Random Graphs from Restricted Classes

Lecture 1

Lecture notes by Konstantinos Panagiotou and Elisabetta Candellero

August 21, 2013

Introduction

Suppose we are given a class $\mathcal{G}$ of *labeled graphs*, i.e.,

$$\mathcal{G} = \bigcup_{n \geq 0} \mathcal{G}_n,$$

where $\mathcal{G}_n$ contains all elements of $\mathcal{G}$ that have n vertices.

Question. *Consider a **random graph** G_n from the set $\mathcal{G}_n$. How does it look like?*

An example of such graph is the *Erdős-Renyi random graph*, typically denoted by $G_{n,p}$. Here n is the number of vertices, and $0 < p < 1$ is the probability that any two vertices are connected by an edge. One of the key properties of $G_{n,p}$ is that each edge appears *independently* of all the others.

In this course we will consider other classes of graphs $\mathcal{G}$, for example:

(i) trees and families of *planar* graphs.

(ii) *Outerplanar* graphs[1], i.e., those which do not contain the complete graph K_4 nor the complete bipartite graph $K_{2,3}$ as minors. (Recall that a graph H is said to be a *minor* of G, if H can be constructed from G only by deletion of edges –and vertices– and by contraction of edges.)

(iii) *Series-parallel* graphs, which do not contain K_4 as minor.

In particular, we will study the *global structure* and some *local properties* (for example the distribution of the degree sequence) of such families of graphs.

[1] An **outerplanar graph** can be embedded in the plane in such a way that there is a face containing all the vertices.

1 Trees

Denote by $\mathcal{T}_n$ the set of labeled trees with n vertices. By **Cayley's formula** the number of such trees is given by $|\mathcal{T}_n| = n^{n-2}$. We denote by

$$d_k(T) := \#\{\text{vertices of } T \text{ having degree } k\}.$$

Question. *Let T_n be chosen uniformly at random from $\mathcal{T}_n$. What can we say about the **distribution** of the random variable $d_k(T_n)$?*

In order to answer this question, we need to proceed by steps.

First Step: Rooting. Define the family of rooted trees on n vertices to be

$$\mathcal{T}_n^\bullet := \mathcal{T}_n \times [n],$$

where $[n] = \{1, \ldots, n\}$, and consider an element $T_n^\bullet$ chosen uniformly at random from $\mathcal{T}_n^\bullet$. Then we have that

$$d_k(T_n^\bullet) \stackrel{d}{=} d_k(T_n),$$

meaning that the distribution of the degree sequence (like many other properties of such trees) is not affected by the rooting.

Second Step: Sampler for rooted Trees. In order to create a *sampler*, we proceed as follows:

(i) create a single "untouched" vertex (which will be the root),

(ii) *while there are untouched vertices*, perform the following actions.

   - Select any untouched vertex (arbitrarily), denote it by v.
   - Given a value n_v distributed according to a Poisson random variable of parameter 1, and create n_v untouched vertices.
   - Connect the new vertices to v and declare v "touched".

   This is nothing else than a *branching process* whose offspring distribution follows a Poisson of parameter 1.

(iii) Label the vertices uniformly at random, choosing the labels from a set S (with $|S|$ equal to the number of vertices in the tree).

We denote this sampler by $\Gamma T^\bullet$.

Remark 1.1. *Note that we have **no control over the size of the output**, i.e., the number vertices in the generated tree is a random variable.*

Remark 1.2. *All choices made in the course of the execution of the sampler are independent.*

Now we can state the following result.

Lemma 1.3. *Let $n \in \mathbb{N}$. Then the following is true.*

(a) There is a $c > 0$ such that $\mathbb{P}\left(|\Gamma T^\bullet| = n\right) \sim cn^{-3/2}$.

(b) Let $T^\bullet \in \mathcal{T}_n^\bullet$. Then

$$\mathbb{P}(\Gamma T^\bullet = T^\bullet \mid |\Gamma T^\bullet| = n) = |\mathcal{T}_n^\bullet|^{-1} = n^{-n+1}.$$

Thus, conditional on the event "$|\Gamma T^\bullet| = n$" we get the uniform distribution on $\mathcal{T}_n^\bullet$.

