8500 COMBO TOPICS: RANDOM GRAPHS AND CELL COMPLEXES

MATTHEW KAHLE

Lecture 1: Introductions

Lecture 2: The Probabilistic Method

Lecture 3: Erdős-Rényi Model and Thresholds

Goals:

(1) Introduce a model of random graphs
(2) Basic properties of the model
(3) Thresholds

Definition 1 (Erdős-Rényi Random Graph)**.** An Erdős-Rényi Random Graph $G(n,p)$ has labeled vertex set $[n]$ an each of the $\binom{n}{2}$ possible edges has probability p of appearing, where each edge is jointly independent.

Definition 2 (Jointly Independent)**.** Events are said to be jointly independent if every event is independent of any intersection of the other events.

Definition 3 (Pairwise Independent)**.** Events are said to be pairwise independent if any two events are independent.

Example. Consider rolling two die. Consider the three events of A: the sum is 7, B: die 1 is a 3, and C: die 2 is a 4. Indeed, A and B are independent, A and C are independent, and B and C are as well. However, A, B, and C are not jointly independent since if B and C occur, A has a probability of 1.

Note. The number of edges of an Erdős-Rényi random graph is a binomial random variable. That is, an edge is binary; it is either there or it isn't. The probability of a certain number of edges appearing in $G(n,p)$ is given by the following formula:

$$\mathbb{P}[\#\text{edges} = x] = \binom{N}{x} p^x (1-p)^{N-x}$$

where $N = \binom{n}{2}$.

Models other than the Erdős-Rényi random graph exist. In particular, this models assumptions do not create psychically realistic models for

(1) social networks,
(2) the Internet, or
(3) epidemiology.

It is, however, good for percolation theory in statistical physics.

Technique (2-round Exposure i.e. "Sprinkling")**.** The union of two E.R. graphs $G(n,p_1)$ and $G(n,p_2)$ on the same vertex set has the same distribution as $G(n,p')$ where $p' = p_1 + p_2 - p_1 p_2$.

We can think about this as a way to maintain independence. Usually, p_2 is much smaller than p_1, so we are "sprinkling" a few more edges into the graph. The probability p' is a result of the formula for unions: it is the sum of the probability of the two events minus the probability that they both occur.

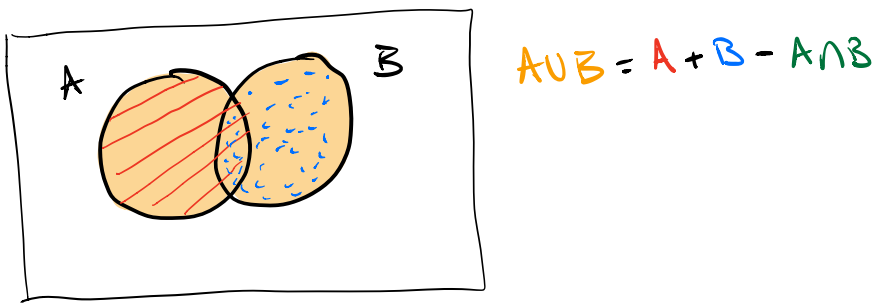


FIGURE 1. A measure theoretic sketch of p'.

Technique (m-round Exposure)**.** The union of m copies of $G(n,p)$ has the same distribution as $G(n,p')$ where $p' = 1-(1-p)^m$.

This p' gives us the probability that we hit an edge at least once; it is the complement of not hitting an edge at all. The complement gives us $1-$, and not hitting the edge at all gives us $(1-p)^m$.

Formula.

$$1-(1-p)^m \leq mp \qquad \text{for } p \in [0,1].$$

One could check this a few ways. If you expand the left with the binomial formula, you will see that it must be less than mp. Alternatively, you could think of mp as the union upper bound of a set of events occurring, as if you added up the probability but did not subtract any of the intersections.

Note. The book Random Graphs is an excellent reference.

We almost always consider the behavior of these Erdős-Rényi graphs as $n \to \infty$. So indeed we are considering a sequence of probability spaces that depend on n. Some texts include this subscript, but by and large, everyone omits it. We often think of the probability p as depending on n. That is, p is a function of n; $p(n)$.

Definition 4 (Asymptotically Almost Surely/With High Probability)**.** We say that something occurs asymptotically almost surely/with high probability (a.a.s./w.h.p) if the probability goes to 1 as n goes to infinity.

Proposition 1. *Let $p = n^\alpha$ for $\alpha \leq 0$. If $\alpha < -1/2$, then w.h.p., $G(n,p)$ contains no K_5 subgraphs. If $\alpha > -1/2$, then w.h.p., $G(n,p)$ contains K_5 subgraphs.*

Note. This proposition will be proved later, and can be sharpened.

Now that we have our model for random graphs, we can talk about properties.

Definition 5 (Graph Property)**.** A graph property is a set of graphs of vertex set $[n]$.

Definition 6 (Monotone Graph Property)**.** A graph property is monotone increasing if it is closed under the addition of edges. It is monotone decreasing if its complement is an increasing property. [n].

Example. (1) "Has at least 15 edges" is monotone increasing
(2) "Is connected" is monotone increasing
(3) "Contains a K_5 subgraph" is monotone increasing
(4) "No graphs/all graphs" is both monotone increasing and decreasing
(5) "Is planar" is monotone decreasing

Non-Example. "Has an even # of edges."

Definition 7 (Convex Property)**.** A graph property is convex if it is closed between two graphs in inclusion of edge sets. This is equivalent to saying that it is the intersection of a monotone increasing and a monotone decreasing property.

Example. (1) "Is a tree" is the intersection of being connected (increasing) and having no cycles (decreasing)
(2) Contains a K_3 subgraph (increasing) but no K_4 (decreasing)
(3) Is 3-connected (increasing) and planar (decreasing)

Definition 8 (Threshold)**.** A function $\overline{p}(n)$ is a threshold for a property $\mathcal{P}$ if whenever
(1) $\lim_{n\to\infty} p(n)/\overline{p}(n) = \infty$ then w.h.p. $G(n,p) \in \mathcal{P}$ and
(2) $\lim_{n\to\infty} p(n)/\overline{p}(n) = 0$ then w.h.p. $G(n,p) \notin \mathcal{P}$.

Note. This definition of threshold is somewhat coarse. Indeed, if $\overline{p}$ is a threshold for $\mathcal{P}$ then so is $c\overline{p}$ for $c > 0$. In other words, thresholds are only defined up to a constant factor.

Theorem 1. *Every nontrivial monotone graph property (not all/no graphs) has a threshold.*

Note. Convex properties can have two thresholds.

Proof. Let $\mathcal{P}$ be a non-trivial monotone increasing graph property. Then $\mathbb{P}[G(n,p) \in \mathcal{P}] = 0$ when $p = 0$ and $\mathbb{P}[G(n,p) \in \mathcal{P}] = 1$ when $p = 1$. Moreover, we claim that $\mathbb{P}[G(n,p) \in \mathcal{P}]$ is a continuous and strictly increasing function.

(Continuous) For each graph, the probability is given by the binomial distribution formula

$$p^e(1-p)^{\binom{n}{2}-e}.$$

This is polynomial in p, so the sum of these polynomials is continuous and smooth!

(S. Increasing) **TODO: Exercise**

Now given $0 < a < 1$, let $p(a)$ denote the unique p such that $\mathbb{P}[G(n,p) \in \mathcal{P}] = a$. For $0 < \varepsilon < 1/2$, let $p = p(\varepsilon)$. Let $m \in \mathbb{Z}$ be large enough so that $(1-\varepsilon)^m < \varepsilon$.

Let $p' = 1-(1-p)^m$. Then $G(n,p')$ has the same distribution as the union of m independent copies of $G(n,p)$. Since $1-(1-p)^m < mp$ by this formula, and $\mathcal{P}$ is an increasing property, we have that

$$\mathbb{P}[G(n,p') \in \mathcal{P}] \leq \mathbb{P}[G(n,mp) \in \mathcal{P}].$$

On the other hand,

$$\mathbb{P}[G(n,p') \notin \mathcal{P}] \leq (1-\varepsilon)^m < \varepsilon,$$

so $\mathbb{P}[G(n,p') \in \mathcal{P}] > 1-\varepsilon$. Now we have that $p(1-\varepsilon) < mp(\varepsilon)$.

Putting it all together, we see that

$$p(\varepsilon) < p(1/2) < p(1-\varepsilon) < mp(\varepsilon).$$

Since m depends on ε but not on n, and since ε can be made arbitrarily small, this implies that $p(1/2)$ is a threshold for $\mathcal{P}$. $\square$

Definition 9 (Sharp)**.** We say that $\overline{p}$ is a <u>sharp</u> threshold for P if for every fixed $\varepsilon > 0$, we have the following:

(1) If $p \leq (1-\varepsilon)\overline{p}$, then $\mathbb{P}[G(n,p) \in \mathcal{P}] \to 0$ as $n \to \infty$ and
(2) if $p \geq (1+\varepsilon)\overline{p}$, then $\mathbb{P}[G(n,p) \in \mathcal{P}] \to 1$ as $n \to \infty$.

Example. (1) Let $m \geq 2$ be fixed. Then $p = n^{-2/(m-1)}$ is a coarse threshold for the containment of K_m subgraphs. If $p = cn^{-2/(m-1)}$, where $c > 0$ is fixed, then

$$\mathbb{P}[G(n,p) \text{ contains a } K_m \text{ subgraph}] \to 1 - e^{-c^{\binom{m}{2}}/m!} \quad \text{as } n \to \infty.$$

(2) In general, for a "nice" enough graph H with v vertices and e edges, there is a coarse threshold for H subgraphs at $\overline{p} = n^{-v/e}$.
(3) Erdős and Rényi's sharp threshold for connectivity. Let $\overline{p} = \log(n)/n$ and let $\varepsilon > 0$. If $p \leq (1-\varepsilon)\overline{p}$, then w.h.p. $G(n,p)$ is disconnected. If $p \geq (1+\varepsilon)\overline{p}$, then w.h.p. $G(n,p)$ is connected.
(4) An even sharper threshold for connectivity: if $p = \frac{\log(n)+c}{n}$, where $c \in \mathbb{R}$ is fixed, then

$$\mathbb{P}[G(n,p) \text{ is connected}] \to e^{-e^{-c}} \quad \text{as } n \to \infty.$$

(5) The birth of the giant component. Let $p = c/n$ where $c > 0$ is fixed.
   (a) If $c < 1$ then w.h.p. the largest connected component has order $\Theta(\log(n))$.
   (b) If $c = 1$ then w.h.p. the largest connected component has order $\Theta(n^{2/3})$.
   (c) If $c > 1$ then w.h.p. the largest connected component has order $\Theta(n)$.
(6) The appearance of cycles is a one-sided sharp threshold. Let $p = c/n$ where $c > 0$ is fixed.
   (a) If $c < 1$ then $\mathbb{P}[\text{there exist no cycles}] \to \sqrt{1-c}e^{c/2+c^2/4}$.
   (b) If $c > 1$ then w.h.p. there exist cycles.
   (c) If $c = 1$ then $\mathbb{P}[\text{there exist cycles}] \approx Cn^{-1/6}$, where C is an explicitly defined constant:

$$C = \frac{1}{\sqrt{2\pi}} e^{3/4} \int_0^\infty e^{-4t^{3/2}} \cos(4t^{3/2} dt.$$

Lecture 4: Linearity of expectation

Definition 10 (Expectation)**.** If X is a discrete random variable taking non-negative integer values, the <u>expectation of X</u> is given by

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i\mathbb{P}[X = i].$$

Note. The expectation of a random variable need not exist.

Note. In this lecture, we'll mainly be working with discrete random variables taking non-negative integer values, so I'll omit the qualifications and write only "random variable" to save time.

Proposition 2. *Let A and B be random variables. When the expectation exists for all random variables below, we have* <u>*linearity of expectation*</u>*:*

$$\mathbb{E}[A] + \mathbb{E}[B] = \mathbb{E}[A + B]$$
$$\mathbb{E}[cA] = c\mathbb{E}[A].$$

Note. This is true whenever the expectations in question exist, not only in the case of independent random variables.

Example. Say you roll a six-sided die until every number from one to six comes up at least once. The expected wait time (number of rolls it takes) turns out to be $\sum_{i=1}^{6} \frac{6}{i} = 14.7$.

To understand this, consider a simpler scenario: say you flip a coin which has probability p of coming up heads. Then intuitively, we expect to wait about $1/p$ flips until we have seen heads. To prove this, we need a short lemma:

Lemma. If $-1 < x < 1$, we have

$$1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2}.$$

Proof. From our understanding of geometric series,

$$f(x) := \frac{1}{1-x} - 1 = x + x^2 + \cdots .$$

Taking a derivative,

$$f'(x) = \frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \cdots ,$$

as desired. □

Now, note that the probability of seeing heads on the nth flip is $(1-p)^{n-1}p$, so the expected wait time to see heads is

$$p + 2(1-p)p + 3(1-p)^2 p + \cdots = \frac{p}{p^2} = \frac{1}{p}.$$

Now we can return to the question about dice. If you've hit m different numbers, the probability you hit a number you haven't hit before is $\frac{6-m}{6}$, so the expected wait time is $\frac{6}{6-m}$. Using linearity of expectation to sum over m, we obtain the desired result.

This is an example of the "coupon collecting problem", in which there are n distinct coupons, and we ask the expected wait time to collect all n. The expectation is

$$n\left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}\right) \approx n \log n.$$

Definition 11 (Permutations and fixed points.)**.** A <u>permutation</u> in S_n is a bijection $[n] \to [n]$. A <u>fixed point</u> of a permutation σ is some i such that $\sigma(i) = i$.

Example. What is the expected number of fixed points of a permutation in S_n chosen uniformly at random?

Let X be a random variable, the number of fixed points in a uniform random permutation in S_n. If X_i is a random variable taking 1 when i is a fixed point and 0 otherwise, then

$X = \sum_{i=1}^{n} X_i$, and we have a decomposition of X as a sum of what we call indicator random variables.[1]

Now, using linearity of expectation and the fact that $\mathbb{E}[X_i] = 1/n$ for all i, we can compute that $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = 1$.

Notation. We say $f \ll g$ or $f = o(g)$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. We say $f \approx g$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

Example. We now provide a few comments on the expected number of triangle (K_3) subgraphs of $G(n,p)$. For a given set of three vertices in $G(n,p)$, the expected number of triangles is p^3. Summing over the $\binom{n}{3}$ such sets of three vertices, the expectation we want is $\binom{n}{3}p^3 \approx \frac{n^3p^3}{6}$.

