

Lecture 27: Graph Models

*Lecturer: John Hopcroft**Scribe: Chris Provan, Shaomei Wu*

1 Graph Models with Power Law Degree Distributions

At the end of last lecture we introduced a very practical problem: how to generate a random graph with power law degree distribution. It turns out to be a nontrivial problem, so we will briefly look at several possible models here.

In a random graph with a power law degree distribution, the expected number of degree k vertices is $c \frac{n}{k^\alpha}$, where α is a parameter of the distribution (typically $\alpha \in [2, 3]$), and c is a constant chosen so that these expectations over all k add up to the total number of vertices, i.e.

$$\sum_{k=1}^{\infty} c \frac{n}{k^\alpha} = n \iff c = \left(\sum_{k=1}^{\infty} \frac{1}{k^\alpha} \right)^{-1}.$$

For $i \in \{1, 2, \dots, n\}$, define h_i to be the largest integer h such that

$$\sum_{k=h_i}^{\infty} c \frac{n}{k^\alpha} \geq i \implies h_i \approx \left(\frac{c}{\alpha - 1} \right)^{\frac{1}{\alpha-1}} \left(\frac{n}{i} \right)^{\frac{1}{\alpha-1}}.$$

Consider the degree sequence h_1, h_2, \dots, h_n . There will be gaps early in the sequence ($h_1 \gg h_2 \gg \dots$) and repeated smaller values at the end of the sequence ($1 = h_n = h_{n-1} = \dots$). For $\alpha = 2$, we have $h_i \approx (c) \left(\frac{n}{i} \right)$, thus the total number of edges under this degree sequence is $\Theta(n \ln n)$ and the highest degree vertex has degree $h_1 \approx n$. For $\alpha > 2$, in the same way we can infer that the total number of edges is $\Theta(n)$. And when $\alpha = 3$, the highest degree vertex has degree $h_1 \approx \sqrt{n}$.

Consider the following four models for generating power law random graphs:

Model 1: “Coin-toss”

For every pair of vertices i and j , toss a coin and create an edge between i and j with probability

$$\frac{h_i h_j}{\sum_{k=1}^n h_k}.$$

The key advantage of this model is that the edges are statistically independent from one another. But there is a problem. Suppose $\alpha = \frac{5}{2}$. Then the highest-degree vertex in the sequence has degree $h_1 = n^{\frac{2}{3}}$ and the probability for the edge between the two highest-degree vertices is thus

$$\frac{h_1 h_2}{\sum_k h_k} \approx \frac{n^{\frac{2}{3}} n^{\frac{2}{3}}}{n} = n^{\frac{1}{3}} > 1,$$

which is not a valid probability. What went wrong? It turns out that there is no simple graph with the associated degree sequence.

Possible solution: Instead of flipping a coin, create a multigraph by generating zero or more edges between two vertices i and j according to a Poisson process with mean $\lambda = \frac{h_i h_j}{\sum_k h_k}$. Under this solution, we will get cliques among the high degree vertices, possibly with duplicated edges in the cliques.

Model 2: “ m edges”

Randomly generate $m = \frac{1}{2} \sum_k h_k$ edges by selecting vertices for endpoints with probability proportional to their degrees. There will be exactly m edges, but the edges will *not* be statistically independent from one another since, for example, the conditional probability of an edge being selected given that some other set of m edges have been selected is zero.

Model 3: “Permutation”

Create a sequence of vertices in which each vertex appears a number of times equal to its degree. Randomly permute entries in the vertex sequence and pair the adjacent vertices in the permuted sequence by adding an edge between them.

For example, suppose we have six vertices having degrees 5, 4, 2, 1, 1, and 1. Then the corresponding vertex sequence is

1 1 1 1 1 2 2 2 2 3 3 4 5 6.

After permutation, it might become

1 2 2 6 1 3 1 4 2 3 1 5 2 1.

Pair the adjacent vertices:

[1 2][2 6][1 3][1 4][2 3][1 5][2 1],

and add each of the 7 edges corresponding to one of the pairs of vertices.

