Lecture 32: The degree of a randomly chosen vertex of preferential attachment graphs

We have studied the (empirical) degree distribution of the preferential attachment tree. Here we look at a closely related quantity, the distribution of the degree of a randomly chosen vertex of the preferential attachment tree on n vertices. To define this object, recall the previous notation: Let $G_n \sim \mathrm{PA}_n$ and for $1 \le i \le n$, let $D_n(i)$ be the degree of vertex i in G_n. Let U_n be uniform on $\{1, \dots, n\}$ independent of G_n and then set $W_n = D_n(U_n)$. Our first result about W_n is an expression for its probability mass function. Let

$$P_n(k) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[D_n(i) = k],$$

be the proportion of vertices with degree k in G_n.

Lemma 32.1. *Let W_n be the degree of a randomly chosen vertex in $G_n \sim \mathrm{PA}_n$. Then*

$$\mathbb{P}(W_n = k) = \mathbb{E}[P_n(k)].$$

Proof. We condition on G_n and then take expectation.

$$\begin{aligned} \mathbb{P}(W_n = k) &= \mathbb{E}\big[\mathbb{P}(D_n(U_n) = k | G_n)\big] \\ &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[D_n(i) = k]\right] \\ &= \mathbb{E}[P_n(k)]. \end{aligned}$$

□

With this result and the estimates in Lecture 31, we can prove the following.

Theorem 32.2. *Let W_n be the degree of a randomly chosen vertex in $G_n \sim \mathrm{PA}_n$ and for $k \ge 1$, let $p_k := 4/(k(k+1)(k+2))$. Then for any $\alpha \in (0,1)$ and $n \ge 2$,*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), (p_k)_{k \ge 1}\big) \le (1-\alpha)^{-2} n^{-\alpha} + n^{-2}.$$

Proof. By the definition of total variation distance and Lemma 32.1, we have

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), (p_k)_{k \ge 1}\big) = \frac{1}{2} \sum_{k \ge 1} \big|\mathbb{P}(W_n = k) - p_k\big| = \frac{1}{2} \sum_{k \ge 1} \big|\mathbb{E}[P_n(k)] - p_k\big|.$$

Lemma 31.1 from Lecture 31 implies that for $1 \le k \le n$,

$$\big|\mathbb{E}[P_n(k)] - p_k\big| \le \left(\frac{2}{1-\alpha}\right) k^{-\alpha} n^{-1},$$

and for $k > n$, $P_n(k) = 0$, so that

$$\big|\mathbb{E}[P_n(k)] - p_k\big| = p_k.$$

Combining the last three displays yields

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), (p_k)_{k \ge 1}\big) \le \left(\frac{1}{1-\alpha}\right) n^{-1} \sum_{k=1}^{n} k^{-\alpha} + \frac{1}{2} \sum_{k=n+1}^{\infty} p_k.$$

For the first term on the right hand side, we have the usual bound

$$\sum_{k=1}^{n} k^{-\alpha} \le 1 + \int_1^n x^{-\alpha} dx = \frac{n^{1-\alpha} - \alpha}{1-\alpha} \le \frac{n^{1-\alpha}}{1-\alpha}.$$

For the second term, we have

$$\sum_{k=n+1}^{\infty} p_k = \frac{2}{(n+2)(n+1)}.$$

Combining the last three displays easily gives the result. □

The next result gives an alternative representation for the limiting degree distribution $(p_k)_{k\geq 1}$.

Proposition 32.3. *If U is uniform on the interval $(0,1)$, and X is a random variable with $\mathscr{L}(X|U) = \mathrm{Geo}_1(\sqrt{U})$, then*

$$\mathbb{P}(X = k) = p_k.$$

Proof. The proof is by conditioning on U.

$$\mathbb{P}(X=k) = \mathbb{E}\big[\mathbb{P}(X=k|U)\big] = \int_0^1 (1-\sqrt{u})^{k-1}\sqrt{u}du = 2\int_0^1 (1-v)^{k-1}v^2 dv = \frac{2\Gamma(3)\Gamma(k)}{\Gamma(k+3)},$$

where in the third equality we have made the substitution $v = \sqrt{u}$, and in the last, we have used two alternative expressions for the beta function. The result follows since for $j \in \mathbb{N}$, $\Gamma(j) = (j-1)!$. □

Comparing this proposition to Theorem 32.2, we find that

$$d_{\mathrm{TV}}\big(D_n(U_n), \mathrm{Geo}_1(\sqrt{U})\big) \xrightarrow[n\to\infty]{} 0.$$

Since $\mathscr{L}(U_n) = \mathscr{L}(\lceil nU \rceil)$, we may suspect that for large i,

$$\mathscr{L}\big(D_n(i)\big) \approx \mathrm{Geo}_1(\sqrt{i/n}).$$

This is in fact true, but requires more space than we have to prove. Instead we just state a precise version of the result we need, which is proven using Stein's method for approximation by the geometric distribution.

Proposition 32.4. *If $D_n(i)$ is the degree of vertex i in $G_n \sim \mathrm{PA}_n$, then there is a constant C not depending on n or i such that*

$$d_{\mathrm{TV}}\big(\mathscr{L}(D_n(i)), \mathrm{Geo}_1(\sqrt{i/n})\big) \leq \frac{C}{i}.$$

With this result, we can improve the rate in Theorem 32.2, by using the following useful property of the total variation distance.

