Stat206: Random Graphs and Complex Networks **Spring 2003**

Lecture 8: Graph Diameter in the Proportional Attachment Model

Lecturer: David Aldous *Scribe: Lara Dolecek*

We will consider a directed graph $\mathcal{G}_n^{(m)}$. An incoming vertex i is attached to any m of the existing $1, 2, ..., i-1$ vertices, with the probability proportional to the degree of the vertex. An example of a graph for $m = 3D1$, and $n = 3D10$ is shown below.

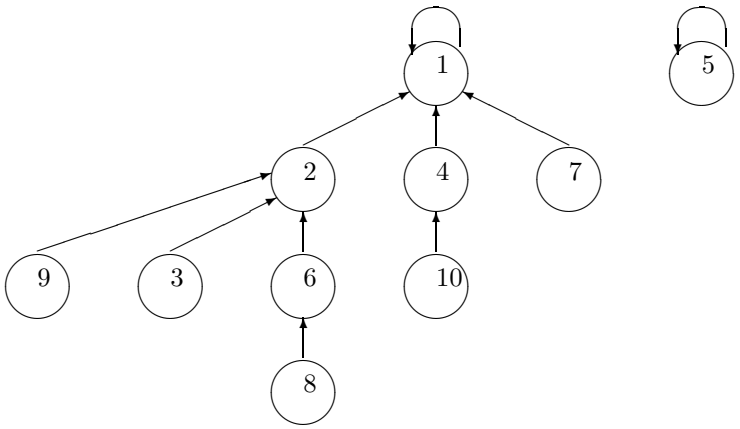

Figure 1: A possible realization of $\mathcal{G}_{10}^{(1)}$

We can construct a $m = 3D2$ realization from the $m = 3D1$ realization by grouping the old nodes i and $i+1$ into a new node $(i, i+1)$. The edges between these new nodes, $(i, i+1)$, and $(j, j+1)$, exist when $i \neq j$, and there is an edge between i and j, or between i and $j+1$, or between $i+1$ and j, or between $i+1$ and $j+1$, in the $m = 3D1$ realization. This is illustrated in the following figure.

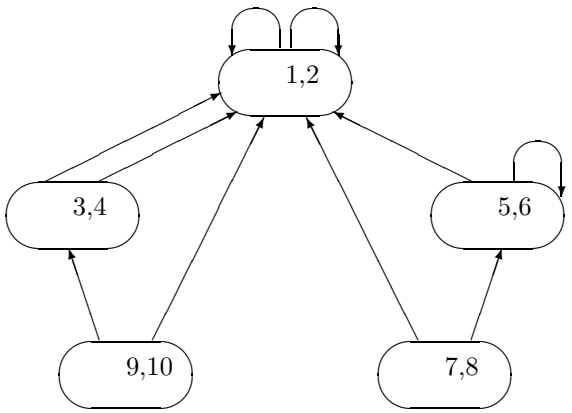

Figure 2: A possible realization of $\mathcal{G}_{10}^{(2)}$

Similarly, for the general $m, m \geq 2$, we group the nodes $(1, 2, ..., m), (m+1, m+2, ..., m+m)$, and so on, and create the edges between $(km+1, km+2, ...(k+1)m)$ and $(lm+1, lm+2, ..., (l+1)m)$, if $k \neq l$, and $\exists$ k_0, l_0 such that $km+1 \leq k_0 \leq (k+1)m$, and $lm+1 \leq l_0 \leq (l+1)m$, and there is an edge between k_0 and l_0 in $\mathcal{G}_n^{(1)}$.

Definition: (for $m = 3D1$) of the process $\mathcal{G}_1, \mathcal{G}_2, ...$ (the sequence of random graphs)

Vertex i has an edge $i \to j$ (where $j \leq i$) with probability:

$$P_{i \to j} = 3D \begin{cases} \frac{\text{degree of j in } \mathcal{G}_{i-1}}{2(i-1)+1} & \text{for } j =\neq i \\ \frac{1}{2(i-1)+1} & \text{for } j = 3Di \end{cases}$$

Note that we allow self loops in this construction.

