September 26, 2005
Paul K Horn

Random Walks on Graphs and Directed Graphs

Introduction

1 Introduction

In this course we will study random walks on graphs and directed graphs (or reversible Markov chains and non-reversible Markov chains). In order to accomplish this, we use at both probabilistic and spectral methods. While it may seem somewhat surprising, spectral methods provide a powerful tool to study random walks, as many related invariants such as mixing rate are controlled by a graph's eigenvalues.

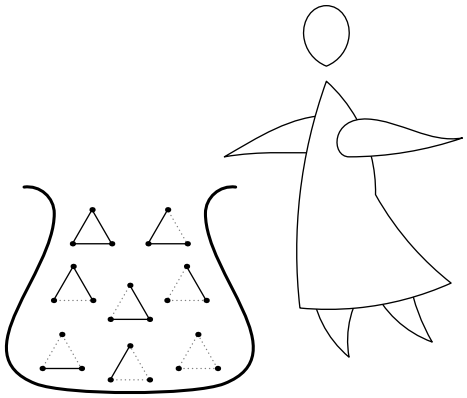
Much of the current literature about random walks concerns (undirected) graphs. One reason for this is that spectral methods having a symmetric matrix gives a great deal of control of the eigenvalues and eigenvectors; while the directed case lacks much of this control. Since graphs tend to be very large and complicated, spectral methods help immensely in allowing us to measure the shape and properties of the graph with just a few invariants, as opposed to trying to keep track of the entire structure.

It is desirable, however, to consider the more complicated directed case. Directed graphs occur often in the real world; the web-graph, for example, is directed. Recently there have been some developments in the directed graph case.

Essentially, our focus boils down to studying three things: random walks, random graphs, and spectral methods.

2 Random Graphs

What is meant by "random graph"? When we consider random graphs, we consider a probability distribution function on Ω, the span of all graphs on n vertices. A way to think about random graphs is that we put all graphs on n vertices into a pot, and we choose graphs out of the pot according to their probability distribution. The easiest probability distribution to think about is just the uniform distribution, that is we are equally likely to chose any of the graphs.



2.1 Erdős-Rényi Model

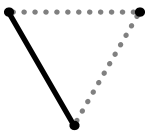
Erdös and Rényi pioneered the field of random graphs in 1960. They popularized the $G(n,p)$ model of random graphs on n vertices with a probability parameter p, with $0 \leq p \leq 1$. Here we give two definitions.

Definition 1 For a graph G, each pair (u,v) is *independently* chosen to be an edge with probability p.

Definition 2 $Pr(X = G) = p^{\#\text{of edges in } G}(1-p)^{\#\text{of non-edges}}$

In the second definition, X is a **random variable**. This is like the arm that reaches into our pot, and plucks out our graph. Also of importance in the independence in Definition 1; this means that selection of any given edge does not depend on any of the other pairs being selected or not selected. This assumption is vital to much of the analysis that is done, since the probability of two independent events is equal to the product of the probabilities of the individual events.

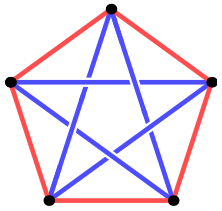
As an example, consider the eight (labeled) graphs on 3 vertices. If we take $p = \frac{1}{2}$, then all graphs have a probability of $\frac{1}{8}$ of being chosen, giving a uniform distribution. If we change p, however, we no longer have a uniform distribution. For example the graph pictured below would be chosen with probability $p(1-p)^2$.



2.2 Ramsey Numbers

As an example of the power of random graphs, we briefly consider Ramsey numbers. The Ramsey number $R(k,k)$ is the smallest n such that any two coloring of K_n contains a monochromatic K_k.

As an example, $R(3,3) = 6$. The following diagram demonstrates that $R(3,3) > 5$; we give a two coloring of K_5 with no monochromatic K_3. It is an easy excercise to show that if there are 6 vertices there must be a monochromatic K_3.



We have the following theorem [2]

Theorem 1. *If* $\binom{n}{k}2^{1-\binom{k}{2}} < 1$ *then* $R(k,k) > n$.

We would like to find the largest n satisfying this theorem. Stirling's approximation [4] tells us

$$\sqrt{2\pi n}(n/e)^n \leq n! \leq e^{1/12n}\sqrt{2\pi n}(n/e)^n$$

In particular we can use this to say that

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

and in order for

$$\left(\frac{en}{k}\right)^k 2^{1-\binom{k}{2}} < 1$$

we have

$$k(\log_2(n) + \log_2(e) + \log_2(k)) + 1 - \binom{k}{2} \quad < \quad 0.$$

This tells us

$$2\log_2 n < k$$

or $n \approx 2^{k/2}$. The best known bounds for the Ramsey number $R(k,k)$ are

$$\frac{\sqrt{2}}{e}(1+o(1))k2^{k/2} \leq R(k,k) \leq \binom{2k-2}{k-1} \approx 4^k$$

The lower bound is due to Spencer, and uses the Lovász Local Lemma [1]. This bound beats the original lower bound, $R(k,k) > \frac{1}{\sqrt{2}e}(1+o(1))k2^{k/2}$, by a factor of two. This original bound was given by Erdős and Rényi [2]. The proof of this is similar to the proof we give for Theorem 1, but requires more careful analysis. Information on upper bounds can be found in Graham, Rothschild, and Spencer [3]. Note that although the inequality $R(k,k) \leq \binom{2k-2}{k-1}$ is often proper (by a parity argument), getting even a $(1-\epsilon)$ improvement is open.

