Random Graphs and Hypergraphs

Based on the course *Random Graphs and Hypergraphs* (049014) by Omer Bobrowski, Technion, 2021.

The main source for the course is *Introduction to Random Graphs* by Frieze & Karonski.

The summary covers the first 8 weeks of the course, focusing on random graphs under the simple model of edges drawn independently according to $Bernoulli(p)$. The (un-summarized) rest of the course focused on random *hyper*graphs (i.e. with hyper-edges, which can connect more than 2 nodes together).

Summarized by Ido Greenberg in 2021.

Contents

Notations:

- $G=(V,E)$, $e_G=|E|$, $e_v=|V|$
- $[n]=\{1,\dots,n\}$
- $a_n \ll b_n \Leftrightarrow a_n=o(b_n)$, $\quad \gg \Leftrightarrow \omega$, $\quad \sim \Leftrightarrow \Theta$, $\quad \approx \Leftrightarrow \mathrm{a_n}/b_n \to 1$
- w.h.p = with high probability (asymptotically)

Introduction

1. Default scope: undirected & unweighted graphs.
2. **Random graphs models:**
   a. $\boldsymbol{G \sim G(n,p)}$: n vertices; edges randomly placed independently w.p. p.
      i. Ignores structures in the graph (e.g. if (x,y),(y,z) are edges, it often should increase the probability of the edge (x,z)).
      ii. Probability of a specific graph is $P(G) = p^{|E|}(1-p)^{N-|E|}$ ($N := \binom{n}{2}$).
   b. $\boldsymbol{G \sim G(n,M)}$: M random edges (equivalently: $G(n,p)$ conditioned on $|E| = M$).
      i. Asymptotically equivalent to $G(n,p)$ with $p = M/N$.
   c. $\boldsymbol{G \sim G(n,d)}$ – ***random d-regular graph***: uniform dist. over n-sized d-regular graphs.
   d. $\boldsymbol{G \sim G(n,r)}$ – ***geometric rand. graph***: nodes are i.i.d in $[0,1]^d$, connected if distance$\leq r$.
   e. ***Preferential attachment*** model: generate by induction – iteratively add a new node and connect it to existing nodes with probabilities proportional to their current degrees.
      i. Respects the structure of Hubs (well-connected nodes get more new edges).
3. The course focuses on the simple $G(n,p)$, despite its lack of structure.
4. Applications examples:
   a. Network analysis, e.g. epidemic spread.
   b. The probabilistic methods: proving existence of objects based on their probability (in a suitable probabilistic model) being positive. For example: existence of certain colorings in graphs.
5. **Extended graph models**:
   a. ***Hypergraph***: $E \subset 2^V$ rather than V^2 (*hyper-edges* are subsets of nodes rather than pairs).
      i. ***Simplicial Complex***: hypergraph that is closed to inclusion, i.e. $\forall e \in E$, all the subsets of e are also in E.
   b. Applications examples:
      i. High-dimensional networks (connections are more general than pairs).
      ii. Triangulation – connections come in triplets.

Basic tools

6. Some bounds:
   a. $1 + x \leq e^x$
   b. $(1-\epsilon)^n \approx e^{-\epsilon n}$ $\quad (\epsilon \to 0, n\epsilon^2 \to 0)$
   c. $\binom{n}{k} \approx \frac{n^k}{k!}$ $\quad (k = o(\sqrt{n}))$
   d. $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$
7. Probabilistic tools:
   a. ***Markov inequality*** (1st moment argument): $X \geq 0 \Rightarrow P(X \geq a) \leq \mu/a$
      i. For integer X and $a := 1$: $P(X > 0) \leq \mu$
   b. ***Chebyshev inequality*** (2nd moment argument): $\forall a > 0: P(|X - \mu| \geq a) \leq \sigma^2/a^2$
      i. Private case for $X \geq 0$: $P(X = 0) \leq \frac{\sigma^2}{\mu^2}$
   c. $Binom(n, \gamma/n) \to^D Poisson(\gamma)$ (convergence in distribution).
8. ***Method of Moments***: $[\forall k: EX_n^k \to EX^k] \Rightarrow [X_n \to^D X]$ (assuming all moments are finite).

9. ***Factorial moment***: $E((S_n)_k) \coloneqq E[S_n(S_n-1)\dots(S_n-k+1)]$
   a. $E((S_n)_k) \to \lambda^k \;\Rightarrow\; S_n \to^D Poisson(\lambda)$.