Definition: (for $m \geq 2$) of the process $\mathcal{G}_1^{(m)}, \mathcal{G}_2^{(m)}, ...$ (the sequence of random graphs)

Vertex i has its u^{th} edge to j (where $j \leq i$) with probability:

$$P_{i \to j} = 3D \begin{cases} \frac{\text{degree of j in } \mathcal{G}_{i-1}^{(m)} \bigcup \{\text{u-1 = edges}\}}{2m(i-1)+2(u-1)+1} & \text{for } j \neq i \text{ and } 1 \leq u \leq= m \\ \frac{1}{2m(i-1)+2(u-1)+1} & \text{for } j = 3Di \text{ and } 1 \leq u \leq m \end{cases}$$

Here the term $\mathcal{G}_{i-1}^{(m)} \bigcup \{\text{u-1 edges}\}$ refers to the graph constructed using all of the edges added so far. Note that self loops contribute two times to the degree of the vertex.

Graph Diameter and Average Length

For the deterministic graph G we define $Diameter(G)$ and $\overline{l}(G)$ as:

$Diameter(G) = 3D\max\ d(v, w)$, and

$\overline{l}(G) = 3D\text{average}\ d(v, w)$,

where $d(v, w)$ denotes the minimum path length between v and w in G. We assume that G is undirected in computing $d(v, w)$.

Similarly, for a random graph $\mathcal{G}_n$ we can define $E[Diameter(\mathcal{G}_n)]$, and $E[\overline{l}(\mathcal{G}_n)]$.

Let us now try to compute the value of these two parameters, by first considering $m = 3D1$.

$m = 3D1$ case

For this case we will use a heuristic argument to quantify the parameters. $\mathcal{G}_n$ is essentially a tree (it can also be a forest). Let X be a random vertex of $\mathcal{G}_n$.

Claim: $d(X, root) \sim \frac{1}{2} \log n$ as $n \to \infty$.

The consequence of the claim is that, $E[\overline{l}(\mathcal{G}_n)] \sim \log n =3\text{D}\frac{1}{2} \log n + \frac{1}{2} \log n$ as $n \to \infty$, since for two random vertices the paths will meet close to the root.

To "prove" the claim, let us first recall the Yule Process, which was discussed in Lecture 4. e^t vertices arrive by time t, so that $t = 3D \log n$, where n denotes the number of vertices. This means that $d(X, root) \approx \frac{\log n}{E[Y]}$,

where Y denotes the time between two successive attachments. This is shown in Figure 3. Note that $Y \sim Y_i$. Prior probability the vertex is born y ago is $e^{(-y)}$. If the vertex is born y ago, then E [number of children now]=3D$e^{y/2}$. The relative probability that the incoming vertex gets attached to a potential parent at time y is proportional to $e^{y/2}$. By the Bayes rule, the posterior probability of y is proportional to $e^{-y}e^{y/2}$=3D$e^{-y/2}$, so that $Y \sim Exp(2)$ and $E[Y] = 3D2$, from which the claim "follows".

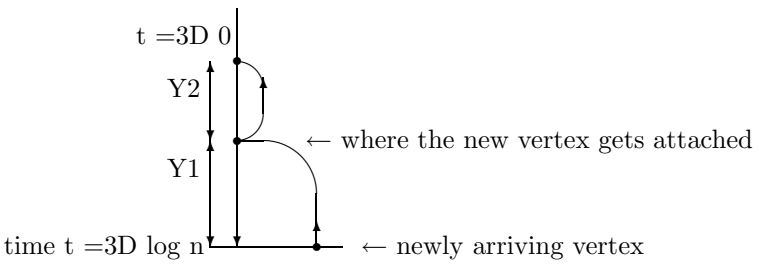

Figure 3: Illustration of vertex arrival and attachment

Remark: In the case $m \geq 2$, if we only consider paths going towards the root, then what is $d(X, root)$?

For $m = 3D2$, let M_k denote the time it takes for 2^k arrivals. ¿From the branching random walk theory, we have that $M_k \sim c_m \cdot k$. If we attempt to use the analogy with the $m = 3D1$ case, the answer to our question might be that $d(X, root) \sim \frac{\log n}{c_m}$. However, this is not quite right as it will be seen in the next section.

