

1 Reminders

Paper Topics The preliminary list of topics for research papers and target conferences has been posted at <http://www.divms.uiowa.edu/~sriram/196/spring12/researchPapers.html>. You are not limited to these topics. Once the topic choosing period has expired you cannot change so think carefully! The list of conferences will be updated as more suggestions are sent in.

Our Focus We don't want to know only the results, but also the techniques so we can use these in the future in other applications.

2 “Small World” Property of $G(n, p)$

2.1 Review of Heuristic Proof

Recall the claim that the expected diameter of a graph $G(n, p)$ is

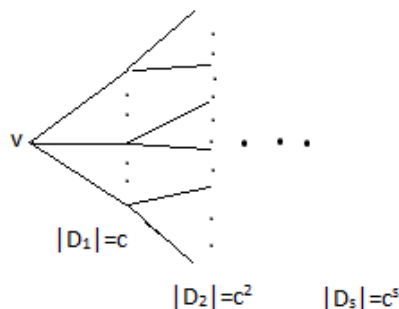
$$\frac{\ln n}{\ln c}, \text{ where } c = p(n-1), \text{ the expected degree}$$

We want to prove the stronger claim,

$$\text{Prob}[\text{distance}(u, v) > \frac{\ln n}{\ln c}] \rightarrow 0 \text{ as } n \rightarrow \infty \forall u, v \in v$$

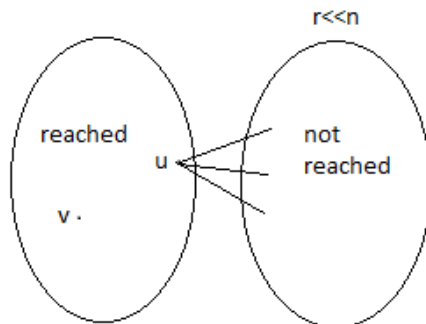
This second claim is stronger because the first allowed for outlier graphs of very small or large diameter—only the average diameter needed to be equal to $\frac{\ln n}{\ln c}$. In the second claim the probability of outlier graphs goes down as the number of vertices rises.

Proof: Heuristic Proof Let v be an arbitrary vertex in $G(n, p)$. We can describe a set D_s as the set of vertices which are s hops from v . If we assume that v has degree c then we can say that $|D_1| = c$. Following this logic we see that at distance s , $|D_s| = c^s$. If $c^s = n$ then all vertices in $G(n, p)$ are reachable in s or fewer hops from v . $s = \log_c n = \frac{\ln n}{\ln c}$ \square



Problems with this proof

1. As c is the expected degree, actual degree may deviate from this.
2. As $c^s \rightarrow n$ the expected number of edges between a vertex u in the visited set and a vertex w in the not yet visited set of size r is $p(r) \ll p(n) \approx c$



3. As $c^s \rightarrow n$, even the expected u, w edges may result in finding far fewer new vertices, and the size of c^s may be much smaller than c .

2.2 Second Heuristic Proof

Because of the problems with the first proof, we need a better and more detailed proof with better theoretical support. A more perfect proof follows.

Proof: Let v be an arbitrary vertex in $G(n, p)$ and let S_i denote the set of vertices at distance i from v . Let $T_i = V \setminus (S_0 \cup S_1 \cup \dots \cup S_i)$. Now let $p \geq C \frac{\ln n}{n}$, where C is a large constant. How is this p value chosen?

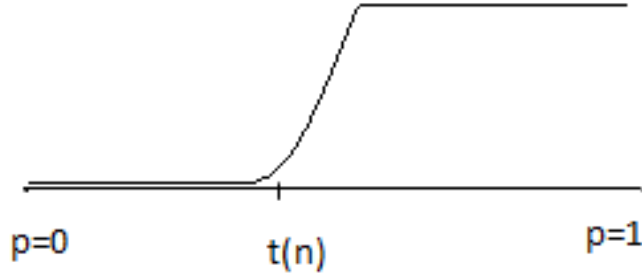
Phase Transitions

Definition 1 Property φ of graph $G(V, E)$ is monotone if G has $\varphi \Rightarrow G'(V, E \cup e)$ also has φ where $e \notin E$ connects two vertices in V .