Proposition 32.5. *If D, X, U are random variables on the same space, then*

$$d_{\mathrm{TV}}\big(\mathscr{L}(D), \mathscr{L}(X)\big) \leq \mathbb{E}\big[d_{\mathrm{TV}}\big(\mathscr{L}(D|U), \mathscr{L}(X|U)\big)\big].$$

Proof. For any event A,

$$\begin{aligned}\big|\mathbb{P}(D\in A) - \mathbb{P}(X\in A)\big| &= \big|\mathbb{E}\big[\mathbb{P}(D\in A|U) - \mathbb{P}(X\in A|U)\big]\big| \\ &\leq \mathbb{E}\big[\big|\mathbb{P}(D\in A|U) - \mathbb{P}(X\in A|U)\big|\big] \\ &\leq \mathbb{E}\big[d_{\mathrm{TV}}\big(\mathscr{L}(D|U), \mathscr{L}(X|U)\big)\big].\end{aligned}$$

Taking the supremum over events A gives the result. □

Theorem 32.6. *Let $n \geq 2$, W_n be the degree of a randomly chosen vertex in $G_n \sim \mathrm{PA}_n$, and let $p_k := 4/(k(k+1)(k+2))$, $k \geq 1$. Then for C equal to the constant of Proposition 32.4, we have*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_n), (p_k)_{k\geq 1}\big) \leq (C+1)\frac{\log(n)+1}{n}.$$

Proof. Let U be uniform on the interval $(0,1)$, set $W_n = D_n(\lceil nU \rceil)$, where $D_n(i)$ denotes the degree of vertex i in G_n, and let $\mathscr{L}(X|U) = \mathrm{Geo}_1(\sqrt{U})$ and $\mathscr{L}(X_n|U) = \mathrm{Geo}_1(\sqrt{\lceil nU \rceil / n})$. Then Propositions 32.3 and 32.5 and the triangle inequality imply

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W_n), (p_k)_{k\geq 1}\big) &\leq \mathbb{E}\big[d_{\mathrm{TV}}\big(\mathscr{L}\big(D_n(\lceil nU \rceil)\,|U\big), \mathscr{L}(X|U)\big)\big] \\ &\leq \mathbb{E}\big[d_{\mathrm{TV}}\big(\mathscr{L}\big(D_n(\lceil nU \rceil)\,|U\big), \mathscr{L}(X_n|U)\big)\big] + \mathbb{E}\big[d_{\mathrm{TV}}\big(\mathscr{L}(X_n|U), \mathscr{L}(X|U)\big)\big] \\ &\leq C\mathbb{E}\big[\lceil nU \rceil^{-1}\big] + \mathbb{E}\left[\sqrt{\frac{n}{\lceil nU \rceil}}\left(\sqrt{\frac{\lceil nU \rceil}{n}} - \sqrt{U}\right)\right], \end{aligned} \tag{32.1}$$

where in the last inequality we have used Proposition 32.4 and that $d_{\mathrm{TV}}\big(\mathrm{Geo}_1(p), \mathrm{Geo}_1(p')\big) \leq p^{-1}|p - p'|$ (left as an exercise), noting $\lceil nU \rceil / n = U + \delta/n \geq U$, where $\delta := \delta(n, U) \in [0, 1)$. For the first term of (32.1), using that $\lceil nU \rceil$ is uniform on $\{1, \ldots, n\}$, we have

$$\mathbb{E}\big[\lceil nU \rceil^{-1}\big] = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{i} \leq \frac{1 + \log(n)}{n}. \tag{32.2}$$

For the second term of (32.1), first note that the concavity of the square root implies

$$\sqrt{\frac{\lceil nU \rceil}{n}} - \sqrt{U} = \sqrt{U + \frac{\delta}{n}} - \sqrt{U} \leq \left(\frac{\delta}{n}\right)\frac{U^{-1/2}}{2} \leq \frac{1}{2n}U^{-1/2}.$$

Thus the second term of (32.1) is bounded

$$\begin{aligned} \mathbb{E}\left[\sqrt{\frac{n}{\lceil nU \rceil}}\left(\sqrt{\frac{\lceil nU \rceil}{n}} - \sqrt{U}\right)\right] &\leq \frac{1}{2n}\mathbb{E}\left[\sqrt{\frac{n}{\lceil nU \rceil}}\,U^{-1/2}\right] \\ &= \frac{1}{2n}\int_0^1 \sqrt{\frac{n}{\lceil nu \rceil}}\,u^{-1/2}du \\ &= \frac{1}{2n}\sum_{i=1}^{n}\sqrt{\frac{n}{i}}\int_{(i-1)/n}^{i/n} u^{-1/2}du \\ &= \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{n}{i}}\left(\sqrt{\frac{i}{n}} - \sqrt{\frac{(i-1)}{n}}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\big(1 - \sqrt{1 - i^{-1}}\big). \end{aligned}$$

Again using the concavity of square root, we have for $i \geq 2$,

$$1 - \sqrt{1 - i^{-1}} \leq \frac{1}{2}\big(1 - i^{-1}\big)^{-1/2} i^{-1} \leq i^{-1},$$

and for $i = 1$, $1 - \sqrt{1 - i^{-1}} = i^{-1}$, so that the last two displays imply

$$\mathbb{E}\left[\sqrt{\frac{n}{\lceil nU \rceil}}\left(\sqrt{\frac{\lceil nU \rceil}{n}} - \sqrt{U}\right)\right] \leq \frac{1}{n}\sum_{i=1}^{n}\frac{1}{i} \leq \frac{1 + \log(n)}{n}.$$

Combining this with (32.2) and (32.1) proves the theorem. □