Proposition 3. *If $p \ll \frac{1}{n}$, then $\mathbb{E}[\#$ of K_3 subgraphs$] \to 0$ as $n \to \infty$.*

Proof. From above, $\mathbb{E}[\#$ of K_3 subgraphs$] \approx \frac{n^3p^3}{6}$. If $p \ll n^{-1}$, $p^3 \ll n^{-3}$, so $n^3p^3 \to 0$ as $n \to \infty$. □

Corollary 1. *If $p \ll \frac{1}{n}$, $\mathbb{P}[\exists K_3$ subgraphs$] \to 0$ as $n \to \infty$.*

Proof. The proof uses Markov's inequality below. □

Theorem 2 (Markov's inequality)**.** *If X is a random variable taking non-negative integer values and a is a non-negative real number,*

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Proof.

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i\mathbb{P}[X = i] \geq \sum_{i=a}^{\infty} i\mathbb{P}[X = i] \geq \sum_{i=a}^{\infty} a\mathbb{P}[X = i] = a\sum_{i=a}^{\infty} \mathbb{P}[X = i] = a\mathbb{P}[X \geq a].$$

□

Corollary 2. *If X_n is a sequence of non-negative integer-valued random variables such that $\mathbb{E}[X_n] \to 0$, $\mathbb{P}[X_n > 0] \to 0$.*

Returning to the questions of K_3 subgraphs of $G(n,p)$, we can also ask a related question: if $p \gg \frac{1}{n}$, does $\mathbb{P}[\exists K_3$ subgraphs$] \to 1$? The answer turns out to be yes, but not because of expectation.

In general, we can ask: if $\mathbb{E}[X_n] \to \infty$, does $\mathbb{P}[X_n > 0] \to 1$? The answer is no!

Say X_n is a random variable taking the value n^2 with probability $1/n$ and the value 0 with probability $1 - 1/n$. Then, $\mathbb{E}[X_n] = n \to \infty$ and $\mathbb{P}[X_n = 0] \to 1$.

Example. What about K_4 subgraphs of $G(n,p)$? Note that $\mathbb{E}[\#$ of K_4 subgraphs$] = \binom{n}{4}p^6$. We can prove that if $p \ll n^{-2/3}$, w.h.p. there are no K_4 subgraphs. It's also true that $p = n^{-2/3}$ is the threshold for the appearance of K_4 subgraphs.

Note. Union bounds can also be used to show that there are no subgraphs of a certain type:

$$\mathbb{P}[A \text{ or } B] \leq \mathbb{P}[A] + \mathbb{P}[B].$$

[1] This is the only idea in the whole subject.

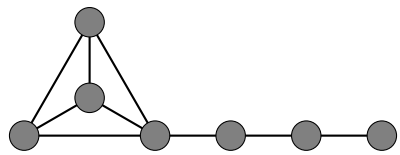


FIGURE 2. The Graph H (credit to Dan).

If $m = \binom{n}{4}$ and $E_1, E_2, \ldots, E_m$ are the events of all possible K_4 subgraphs, $\mathbb{P}[E_i] = p^6$ for all i, so by a union bound, the probability that any of them happen is $\binom{n}{4}p^6$.

Example. We denote by C_m the cycle on m vertices. What is the expected number of C_8 subgraphs in $G(n,p)$?

There are $\frac{n(n-1)\cdots(n-7)}{16}$ C_8 subgraphs in K_n: there are $n(n-1)\cdots(n-7)$ ways to choose the vertices, and we count each cycle 16 times. Thus, $\mathbb{E}[\#\text{ of } C_8 \text{ subgraphs}] = \frac{8!}{16}\binom{n}{8}p^8 \approx \frac{n^8p^8}{16}$.

A good guess for the threshold of the appearance of C_8 subgraphs is $p = n^{-1}$, which is true.

Example. Consider the graph H above. The expected number of H subgraphs in $G(n,p)$ is $\approx \frac{n^7p^9}{6}$. We might think the threshold for the appearance of H subgraphs is $p = n^{-7/9}$, but it's actually $p = n^{-2/3}$.

LECTURE 5: THE SECOND ~~MITTEN~~ MOMENT METHOD.

(Chapter 4 in The Probabilistic Method)

Motivation: we'd like to find some nice way of being able to prove that a variable is nonzero with high probability.

Definition 12. The variance of a random variable, X is defined by

$$\mathrm{Var}[X] := \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

or, using linearity of expectation:

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

"Measures how spread out a variable is".

Definition 13. The standard deviation of X is defined by $\sigma := \sqrt{\mathrm{Var}[X]}$

Notation. We sometimes denote the expectation, $\mu = \mathbb{E}[X]$.

The second moment method is the application of:

Theorem 3 (Chebyshev's Inequality)**.** *For any random variable X, we have*

$$\mathbb{P}\left[|X - \mu| > \lambda\sigma\right] < \frac{1}{\lambda^2}.$$

For $\lambda \geq 1$.

Proof. We have:

$$\sigma^2 = \mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2] \geq \lambda^2\sigma^2\mathbb{P}\left[|X - \mu| > \lambda\sigma\right].$$

Where the last step is an application of Markov's inequality to the random variable $(X-\mu)^2$ with $a = \lambda^2\sigma^2$. $\square$

Computing The Variance. Given a decomposition of a random variable,

$$X = X_1 + X_2 + \cdots + X_m.$$

Then,

$$\mathrm{Var}[X] = \sum_{i=1}^{m} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j].$$

Where the second sum is over ordered pairs, (i, j).

Definition 14. We define the <u>covariance</u> of two random variables as:

$$\mathrm{Cov}[Y, Z] := \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z].$$

"Measures how correlated Y and Z are."

Note. If Y, Z are independent, then $\mathrm{Cov}[Y, Z] = 0$.

Theorem 4.

$$\mathbb{P}[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} = \frac{\sigma^2}{\mu^2}.$$

Proof. Set $\lambda = \mu/\sigma$, then apply Chebyshev:

$$\mathbb{P}[X = 0] \leq \mathbb{P}\left[|X - \mu| > \lambda\sigma\right] < \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}.$$

□

Corollary 3. *If* $\mathrm{Var}[X] = o(\mathbb{E}[X]^2)$, *then* $X > 0$ *with high probability.*

And, in fact:

Corollary 4. *If* $\mathrm{Var}[X] = o(\mathbb{E}[X]^2)$, *then* $X \approx \mathbb{E}[X]$ *w.h.p. Intuitively, we mean, "X is concentrated around its mean." And rigorously, for every* $\varepsilon > 0$, *we have that*

$$\mathbb{P}\left[1 - \varepsilon < \frac{X}{\mathbb{E}[X]} < 1 + \varepsilon\right] \to 1 \qquad \text{as } n \to \infty.$$

Suppose now that $X = X_1 + \cdots + X_m$ is the sum of indicator random variables for the event A_i.

Notation. For indices i, j write $i \sim j$, if $i \neq j$ and the events A_i, A_j are not independent.

Definition 15. We define the interaction terms between non-independent events as:

$$\Delta = \sum_{i \sim j} \mathbb{P}[A_i \cap A_j].$$

Then,

$$\mathrm{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X_i X_j] = \mathbb{P}[A_i \cap A_j].$$

And further, $\mathrm{Cov}[X_i, X_j] = 0$ when $i \neq j$ and $i \nsim j$, so:

Proposition 4.

$$\mathrm{Var}[X] \leq \mathbb{E}[X] + \Delta.$$

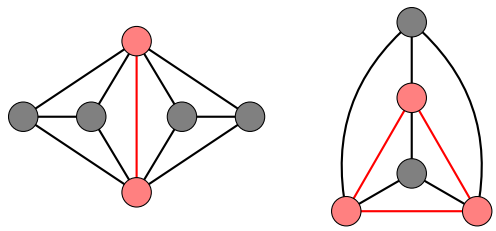


FIGURE 3. Cases of 2 and 3 vertex overlap.

Corollary 5. *If $\mathbb{E}[X] \to \infty$ and $\Delta = o(\mathbb{E}[X]^2)$, then $X > 0$ w.h.p. Moreover, $X \approx \mathbb{E}[X]$ w.h.p.*[2]

Example. Let $X = \#\{K_4 \subset G(n,p)\}$. We know that

$$E[X] = \binom{n}{4} p^6.$$

And if $p \ll n^{-2/3}$, then w.h.p. there are no K_4 subgraphs (from last class).

Now assume that $p \gg n^{-2/3}$ so that $\mathbb{E}[X] \to \infty$ (we want to show that this is a threshold). Write $X = X_1 + \cdots + X_m$, $m = \binom{n}{4}$, where X_i are indicator random variables for each possible occurence of a K_4. We want to compute $\Delta = \sum_{i \sim j} \mathbb{P}[A_i \cap A_j]$, where the sum is over non-independent events.

Question. How can two 4 vertex subsets of V correspond to non-independent events?

Answer. If they overlap!

We consider the cases where they do overlap:

- 1 Vertex of overlap.
  Then they are actually independent events, since they share no edges (which are the only random events).
- 2 vertices overlap.
  Then the contribution to Δ is $O(n^6 p^{11})$, since there are 6 vertices and there are $11 = 6 + 6 - 1$ edges in the union of the two graphs. (See Figure (3)).
- 3 vertices overlap.
  Then the contribution to Δ is $O(n^5 p^9)$, since there are 5 vertices and there are $9 = 6 + 6 - 3$ edges.
- 4 vertices overlap.
  Then they are actually the same graph, so they don't contribute to Δ—the sum is taken over distinct graphs.

Thus, $\Delta = O(n^6 p^{11} + n^5 p^9)$. We want to compare this to $\mathbb{E}[X]^2 = \Theta(n^8 p^{12})$. Since $p \gg n^{-2/3}$,

$$n^8 p^{12} \gg n^8 p^9 n^{-2} = n^6 p^9 > n^6 p^{11}, n^5 p^9.$$

Thus, in the regime, $p \gg n^{-2/3}$, we have $n^6 p^{11} = o(\mathbb{E}[X]^2)$ and $n^5 p^9 = o(\mathbb{E}[X]^2)$. So $\Delta = o(\mathbb{E}[X]^2)$, and $X \approx \mathbb{E}[X]$ w.h.p.

In general, we have the following:

[2]Note that the lack of n subscripts to measure the sequence of X distributions is convenient notationally here.

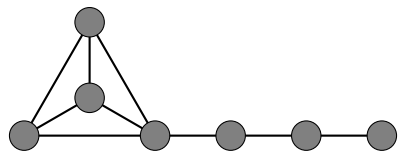


FIGURE 4. The Graph H.

Proposition 5. *For any fixed $m \geq 2$, $\bar{p} = n^{-2/(m-1)}$ is the threshold for the appearance of K_m subgraphs in $G(n,p)$.*

Note. For a K_m with vertex set V and edge set E, we can rewrite this threshold as:

$$n^{-2/(m-1)} = n^{-m/\binom{m}{2}} = n^{-|V|/|E|}.$$

Definition 16. For a graph H with ν vertices and e edges, we define the the <u>density</u> of H by $\rho(H) = e/\nu$.

Definition 17. A graph H is <u>balanced</u> if for every subgraph $H' \subset H$, we have that

$$\rho(H') \leq \rho(H).$$

"Every vertex carries more edges with it."

Theorem 5. *Let H be a balanced graph with ν vertices, e edges. Then the threshold for $G(n,p)$ containing H subgraphs is $n^{-\nu/e}$.*

Theorem 6. *Let H be a balanced graph with ν vertices, e edges and a automorphisms. If $p \gg n^{-\nu/e}$, then with high probability:*

$$\#\{H \subset G(n,p)\} \approx \frac{n^\nu p^e}{a}.$$

What happens for not balanced graphs? In general, the subgraph with the highest density determines the threshold.

Example. The graph H in Figure (4) has $\rho(H) = \frac{9}{7}$, but the density over all subgraphs is maximized by the K_4 subgraph, which has $\rho(K_4) = \frac{3}{2}$. Thus the threshold for H subgraphs is $n^{-2/3}$.

LECTURE 6: THE POISSON PARADIGM

For further notes and details, see Chapter 8 of Alon and Spencer.

Definition 18. Let $\mu > 0$. A <u>Poisson distribution</u> with mean μ is the probability distribution with law ν defined

$$\nu(\{k\}) = e^{-\mu}\frac{\mu^k}{k!}$$

for all integers $k \geq 0$. In other words, X is Poisson distributed when $\mathbb{P}[X = k] = e^{-\mu}\frac{\mu^k}{k!}$ for all integers $k \geq 0$. We will denote this by $X \sim \text{Pois}[\mu]$.

Poisson distributions appear in nature as the laws of occurrences of "rare events." For example, the number of emails your inbox receives from 3:00 - 4:00 PM on a Friday is roughly Poisson distributed, and you can estimate its mean by averaging how many emails you get during that hour each week.

Now that we've called this object $\mathrm{Pois}[\mu]$ a "distribution," we should back up our suggestive naming.

Proposition 6. *For any $\mu > 0$, $\mathrm{Pois}[\mu]$ is a probability distribution with mean μ.*

Proof. Let $X \sim \mathrm{Pois}[\mu]$. Using that X takes values only on nonnegative integers and abusing the letter $\mathbb{P}$, we observe that

$$\sum_{k=0}^{\infty} \mathbb{P}[X = k] = \sum_{k=0}^{\infty} e^{-\mu} \frac{\mu^k}{k!} = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{-\mu} e^{\mu} = 1$$

as we needed for $\mathrm{Pois}[\mu]$ to be a probability distribution. Calculating the mean now,

$$\sum_{k=0}^{\infty} k\mathbb{P}[X = k] = \sum_{k=1}^{\infty} k \left(e^{-\mu} \frac{\mu^k}{k!} \right) = e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu \cdot \mu^{k-1}}{(k-1)!} = \mu e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = \mu e^{-\mu} e^{\mu} = \mu$$

as we predicted. □

The above proof provides a way to remember the law for Poisson random variables: $e^{-\mu}$ followed by the k-th term of the power series expansion for the exponential e^{μ} to cancel out in the sum.

We now state a result which we will use several times to show that random variables are Poisson distributed.