Proof. Let $T^\bullet \in \mathcal{T}_n^\bullet$. By induction we will show that

$$\mathbb{P}(\Gamma T^\bullet = T^\bullet) = \frac{e^{-n}}{n!}, \tag{1}$$

from which (a) and (b) will follow rather straightforwardly. To see that (1) implies (b), it suffices to notice that the right-hand-side does not depend on the tree itself, but only on its size. To see that (a) is a consequence of (1), it suffices to observe that

$$\mathbb{P}\left(|\Gamma T^\bullet| = n\right) = |T^\bullet| \frac{e^{-n}}{n!},$$

and the claim follows by applying Stirling's formula.

Now we prove (1) inductively. It is straightforward to verify that (1) holds for $= 1$. Now suppose that the root of $T^\bullet$ has degree d. Denote by $T_1^\bullet, \ldots, T_d^\bullet$ the d subtrees, where the chosen order is arbitrary. Then

$$\mathbb{P}(\Gamma T^\bullet = T^\bullet) \;=\; \frac{1}{n}\,\mathbb{P}(\mathrm{Po}(1) = d)\,\mathbb{P}(T_1^\bullet, \ldots, T_d^\bullet).$$

The factor $1/n$ corresponds to the probability that the root is assigned the correct label. The quantity $\mathbb{P}(\mathrm{Po}(1) = d)$ is the probability that the root has degree d. Finally, $\mathbb{P}(T_1^\bullet, \ldots, T_d^\bullet)$ stands for the probability that the rest is correct, in the following sense. Since the $T_i^\bullet$'s can be generated in any order, we have $d!$ possible choices. Furthermore, the labels have to be distributed correctly, which happens with probability

$$\frac{1}{\binom{n-1}{|T_1^\bullet|, \ldots, |T_d^\bullet|}} = \frac{\prod_{i=1}^d |T_i^\bullet|!}{(n-1)!}$$

By induction assumption on the subtrees, one gets the statement. □

Denote by $Z_1, Z_2, \ldots$ the random choices made to build up $\Gamma T^\bullet$. Then the Z_i's are iid $\mathrm{Po}(1)$ random variables.

Let $d_k^\bullet(T^\bullet)$ denote the number of vertices *different from the root* that have degree k. Then,

$$d_k^\bullet(T^\bullet) \in \{d_k(T^\bullet), d_k(T^\bullet) - 1\}.$$

Remark 1.4. *Conditional on the event* $\{|\Gamma T^\bullet| = n\}$, *we have*

$$d_k^\bullet(\Gamma T^\bullet) = \sum_{i=2}^{n} \mathbf{1}_{\{Z_i = k-1\}}.$$

The following is a useful a concentration result, known as *Chernoff bounds*.

Theorem 1.5 (Chernoff bounds)**.** *Let* $X \sim \mathrm{Bin}(n,p)$, *with* $\mu := \mathbb{E}(X) = np$. *Then, for all* $t \geq 0$, *we have:*

$$\mathbb{P}(X \geq \mu + t) \leq \exp\left\{-\frac{t^2}{2(\mu + t/3)}\right\},$$

as well as

$$\mathbb{P}(X \leq \mu - t) \leq \exp\left\{-\frac{t^2}{2\mu}\right\}.$$

By definition of the Z_i's we have

$$\mathbb{P}(Z_i = k-1) = \frac{e^{-1}}{(k-1)!}, \tag{2}$$

hence denote by $\mu_k := \frac{n-1}{e(k-1)!}$. Then by (2), for every $0 < \varepsilon < 1$ we obtain

$$\mathbb{P}\left(\Big|\sum_{i=2}^{n} \mathbf{1}_{\{Z_i=k-1\}} - \mu_k\Big| \geq \varepsilon\mu_k\right) \leq 2e^{-\varepsilon^2\mu_k/4}.$$

Such a statement should also hold for a random tree, since the conditioning in Lemma 1.3 does not seem too "severe". This is indeed the case.

