CS 4850 Lecture 4
Spring 2009

Properties of Random Graphs, Continued

Notes by Christie Brandt (ccb35) and Scott Rogoff (scr32)

Degree distribution for G(n, 1/n)

Expected degree = 1 (sparse graph)

Largest degree vertex has degree log(n)/log(log(n)).
Proof:

$$\Pr(\deg = k) = \binom{n}{k}\left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^{n-k}$$

If n is large and k is small (in other words, k = o(n)), then

$$\binom{n}{k} = \left(\frac{n(n-1)(n-2)...(n-k)}{k!}\right)$$

We can then write the probability that deg=k as:

$$\Pr(\deg = k) = \left(\frac{n(n-1)...(n-k)}{k!}\right)\left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^n \left(1-\frac{1}{n}\right)^{-k}$$

$$\approx \left(\frac{n^k}{k!}\right)\left(\frac{1}{n^k}\right)\left(1-\frac{1}{n}\right)^n \left(1-\frac{1}{n}\right)^{-k}$$

$$\approx \left(\frac{1}{k!}\right)\left(1-\frac{1}{n}\right)^n$$ (for n >> k, the last term is approximately 1)

$$\approx \left(\frac{1}{k!}\right)e^{-1}$$ (approximation for e when n is large)

$$= \left(\frac{1}{e \cdot k!}\right)$$

Using the fact that Pr(deg=k) $\approx \frac{1}{e \cdot k!}$, we can calculated the expected value for $k = \frac{\log(n)}{\log(\log(n))}$.

We will begin by examining k^k:

$$\log(k^k) = k\log(k) = \frac{\log(n)}{\log(\log(n))}\left(\log\left(\frac{\log(n)}{\log(\log(n))}\right)\right) = \frac{\log(n)}{\log(\log(n))}\left(\log(\log(n)) - \log(\log(\log(n)))\right)$$

$$\approx \frac{\log(n)}{\log(\log(n))}\log(\log(n))$$ (because log(log(log(n))) is a very small number)

$$\approx \log(n).$$

$$k^k \approx n$$

Thus since $(k!) \le k^k$, we know that $(k!) \le k^k \approx n$.

Therefore, $\Pr\left(\deg = \frac{\log(n)}{\log(\log(n))}\right) = \Pr(\deg = k) \approx \frac{1}{e \cdot k!} \le \frac{1}{e \cdot k^k} \approx \frac{1}{e \cdot n}$

The probability that a chosen vertex has degree $\frac{\log(n)}{\log(\log(n))}$ is $\frac{1}{e \cdot n}$.

We can use this to calculate the probability, with n vertices, of having one with degree k:

$$\Pr\left(\exists \text{vertex of deg} = \frac{\log(n)}{\log(\log(n))}\right) = 1 - \left(1 - \frac{1}{e \cdot n}\right)^n = 1 - \left(1 - \frac{1}{e \cdot n}\right)^{en/e} = 1 - (e^{-1})^{1/e} = 1 - e^{-1/e} \approx 0.3$$

Probability that a Generated Graph has a Triangle

Let us examine the graph $G(n, \frac{p}{n})$. What is the expected number of triangles in this graph?

A triangle is created when three vertices are chosen that have edges between each pair. Therefore, there are $\binom{n}{3} = \frac{n^3}{6}$ sets of possible "triangle candidates" in the graph.

The probability of forming a triangle = $\Pr(\textit{given triple is a triangle}) = \left(\frac{d}{n}\right)^3$.

Therefore, the expected number of triangles in the graph = $\left(\frac{n^3}{6}\right)\left(\frac{d^3}{n^3}\right) = \left(\frac{d^3}{6}\right)$

This means that the number of triangles in the graph is independent of the number of nodes in the graph—increasing the nodes will not increase the number of triangles.

Diameters of Dense Graphs

The **diameter** of a graph = the maximum distance between any two vertices.

Let us consider a denser graph, G(n,p) where p = a constant.

