CS 134: Networks Section 2
Prof. Yaron Singer Spring 2017

1 The configuration model

Another model for constructing graphs is the configuration model. We start by specifying degrees $d_1, d_2, \ldots, d_n$ for each of our n nodes. Each node i has d_i edge "stubs" coming out of it. We now connect these stubs uniformly at random together to form edges until all stubs are connected (note that $\sum_i d_i$ must be even). We allow self-loops to form, as with high enough n this occurs with low enough probability.

When all d_i are equal, we get what is known as a random d-regular graph, for some $d > 0$. Consider what happens with various d. When $d = 1$, we get a collection of disjoint edges. At $d = 2$, the only connected (sub)graphs that can form are disjoint cycles. For $d \geq 3$, we finally get good network properties, including a good expansion, as discussed in class:

$$\alpha = \min_{|S| \leq n/2} \frac{|e^{out}(S)|}{|S|}$$

Formally, for each $d \geq 3$, there exists a constant α depending only on d such that a random d-regular graph of any size has expansion at least α.

2 Clustering coefficients on ring and square lattices.

Consider $v \in V$. We define the set S to be v and its neighbors. $S = \mathcal{N}(v) \cup v$. Then v has d neighbors in S, the points that are 1 "hop" away from v have $d-1$ neighbors in S, and in general the points that are i "hops" away from v have $d-i$ neighbors in S.

If we add them together, there are $d + 2 * \sum_{i=1}^{d/2} d - i$. However, every pair of neighbors is counted twice. So the number of pairs of neighbors in S is $\frac{d}{2} + \sum_{i=1}^{d/2} d - i$. Then we subtract the number of pairs that include v, because we want to get the number of pairs of node v's neighbors, which do not include v. So the number of pairs of v's neighbors is $\frac{d}{2} + \sum_{i=1}^{d/2}(d-i) - d$. Therefore, the clustering coefficient is

$$C(v) = \frac{\frac{d}{2} + \sum_{i=1}^{d/2}(d-i) - d}{\binom{d(v)}{2}}.$$

By simplifying the right hand side[1], we get $C(v) = \frac{3}{4}(\frac{d-2}{d-1})$.

3 Calculations on $G(n, p)$

Proposition: with high probabilty we have an isolated node in $G(n, p)$ when the average degree is sublogarithmic. Let X_v be a random variable denoting the degree of node v, and for every

[1] Recall that $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$.

other node w, let $X_{v,w}$ be a random variable equal to 1 if there's an edge between v and w, and 0 otherwise. Thus,

$$\begin{aligned} X_v &= \sum_w X_{v,w} \\ E[X_v] &= \sum_w E[X_{v,w}] \\ &= \sum_w p = (n-1)p \end{aligned}$$

Let $p = \frac{c}{n-1}$ for some constant c. For a node to be isolated, it must have no incident edges. Let $\mathcal{E}_v$ be the probability that a node is isolated. We get

$$\begin{aligned} \Pr[\mathcal{E}_v] &= (1-p)^{n-1} \\ &= (1 - \frac{c}{n-1})^{n-1} \\ &= ((1 - \frac{c}{n-1})^{\frac{n-1}{c}})^c \end{aligned}$$

The quantity in the outermost parentheses is in the range $[1/4, 1/e]$ as we let $p = \frac{c}{n-1}$ range from 1/2 down to 0. Hence, $\Pr[\mathcal{E}_v]$ is bounded between 4^{-c} and e^{-c} for our constant c. How big must c be to get with high probability no isolated nodes? Let $\mathcal{E}$ be the event that there a at least one isolated node. Using a union bound:

$$\begin{aligned} \mathcal{E} &= \bigcup_v \mathcal{E}_v \\ \Pr[\mathcal{E}] &\leq \sum_v \Pr[\mathcal{E}_v] \\ &\leq ne^{-c} \end{aligned}$$

We now simply choose c large enough so that e^{-c} is small enough to cancel n. $c = \ln n$ is too small, but $c = 2\ln n$ does the job:

$$\Pr[\mathcal{E}] \leq ne^{-2\ln n} = n \cdot n^{-2} = n^{-1}$$

4 Diameter on d-regular random graphs

A d-regular random graph is a graph where all nodes have degree d and every two nodes have the same likelihood to be connected. That is, fix some number of vertices n, consider all graphs on n vertices that are d-regular, and pick one of these uniformly at random.

For any $d \leq n-1$, it is always possible to construct a d-regular random graph and we will talk more about that in a few lectures. We recall the *expansion* of a graph from assignment 2. [2]

The crucial property of d-regular is that they are likely to be good expanders. That is, a d-regular random graph is likely to have an expansion of α. A graph G is an α-expander if each

[2] More precisely, these are multi graphs, i.e. these are graphs that can also have self loops and multiple edges between the same pair of nodes; as we will see when we discuss these graphs, in these muligraphs, it is unlikely to have more than a negligible number of self loops and multiple edges.

vertex set $S \subset V$ with $|S| \leq n/2$ has at least $\alpha|S|$ edges leaving. Intuitively, this means that no two "large chunks" of the graph can be disconnected by removing few vertices or edges.

The *expansion* of a graph $G = (V, E)$ is the minimum, over all cuts we can make (dividing the graph in two pieces), of the number of edges crossing the cut divided by the number of vertices in the smaller half of the cut. Formally, it is

$$\alpha = \min_{S \subseteq V, 1 \leq |S| \leq \frac{|V|}{2}} \frac{|e(S)|}{|S|}$$

where $e(S)$ is the number of edges leaving the set S.

