calculations shown in the course.

The End!