Random graph process & thresholds

10. ***Random graph process***: $T_1 \dots T_N \sim unif[0,1]$ i.i.d ($N \coloneqq \binom{n}{2}$). $\boldsymbol{E_t} \coloneqq \{\boldsymbol{e} \in [\boldsymbol{N}]: \boldsymbol{T_e} \le \boldsymbol{t}\}$.
    a. Adding edges as t increases; $\forall t \in [0,1]$: $G_t \sim G(n,t)$.
11. A set A of graphs is ***monotone increasing*** if $G \in A \Rightarrow G + \{e\} \in A$ (adding edges keeps us in the set; e.g. "all graphs with min degree 5"). A is also termed *property*.
    a. ***Monotone decreasing*** is defined similarly.
    b. Claim: A is increasing & $t_1 < t_2$ ➔ $P_{t_1}(A) \le P_{t_2}(A)$. (since $G_{t_1} \subset G_{t_2}$)
12. $p^*(n)$ is a ***threshold*** for A if: $\lim_n P_t(A) = \begin{cases} 0 \; if \; t \ll p^* \\ 1 \; if \; t \gg p^* \end{cases}$
    a. A ***sharp threshold***: $\lim_n P_t(A) = \begin{cases} 0 \; if \; t \le (1-\epsilon)p^* \\ 1 \; if \; t \ge (1+\epsilon)p^* \end{cases}$
    b. Intuition: a threshold on p beyond which, the RG suddenly satisfies A's condition.
    c. Example: A={graphs with at least one edge} $\Rightarrow$ #edges$\sim Binom\left(\binom{n}{2}, p\right)$ $\Rightarrow$ $P(no\ edges) \approx (1-p)^{n^2/2} \approx e^{-pn^2/2} \;\Rightarrow\; p^* = n^{-2}$ is a threshold.
    d. Example: A={graphs with isolated node} (decreasing) $\Rightarrow p^* = \frac{\log n}{n}$ (sharp thresh).
13. **Theorem: every nontrivial monotone increasing property ($\boldsymbol{\phi \notin A}$) has a threshold.**
    a. Proof: $P_p(A)$ increases polynomially in p from 0 to 1, so $\exists p^*: P_{p^*}(A) = 0.5$, and we can show that it satisfies the threshold property.
    b. **Kalai & Friedgut**: (almost) every nontrivial A has a **sharp threshold**.

Vertex degrees

14. Denote $\boldsymbol{X_d} \coloneqq$ **number of vertices with degree $\boldsymbol{d}$**.
15. Example – **non-isolated nodes ($\boldsymbol{X_0}$)**:
    a. $E[X_0] = \sum_{v \in V}(1-p)^{n-1} = n(1-p)^{n-1} \approx ne^{-np}$.
    b. **Theorem**:
       i. $np = \log n - \omega(n)$ ➔ $\frac{X_0 - EX_0}{\sqrt{Var(X_0)}} \to^D N(0,1)$
       ii. $np = \log n + c$ ➔ $X_0 \to^D Poisson(e^{-c})$
       iii. $np = \log n + \omega(n)$ ➔ $X_0 \to^D 0$
    c. Proof relies on Stein's method (i), method of moments (ii), and Markov inequality (iii).
16. While $X_0(p)$ decreases with p from n to 0, $X_d(p)$ isn't monotone, since $X_d(0) = X_d(1) \approx 0$.
17. **Behavior of $\boldsymbol{X_d}$**: $E[X_d] = n\binom{n-1}{d}(1-p)^{n-d-1}p^d \approx \frac{n(np)^d}{d!}e^{-np}$. In particular:
    a. $np \ll n^{-1/d}$ ➔ $EX_d \to 0$, $X_d \to 0$
    b. $np = cn^{-1/d}$ ➔ $EX_d \to \frac{c^d}{d!}$, $X_d \to Poisson$
    c. $n^{-1/d} \ll np \le \log n + d\log^2 n + \omega(n)$ ➔ $EX_d \to \infty$, $\frac{X_D - EX_D}{\sqrt{Var(X_D)}} \to N(0,1)$
    d. $np = \log n + d\log^2 n + c$ ➔ $EX_d \to \frac{e^{-c}}{d!}$, $X_d \to Poisson$
    e. $np = \log n + d\log^2 n + \omega(n)$ ➔ $EX_d \to 0$, $X_d \to 0$ $(log^2 = loglog)$