Proposition 7. *Let $\mu > 0$ and let $X \sim \mathrm{Pois}[\mu]$. For any $r \geq 1$, we have*

$$\mathbb{E}\left[\binom{X}{r} \right] = \frac{\mu^r}{r!}.$$

Proof. We will prove the equivalent statement that

$$\mathbb{E}[X_{(r)}] := \mathbb{E}[X(X-1)\cdots(X-r+1)] = \mu^r$$

obtained by multiplying the desired result by $r!$. Rolling up our sleeves,

$$\begin{aligned}
\mathbb{E}[X(X-1)\cdots(X-r+1)] &= \sum_{k=0}^{\infty} k(k-1)\cdots(k-r+1) e^{-\mu} \frac{\mu^k}{k!} \\
&= e^{-\mu} \sum_{k=r}^{\infty} \frac{\mu^r \cdot \mu^{k-r}}{(k-r)!} \\
&= \mu^r e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} \\
&= \mu^r
\end{aligned}$$

by exactly the same trick we're now applying for the third time. □

One of the most useful features of Poisson distributions is showing that sequences of distributions converge to Poisson via the method of moments. Of course, to talk about this, we need to know what it means for a distribution to converge. Recall that for a real-valued random variable (a.k.a. rvrv) X, we may assign a <u>cumulative distribution function</u> $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = \mathbb{P}(X \leq x)$.

Definition 19. Suppose $\{X_n\}_n$ is a sequence of rvrv's with cumulative distribution functions $\{F_n\}_n$, and similarly for X and F. We say that X_n converges in law, converges in distribution, or converges weakly to X, denoted $X_n \to X$ in law etc., if for every $x \in \mathbb{R}$ at which F is continuous, we have $\lim_{n\to\infty} F_n(x) = F(x)$.

Note. The convergence only at points of continuity of F, which is only right-continuous, is important. To motivate this, consider the example $X_n \sim \text{Unif}((0, \frac{1}{n}))$ and $X = 0$ a.s.. Any sane person will tell you that $X_n \to 0$, however $F_n(0) = 0$ for all n while $F(0) = 1$, so $\lim_{n\to\infty} F_n(0) \neq 0$.

If $X \sim \text{Pois}[\mu]$ and $\{X_n\}$ is a sequence of integer-valud random variables, we see that $X_n \to X$ in law when $\mathbb{P}[X_n = k] \to e^{-\mu}\frac{\mu^k}{k!}$ for each integer $k \geq 0$.

Theorem 7 (Method of Moments)**.** *Let $\{X_n\}$ be a sequence of integer-valued random variables. Let $\mu > 0$. If*

$$\lim_{n\to\infty} \mathbb{E}\left[\binom{X_n}{r}\right] = \frac{\mu^r}{r!}$$

for all fixed $r \geq 0$, then $X_n \to \text{Pois}[\mu]$ in law. In other words, $\mathbb{P}[X_n = k] \to e^{-\mu}\frac{\mu^k}{k!}$ for all integers $k \geq 0$.

Example. For an integer $n \geq 1$ and $p \in [0, 1]$, define a binomial random variable $X \sim B(n, p)$ by letting

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

This is the number of heads appearing in n tosses of a biased coin which comes up heads with probability p. Fix $\mu > 0$. Now, define a sequence $\{X_n\}$ of binomial random variables $X_n \sim B(n, \mu/n)$ for $n > \mu^{-1}$. Notice that $\mathbb{E}[X_n] = \mu$ for each n, and further that for any r,

$$\mathbb{E}\left[\binom{X_n}{r}\right] = \binom{n}{r}\left(\frac{\mu}{n}\right)^r = \frac{n(n-1)\cdots(n-r+1)}{n^r}\frac{\mu^r}{r!} \to \frac{\mu^r}{r!}$$

meaning that $X_n \to \text{Pois}[\mu]$ in law by the method of moments.

Example. Let X_n be distributed as the number of fixed points in a uniformly randomly chosen permutation from S_n. We have for any r,

$$\mathbb{E}\left[\binom{X_n}{r}\right] = \binom{n}{r}\frac{(n-r)!}{n!} = \frac{1}{r!}$$

meaning that $X_n \to \text{Pois}(1)$ in law. As a result, for large n, the number of derangements in S_n is approximately equal to $n!\mathbb{P}(\text{Pois}(1) = 0) = n!/e$. This approximation is as good as it gets; the number of derangements is in fact the closest integer to $n!/e$.

Example. We have recently seen that $p = n^{-2/3}$ is the (coarse) threshold for asymptotic appearance of K_4 subgraphs in $G(n, p)$. In other words, if $p \ll n^{-2/3}$, then w.h.p. there are no K_4 subgraphs in $G(n, p)$, and if $p \gg n^{-2/3}$, then w.h.p. there are K_4 subgraphs in $G(n, p)$. So what happens when $p = cn^{-2/3}$? Note that, in this case, $\mathbb{E}[\text{number of } K_4 \text{ subgraphs}] = c^6/24$, so you might guess our next result.

Proposition 8. *Let X be the number of K_4 subgraphs in $G(n, p)$ and fix $c > 0$. If $p = cn^{-2/3}$, then $X \to \text{Pois}(c^6/24)$ in law.*

Sketch. Let's go ahead and write $\mu = c^6/24$. We want to get

$$\mathbb{E}\left[\binom{X}{r}\right] \to \frac{\mu^r}{r!}.$$

Let's try $r = 2$ first. We see that

$$\mathbb{E}\left[\binom{X}{2}\right] \approx \frac{1}{2}\binom{n}{4}\binom{n-4}{4}p^{12} \approx \frac{\mu^2}{2!}.$$

There are other terms coming from cases where the K_4 subgraphs intersect, but they are $O(n^6p^{11} + n^5p^9)$ and thus dominated by the n^8p^12 of the given term when $p = cn^{-2/3}$. As an exercise, you can see the dominant term for general $r \geq 1$:

$$\mathbb{E}\left[\binom{X}{r}\right] \approx \frac{1}{r!}\binom{n}{4}\binom{n-4}{4}\cdots\binom{n-4(r-1)}{4}p^{6r} \approx \frac{\mu^r}{r!}.$$

In each case, the disjoint K_4 subgraphs term dominates all terms where the graphs intersect. □

Corollary 6. *Let $p = cn^{-2/3}$, where $c > 0$ is fixed. Then*

$$\mathbb{P}[\textit{There are no } K_4 \textit{ subgraphs}] \to e^{-c^6/24}$$

as $n \to \infty$, and hence

$$\mathbb{P}[\textit{There is a } K_4 \textit{ subgraph}] \to 1 - e^{-c^6/24}.$$

The Poisson paradigm can give "concentration of measure"-type results, allwoing you to get much tighter bounds on how spread-out a distribution is than Chebshev's inequality would give.

Definition 20. The <u>normal distribution</u> $\mathcal{N}(0,1)$ is the real-valued probability distribution with probability density function $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. In general, the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2 has probability density $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$.

Proposition 9. *Let $X \sim \mathrm{Pois}[\mu]$ where $\mu \to \infty$. Then*

$$\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}[X]}} \to \mathcal{N}(0,1)$$

in law.

Lecture 7: The threshold for connectivity

Our main goal is to show that $p = \frac{\log n}{n}$ is a sharp threshold for connectivity of $G(n,p)$.

Question. Which graph properties have sharp thresholds and which have coarse thresholds?

The rule of thumb (made precise through various theorems) seems to be that "local properties" have coarse thresholds and "global properties" have sharp thresholds.

Theorem 8 (Erdős and Rényi)**.** *Let $\epsilon > 0$ be fixed.*

- *If $p \geq \frac{(1+\epsilon)\log n}{n}$ then w.h.p. $G(n,p)$ is connected.*
- *If $p \leq \frac{(1+\epsilon)\log n}{n}$ then w.h.p. $G(n,p)$ is disconnected.*

Furthermore, for any $c \in \mathbb{R}$ fixed, if $p = \frac{\log n + c}{n}$, then

$$\mathbb{P}[G(n,p) \text{ is connected}] \to e^{-e^{-c}}$$

as $n \to \infty$.

Useful Estimate. For every real x, we have $1-x \leq e^{-x}$. Moreover, if $x \approx 0$ then $1-x \approx e^{-x}$.

Proof. Calculus. □

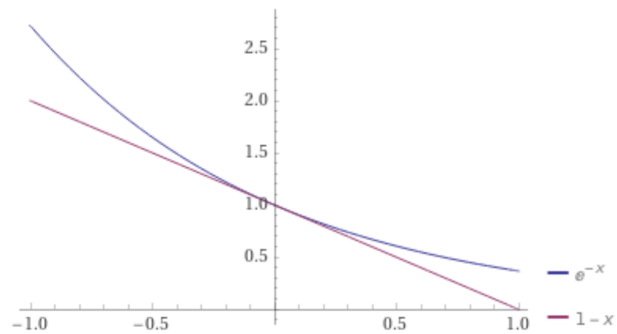


FIGURE 5. The curves $y = e^{-x}$ and $y = 1 - x$.

Useful Estimate. For every $n \geq 1$, we have

$$e\left(\frac{n}{e}\right)^n \leq n! \leq en\left(\frac{n}{e}\right)^n.$$

Proof. (Lower bound) Proof by induction on n. The case $n = 1$ is trivial. Assume $n! \geq e\left(\frac{n}{e}\right)^n$. Then by assumption

$$(n+1)! = (n+1)n! \geq (n+1)e\left(\frac{n}{e}\right)^n,$$

so it suffices to show that

$$(n+1)\left(\frac{n}{e}\right)^n \geq \left(\frac{n+1}{e}\right)^{n+1}.$$

But then this can be rearranged to

$$e \geq \left(\frac{n+1}{n}\right)^n = \left(1 + \frac{1}{n}\right)^n,$$

which can be obtained from the previous useful estimate by substituting $x = -1/n$. □

Proposition 10. *If $p \approx \frac{\log n}{n}$ then w.h.p. $G(n,p)$ consists of a unique giant component containing "most" of the vertices, together with isolated vertices.*

Note. This "structure result" holds once $p \geq \frac{\log n}{2n}$. The "giant component" emerges at $p = \frac{1}{n}$. By the time $p \approx \frac{\log n}{n}$, the giant has gobbled up all the other components except isolated vertices.

Proof. It suffices to show that there are no connected components of order k, for $2 \leq k \leq (2/3)n$ (We could replace $2/3$ with any $\lambda \in (1/2, 1)$). Let's consider first components of order 2. By a union bound, we have that

$$\begin{aligned}
\mathbb{P}[\text{there exists a component of order 2}] &\leq \binom{n}{2} p(1-p)^{2(n-2)} \\
&\leq \frac{n^2}{2} \frac{\log n}{n} \left(1 - \frac{\log n}{n}\right)^{2(n-2)} \\
&\leq \frac{n \log n}{2} \exp\left(\frac{-2(n-2)\log n}{n}\right) \\
&= O\left(\frac{\log n}{n}\right).
\end{aligned}$$

Now let's consider components of order k. By a union bound, we have that

$$\mathbb{P}[\exists \text{ a component of order } k] \leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k}.$$

This estimate uses Cayley's Theorem, which says that the number of spanning trees on k labeled vertices is k^{k-2}. Then it suffices to show that

$$\sum_{k=3}^{(2/3)n} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \to 0$$

as $n \to \infty$. We can bound

$$\begin{aligned}
\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} &\leq \frac{n^k}{k!} k^{k-2} p^{k-1} \exp(-pk(n-k)) \\
&\leq \frac{1}{ek^2 p} (enp)^k \exp(-pk(n-k)) \\
&\leq \frac{1}{ek^2 p} \left(\frac{enp}{\exp(p(n-k))}\right)^k.
\end{aligned}$$

Since $p \approx \frac{\log n}{n}$, we have

$$\frac{1}{ek^2 p} \left(\frac{enp}{\exp(p(n-k))}\right)^k \approx \frac{n}{ek^2 \log n} \left(\frac{e \log n}{n^{1-k/n}}\right)^k.$$

Now it suffices to show that

$$\sum_{k=3}^{(2/3)n} \frac{n}{ek^2 \log n} \left(\frac{e \log n}{n^{1-k/n}}\right)^k \leq \sum_{k=3}^{(2/3)n} n \left(\frac{e \log n}{n^{1-k/n}}\right)^k \to 0.$$

Note that

$$n \left(\frac{e \log n}{n^{1-3/n}}\right)^k \leq \frac{1}{n}$$

for large n and since $k \leq (2/3)n$, we have

$$\frac{e \log n}{n^{1-k/n}} \leq \frac{e \log n}{n^{1/3}}.$$

Set $a = 1/n$ and $r = \frac{e \log n}{n^{1/3}}$. Then we can bound the above series by the geometric series $a(1 + r + r^2 + \ldots) = \frac{a}{1-r}$, where $a \to 0$ and $r \to 0$. $\square$

Proposition 11. *If $p = \frac{\log n + c}{n}$, then $\mathbb{E}[\#$ isolated vertices$] \to e^{-c}$, and $\#$ isolated vertices $\to$ Pois$[e^{-c}]$. In particular, $\mathbb{P}[$there are no isolated vertices$] \to e^{-e^{-c}}$.*

Proof. For $r \geq 2$, we have

$$\begin{aligned}
\mathbb{E}\left[\binom{X}{r}\right] &= \binom{n}{r}\left(1 - \frac{\log n + c}{n}\right)^{(n-r)r+\binom{r}{2}} \\
&\approx \frac{n^r}{r!} \exp\left(-\frac{(\log n + c)rn}{n}\right) \\
&\to \frac{e^{-cr}}{r!}.
\end{aligned}$$

So if we set $\mu = e^{-c}$, we have $\mathbb{E}\left[\binom{X}{r}\right] \to \frac{\mu^r}{r!}$ for every fixed $r \geq 1$. Therefore, we conclude that $X \to \text{Pois}[\mu]$ in law. $\square$

Putting it all together, we get the following theorem.

Theorem 9. *If $p = \frac{\log n + c}{n}$ then $\mathbb{P}[G(n,p)$ is connected$] \to e^{-e^{-c}}$ as $n \to \infty$.*

Proof. By Proposition 10, $G(n,p)$ consists of a unique giant component and isolated vertices. By Proposition 11, $\mathbb{P}[$there are no isolated vertices$] \to e^{-e^{-c}}$. $\square$

Theorem 9 has since been sharpened in several ways, for example in the following results:

Theorem 10 (Bollobás and Thomason, 1985)**.** *Consider the random graph process where the edges are added uniformly at random one-at-a-time. With high probability, the stopping time for becoming connected is equal to the stopping time for having minimal degree one.*

Theorem 11 (Bollobás and Thomason, 1985)**.** *Consider the random graph process where the edges are added uniformly at random one-at-a-time. For each $k = 1, 2, \ldots, n-1$, with high probability the stopping time for becoming k-connected is equal to the stopping time for having minimal degree k.*

The first paper about the topology of random simplicial complexes described a generalization of Theorem 8 to higher dimensions. We now know of several higher-dimensional generalizations, and we will discuss some of these in class this semester. There are still some things we don't know, and one "missing" high-dimensional generalization of the Erdős-Rényi result is what we need to prove the "bouquet-of-spheres conjecture."