Examples

1. Having a giant component
2. Having diameter less than or equal to a constant value d
3. Connectivity

Definition 2 Function $t(n)$ is a threshold function for monotone property φ if $\text{Prob}[G(n, p(n)) \text{ has property } \varphi] \rightarrow 0$ as $\frac{p(n)}{t(n)} \rightarrow \infty$ ($\lim_{n \rightarrow \infty} \frac{p(n)}{t(n)} = 0$)



Examples

1. For connectivity, $t(n) = \frac{\ln n}{n}$
2. For a giant component, $t(n) = \frac{1}{n}$. Before this threshold, all components have $O(\log n)$, afterwards they have $\Omega(n)$. This is a double jump point.

Note that a threshold will exist for any monotone property, not only ones with a phase transition, and proving that a given property has a phase transition can be a tricky process. However, for properties that we know have phase transitions (like the ones above) the threshold indicates the point where the phase transition occurs. Now that we know this, we want to work in the region of connected graphs by setting $p \geq C \frac{\ln n}{n}$, $C > 1$, placing us after the phase transition of connectivity in the region where nearly every graph is connected. Clearly if the graph is not connected we cannot prove the small world property for the whole graph, so that is why we take this step.

Lemma 3 Suppose $|T_i| \geq \frac{n-1}{2}$. Then for $|S_{i+1}| \approx C |S_i|$, $\text{Prob}[\frac{c}{4} |S_i| \leq |S_{i+1}| \leq \frac{5}{4} |S_i|] \geq 1 - \frac{1}{n}$

Proof: Let $u \in T_i$. Let $X_u = \begin{cases} 1 & \text{if } u \text{ is connected to } S_i \\ 0 & \text{otherwise} \end{cases}$

Then $\text{Prob}[X_u = 1] = p |S_i|$. Since the expectation of such a function is the probability of the positive outcome, $E[X_u] = \text{Prob}[X_u = 1] = p |S_i|$.

Let $X = \sum_{u \in T_i} X_u$ Note that $S_{i+1} = X$.

Then $E[X] = E[\sum_{u \in T_i} X_u] = \sum_{u \in T_i} E[X_u] = |T_i| p |S_i|$

The size of X is the number of vertices that are connected to S_i . Those vertices will be in S_{i+1} , so we can say that $E[|S_{i+1}|] = p |S_i| |T_i|$.

Since for $|T_i| \geq \frac{n-1}{2}$, $E[|S_{i+1}|] \geq p |S_i| \frac{n-1}{2}$

and for $|T_i| \leq n-1$, $E[|S_{i+1}|] \leq p |S_i| (n-1) = c |S_i|$,

we can conclude that $p |S_i| \frac{n-1}{2} \leq E[|S_{i+1}|] \leq c |S_i|$.

However, this is just the expectation—in real life an example could fall outside of these bounds. How can we increase these bounds so they contain the vast majority of probable graphs? We can use “tail inequalities”.

Tail Inequalities There are many types of tail inequalities, e.g. Markov’s inequality, Chebyshev’s inequality, and Chernoff bounds[1]. We will use Chernoff bounds here.

Let $X_i \in 0, 1$ with $Prob[X_i = 1] = p_i$, therefore $E[X_i] = p_i$.

Let $X = \sum_{i=1}^n X_i$, therefore $E[X] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p_i$.

Let us denote $E[X]$ as μ .

If X_i are independent, we get excellent tail bounds. Using Chernoff bounds (C.B.), we can say the following:

Upper tail: $Prob[X > (1 + \sigma)\mu] \leq C.B.$

Lower tail: $Prob[X < (1 - \sigma)\mu] \leq C.B.$

The upper tail bound in this case is $(\frac{e^\sigma}{(1+\sigma)^{1+\sigma}})^\mu$

We can simplify this as:

$$\sigma > 2e - 1 : 2^{-\sigma\mu}$$

$$\sigma \leq 2e - 1 : e^{-\frac{\mu\sigma^2}{4}}$$

So for $|S_{i+1}| = X = \sum_{u \in T_i} E[X_u]$, $Prob[X_u = 1] = p | S_i |$ we apply Chernoff bounds on $|S_{i+1}| = X$.

In particular, $Prob[|S_{i+1}| \geq \frac{5}{4}c | S_i |]$ can be seen as $Prob[|S_{i+1}| \geq \frac{1}{4}c | S_i | + c | S_i |]$, and we know that $E[|S_{i+1}|] = c | S_i |$

□ □

Next Time Finish phase transitions, move on to the Watts-Strogatz model.

References

- [1] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23(4):493–507, 1952.