Theorem 1.6. *Let* $k \in \mathbb{N}$ *and* $\varepsilon > 0$. *Then, for* n *sufficiently large*

$$\mathbb{P}(d_k^\bullet(T_n^\bullet) \in (1 \pm \varepsilon)\mu_k) \geq 1 - e^{-\varepsilon^2\mu_k/5}.$$

Proof. As a shorthand for the notation, let us define

$$I = I(\varepsilon, k, n) := (1 \pm \varepsilon)\mu_k.$$

We observe that

$$\begin{aligned}\mathbb{P}\big(d_k^\bullet(T_n^\bullet) \notin I\big) &= \mathbb{P}\big(d_k^\bullet(\Gamma T^\bullet) \notin I \mid |\Gamma T^\bullet| = n\big) \\ &= \big(\mathbb{P}\big(|\Gamma T^\bullet| = n\big)\big)^{-1}\,\mathbb{P}\big(d_k^\bullet(\Gamma T^\bullet) \notin I,\ |\Gamma T^\bullet| = n\big).\end{aligned}$$

By Lemma 1.3 we know that $\mathbb{P}\big(|\Gamma T^\bullet| = n\big) \sim cn^{-3/2}$, for some $c > 0$. Hence

$$\begin{aligned}\big(\mathbb{P}\big(|\Gamma T^\bullet| = n\big)\big)^{-1}\,\mathbb{P}\big(d_k^\bullet(\Gamma T^\bullet) \notin I,\ |\Gamma T^\bullet| = n\big) &\sim cn^{3/2}\mathbb{P}\big(d_k^\bullet(\Gamma T^\bullet) \notin I,\ |\Gamma T^\bullet| = n\big) \\ &\leq n^{3/2}\mathbb{P}\left(\sum_{i=2}^{n} \mathbf{1}_{\{Z_i=k-1\}} \notin I\right).\end{aligned}$$

The last equality is due to Remark 1.4. At this point the Chernoff bounds yield

$$\mathbb{P}\big(d_k^\bullet(T_n^\bullet) \notin I\big) \le n^{3/2}\mathbb{P}\big(\sum_{i=2}^{n} \mathbf{1}_{\{Z_i=k-1\}} \notin I\big) \le 2n^{3/2} \exp\left\{-\frac{\varepsilon^2}{4}\mu_k\right\}.$$

Now we can choose n so large that

$$\frac{\varepsilon^2}{100}\mu_k = \frac{\varepsilon^2(n-1)}{100e(k-1)!} > \log(2n^{3/2}).$$

Hence the claim follows. □

Remark 1.7. *Remarks about main ideas that have been used.*

(i) Sequence of iid random variables relates to a combinatorial object;

(ii) the probability to have a sampler of size n is not too small;

(iii) exploit exponentially rare events.

We will see in the forthcoming lectures how this can be generalized to many other classes of graphs.

Exercises

1. Prove Lemma 1.3 (a) probabilistically by performing the following steps:

   (i) Show that

   $$\{|\Gamma T^\bullet| = n\} = \left\{\sum_{i=1}^{n} Z_i = n-1,\ \forall 1 \le N \le n-1 : \sum_{i=1}^{N} Z_i \ge N\right\}.$$

   (ii) Let $z_1, \dots, z_n \in \mathbb{N}_0$ be such that $\sum_{i=1}^{n} z_i = n-1$. Show that there is exactly one index $1 \le I \le n$ such that the rotated sequence

   $$y_1, \dots y_n = z_{I+1} \dots z_n z_1 \dots z_I$$

   has the property $\forall 1 \le N \le n-1 : \sum_{i=1}^{N} y_i \ge N$.

   (iii) Use (i) and (ii) to show that there is a $c > 0$ such that $P(|\Gamma T^\bullet| = n) \sim cn^{-3/2}$.

2. Extend Theorem 1.6 to the case where we deal with degrees $k = k(n)$. In particular, show that the conclusion still holds if $k \le (1-\varepsilon)\log(n)/\log\log(n)$, for any $\varepsilon > 0$.

3. Let $\Delta(G)$ be the maximum degree of G. Show that there is a constant $C > 0$ such that with high probability

   $$\Delta(T_n) \le C\log(n)/\log\log(n).$$