$G(n,p)$ also exhibits a "phase transition."

Theorem 12 (Erdős and Rényi)**.** *Let $p = \frac{c}{n}$, where $c > 0$ is fixed.*

- *If $c < 1$ then w.h.p. all connected components are of order $O(\log n)$.*
- *If $c > 1$ then w.h.p. there is a unique giant component of order $\Omega(n)$. All other components are of order $O(\log n)$.*
- *If $c = 1$ then w.h.p. the largest component has order $\Theta(n^{2/3})$.*

The higher dimensional picture of the phase transition is still poorly understood. However, a breakthrough paper by Linial and Peled sheds the first light on higher-dimensional analogues of the "giant component."

Lecture 8: Expander Graphs

Definition 21. Let $G = (V, E)$ be a graph. For a collection of vertices $A \subseteq V$ we define the boundary of A by

$$\partial A = \{(x, y) \in E \mid x \in A, y \notin A\}.$$

Definition 22. Let $G = (V, E)$ be a graph. We define the Cheeger number (or isoperimetric constant) of G by

$$h(G) = \min\{\frac{|\partial A|}{|A|} : A \subseteq V(G), 0 < |A| < |V(G)|/2\}.$$

Example. Let $G = K_n$. For any subset of vertices A with $|A| = k$, we have $\partial A = k(n-k)$, so

$$\begin{aligned} h(K_n) &= \min\{\frac{k(n-k)}{k} : 0 < k < n/2\} \\ &= \lfloor n \rfloor/2 \lfloor . \rfloor \end{aligned}$$

Example. Let $G = C_m$, the cycle of length m. For any nontrivial subset of vertices A, we have $|\partial A| \geq 2$, so

$$h(K_n) = \min\left\{\frac{|\partial A|}{|A|} : A \subseteq V(G), 0 < |A| < |V(G)|/2\right\} \leq 2/(m/2) = 4/m$$

Definition 23. A graph is k-regular if every vertex has degree k.

Definition 24. We say that a sequence of k-regular graphs $\{G_n\}$ is an expander family if the number of vertices $\#V(G_n) \to \infty$, and $\liminf_{n\to\infty} h(G_n) > 0$.

Definition 25. The adjacency matrix A of a graph G has rows and columns indexed by the set of vertices, and $A_{ij} = A_{ji} = 1$ if $v_i \sim v_j$ and $A_{ij} = A_{ji} = 0$ otherwise.

Note. The adjacency matrix A is symmetric, so it is diagonalizable. In particular, A has n real eigenvalues (counted with multiplicity). If G is a k-regular graph, then $k = \lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq -k$. The quantity $k - \lambda_2$ is called the spectral gap of G.

Definition 26. We say that a sequence of k-regular graphs $\{G_n\}$ is an expander family if the number of vertices $\#V(G_n) \to \infty$, and

$$\lim \inf_{n\to\infty} (k - \lambda_2(G_n) > 0$$

Note. The two definitions of expander family turn out to be equivalent. Indeed we have the following inequalities relating Cheeger number and spectral gap for a k-regular graph:

$$\frac{1}{2}(k - \lambda_2) \leq h(G) \leq \sqrt{2k(k - \lambda_2}.$$

(Due to Dodziuk and Alon-Milman, independently).

Definition 27. A Ramanujan graph is a k-regular graph satisfying $\lambda_2 \leq 2\sqrt{k-1}$

Theorem 13. *Let $G \sim G(n, p)$ where $p \geq \frac{(1+\varepsilon)\log n}{n}$. Then w.h.p. $\lambda_1 \approx 1$.*

Note. The Erdos-Renyi random graphs aren't typically "regular graphs" but they are really good expanders, in this sense.

Theorem 14. *Let $\tilde{G}$ denote the largest component of $G(n,p)$. Fix $\delta > 0$ and let $p \geq \left(\frac{1}{2}+\delta\right)\frac{\log n}{n}$. Set $d = p(n-1)$. For every $\varepsilon > 0$, there exists a constant $C = C(\delta,\varepsilon)$ such that $|1-\lambda_2(\tilde{G}) < \frac{C}{\sqrt{d}}|$ with probability at least $1 - C_n \exp(-2(2-\varepsilon)d) - C\exp(-d^{1/4}\log n)$.*

Lecture 9: Concentration of Measure

The Second Moment Method and Chebyshev's inequality allows us to show that any random variable (with mean and variance) is not likely to be too far from its mean. If $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$, then

$$\mathbb{P}[|X-\mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}$$

But in many situations, we can do much better than this. I.e. for many random variables that you might care about, the probability that it is far from its mean is much smaller than suggested by the second moment method.

Chernoff bounds: Suppose $X = X_1 + X_2 + \ldots + X_n$ is the sum of n independent and identically distributed (i.i.d.) random variables taking values in {0,1} i.e, $X = B(n,p)$ with $\mu = np$. Then for any $0 \leq \delta \leq 1$, we have

$$\mathbb{P}[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \leq \exp\left(-\delta^2\mu/2\right), \text{ and}$$
$$\mathbb{P}[X \geq (1+\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \leq \exp\left(-\delta^2\mu/3\right)$$

As a corollary, $\mathbb{P}[|X-\mu| \geq \delta\mu] \leq 2\exp\left(-\delta^2\mu/3\right)$.

Tail bounds for Poisson variables. Let $P \sim \mathrm{Pois}[\mu]$. For $\epsilon > 0$ we have,

$$\mathbb{P}[P \leq \mu(1-\epsilon)] \leq \exp\left(-\epsilon^2\mu/2\right), \text{ and}$$
$$\mathbb{P}[P \geq \mu(1+\epsilon)] \leq \left[e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}\right]^{\mu}.$$

Janson's inequalities. Let Ω be a finite universal set, and let R be a random subset of Ω given by $\mathbb{P}[r \in R] = p_r$, these events being mutually independent over R.

Let $\{A_i\}_{i\in I}$ be subsets of Ω, and I a finite index set.

Let B_i be the event $A_i \subseteq R$, let X_i be the indicator random variable for B_i, and set $X = \sum_{i\in I} X_i$.

Write $i \sim j$ if $i \neq j$ and $A_i \cap A_j \neq \varnothing$.

Define $\Delta = \sum_{i\sim j} \mathbb{P}\left[B_i \text{ and } B_j\right]$, where the sum is over ordered pairs.

We set $M = \prod_{i\in I} \mathbb{P}\left[\bar{B}_i\right]$, where $\bar{B}_i$ is the event that B_i does not happen.

Note that M is the probability that none of the B_i happen if they are all independent.

Finally, set $\mu = \mathbb{E}[X] = \sum_{i\in I} \mathbb{P}\left[B_i\right]$.

Janson's inequality: Assume that for every $i \in I$, $\mathbb{P}\left[B_i\right] \leq \epsilon$.

Then $M \leq \mathbb{P}$ [none of the B_i occur] $\leq M\exp[(1/(1-\epsilon))\Delta/2]$. Furthermore, $\mathbb{P}$ [none of the B_i occur] $\leq \exp(-\mu + \Delta/2)$.

Comment. This is useless if $\Delta \geq 2\mu$. In that case we should use the Extended Janson's Inequality.

Extended Janson's inequality. With the same assumptions as before, but with the additional assumption that $\Delta \geq \mu$, $\mathbb{P}$ [none of the B_i occur] $\leq \exp(-\mu^2/2\Delta)$.

Applications of concentration of measure to random graphs.

Proposition 12. *Let $p \gg \log n/n$. set $\mu = p(n-1)$. Then W.h.p. every vertex of $G(n,p)$ has degree $\approx \mu$.*
In other words, for every $\epsilon > 0$ we have, $\mathbb{P}[(1-\epsilon)\mu \leq \deg(v) \leq (1+\epsilon)\mu] \geq 1-\epsilon$, for sufficiently large n

Proof. Let $p \geq C \log n/n$.
Let v be a vertex in $G(n,p)$. We observe that $\deg(v) \sim B(n-1,p)$, and $\mu \geq C \log n$.
By Chernoff bounds, we have that

$$\mathbb{P}[|\deg(v) - \mu| \geq \epsilon\mu] \leq 2\exp\left(-\epsilon^2\mu/3\right) \leq 2n^{-\epsilon^2 C/3}$$

□

Proposition 13. *Let $\delta > 0$ and suppose $p \gg \log n/n$. Set $\mu = p(n-1)$. Then w.h.p. every vertex of $G(n,p)$ has degree $\approx \mu$.*
In other words, for every $\epsilon > 0$ we have, $\mathbb{P}[(1-\epsilon)\mu \leq \deg(v) \leq (1+\epsilon)\mu] \geq 1-\epsilon$, for sufficiently large n.

Proof. By Chernoff bounds, we have that

$$\mathbb{P}[|\deg(v) - \mu| \geq \epsilon\mu] \leq 2\exp\left(-\epsilon^2\mu/3\right) \leq 2n^{-\epsilon^2 C/3}$$

Since $\epsilon > 0$ is fixed and $C \to \infty$ by assumption, we have $\mathbb{P}[|\deg(v) - \mu| \geq \epsilon\mu] \ll n^{-1}$ for large enough n.
Then applying a union bound over all n vertices, the total probability that any vertex has degree too far from its mean is still tending to zero.

□

Proposition 14. *Let $\delta > 0$ be fixed and $G \sim G(n,p)$. If $p \leq \frac{n^{-\delta}}{n^{2/3}}$ then w.h.p. there exists a vertex of G not contained in any K_3 subgraph. If $p \geq \frac{n^{\delta}}{n^{2/3}}$ then* w · h · p· *every vertex of G is contained in a K_3 subgraph.*

Comment. We state it this way for simplicity, but the threshold is much sharper than this. We leave stating and proving a sharper statement as an exercise.

Proof. Let v be any vertex of $G(n,p)$.
We have $\mathbb{E}[\#\text{ triangles containing } v] = \binom{n-1}{2} p^3$.
So if $p \ll n^{-2/3}$ then $\mathbb{E}[\#] \to 0$ and w.h.p. v is not contained in any triangles, by Markov's inequality.
So suppose $p = n^{\delta - 2/3}$ for some $\delta > 0$.
Then $\mu = \mathbb{E}[\#\text{ triangles containing } v] \approx (1/2)n^{3\delta}$.
If I indexes all triangles containing vertex v, then $\Delta = \sum_{i \sim j} \mathbb{P}[B_i \text{ and } B_j] \approx (1/6)n^{(-1/3)+5\delta}$.
We may assume WLOG that $\delta > 0$ is small. Then $\Delta = o(1)$.
Applying Janson's Inequality, we have that
$\mathbb{P}[v \text{ not contained in any triangle }] \leq \exp(-\mu + \Delta/2) \leq \exp\left(-n^{3\delta + o(1)}\right)$.
Applying a union bound, summing over all n vertices, we see that the total probability that there exists a vertex not contained in a triangle is at most $n\exp\left(-n^{2\delta}\right) \to 0$. □

Lecture 10: Correlation Inequalities

Definition 28. Let $Q^n = \{0,1\}^n$ denote the n-dimensional cube.

We then let $\mathbb{P}(n)$ be the power set of $[n]$, i.e. the set of all subsets if $\{1, 2, ..., n\}$. It is easy to see that the elements of $\mathbb{P}(n)$ are in natural 1-1 correspondence with vertices of Q^n. A set $A \subset [n]$ correspodss with its characteristic function $1_A : S \to \{0,1\}$, where $1_A(x) = 1$ if $x \in A$ and 0 otherwise.

Notation. For $A, B \in \mathcal{P}(n)$, we write $A \leq B$ if $A \subset B$.

Lecture 11: Cycles in $G(n,p)$

This section is about some theory of cycles in random graphs.

Question. What is the threshold for the appearance of cycles?

Suppose $c > 0$ and $p = \frac{c}{n}$. Write $G \sim G(n,p)$. Recall that for a fixed balanced graph H, with v vertices, e edges and a automorphisms,

$$\mathbb{E}[\#H \text{ subgraphs in } G] \approx \frac{n^v p^e}{a}.$$

Since cycles are balanced,

$$\mathbb{E}[\#C_m] \approx \frac{n^m p^m}{2m} = \frac{c^m}{2m}.$$

Thus,

$$\mathbb{E}[\# \text{ cycles in total}] \to \frac{c^3}{6} + \frac{c^4}{8} + \frac{c^5}{10} + \dots.$$

Recall that $-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots$ for $0 \leq x < 1$. Therefore, for $0 \leq c < 1$,

$$\mu = \mathbb{E}[\# \text{ cycles}] \to \frac{1}{2}(-\log(1-c) - (c + c^2/2)).$$

It follows immediately that if $p = o(1/n)$, then w.h.p. there are no cycles, because for $p = \frac{c}{n}$ we have

$$\mu = \mathbb{E}[\# \text{ cycles}] = \frac{1}{2}(-\log(1-c) - (c + c^2/2)) \to 0, \text{ for } c \to 0.$$

If we already know there is a phase transition at $p = \frac{1}{n}$, then it is not hard to prove that there must be cycles if $p \geq \frac{c}{n}$ and $c > 1$. Recall Theorem 12 the Phase Transition Theorem.

Proposition 15. *Let $p = \frac{c}{n}$, where $c > 1$ is fixed. Then the largest component of $G(n,p)$ has order $\approx \lambda n$ with high probability, where $\lambda = \lambda(c)$ is a positive constant that only depends on c.*

Taking the above proposition for granted, let us see that there must by cycles w.h.p. for $c > 1$. Of course we know that the expected number of cycles is tending to infinity, but this approach will be easier than computing a second moment, etc.

Proposition 16. *If $p \geq \frac{c}{n}$ and $c > 1$ is fixed. Then w.h.p. $G(n,p)$ contains at least one cycle.*

Proof. Set $c = 1+\epsilon$, where $\epsilon > 0$ is fixed. Consider $G(n, p_1)$ where $p_1 = \frac{1+\epsilon/2}{n}$. By Proposition 15, w.h.p. there exists a connected component with λn vertices. If this component contains a cycle, we are done. Otherwise, this component is a tree.