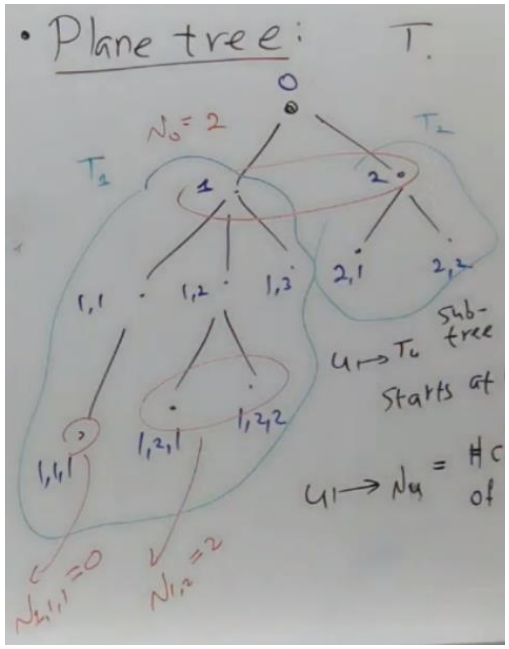
18. **Min/max degree** ($\delta(G), \Delta(G)$):
    a. Note: $E[d_v] = (n-1)p \approx np$.
    b. Theorem:
        i. $np = c \in (0, \infty)$ ➔ $\Delta(G) \approx \log n \,/ \log^2 n$
        ii. $np \gg \log n$ ➔ $\delta(G) \approx \Delta(G) \approx np$
            1. The proof uses Chernoff inequality.
            2. Convergence in probability: $\frac{\delta(G)}{np}, \frac{\Delta(G)}{np} \to^P 1$

Connectivity

19. What is the smallest p s.t. $G(n,p)$ is connected?
    a. If $np = \log n - \omega(n)$ then $X_0 > 0$ w.h.p ➔ ∃isolated nodes ➔ disconnected.
        i. Otherwise ➔ no isolated nodes. Non-trivially, in this case we also have connectivity w.h.p. That is, dis-connectivity comes only form isolated nodes.
20. **Theorem (Erdos & Renxi, 1959)**: $\boldsymbol{\lim_n P(connected) = \begin{cases} 1 & np = \log n + \omega(n) \\ e^{-e^{-c}} & np = \log n + c \\ 0 & np = \log n - \omega(n) \end{cases}}$
    a. Also if $np = \log n + c$: $\boldsymbol{n_{components} \approx 1 + X_0 \to Poisson(e^{-c})}$ (i.e. 1 large component + X_0 isolated nodes).
    b. The proof uses Cayley's formula for the number of spanning trees on k nodes (k^{k-2}), and calculates the probability of each spanning tree (which is essentially a connected graph).
21. **Theorem – connected component size**: for $np = c \in (0, \infty)$ (i.e. the "disconnected limit"):
    a. If $c < 1$: all connected components are of size $O(\log n)$.
    b. If $c = 1$: largest component is $\approx n^{2/3}$.
    c. If $c > 1$: ∃ a component of $\approx \left(1 - \frac{x^*}{c}\right) n$ nodes, and the other components are $O(\log n)$.
        i. $xe^{-x} = ce^{-c}$ has one solution $x = c$ and one $x < 1$. x^* is the smaller solution.
        ii. The proof shows that w.h.p, no components exist of size $[a \log n, bn]$; and that only $\frac{x^*}{c} n$ nodes reside on smaller components; thus all the other nodes are in larger components.
        iii. Then the proof shows that this "giant" component is unique. Note that if $bn > 0.5n$, the uniqueness is trivial since there are only n nodes. In general, we can assume multiple giant components at $c_1 = c - \epsilon > 1$, and show that when adding components to move from c_1 to c, then w.h.p we connect all the giant components together.

Local limits

22. What do we see when we explore a random graph – from an arbitrary node v, up to a distance r?
23. ***Pointed graph***: (G, v^*) (for some $v^* \in V$); ***Neighborhood*** $[(G, v^*)]_r$: all nodes of distance $\leq r$ from v^*, along with the edges that connect them to v^*.
24. ***Plane tree***: labeling the nodes, subtrees & neighborhoods in a tree according to the corresponding paths from the root.



   a. ***Galton-Watson Tree***: a plane tree with a distribution $\pi = (\pi_0, \pi_1, \dots)$, s.t. for each node in the tree, its number of children is drawn according to π. It is required that $\sum_{i=1}^{\infty} \pi_i = 1$ and $\pi_1 \leq 1$ ($\pi_1 = 1$ is the trivial case of an infinite path graph).
   b. **Theorem – finiteness of GW trees**: let $T \sim GW_\pi$, and denote $z_0 \coloneqq \min\{z \mid \sum_{k=0}^{\infty} \pi_k z^k = z\} \leq 1$. Then $P(|T| < \infty) = z_0$ and $[z_0 < 1 \iff E[\pi] = \sum_{k=0}^{\infty} k\pi_k > 1]$.
      i. The last claim relies on a straightforward function exploration of $\sum_{k=0}^{\infty} \pi_k z^k - z$.