Proof of Theorem 1. We color the edges of a complete graph on n vertices randomly, so that edges are chosen to be blue independently with probability $1/2$. Note that the blue (and red) edges are random graphs $G(n, 1/2)$. Consider a subset S of k vertices. The probability that the induced K_k on S is blue is

$$P(K_k \text{ is blue}) = \left(\frac{1}{2}\right)^{\binom{k}{2}}$$

thus as there are two possible colors,

$$P(K_k \text{ is monochromatic}) = \frac{1}{2^{\binom{k}{2}-1}}.$$

Therefore

$$P(\text{There exists a monochromatic } K_k) \leq \binom{n}{k} 2^{1-\binom{k}{2}}$$

If this probability is less than 1, then there exists a graph on n vertices containing no monochromatic K_k; proving the theorem. □

3 Random Walks

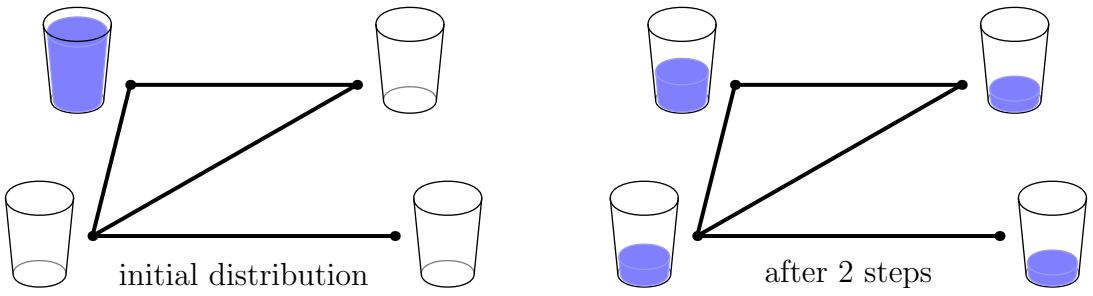
We first consider walks on undirected graphs. A **walk** is a sequence of vertices $v_0, v_1, \ldots, v_t$ where $v_i \sim v_{i+1}$. (that is $\{v_i, v_{i+1}\} \in E(G)$) for $i = 0, \ldots, t-1$.

One way to think about random walks is a pebble starts at vertex v_0, and moves from there. The probability it moves from vertex u to v (assuming it sits at u) is given by

$$P(u, v) = \begin{cases} \frac{1}{d_u} & \text{if } v \sim u \\ 0 & \text{else} \end{cases}$$

(where d_u is the degree of vertex u). This is a walk using a transition probability matrix, P.

A second way of looking at random walks is to look at the probability distribution of destinations after t steps. In this case, we think of a mass of size one starting at the initial vertex v_0, then spreading itself out over the vertices according to the transition rules; the mass at a particular vertex at time t represents the probability that a walk would end there at time t.



Formally we can define an initial distribution

$$f_0 = \begin{cases} 1 & \text{at } v_0 \\ 0 & \text{elsewhere} \end{cases}$$

We can actually use any initial distribution so long as $f(v) \geq 0$ for all $v \in V(G)$, and $\sum f(v) = 1$. Then we can recursively define the distribution at time t by

$$f_t(v) = \sum_{u \sim v} f_{t-1}(u) P(u, v)$$

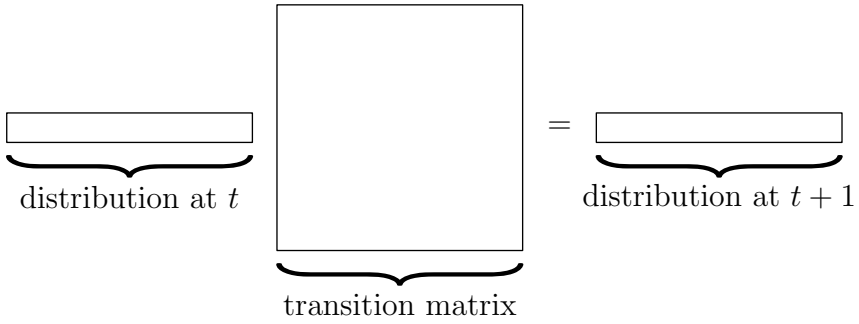
by representing f_t as a vector, and P as a transition matrix we get the following nice representation:

$$f_t = f_{t-1} P$$

and in general

$$f_t = f_0 P^t$$

It is useful to note that multiplying on the left by f (as opposed to the more traditional multiplication on the right) allows a more convenient multiplication. This will be our convention.



References

[1] N. Alon, J. Spencer, The Probabilistic Method, second edition, John Wiley and Sons, 2000

[2] P. Erdős, A. Rényi, "On the evolution of random graphs" n, Maygyar. Tud. Akad. Mat. Kutató Int. Közl. 5 (1960) 17-61.

[3] R.L. Graham, B.L. Rothschild, and J.H. Spencer, Ramsey Theory, second edition, Wiley, New York, 1990.

[4] B. Bollobás, Modern Graph Theory, Springer 1998.