The diameter of the dense graph = 2. If G has diameter 2, this means that every pair of vertices (u, v) is either directly connected by an edge or there exists a vertex w such that the edges (u, w) and (v, w) are in G.

We will call a pair of vertices (u,v) **"bad"** if u, v are not connected by an edge and for all vertices w in the graph, w is not connected to both u and v. In other words, the distance between a "bad" pair of vertices (u,v)>2.

$$\Pr((u, v)\ \textit{is a bad pair}) = (1 - p)(1 - p^2)^{n-2}$$

(The first part is the probability that there is no edge between u and v. The second part is the probability that no w exists that has an edge to both u and v.)

Let x = the number of bad pairs.

$$\lim_{n\to\infty}(E(x)) = \lim_{n\to\infty}\left(\binom{n}{2}(1 - p)(1 - p^2)^{n-2}\right)$$

(In other words, it is the limit of (the number of vertices)*(the probability a vertex is bad))
Since p is a constant <1,

$$\lim_{n\to\infty}(E(x)) = \lim_{n\to\infty}(n^2)\,(some\ fraction)^{n-2} = 0$$

Therefore, since

$$\lim_{n\to\infty}(E(x)) = 0$$

P(diameter ≤ 2) = 1.

Digression: Expected Values and Probabilities

Let x = the number of bad pairs. What if $\lim_{n\to\infty} E(x) = 0$?

This means that Pr(G has a bad pair) = 0,
Pr(G has no bad pairs) = 1.

Why does this follow? In other words, why does the expected value being zero assure us that Pr(G has no bad pairs) = 1?

Example 1 - Say that Pr(bad pair) = 1/10.
Then E(x) = (0.1)(1) + (0.9)(0) = 0.1; in other words, the expected value is not zero.

If any constant number of bad pairs exists, then the limit of the expected value will not be zero.

On the other hand, if $\lim_{n\to\infty}(E(x)) = \infty$, we can't conclude anything.

Example 2 - Say that we examine a set of graphs. Half of the graphs have an infinite number of bad pairs. The other half of the graphs has zero bad pairs. E(x) = ∞, but the probability of getting a graph with a bad pair is NOT 1.
This is because the probability is not concentrated. Without looking at the variance (the second moment), we can't conclude anything.

Phase transitions

Consider G(n,p) while p increases: what happens?

- p = 0: all vertices are isolated
- p is o($\frac{1}{n}$): G is a collection of trees, each of size ≤ log(n)
- p = $\frac{1}{n}$: a phase transition occurs: ∃ a cycle
- at some larger p: a phase transition occurs: a giant component forms
- at some larger p: a phase transition occurs: the graph becomes connected
- at some larger p: a phase transition occurs: the graph's diameter becomes 2

For any increasing property of a combinatorial structure, a phase transition exists.

Formal Definition: If ∃ $p(n)$ such that for $p_1(n)$ where $\lim_{n\to\infty}\left(\frac{p_1(n)}{p(n)}\right) = 0$, $G(n, p_1(n))$ does not have a property (almost surely, with probability 1) and for $p_2(n)$, where $\lim_{n\to\infty}\left(\frac{p_2(n)}{p(n)}\right) = \infty$, $G(n, p_2(n))$ almost surely *does* have that property, we say that $p(n)$ is the threshold for the property and the property has a phase transition at $p(n)$.

Example 1 – Consider a randomly generated Boolean expression, such as

$$f(x_1, x_2, \dots, x_n) = (x_1 + x_2 + x_3)(\overline{x_1} + x_4 + x_5)\cdots$$