Now we take the union of $G(n, p_1)$ and $G(n, p_2)$, where $p_2 = \frac{\epsilon/2}{n}$. The giant component in $G(n, p_1)$ has at least λn vertices, so we consider the $\binom{\lambda n}{2} \approx \frac{\lambda^2 n^2}{2}$ possible edges between these vertices. The probability that none of these edges get included in $G(n, p_2)$ is at most

$$\left(1 - \frac{\epsilon/2}{n}\right)^{\lambda^2 n^2/2} \leq \exp\left(-\frac{\epsilon/2}{n}\frac{\lambda^2 n^2}{2}\right) = \exp\left(-Cn\right),$$

where $C = \epsilon\lambda^2/4$. □

It is not just cycles that appear when $p \geq \frac{c}{n}$ and $c > 1$. Topologically speaking, everything appears.

Theorem 15 (Ajtai, Komlós and Szemeródi, 1979)**.** *Let H be any fixed graph. If $p \geq \frac{c}{n}$ and $c > 1$, then w.h.p. $G(n, p)$ contains H as a topological minor, i.e. $G(n, p)$ contains a subdivision of H as a subgraph.*

Corollary 7. *If $p \geq \frac{c}{n}$ and $c > 1$, then w.h.p. $G(n, p)$ is not planar.*

What if $p = \frac{c}{n}$ and $0 < c < 1$? Recall that $\mu = \mathbb{E}[\#\text{ cycles}] \to \frac{1}{2}(-\log(1-c) - (c + c^2/2))$. We can use the method of moments to show that # cycles $\to \text{Pois}(\mu)$. Thus,

$$\mathbb{P}[\text{there exists no cycles}] \to e^{-\mu} = \sqrt{1-c}\, e^{c/2+c^2/4}.$$

It is often useful to consider the random graph as a stochastic process. Assign each of the $\binom{n}{2}$ possible edges a uniform random number in the interval $[0, 1]$. For $0 \leq t \leq 1$, at time t include all the edges that got assigned a number at most t. In this way, the random graph "evolves" from an edge-empty graph to a complete graph.

The stochastic process point of view allows for more refined questions.

Question. what can be said about the size of the first cycle to appear in the random graph process?

Let L_n denote the length of the first cycle in the random graph process. The follow theorem tells us that L_n approaches a limiting distribution, albeit one without a mean.

Theorem 16. *For every fixed $l \geq 3$, we have*

$$\lim_{n\to\infty} \mathbb{P}[L_n = l] = \frac{1}{2}\int_0^1 x^{l-1}(1-x)^2 e^{x/2+x^2/4} dx.$$

Theorem 17 (Flajolet, Knuth and Pittel, 1989)**.** *The first cycle has expected length $\approx Kn^{1/6}$, where*

$$K = \frac{1}{\sqrt{8\pi i}}\int_{-\infty}^{\infty}\int_{\Gamma} \exp\left(\frac{(\mu+2s)(\mu-s)^2}{6}\right)\frac{s}{ds} d\mu \approx 2.03337$$

for a certain contour Γ. The standard deviation if of order $n^{7/12}$, and the average length of the kth cycle if proportional to $n^{1/6}(\log n)^{k-1}$.

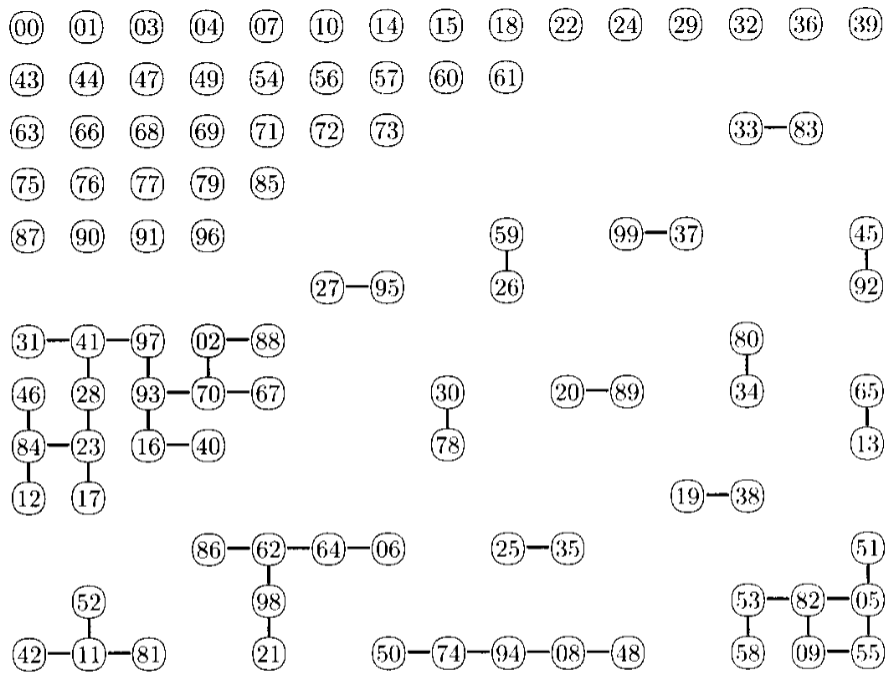


Fig. 1. The final state of a graph on 100 vertices that has evolved until a cycle first appears. The successive ordered pairs $\langle 31,41\rangle$ $\langle 59,26\rangle$ $\langle 53,58\rangle$ $\langle 97,93\rangle$ $\langle 23,84\rangle$ $\langle 62,64\rangle$ $\langle 33,83\rangle$ $\langle 27,95\rangle$ $\langle 02,88\rangle$ $\langle 41,97\rangle$ $\langle 16,93\rangle$ $\langle 99,37\rangle$ $\langle 51,05\rangle$ $\langle 82,09\rangle$ $\langle 74,94\rangle$ $\langle 45,92\rangle$ $\langle 30,78\rangle$ $\langle 16,40\rangle$ $\langle 62,86\rangle$ $\langle 20,89\rangle$ $\langle 98,62\rangle$ $\langle 80,34\rangle$ $\langle 82,53\rangle$ $\langle 42,11\rangle$ $\langle 70,67\rangle$ $\langle 98,21\rangle$ $\langle 48,08\rangle$ $\langle 65,13\rangle$ $\langle 28,23\rangle$ $\langle 06,64\rangle$ $\langle 70,93\rangle$ $\langle 84,46\rangle$ $\langle 09,55\rangle$ $\langle 05,82\rangle$ $\langle 23,17\rangle$ $\langle 25,35\rangle$ $\langle 94,08\rangle$ $\langle 12,84\rangle$ $\langle 81,11\rangle$ $\langle 74,50\rangle$ $\langle 28,41\rangle$ $\langle 02,70\rangle$ $\langle 19,38\rangle$ $\langle 52,11\rangle$ produce nothing but free trees in the initially empty graph, but then $\langle 05,55\rangle$ yields a cycle of length 4. (This cycle appears in the lower right corner.) At this point there still are 40 isolated vertices (shown at the upper left) that have not yet been mentioned.

Lecture 12: The Phase Transition

Comment. The previous lecture describes a 1-sided sharp threshold for Cycles.

Comment. We don't need all the techniques from the previous lecture to prove that $G(n,p)$ has cycles when $p = c/n$ with $c > 2$, since in that case with high probability $G(n,p)$ has more than n edges, so it must have cycles.

This lecture discusses the "Birth of the Giant" in $G(n,p)$, that is, the precise threshold for p at which a giant connected component (of size $\Theta(n)$) emerges. This discussion is based on Chapter 11 from *The Probabilistic Method* by Alon and Spencer.

In what follows we will describe the structure of $G(n,p)$ for p around $1/n$. To make the exposition cleaner, we will parametrize p in two different ways, choosing one or the other depending on convenience:

- **Coarse Parameterization:** $p = c/n$.
- **Fine Parameterization:** $p = 1/n + \lambda n^{-4/3}$.
  We may also write $p = \frac{1+\varepsilon}{n}$, where $\varepsilon = \lambda n^{-1/3}$.

Definition 29. Given a connected component C of a graph G with $v = |V(C)|$ vertices and $e = |E(C)|$ edges, its **complexity** is the number $e - v + 1$.

Connected components of complexity one or zero are called **simple**.
Connected components with greater complexity are called **complex**.

Comment. Note the complexity counts the number of linearly independent cycles of a connected component. Thus, simple components are either trees or have exactly one cycle.

Notation. Let L_i denote the number of vertices in the i-th largest component of $G(n,p)$.

The Birth of the Giant. We identify the following regimes:

- **Very Subcritical:** $p = \frac{c}{n}$, with $c < 1$ a constant.
  - All components are simple.
    From last lecture we now there is a positive probability that there are cycles, however w.h.p. there is at most one per connected component.
  - $L_1 = \Theta(\log(n))$.
  - $L_k \approx L_1$, fore every fixed k.
    I.e. the first k components are roughly the same size (up to a constant), for each fixed k.
- **Barely Subcritical:** $p = \frac{1-\varepsilon}{n}$ with $\varepsilon = \lambda n^{-1/3}$. We assume $\varepsilon \to 0$ slowly enough so that $\lambda \to \infty$.
  - All components are simple.
  - $L_1 = \Theta(\epsilon^{-2}\log\lambda) = \Theta(n^{2/3}\lambda^{-2}\log\lambda)$.
  - $L_k \approx L_1$ for every fixed k.
- **Critical:** $p = \frac{1+\lambda n^{-1/3}}{n}$, with λ a constant.
  - For k fixed, all the largest k components have size $L_k = \Theta(n^{2/3})$. We have here many components *competing* to be *the giant*.
  - Paremetrizing $L_k = c_k n^{2/3}$ and d_k the complexity of L_k, the joint distribution $(c_1, \ldots, c_k, d_1, \ldots, d_k)$ converges to a known and non-trivial probability distribution.
  - All these properties are true for any constant λ, including the case $\lambda = 0$. Meaning that the joint distribution referenced to above, varies continuously with λ.
- **Barely Supercritical:** $p = \frac{1+\varepsilon}{n}$, with $\varepsilon = \lambda n^{1/3}$. We assume $\epsilon \to 0$ slowly so that $\lambda \to \infty$.
  - $L_1 \approx 2\varepsilon n = 2\lambda n^{2/3}$.
  - $L_2 = \Theta\left(\varepsilon^{-2}\log\lambda\right) = \Theta\big(\big(n^{2/3}\lambda^{-2}\log\lambda\big)$.
  - L_1 has complexity tending to ∞ as $n \to \infty$.
  - All other connected components are simple.

  Notice
  $$\frac{L_2}{L_1} \approx \frac{2\lambda n^{2/3}}{n^{2/3}\lambda^{-2}\log\lambda} = \frac{2\lambda^3}{\log\lambda} \to \infty$$
  Since $\lambda \to \infty$. So the largest component begins to grow larger in comparison with the rest.
- **Very Supercritical:** $p = \frac{c}{n}$, $c > 1$ constant. The Giant component is born.
  - $L_1 \approx yn$, where $y = y(c)$ is the positive solution to the equation
    $$e^{-cy} = 1 - y$$
  - $L_2 = \Theta(\log n)$

- L_1 has complexity tending to ∞ as $n \to \infty$.
- All other components are simple.

It's interesting to notice that this behavior is the same even if c is large. We can see the process of increasing c from one constant value to another as adding extra edges to the graph. This theorem thus tells us that doing so will not increase the complexity of the smaller components, it will either combine components or (most likely) just increase the complexity of the largest component.

One of the most fundamental techniques to study the phase transition of $G(n, p)$ is to compare the "graph exploration process" to the "Galton-Watson branching process". We describe the later and its relationship with the former in the following lecture.

Lecture 13: The Galton-Watson process

resource: Chapter 11: The ErdÖs-Rényi Phase Transition from the probabilistic method.

Let Z be a probability distribution over the nonnegative integers, say Poisson. The Galton-Watson process begins with a single root note, we call her Eve. Eve has Z children. Each of her children (if there are any) now independently has Z children, and so on. Let T be the total number of nodes (including Eve herself) created in the process. We write $T = \infty$ if the process goes on forever.

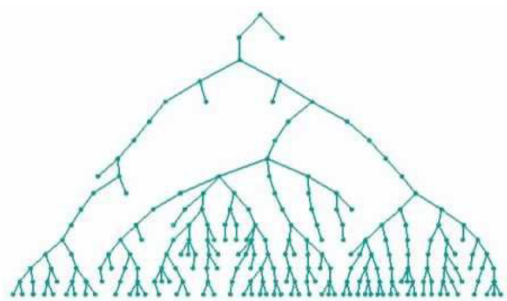


FIGURE 6. A super-critical Galton-Watson branching process

Fix $c > 0$. Let $(Z_t)_{t=1}^{\infty}$ be an infinite sequence of i.i.d. random variables, each having distribution Pois(c). Define an auxiliary random sequence recursively by

$$Y_0 = 1 \text{ and } Y_t = Y_{t-1} + Z_t - 1.$$

Let T be the minimal t with $Y_t = 0$. If no such t exists, then we write $T = \infty$.

Interpretation: Z_t is the number of nodes born at time t, Y_t is the queue size at time t, and T is the total size of the Galton-Watson process.

Theorem 18.
- *If $c < 1$, then T is finite with probability 1.*
- *If $c = 1$, then T is finite with probability 1.*
- *If $c > 1$, then T is infinite with probability $y = y(c)$, where y is the positive real number satisfying $e^{cy} = 1 - y$.*

Compare with the statement about the giant component in $G(n, p)$. Suppose $p = \frac{c}{n}$ where $c > 1$ is constant. Then $L_1 \approx yn$ where $y = y(c)$ is that positive real number satisfying the implicit equation $e^{cy} = 1 - y$.

Proof. Now suppose that $c \geq 1$. Let y be the probability that $T = \infty$, and set $z = 1 - y = \mathbb{P}[T < \infty]$. Given that Eve has i children, the probability that the branching process is z^i, as all i branches must be finite. So

$$z = \sum_{i=0}^{\infty} \mathbb{P}[Z_t = i] z^i = \sum_{i=0}^{\infty} e^{-c} \frac{c^i z^i}{i!} = e^{c(z-1)}.$$

Since $z = 1 - y$, we have $e^{cy} = 1 - y$.

If $c = 1$, then $e^{-y} > 1 - y$ for $y > 0$, so the solution must be $y = 0$.

For $c > 1$, the function $f(y) = 1 - y - e^{-cy}$ has $f(0) = 0, f(1) < 0$, and $f'(0) = c - 1 > 0$, so there exists a $y \in (0, 1)$ with $f(y) = 0$. Furthermore, as $f''(y) \leq 0$ for all y, there is precisely one such y. We have shown that either $\mathbb{P}[T < \infty] = 1$ or $\mathbb{P}[T < \infty] = 1 - y > 0$.