The random graph generated by this model has exactly m edges and the exact degree sequence. However, as in Model 2, the edges here will not be statistically independent.

Model 4: “Full graph selection”

Consider the set of all graphs with the given degree sequence and select one uniformly at random. This model has nice mathematical properties (e.g., the required number of edges, the exact degree sequence), but is difficult to use in practice.

Suggested exercises for this part:

- (1) What is the probability of a clique among the ten (or any constant) highest degree vertices (especially for Models 1 and 2)?
- (2) What are the observable differences and similarities between the different models? In other words, given a graph generated by one of four models, can you identify which one? (e.g. some models have exactly m edges, so they can be ruled out if the graph has some other number.)
- (3) Are Models 2 and 3 different? What about models 3 and 4?

2 Eigenvalues of d -regular Graphs

We now return to eigenvalues and eigenvectors of adjacency matrices for some specific graphs.

Lemma 1. *Let G be an undirected, connected, regular degree d graph with adjacency matrix A . Then the eigenvalues of A satisfy*

$$d = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq -d,$$

with $\lambda_n = -d$ if and only if G is bipartite.

Proof. We first want to show that d is an eigenvalue and all the other eigenvalues are less than d . Let μ be the column vector of ones, i.e., $\mu = (1, 1, \dots, 1)^T$. Since every vertex in G is of degree d , $A\mu = d\mu$, so d is an eigenvalue of A . Let x be an eigenvector not proportional to μ , and define x_{max} to be the largest coordinate of x . Define the set $S = \{i : x_i = x_{max}\}$.

Since x is not proportional to μ , $|S| < n$, and since G is connected, there must be some $j \in S$ adjacent to a vertex $i \notin S$ with $x_i < x_{max}$. Since A is the adjacency matrix for a d -regular graph, row j of A has d 1's, including a 1 in the i -th column because vertex i and j are connected. So the j -th element of the vector Ax is

$$(Ax)_j = \sum_{k:A_{jk}=1} x_k = \sum_{k \neq i:A_{jk}=1} x_k + x_i \leq (d-1)x_{max} + x_i < dx_{max}.$$

Thus if λ is the eigenvalue corresponding to x , $Ax = \lambda x$ hence $(Ax)_j = \lambda x_j < dx_{max}$. But since $j \in S$, $x_j = x_{max}$, so $\lambda x_{max} = \lambda x_j < dx_{max}$, and $\lambda < d$.

The analogous proof for $\lambda \geq -d$ is left as a possible exercise.

For the final part of the lemma, note that G is bipartite if and only if its adjacency matrix can be written as a block matrix

$$A = \left(\begin{array}{c|c} 0 & A_2 \\ \hline A_1 & 0 \end{array} \right).$$

Then using a similar analysis as in the beginning of the proof with an eigenvector of -1 's and 1 's in the elements corresponding to A_1 and A_2 , respectively (or the reverse) gives the desired result. \square

Note that the proof requires that G be connected in order to show $d > \lambda_2$. If instead G is an undirected d -regular graph with k connected components, then its adjacency matrix can be written as a block matrix

$$A = \left(\begin{array}{c|c|c|c|c} A_1 & 0 & 0 & \dots & 0 \\ \hline 0 & A_2 & 0 & \dots & 0 \\ & \vdots & & \ddots & \vdots \\ \hline 0 & 0 & 0 & \dots & A_k \end{array} \right),$$

where A_1, A_2, \dots, A_k are degree d adjacency matrices. Thus each of the k vectors with 1 's in the elements corresponding to any subset of the A_i 's and 0 's everywhere else is an eigenvector corresponding to an eigenvalue $\lambda = d$. These eigenvectors form a k -dimensional space, and thus the modified result for an unconnected graph is that the eigenvalues of A satisfy

$$d = \lambda_1 = \lambda_2 = \dots = \lambda_k > \lambda_{k+1} \geq \lambda_{k+2} \geq \dots \geq \lambda_n \geq -d.$$

A more detailed proof is presented in the next lecture.