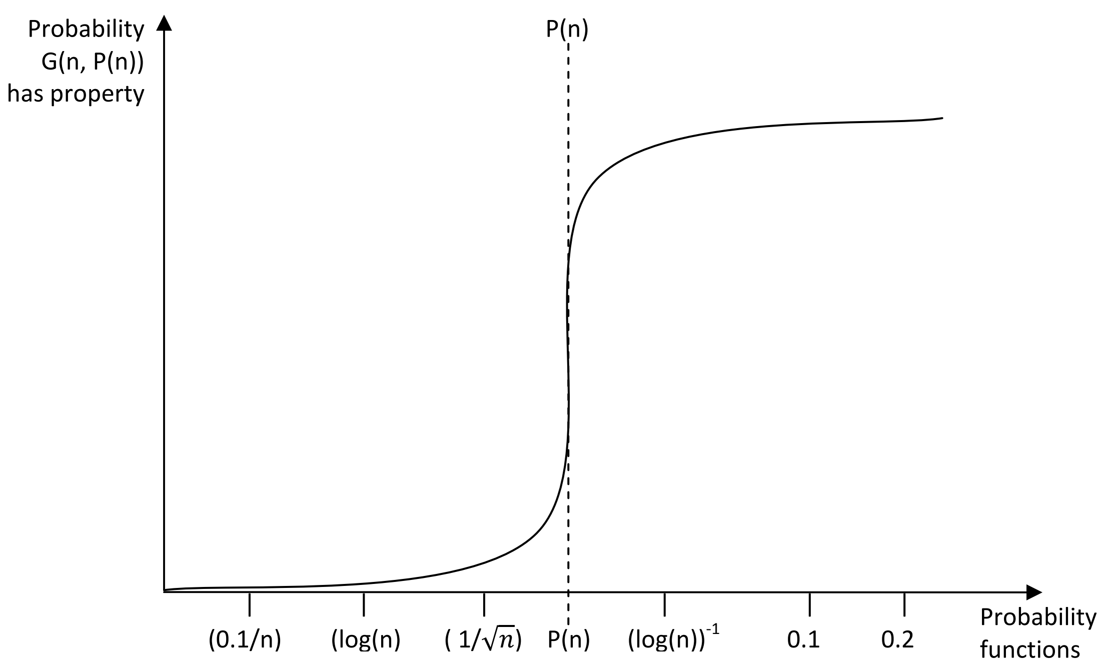
Question: Is this expression satisfiable?

Actually, we can determine this based on the number of clauses: If this number is less than a certain phase transition value, the Boolean expression will almost certainly be satisfiable. If the number of clauses exceeds this value, it will almost certainly not be satisfiable.

Example 2 – One can generate a Sudoku game by starting with a filled grid and erasing some of the numbers. How many can be removed before there ceases to be a unique solution? (A phase transition will be more apparent if we consider larger than just the standard 9x9 grid.)

We can represent a phase transition with a graph. Across the x axis, we consider a subset of functions with a linear ordering. The y axis represents the probability that G(n,p(n)) has the desired property. To the left of the function at which the phase transition occurs, the probability that the property occurs will be essentially 0.

Phase Transition Diagram



Sharp Phase Transitions

A property with a sharp phase transition is even easier to think about because you don't even need to worry about asymptotic behavior.

A property has a **sharp phase transition** at P(n) if, when c<1, a graph G(n, c P(n)) does not have the property, but when c>1, a graph G(n, c P(n)) does have the property.

Disappearance of Isolated Vertices (aka graph becomes connected)

Claim: there exists a sharp threshold at $\frac{\ln(n)}{n}$.
Let x = number of isolated vertices.
E(x) = (# vertices) (probability no edges are present) = $(n)(1-p)^{n-1}$

Let $P(n) = \frac{c(\ln(n))}{n}$.

$$
\begin{aligned}
\text{Then } E(x) &= n\left(1-\frac{c(\ln(n))}{n}\right)^{n-1} \\
&\approx n\left(1-\frac{c(\ln(n))}{n}\right)^{n} \\
&= n\left(1-\frac{c(\ln(n))}{n}\right)^{c(\ln(n))\left(\frac{n}{c(\ln(n))}\right)} \\
&\approx n\left(\frac{1}{e}\right)^{c(\ln(n))} \\
&= n\left(\frac{1}{n}\right)^{c} \\
&= n^{1-c}
\end{aligned}
$$

Therefore, $\lim_{n\to\infty}(E(x)) = \begin{cases} \infty, c < 1 \ (using\ second\ moment) \\ 0, c > 1 \end{cases}$