So how does the expansion property help us? The following theorem shows that the expansion of a d-regular graph provides a guarantee on its diameter. Recall that a d-regular graph is a graph where all nodes have degree d.

Theorem 1. *Suppose a graph G is d-regular, for some constant $d \geq 3$, and has constant expansion α. Then the diameter of G is $O\left(\frac{d}{\alpha} \log n\right)$.*

Proof. We show that any two vertices s and t are a distance at most $O\left(\frac{d}{\alpha} \log n\right)$ apart. Let S_j be the set of vertices reachable from s in at most j steps (we can think of S_j as being formed by a breadth-first-search that starts at s).

Consider some j such that $|S_j| \leq n$. Because G has expansion α, there are at least $\alpha|S_j|$ edges leaving S_j. Since G is d-regular, there are at least $\alpha|S_j|/d$ vertices outside of S_j connected to S_j.

Therefore,

$$|S_{j+1}| \geq |S_j| + \frac{\alpha}{d}|S_j| = \left(1 + \frac{\alpha}{d}\right)|S_j|.$$

Because $S_0 = \{s\}$, we get

$$|S_j| \geq \left(1 + \frac{\alpha}{d}\right)^j.$$

Now pick $j = \frac{d}{\alpha} \log n$. Then we have

$$|S_{\frac{d}{\alpha}\log n}| \geq \left(1 + \frac{\alpha}{d}\right)^{\frac{d}{\alpha}\log n}.$$

We use the well-known fact that $\left(1 + \frac{1}{k}\right)^k \geq 2$ for $k \geq 1$ to get that $|S_{\frac{d}{\alpha}\log n}| \geq 2^{\log n} = n.$ [3]

Therefore, the size of S_j reaches at least $\frac{n}{2}$ before this point, *i.e.* before $j = \frac{d}{\alpha} \log n$. Now, by the exact same reasoning, if we start at t and consider the sets T_j, we find that the size of T_j reaches at least $\frac{n}{2}$ before $j = \frac{d}{\alpha} \log n$. But then $T_{\frac{d}{\alpha}\log n}$ and $S_{\frac{d}{\alpha}\log n}$ must have some vertex in common (since both are larger than $\frac{n}{2}$). That means there is a path from s to t of length at most $2\frac{d}{\alpha} \log n$, because we can go from s to this common vertex in at most $\frac{d}{\alpha} \log n$ steps, and then similarly to get from this vertex to t. □

So the above theorem implies that for $d \geq 3$, a d-regular graph with constant expansion α gives us the small world property that we want: it has small (logarithmic) diameter. Using randomness, we can construct d-regular graphs with constant expansion.

[3]Note: This assumes that $k = \frac{d}{\alpha}$ is bigger than 1. But if it is less, then we can fix the argument by taking $j = c\frac{d}{\alpha} \log n$ with $c = \alpha/d$. Since α and d are constant, c is constant and the asymptotic result remains the same.

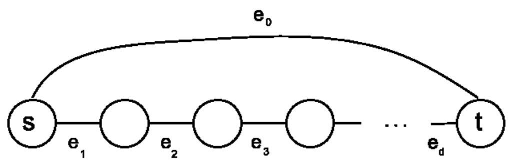

Figure 1: A graph for random routing protocol to act on. We wish to route from s to t; only edges coming out from s are depicted and have weight 1.

For $d \geq 3$, it turns out that d-regular random graphs have expansion that is a function of d only, not of the number of vertices (we skip the proof). While d-regular random graphs have small diameter, they unfortunately have low clustering coefficients (about d/n and we skip the proof as well).

5 Routing on random graphs

Let's be clear about what we mean when we discuss routing on random graphs. For a given distribution of graphs $\mathcal{G} = \{G_1, G_2, ...\}$ we can use $X_{s,t}$ to denote the random variable describing the number of steps the algorithm takes from s to t. This object is a random variable since the graph is randomly drawn from the distribution $\mathcal{G}$. The delivery time from s to t is $\mathbb{E}_{G \sim \mathcal{G}}[X_{s,t}]$.

An important insight that will make the proof much easier to handle is that instead of analyzing a deterministic algorithm routing on a random graph, we can analyze a randomized algorithm routing on a fixed graph. Let's first see an example of such a random protocol. In Figure 1, we depict a graph where a source node has links to a set of nodes u_1, u_2, u_3, $\cdots$ and some target node t. A random routing protocol selects an edge e from s with some probability p_e and lands at some destination node in the set $\{u_1, u_2, \cdots\} \cup \{t\}$. We then repeat with an another edge picked at random from the node where the protocol previously landed until we get to t. Denote by $Y_{s,t}$ the number of steps taken to go from s to t with such a random routing protocol.

Instead of generating a random graph and analyzing the algorithm, we can therefore generate the long-range links from a node u only once the algorithm visits node u since no node is visited more than once. Using the principle of deferred decisions, we have that $\mathbb{E}_{G \sim \mathcal{G}}[X_{s,t}] = \mathbb{E}[Y_{s,t}]$ and that analyzing the deterministic protocol on the random graph is equivalent to analyzing a randomized algorithm on a fixed graph.

References

[1] Jon Kleinberg. *The small-world phenomenon: An algorithmic perspective.* Proc. 32nd ACM Symposium on Theory of Computing, 2000.