$m \geq 2$ case

For the case $m \geq 2$, we will address the following Theorem from [1]:

$$E[Diameter(\mathcal{G}_n)] = 3D\frac{(1 + o(1)) \log n}{\log \log n} =$$

$$E[\overline{l}(\mathcal{G}_n)] = 3D\frac{(1 + o(1)) \log n}{\log \log = n}$$

Basically, the idea is to consider B_k where B_k denotes the number of vertices within distance k from a randomly chosen vertex. It can be shown that $\frac{B_{k+1}}{B_k} \approx \log n$ so that $B_k \approx (\log n)^k$. By solving $(\log n)^k \approx \sqrt{n}$ it follows that $k \approx \frac{1/2 \log n}{\log \log n}$ so that $\overline{l}(\mathcal{G}_n) \approx 2k$.

Each possible path of length L in $\mathcal{G}_n^{(m)}$ corresponds to m^{2L} possible edge-sets S of $\mathcal{G}_{mn}^{(1)}$. We can assume that each edge-set S has property that a vertex j is linked to at most one vertex $i, i < j$. Denote this property by (*).

Lemma [1]:

For a non random edge-set S satisfying property (*), the following holds:

$P(S \subseteq \mathcal{G}_n^{(1)}) \leq 2^{e(S)(\triangle(S)+2)} \prod_{i,j \in S} \frac{1}{\sqrt{ij}}$

where $e(S)$ denotes the number of edges in S, and $\triangle(S)$ is the maximum degree of vertex in S (Note that S is really an edge-set but now think of it as a graph).

The result of the preceding Lemma can be used to establish a lower bound on the diameter.

We will use $\triangle(S) \leq 2$ (valid for a tree-like structure). Let us first fix one possible path of length L in $\mathcal{G}_n^{(m)}$,

and denote it by $v_0, v_1, ..., v_L$. (Renumbered with respect to $\mathcal{G}_n^{(1)}$). Now pick some corresponding edge-set S. Now, from Lemma,

$P(S \subseteq \mathcal{G}_{mn}^{(1)}) \leq 2^{4L} \prod_{t=3D0}^{L-1} \frac{1}{\sqrt{v_t * v_{t-1}}} = 3D16^L \frac{1}{\sqrt{v_0 * v_L}} \prod_{t=3D0}^{L-1} \frac{1}{v_t}$.

So that,

$P(\text{path } v_0, v_1, ..., v_L \text{ exists in } \mathcal{G}_n^{(m)}) \leq m^{2L} \frac{16^L}{\sqrt{(v_0 * v_L)}} \prod_{t=3D0}^{L-1} \frac{1}{v_t}$.

Now, let us fix the end vertices so that $n-1 = 3Dv_0, v_1, ..., v_L = 3Dn$. The above then becomes

$$\begin{aligned} P(\exists \text{ a path in } \mathcal{G}_n^{(m)} \text{of length L from n-1 to n}) &\leq \sum_{v_1, v_2, ..., v_{L-1} \in \{1,2,...,n-2\}} (m^{2L}) \frac{16^L}{\sqrt{v_0 * v_L}} \prod_{t=3D0}^{L-1} \frac{1}{v_t} \\ &\leq \frac{(16m^2)^L}{n} (\sum_{i=3D1}^{n} \frac{1}{v_i})^{L-1} \\ &\leq \frac{(16m^2)^L}{n} (\log n)^L. \end{aligned}$$

We want to find the largest $L = 3DL_n$ which will drive the upper bound (RHS) to zero. Consider

$\frac{\log(RHS)}{\log n} = 3D \frac{const * L}{\log n} - 1 + \frac{L \log\log n}{\log n}$.

So if we choose

$L = 3D \frac{(1-\epsilon)\log n}{\log\log n}$

this ratio will then $\to -\epsilon$. This means that $\log(RHS) \to -\infty$, so that $RHS \to 0$. Now we can conclude that $P(\exists \text{ a path in } \mathcal{G}_n^{(m)} \text{of length } \leq \frac{(1-\epsilon)\log n}{\log\log n}) \to 0$.

Next time we will prove this Lemma.

References

[1] B.BOLLOBAS and O.RIORDAN "The Diameter of a Scale Free Random Graph", to appear