The argument that $\mathbb{P}[T < \infty] \neq 1$ is left as an exercise. $\square$

In fact, if $c = 1 + \varepsilon$, then $\mathbb{P}[T = \infty] \approx 2\varepsilon$, as follows.

$$\begin{aligned} 1 - y &= e^{-cy} = e^{(1+\varepsilon)y} \\ \log(1-y) &= -(1+\varepsilon)y \\ y + \frac{y^2}{2} &\approx (1+\varepsilon)y \\ y &\approx 2\varepsilon \end{aligned}$$

Compare this with the statement we have about the giant component in $G(n,p)$. Suppose $p = \frac{1+\varepsilon}{n}$ with $\varepsilon = \lambda n^{1/3}$. If $\varepsilon \to 0$ and $\lambda \to \infty$, then $L_1 \approx 2\varepsilon n$.

Q. What does Galton-Watson have to do with $G(n,p)$?

Choose an arbitrary vertex v. We will generate the component $C(v)$ containing with a standard "breadth-first search" algorithm. In this procedure, at any moment, all vertices will be live, dead, or neutral. The live vertices are contained in a queue. At time zero, v is live, the queue consists of a single vertex, and all other vertices are neutral. At each time t, we "pop" a live vertex from the top of the queue, and check all pairs $\{w, w'\}$ for adjacency in G. The popped vertex w is now dead, and any neutral adjacent w' are added to the bottom of the queue in any order and are now live. The procedure ends at time T when the queue is empty, and $T = |C(v)|$.

Q. How many neighbors will vertex v have?

This is a binomial random variable $B(n-1,p)$. Since we are assuming that $p \approx c/n$ for some constant $c > 0$, this is well approximated by $\mathrm{Pois}(c)$. This is a hint that the "graph exploration" process (i.e. breadth-first search in $G(n,p)$ can be well approximated by a Galton-Watson process.

Here is one statement that makes the connection precise. (Theorem 11.6.1 in Alon & Spencer)

Theorem 19. *Let $C(v)$ be the connected component of vertex v in $G(n,p)$, where $p = \frac{c}{n}$. Let T_c be the size of the Galton-Watson tree, where every node has Pois(c) descendents. Then*

$$\lim_{n\to\infty} \mathbb{P}[|C(v)| = k \text{ in } G(n, c/n)] = \mathbb{P}[T_c = k].$$

We won't prove this statement, but will instead discuss one of its immediate consequences.

Proposition 1. For any real $c > 0$ and integer $k \geq 1$, we have

$$\mathbb{P}[T_c = k] = \frac{e^{ck}(ck)^{k-1}}{k!}.$$

Proof. Let v be an arbitrary vertex in $G(n, c/n)$. Using the previous proposition, we will show instead that $\mathbb{P}[|C(v)| = k \text{ in } G(n,c/n)] \to \frac{e^{ck}(ck)^{k-1}}{k!}$. There are $\binom{n}{k-1}$ possible choices for the other $k-1$ vertices in $C(v)$. On any particular such set, the probability is $O(p^k) = O(n^{-k})$ that there are more than $k-1$ edges.

So w.h.p. the connected component containing v is a tree. On any given set of k vertices, there are precisely k^{k-2} tree. Each occurs with probability

$$p^{k-1}(1-p)^{\binom{k}{2}-k+1} \approx p^{k-1} = c^{k-1}n^{1-k}.$$

So $\mathbb{P}[C(v) = k] \approx \binom{n}{k-1}k^{k-2}c^{k-1}n^{1-k}e^{-ck} \to \frac{e^{ck}(ck)^{k-1}}{k!}$. $\square$

Note that for large k, by Stirling's approximation we have
$\frac{e^{ck}(ck)^{k-1}}{k!} \approx \frac{1}{\sqrt{2\pi}}k^{3/2}c^{-1}(ce^{1-c})^k$.

For any $c \neq 1$, we have $ce^{-c} < 1$ and therefore $\mathbb{P}[T = k]$ approaches zero at exponential speed in k.

For $c = 1$, $\frac{e^{ck}(ck)^{k-1}}{k!} \approx \frac{1}{\sqrt{2\pi}}k{-}3/2$.

Suppose that $c \to 1^+$ so $\varepsilon \to 0^+$. We have

$$\log(ce^{1-c} = \log(1+\varepsilon) - \varepsilon \approx -\frac{\varepsilon^2}{2}.$$

Thus $\mathbb{P}[T_{1+\varepsilon} = u] \approx \frac{1}{\sqrt{2\pi}}u^{-3/2} for u = o(\varepsilon^{-2})$.

When u reaches order ε^{-2} there is a change. For $u = A\varepsilon^{-2}$ and fixed A,

$$\mathbb{P}[T_{1+\varepsilon} = u] \approx \frac{1}{\sqrt{2\pi}}\varepsilon^3 A^{-3/2}e^{-a/2}.$$

The parameterization explained.

Q. Why is it useful to consider $p = \frac{1\pm\varepsilon}{n}$ where $\varepsilon \approx n^{-1/3}$? What is special about the exponent -1/3 ?

Consider the Poisson Galton-Watson branching process with $c = 1 + \varepsilon$. It is infinite with probability $\approx 2\varepsilon$. Otherwise its probability of exceeding $A\varepsilon^{-2}$ drops exponentially with A. The graph exploration process mimics the Galton-Watson process, as long as it is not too successful.

The cases where the Poisson branching process is finite are mimicked, yielding components of size up to roughly ε^{-2}. The cases where the Poisson branching process is infinite are mimicked by components that "escape" until the ecological limit sets in. These components all join together to form a single dominant component of order $2\varepsilon n$. In order for the above rough picture to hold, there must be a distinction between the small components and the dominant component, so we need $2\varepsilon n \gg \varepsilon^{-2}$. This leads to $\varepsilon = \Theta(v^{-1/3})$ as a breakpoint. If $\varepsilon \gg n^{-1/3}$, then we have the distinction between small and dominant components and are in the supercritical regime.

Comment: the case of a critical Galton-Watson tree (i.e. $c = 1$) is especially interesting, and is known to have a <u>scaling limit</u>.

David Aldous: "Take a critical Galton-Watson branching process and condition on total population until extinction being . This gives a random tree. Rescale edge-lengths to have length $n^{-1/2}$.... In a certain sense this can be formalized, so the $n \to \infty$ weak limit of these random trees is the Continuum Random Tree..."

Lecture 14: Random geometric graphs

Useful Resource: Random geometric graphs by Mathew Penrose.

Define a random geometric graph $G(X_n; r_n)$ as follows. Fix $d \geq 2$ and let the vertices X_n be n iid points chosen uniformly in the cube $[0,1]^d$. The edges are between vertices p, q where $d(p,q) \leq r_n$.

Note. we usually study asymptotics as $r = r(n)$ and $n \to 0$ with $r \to 0$ at some rate

Note. People also study other distributions besides the uniform distribution on $[0,1]^d$. For example, multivariate normal distribution, uniform distributions on the sphere S^d or other Riemannian manifolds, or in general probability distributions with a bounded measurable density function.

Note. Vertices are independent, but now edges aren't.

Question: What is the expected degree of a vertex in $G(X_n; r_n)$?.

Assuming that $r_n \to 0$, $\mathbb{P}[x \sim v] \approx \omega_d r_n^d$ where ω_d is the volume of the unit d-dimensional ball. So $\mathbb{E}[\deg(v)] \approx \omega_d n r_n^d$ (This suggest a big change in the structure of $G(X_n; r_n)$ once (thermodynamic limit).)

Question: For now, set $d = 2$ What's the expected number of triangles? The threshold for the appearance of triangles?

Given 3 points $x, y, z \in X_n$, what's the probability $x \sim y$, $y \sim z$, and $z \sim x$? Assuming $r \to 0$, it's something like $(\pi r^2) \times (\pi r^2) \times C$ where $C = \mathbb{P}[y \sim z | x \sim y \text{ and } x \sim z]$. Then $\mathbb{E}[\# K_3] = \Theta(n^3 r^4)$ and we might expect the threshold for appearance of triangles to be $r = n^{-3/4}$. It's true!

In general, for any fixed subgraph H with k vertices and fixed $d \geq 2$, we have $\mathbb{E}[\# H] = \Theta(n^k r^{d(k-1)})$. (The implied constant depends on H and d, and can be written explicitly as an integral of certain indicator functions over $\mathbb{R}^{dk}$.) So the threshold is

What about the threshold for connectivity? Let's try dividing up $[0,1]^d$ into small cubes of side length ϵr for fixed $0 < \epsilon < 1/2\sqrt{d}$. If every small cube contains at least one point, then $G(X_n; r_n)$ is connected! This is a coupon collector problem, which suggests selecting $r = (clog_n/n)^{1/d}$. For a given small cube Q, $\mathbb{P}[Q \text{ does not contain any points of } X_n] = (1 - (\epsilon r)^d)^n \leq \exp -(\epsilon r)^d n \approx n^{-c\epsilon^d}$.

A sharper result is $r = \left(\frac{2^{d-1}}{d\omega_d} \cdot \frac{\log n + c}{n}\right)^{1/d}$. Then

Theorem 20. *There exists a constant*

There's a "Poisson point process of uniform intensity" on $\mathbb{R}^d$ which give you spatial independence.

Some related references: The Birth of the Giant Component, The Evolution of Random Graphs and The First Cycles in Evolving Graphs.

Lecture 15: Simplicial Complexes

useful resources. algebraic topology book Allen Hatcher: free pdf on his website and our Carmen page. It's very readable if you have some background. It's a more general framework than simplicial complexes.

Anders Björner's Topological methods notes (ch 34): good intro to the subject. It starts with examples and then the tools (consider reading part II first). It's also available on our page.

Abstract simplicial complexes.

Definition 30. A simplicial complex Δ is a collection of subsets of a set V such that :

- for every $v \in V$, we have $\{v\} \in \Delta$, and
- if $X \in \Delta$ and $Y \subset X$ is a non-empty subset, then $Y \in \Delta$.

Terminologies:

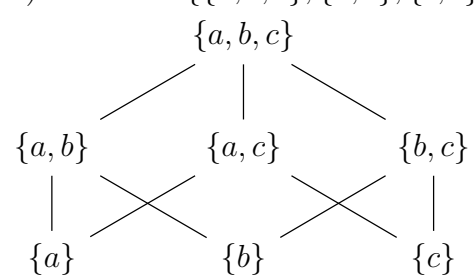
- The elements of the set V ("Ground set") are called vertices.
- The elements of Δ are called faces.
- The dimension of a face X is defined by $\dim(X) = |X| - 1$. (We will be talking about finite X so this definition is fine.)
- Then dimension of the simplicial complex is the largest dimension of any of its faces.
- The complex Δ is said to be *finite* if it has finitely many faces, or equivalently if the vertex set is finite. (This is stronger than finite dimensional)

Examples:

- A 1-dimensional simplicial complex is the same thing as an undirected graph (with no loops of multiple edges, i.e. ground set V and edge set E of unordered pairs).
- Let $\Delta = 2^V \setminus \emptyset$, i.e., Δ is the collection of all nonempty subsets of V. If $|V| = d + 1$, we sometimes call Δ the *simplex of dimension d*. (If you only have two elements in the $V = a, b$, then it's a simplex of dimension 1 because of the line segment. You only have vertices $\{a\}, \{b\}$ and a line segment/edge represented by $\{a, b\}$. If you have three element then you can make a triangle.)

The faces of a simplicial complex Δ naturally have the structure of a poset, where they are partially ordered by inclusion. This is called the *face poset* of Δ.

Example (2-d simplex)**.** Let $\Delta = \{\{a, b, c\}, \{a, b\}, \{b, c\}, \{c, a\}, \{a\}, \{b\}, \{c\}\}$.



A simplicial complex always has a face poset but it's not always a lattice (not always have meets/joins). The face poset in a way *is* the simplicial complex. You can recover a simplicial complex up to some combinatorial isomorphism if given a face poset. You can tell if a face is contained in another face from the face poset.

Question. why are they called *simplicial complexes*?

Because they are made out of simplicies.

Definition 31. A d-dimensional simplex is a d-dimensional polytope which is the convex hull of its $d + 1$ vertices.

E.g. 0-dim simplex is a point; 1-dim simplex is a line segment; 2-dim simplex is a triangle;3-dim simplex is a tetrahedron.

The *geometric realization* $|\Delta|$ of a finite abstract simplicial complex Δ is a topological space, defined as follows:

> Choose an embedding of the vertex set some Euclidean space $\mathbb{R}^N$ of sufficiently high dimension (certainly $N = |V|$ would be sufficient; twice the dimension of this simplicial complex plus one would also work), so that every set of vertices in a face is affinely independent. Then any face in Δ can be identified with the geometric simplex in $\mathbb{R}^N$ spanned by the corresponding embedded vertices. Take the union of all these simplicies.

Note(about $N = 2\dim(\Delta)+1$): if you have 2-dim simplicial complex, you can embed it in 5 dimensions by putting the vertices in general position and choosing the vertices randomly. Because probabilistically, you never have three points on a line, or four points in a plane or five points in a three dimensional space, etc., for points in general position.

Comment. You can defined the geometric realization of an infinite simplicial complex as well, but it is a little bit more subtle. (It involves taking either a direct limit or colimit in the appropriate category so this requires a bit more algebraic topology.)

Definition 32. A topological space X is said to be *triangulable* if there exists a simplicial complex Δ such that $|\Delta|$ is homeomorphic to X. In this case Δ is said to be a *triangulation* of X.

Intuition: can you cut up a topological space into simplicies so that they form a simplicial complex

Note. often we aren't careful about distinguishing a simplicial complex and its geometric realization (unique up to homeomorphism). Two simplicial complexes that are combinatorically the same have geometric realization that are isometric.

Example (triangulable spaces)**.**

- topological manifolds , in dimension up to 3
- smooth manifolds, in any dimension
- semialgebraic sets

topological manifolds: topological spaces that's locally Euclidean, Hausdorff, and second countable; the only 1-d topological manifolds are circles/lines; 2-d compact manifold: sphere/torus/Klein bottle(non orientable); 2-d non-compact ones aren't really classified; 3-d ones are complicated.

For smooth functions, you can talk about not only continuous maps, but smooth maps between them. One way to triangulate a smooth manifold is to take a Morse function on the manifold and reconstruct some kind of cell structure and further subdivide it into a simplicial complex.

Non-Example (non-triangulable spaces)**.**

- the Cantor set
- non-Hausdorff spaces, such as the "line with two origins"

- certain topological manifolds – famous examples given by Casson, Freedman, Rokhlin in the 1980's("the E8 manifold"), and recently by Ciprian Manolescu (in every dimension $d \geq 4$)

Comment. Many topological spaces are "weakly homotopy equivalent" to simplicial complexes, including finite T_0 spaces.

weak homotopy equivalent: there exists a continuous map that induces isomorphism on all the homotopy groups. A corollary of that is that it induces isomorphisms on all the homology and cohomology groups.