25. ***Graph isomorphism*** $\boldsymbol{G_1 \cong G_2}$: a bijection (one-to-one & onto) $f: V_1 \to V_2$ with $(u,v) \in E_1 \Leftrightarrow \big(f(u), f(v)\big) \in E_2$. That's essentially a relabeling of the nodes.
   a. ***Pointed-graph isomorphism***: also require $f(v_1^*) = v_2^*$.
26. **Theorem – *local limit* of a pointed graph**: let $G \sim G(n,p)$ with $np = c$. Then the r-neighborhood satisfies $[(\boldsymbol{G}, \boldsymbol{v}^*)]_{\boldsymbol{r}} \xrightarrow{\boldsymbol{n \to \infty}} \boldsymbol{GW_\pi}$, where $\pi = Poisson(c)$.
   a. Note that $E[\pi] = c \leq 1$, thus $P(|T| < \infty) = 1$ and $[(G, v^*)]_r$ is indeed finite.
   b. The formal way to phrase the limit is through graph isomorphism – for $T \sim GW_\pi$ and $T^* = (T, u^*)$: $\forall \boldsymbol{t} \in \boldsymbol{trees}$: $\boldsymbol{P}([(\boldsymbol{G}, \boldsymbol{v}^*)]_{\boldsymbol{r}} \cong \boldsymbol{t}) = \boldsymbol{P}([\boldsymbol{T}^*] \cong \boldsymbol{t})$.
   c. Conclusion: w.h.p, $[(G, v^*)]_r$ is a tree $\left(\sum_{t \in trees} P([(G, v^*)]_r \cong t) \to 1\right)$.

Cycles

27. Let G and denote by $\boldsymbol{\gamma_k}$ the **number of *Hamiltonian cycles*** of size k within G.
   a. Note: a Hamiltonian cycle forbids repetition of nodes, i.e. $|E| = |V| = k$.
   b. Example: in the clique K_n, there are $\binom{n}{k}$ subsets of k nodes, each with $k!$ different cycles up to 2 reflections and k cyclic shifts, thus $\gamma_k = \binom{n}{k}(k-1)!/2$.
28. **Theorem**: let $G \sim G(n,p)$ with $np = c$. Then $\boldsymbol{\gamma_k} \xrightarrow{\boldsymbol{n \to \infty}} \boldsymbol{Poisson}\big(\boldsymbol{c^k/2k}\big)$ in distribution.
   a. **Total cycles**: $\Gamma \coloneqq \sum_{k=3}^{n} \gamma_k$.
      i. Claim: $E[\Gamma] \xrightarrow{n \to \infty} \Lambda_c \coloneqq \frac{1}{2}\left(\log\frac{1}{1-c} - c - c^2/2\right)$ (assuming $c < 1$).
      ii. In fact, $\Gamma \to Poisson(\Lambda_c)$ (was not proved).

29. **Theorem** (*Erdos-Renyi*, 1960) – **probability of having no cycles**: let $G \sim G(n,p)$ with $np = c$. Then $\lim_{n\to\infty} P(G\ is\ acyclic) = \begin{cases} \sqrt{1-c}e^{\frac{c}{2}+\frac{c^2}{4}} & c < 1 \\ 0 & c \geq 1 \end{cases}$.
30. **Theorem** (**Hamiltonian cycle of size $\boldsymbol{k = n}$**): let $G \sim G(n,p)$, $np = \log n + \log^2 n + c(n)$. Then $\lim_{n\to\infty} P(\gamma_n = 1) = \lim_{n\to\infty} P(\delta(G) \geq 2) = \begin{cases} 1 & c(n) \to \infty \\ e^{-e^{-c}} & c(n) \to c \in (0,\infty) \\ 0 & c(n) \to -\infty \end{cases}$.
    a. Note that necessarily $\gamma_n \in \{0,1\}$, and that $\gamma_n = 1$ requires connectivity (hence $np \geq \log n$) and min degree $\delta \geq 2$.
    b. The ranges don't cover the whole R – possibly a mistake.
    c. Only the first case ($c(n) \to \infty$) was proven.

Other topics

31. ***Diameter*** = longest distance in the graph.
32. **Claims**:
    a. $np \gg \log n$ ➔ $diam(G) \approx \log n \,/\, \log np$.
    b. $np = [n \log n^2 - \log c]^{1/d}$ ➔ $diam(G)$ decreases with c from $d+1$ to d.
33. ***Independent set***: a subset of nodes that contains no edges. $\boldsymbol{\alpha(G)} \coloneqq |\text{largest independent set}|$.
34. **Claim**: $np = c \cdot n$ ➔ $\alpha(G) \approx 2\log_{1/(1-p)} n$.
    a. ***Chromatic number*** ($\boldsymbol{\chi(G)}$): minimal number of colors that allows coloring s.t. neighbors always have different colors.
    b. **Corollary**: $\chi(G) \approx \frac{n}{2\log_{1/(1-p)} n}$.