Almost any space you can think of is at least weakly homotopy equivalent to a simplicial complex.

Given a poset P, one can construct a simplicial complex $\Delta = \Delta(P)$ as follows:

- The vertices of Δ are the elements of P.
- The faces of Δ correspond to chains in P.

Δ is sometimes called the order complex of P.

Question. What if you start with a simplicial complex Δ, take its face poset P and then take the order complex of P, call it Δ'?

It turns out that $|\Delta|$ is homeomorphic to $|\Delta'|$. Δ' is called the *barycentric subdivision* of Δ, sometimes denoted $\mathrm{Bd}\,\Delta$. Let $\mathrm{Bd}^n\,\Delta$ denote the nth barycentric subdivision of the simplicial complex Δ.

In Figure 7 below, the triangle (abc) is filled (2-d simplex). Below it is its face poset and on the right is its barycentric subdivision. Observe on the barycentric subdivision, the nodes are all the elements in the poset, an edge is a chain of length 2, a 2-d face (e.g. triangle (c, bc, abc)) is a chain of length 3.

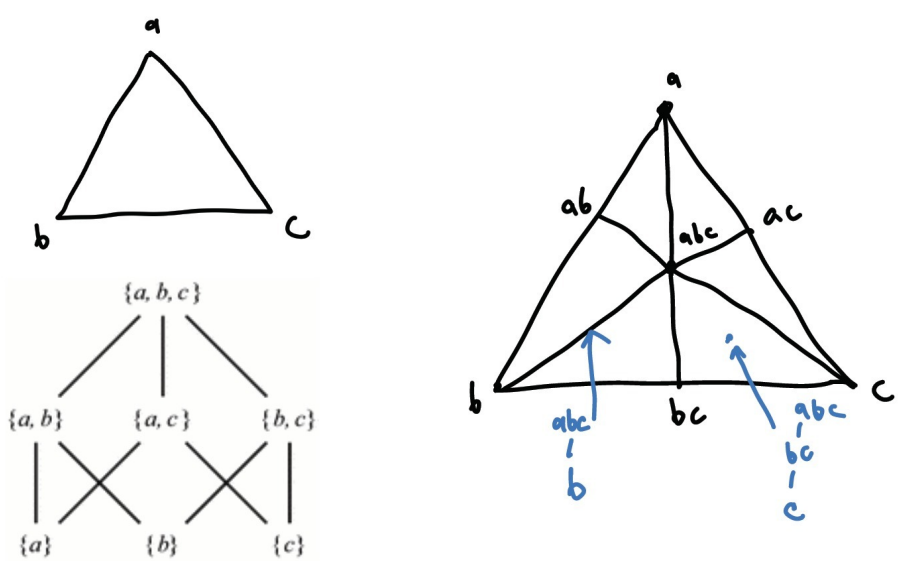


FIGURE 7. barycentric subdivision (right)

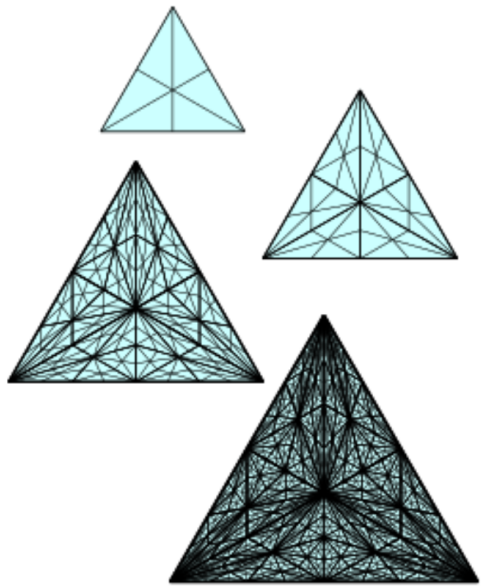


FIGURE 8. 4 stages of barycentric subdivision

Definition 33. Given two simplicial complexes, K and L, a simplicial map is a function from vertices of K to vertices of L, with the property that the images of the vertices of a face of K always span a face of L.

Fact. A simplicial map $K \to L$ induces a continuous map between the geometric realizations $|K| \to |L|$.

Theorem 21 (Simplicial approximation theorem)**.** *Suppose that K and L are simplicial complexes. If $F : |K| \to |L|$ is any continuous map, then there exists a natural number n and a simplicial map $f : \mathrm{Bd}^n K \to L$, such that f is homotopic to F.*

Here when we say f is homotopic to F we are actually talking about $\tilde{f}$, the induced map between geometric interpretations:

$$\tilde{f} : |\mathrm{Bd}^n K| \to |L|$$

Recall that $|\mathrm{Bd}^n K|$ is homeomorphic to $|K|$.

LECTURE 16: CW COMPLEXES AND HOMOTOPY GROUPS

Our crash course in algebraic topology continues. Again, a useful external resource is Hatcher's Algebraic Topology. Chapter 0 assumes only familiarity with point set topology, defining CW complexes and other geometric notions. Available online here. Another excellent book is Introduction to Topological Manifolds by John Lee, which is a Springer book so it is available online through the OSU library.

More Simplicial Complex Examples. The boundary of a tetrahedron. Begin with a simplicial complex with 4 vertices and every possible simplex, then remove the sole 3-simplex, leaving all simplices of lower dimension. The resulting simplicial complex has

14 simplices (one short of the maximum possible $2^4 - 1$) and is homeomorphic to the 2-sphere S^2. In fact, this is the smallest triangulation of the sphere, having only 4 triangles. An analogous construction can be completed for higher dimensional simplices to create a minimal triangulation of the n-sphere with $2^{n+1} - 2$ simplices total.

Torus on 7 vertices. The smallest possible triangulation of the torus requires 7 vertices, 21 edges, and 14 triangles. It can be constructed by embedding the complete graph K_7 on the torus.

Projective plane on 6 vertices. We can construct a simplicial complex homeomorphic to the projective plane as follows. The icosahedron is a simplicial complex homeomorphic to the sphere. The antipodal map on the icosahedron can be used to pair up the simplices. Quotienting by this action gives us a simplicial complex which is homeomorphic to the projective plane. Since the icosahedron has 12 vertices, 30 edges, and 20 faces, this new simplicial complex has 6, 15 and 10 vertices, edges, and faces respectively. It can also be constructed by embedding the complete graph K_6 in the projective plane.

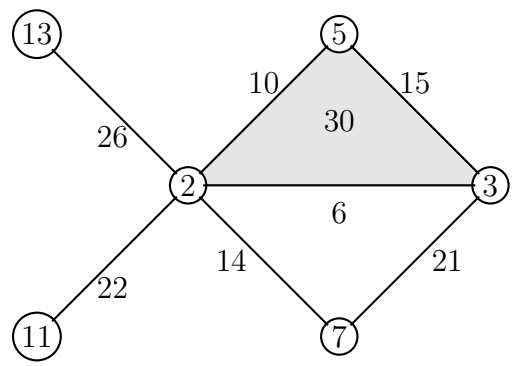


FIGURE 9. The simplicial complex Δ_{30}, when the first 2-simplex appears.

An example from number theory. For any $n \geq 1$, we define a simplicial complex Δ_n as follows. The vertices are the primes $\leq n$. The simplices are the collections of primes whose product is still $\leq n$. That is, the simplices are the sets $\{p_1, \ldots, p_k\}$ of primes such that

$$p_1 p_2 \cdots p_k \leq n.$$

The resulting complex has one simplex for each squarefree number less than or equal to n. An example is depicted in Figure 9.

There is an open question associated with this family of complexes...

The Mertens Function.
If a simplicial complex Δ has f_n cells in dimension n, the Euler Characteristic is defined to be

$$\chi(\Delta) = f_0 - f_1 + f_2 - \ldots$$

The open question relates to the asymptotics of $\chi(\Delta_n)$. There is a conjecture that

$$|\chi(\Delta_n)| = O(n^{1/2+\epsilon}).$$

As it turns out, this conjecture is *equivalent to the Riemann Hypothesis.*
The function $\chi(\Delta_n)$ is sometimes called the Mertens function. It was once conjectured that $|\chi(\Delta_n)| \leq \sqrt{n}$ for all n but a counterexample was proven to exist for some $n \leq e^{1.59\times 10^{40}}$. At the same time, no one has yet explicitly constructed a counterexample. More information can be found in Anders Björner's paper A Cell Complex in Number Theory.

CW Complexes. The notion of a CW complex is similar to that of a simplicial complex in the sense that it is a topological space with a combinatorial flavor. However, the construction allows more degrees of freedom. Additionally, the category of CW complex has more desirable properties.

The following is a loose definition of a CW complex. A 0-dimensional CW complex is a finite discrete set. (One could allow for infinite sets, but this will not be necessary for our purposes.) A 1-dimensional CW complex is a graph, where we allow for multi-edges and loops. In general, and n-dimensional CW complex is a k-dimensional CW complex A (for some $k < n$) along with one or more copies of the n-dimensional ball (called n-cells) which are "attached" at their boundaries. More precisely, a copy of D^n is attached by specifying a continuous function f, called the *attaching map*, from $\partial D^n = S^{n-1}$ to the k-dimensional CW complex, then quotienting by f, thought of as a relation.

For instance, we turn a 0-dimensional CW complex into a 1-dimensional CW complex by attaching copies of $D^1 = [0, 1]$. The attaching maps are functions f from $S^0 = \{0, 1\}$ to the set of discrete points, which has the effect of adding an edge from $f(0)$ to $f(1)$. There is no stipulation that $f(0) \neq f(1)$, so f could attach a loop to a vertex.

Torus CW complex. We take a single 0-cell and attach two 1-cells (call them a and b). Note that there is only one possible attaching map for each 1-cell: the constant map to the 0-cell. The resulting CW complex is a wedge sum of two circles, depicted in Figure 10. To complete the complex, we attach a 2-cell by the following attaching map on $\partial D^2 = S^1$. We define the map piecewise on each quadrant of S^1. Traveling counterclockwise around S^1, the first quadrant makes a clockwise loop around a, the second a clockwise loop around b, the third a counterclockwise loop around a, and finally the fourth a counterclockwise loop around b. Choosing appropriate orientations for a and b, we could notate this path as $aba^{-1}b^{-1}$.

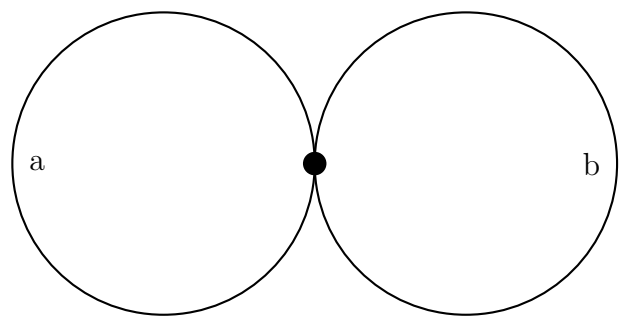


FIGURE 10. A wedge sum of two circles.

While this attaching map may be hard to visualize, the resulting space is homeomorphic to a torus. The more standard depiction of a torus as a square with sides identified encodes the same data: one vertex, two edges, and a 2-cell attached by $aba^{-1}b^{-1}$.

Note that, while a simplicial complex representing the torus needs at least 42 simplices, this construction results in a torus using only 4 cells.

A simpler sphere and projective plane. In a similar vein, a simplicial complex representing an n-sphere needs at least $2^{n+1} - 2$ simplices. With the extra freedom allowed by the construction of a CW complex, however, we can get it done with just two: take a single 0-cell and attach an n-cell via the constant map.

A slightly more sophisticated construction can also be helpful. We can proceed inductively. To construct S^0, simply take two 0-cells. To construct S^n, first construct S^{n-1}, then attach two n-cells via the identity map $\partial D^n = S^{n-1} \to S^{n-1}$.

There are two immediate advantages to this construction over the simpler one. First, with proper care, it can be iterated indefinitely to construct the infinite-dimensional sphere. Second, the antipodal map on S^n respects the CW structure. That is, the antipodal map sends k-cells homeomorphically to k-cells for all k. Therefore we can quotient by the antipodal map to achieve a CW complex which is homeomorphic to the projective space $\mathbb{R}P^n$.

Orientable surface of finite genus. Our construction of the torus above can be generalized to construct the orientable surface of genus g in the following manner. Start with a single 0-cell and attach $2g$ 1-cells $\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g$ via the constant map. Then attach a single 2-cell by the map

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_n\beta_n\alpha_n^{-1}\beta_n^{-1}$$

That is, the attaching map is the loop which travels around α_1, then β_1, then backwards around α_1 and β_1, then moves on to α_2, etc.

Fundamental Groups. Given a pointed space (X, x), The fundamental group $\pi_1(X, x)$ is the the set of all homotopy classes of loops which begin and end at x. This set indeed has a group structure, with multiplication defined by concatenation of paths. Then the identity is the homotopy class of the constant path at x, and the inverse of a path is the reverse path. If X is a connected space, then the isomorphism class of $\pi(X, x)$ does not depend on x. For this reason, if X is connected, we often suppress the x and simply write $\pi_1(X)$.

A couple examples:

- $\pi_1(S^1) \cong \mathbb{Z}$, the isomorphism given by the winding number of the path. Proving this is harder than it looks.
- If X_g is the orientable surface of genus g, then

$$\pi_1(X_g) \cong \langle \alpha_1, \beta_1, \ldots, \alpha_g, \beta_g | \alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_n\beta_n\alpha_n^{-1}\beta_n^{-1}\rangle.$$

The presentation in this second example may look familiar from our construction of X_g as a simplicial complex. Indeed, if X is a CW complex, a presentation for the fundamental group can be found by adding one generator for each 1-cell and one relation for each 2-cell, with the relator matching the path of the attaching map. In particular, every simplicial complex can be thought of as a CW complex, so we can write down a presentation for the fundamental group of any CW complex.

Conversely, for any finitely presented group G, we can reverse this process to create a CW complex X which has G as its fundamental group. We simply begin with a 0-cell, pick a presentation for G, attach a loop the 0-cell for each generator, then attach a 2-cell for each relation with the attaching map given by the terms of the relation.

Lecture 17: Homology

Resources: Allen Hatcher's Algebraic Topology and Anders Björner's Topological Methods.

First, we ask ourselves, "What is homology"? Informally, we can think of homology as an algebraic gadget, such as a sequence of groups or vector spaces, that measures how far a chain complex is from being exact. This gives us a way to extract information from a topological space without having to visualize it. A common metaphor for homology is that homology tells us how many "holes" a topological space has.

Example. For an n-dimensional sphere S^n, we have

$$H_k(S^n) = \begin{cases} \mathbb{Z} & k = 0, n, \\ \{0\} & \text{otherwise.} \end{cases}$$

We see that homology is indexed by the nonnegative integers. For a topological space X, the zeroth homology $H_0(X)$ tells us how many connected components a topological space has. For homology with integer coefficients, $H_0(X)$ is isomorphic to the free abelian group of rank equal to the number of connected components. For $k > 0$, H_k detects the k-dimensional "holes". We may think of the n-dimensional sphere as an n-dimensional hole based on its homology.

Because homology is a topological invariant, knowing the homology of S^n for each n tells us that S^m and S^n are not homeomorphic for $m \neq n$, i.e. spheres of different dimensions are different topological spaces. Another fact that falls out of the homology groups for spheres is that $\mathbb{R}^m$ and $\mathbb{R}^n$ are not homeomorphic when $m \neq n$. One way to see this is to take the one-point compactification of $\mathbb{R}^m$ and $\mathbb{R}^n$ for $m \neq n$ and obtain S^m and S^n, respectively, which are not homeomorphic, so $\mathbb{R}^m$ and $\mathbb{R}^n$ must not be homeomorphic. We can get the same result by puncturing the two Euclidean spaces as well.

The next two examples are examples of topological spaces whose homology is well-known. The homology of more complicated spaces can sometimes be computed from the homology of these spaces.

Example. For a torus $T^n = \underbrace{S^1 \times S^1 \times \cdots \times S^1}_{n}$, we have,

$$H_i(T^n) = \mathbb{Z}^{\binom{n}{i}}.$$

This result can be obtained using the Künneth formula, which allows for the computation of homology of products of spaces, combined with the homology of S^1, which we have just seen.

It is not always true that homology with integer coefficients yields free abelian groups. An example where we have torsion in homology is with the projective plane.

Example. For a projective plane P, its homology is

$$H_k(P) = \begin{cases} \mathbb{Z} & k = 0, \\ \mathbb{Z}_2 & k = 1, \\ \{0\} & \text{otherwise.} \end{cases}$$

In order to calculate the homology of a topological space, we first associate a chain complex to a topological space.

Definition 34. A **chain complex** is a sequence of abelian groups (or more generally R-modules) $C_0, C_1, C_2, \ldots$ connected by homomorphisms $\delta_n : C_n \to C_{n-1}$ called **boundary operators**

$$\cdots \xrightarrow{\delta_{n+1}} C_n \xrightarrow{\delta_n} C_{n-1} \xrightarrow{\delta_{n-1}} \cdots \xrightarrow{\delta_2} C_1 \xrightarrow{\delta_1} C_0 \xrightarrow{\delta_0} 0$$

such that $\delta_n \circ \delta_{n+1} = 0$ for every n. This last condition implies that the image of δ_{n+1} is contained in the kernel of δ_n.

We then define homology on the chain complex:

$$H_n(X) = \ker(\delta_n)/\mathrm{im}(\delta_{n+1}).$$

Note that here R denotes a commutative ring with 1. The subscripts on the boundary operators δ are often omitted. Then the condition $\delta_n \circ \delta_{n+1} = 0$ is written more compactly as $\delta^2 = 0$.

Oftentimes in algebraic topology, R is taken to be $\mathbb{Z}$, in which case the chain complex is a sequence of abelian groups, or a field, in which case the chain complex is a sequence of vector spaces. For the case of R being a field, since each $H_n(X)$ is a vector space, we can compute the dimension of each $H_n(X)$ by calculating the rank and nullity of each boundary operator, which are linear maps in this case. This is something a computer is well-equipped to do.

The most common choices of fields as coefficients over which homology is computed are $\mathbb{R}$ and $\mathbb{Z}_2$. For example, the ith *Betti number* of a topological space X is defined to be $\beta_i = \dim H_i(X, \mathbb{R})$. It turns out that we can actually replace $\mathbb{R}$ with any field of characteristic 0 and still get the same number.

We will see later how to associate a chain complex to a topological space, which will then allow us to compute homology for a topological space. Before we do that, here are some useful properties about homology:

- Homology is a homotopy invariant.
- Homology is a covariant functor.
- The Mayer-Vietoris sequence allows for the computation of homology of a bigger, more complicated space in terms of small, simpler pieces.
- The Künneth formula allows you to compute homology of products of spaces.
- The universal coefficient theorem relates homology with different coefficient rings: if you know homology with $\mathbb{Z}$ coefficients, you can compute (co)homology with any coefficients.

Sometimes we want to show is that homology is trivial. The following are useful tools to that end:

- $H_i(X) = 0$ if $i > \dim(X)$.
- If X is a CW complex with no i-dimensional cells, then $H_i(X) = 0$.
- If $\pi_1(X) = 0$, then $H_1(X) = 0$.

We can use the fact that ith homology is trivial for i larger than the dimension of a space by finding a space of lower dimension Y that is homotopy equivalent to a space of higher dimension X and concluding that the ith homology is trivial for $i > \dim(Y)$.

Relation to homotopy groups. The concepts of homology and homotopy groups are related by Hurewicz Theorem. Before stating the theorem, we will need a definition.

Definition 35. A topological space X is **n-connected** if X is nonempty and path connected, and if $\pi_i(X) = 0$ for $1 \leq i \leq n$. This is equivalent to saying that X is n-connected if for every map $f : S^i \to X$, there exists a map $\tilde{f} : B^{i+1} \to X$ such that $f = \tilde{f} \circ i$, where $i : S^i \to B^{i+1}$

is the inclusion map.

$$\begin{tikzcd} & B^{i+1} \arrow[d, dashed, "\tilde{f}"] \\ S^i \arrow[ur, "i"] \arrow[r, "f"'] & X \end{tikzcd}$$

The canonical example of an n-connected space is S^{n+1}. In fact, S^{n+1} is n-connected but not $(n+1)$-connected.

Theorem 22. *(Hurewicz Theorem) For any path-connected space X and $n \geq 1$, there is a group homomorphism $h_* : \pi_n(X) \to H_n(x)$.*

For $n = 1$, this homomorphism induces an isomorphism between the abelianization of the fundamental group and the first homology group $\tilde{h}_* : \pi_1(X)_{ab} \to H_1(X)$.

If $n \geq 2$ and X is $(n-1)$-connected, then $h_* : \pi_n(X) \to H_n(x)$ is an isomorphism. In addition, $h_* : \pi_{n+1}(X) \to H_{n+1}(x)$ is an epimorphism.

Associating a chain complex to a topological space. There are multiple ways to associate a chain complex to a topological space. One of these is simiplicial homology.

Simplicial homology. Let S be a simplicial complex. The group of k-chains is defined to be

$$C_k = \left\{ \sum_{i=1}^{N} c_i \sigma_i \mid c_i \in \mathbb{Z}, \sigma_i \text{ is an oriented } k\text{-simplex} \right\}.$$

Each element is taken to be a formal linear combination of k-simplices.

As for the boundary operators, given an oriented k-simplex $\sigma = (v_0, v_1, \ldots, v_k)$, we define

$$\delta_k(\sigma) = \sum_{i=0}^{k} (-1)^i (v_0, \ldots, \hat{v}_i, \ldots, v_k),$$

where $(v_0, \ldots, \hat{v}_i, \ldots, v_k)$ is the $(k-1)$-dimensional face obtained by deleting the ith vertex of σ. Then we extend linearly to get $\delta_k : C_k \to C_{k-1}$. We may then also check that $\delta^2 = 0$ by checking this on each simplex.

From the existence of simplicial homology, we get already get a few useful properties. Let f_k denote the number of k-simplices. Then $\beta_k \leq f_k$ for every k because we can see $H_k(X)$ as the quotient of some submodule of C_k and C_k has f_k elements.

Exercise. We also get a lower bound: $-f_{k-1} + f_k - f_{k+1} \leq \beta_k$.

Proposition 17. *For any finite simplicial complex, we have*

$$f_0 - f_1 + f_2 - \cdots = \beta_0 - \beta_1 + \beta_2 - \cdots .$$

Proof. Induct on the number of faces. If there is only one face, then $f_0 = \beta_0 = 1$.

Adding a k-dimensional face increases β_k by one or decreases β_k by one. Either way, the equality continues to hold. □

Corollary 8. *Since homology is a homotopy invariant, then so is the Euler characteristic $(\beta_0 - \beta_1 + \beta_2 - \cdots)$.*

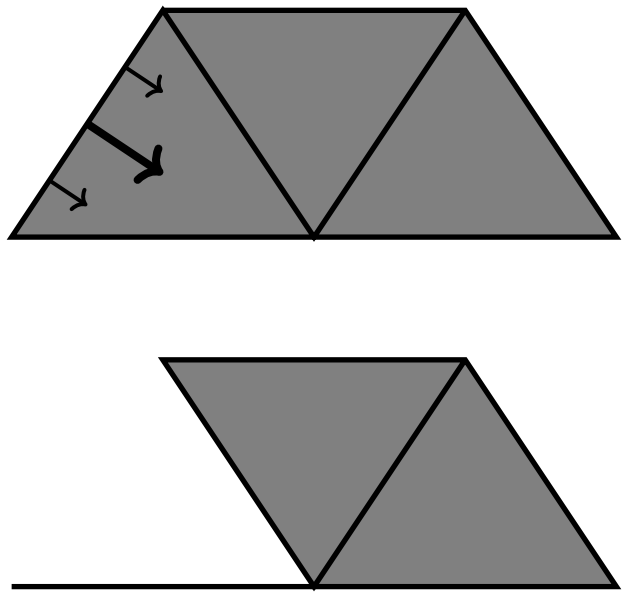


FIGURE 11. Collapsing a single triangle onto a portion of its boundary.

LECTURE 18: COMBINATORIAL HOMOTOPY THEORY

A useful reference for this section and resource for combinatorial homotopy theory in general is A User's Guide to Discrete Morse Theory, allegedly accessible to those without a background in Morse Theory. Also worth looking at is Anders Björner's Topological Methods.

When examining the topology of a simplicial complex X, it can be useful to find another homotopy equivalent complex (simplicial or CW) with fewer faces. To this end, we introduce the notion of a simplicial collapse. In its simplest form, it looks something like Figure 11.

Essentially, we may deformation retract certain cells (along with a piece of their boundary) that are not contributing to the topology of our complex. We notice that we can do this whenever neither the cell we are contracting nor the boundary piece we are stretching is not part of another higher-dimensional face (otherwise we'll be inadvertently stretching another face that doesn't appear in the picture above). Let's make this notion precise now.

Definition 36. Given a simplicial complex Δ, suppose that σ, τ are distinct faces of Δ such that

- $\sigma \subset \tau$ (in particular, $\dim \sigma < \dim \tau$),
- τ is a maximal face of Δ, and
- no other maximal face of Δ contains σ.

A simplicial collapse is the removal of all simplices γ with $\sigma \subseteq \gamma \subseteq \tau$. In the case like our example where $\dim \sigma = \dim \tau - 1$ and we remove only the two faces σ and τ, we call this an elementary collapse. One can check that these collapses are homotopy equivalences, and that any simplicial collapse arises as a sequence of elementary collapses. We call a simplicial complex collapsible if it can be reduced to a single vertex by a sequence of collapses.

Notice that the example we showed earlier is collapsible, as shown in Figure 12.

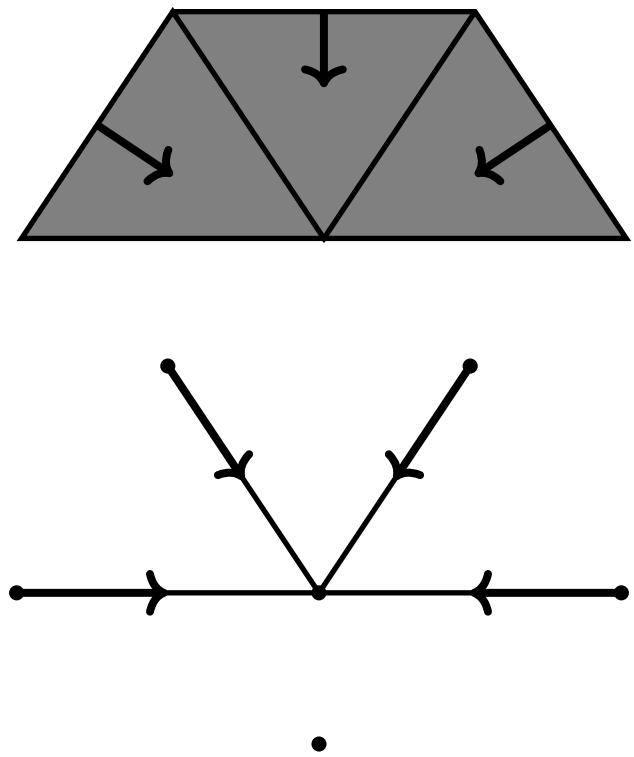


FIGURE 12. Performing a series of seven elementary collapses to collapse the complex to a point.

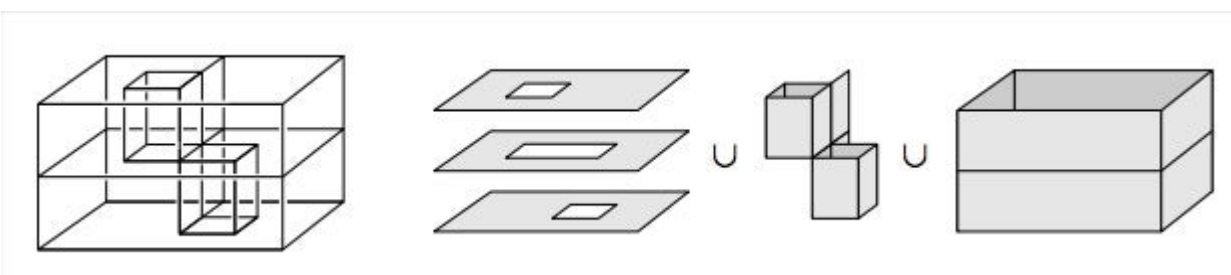


FIGURE 13. Bing's House with Two Rooms, a famous topological counterexample.

It is worth commenting that we don't know that our example complex is collapsible simple because it is contractible; there are non-collapsible complexes which are contractible (see Figures 13 and 14).

Question. Is it true that if two complexes X and Y are homotopy equivalent, then they can be related be a series of collapses and "anti-collapses" (also called "expansions")?

The answer turns out to be No, in general, so collapses don't play quite as well with homotopy as deformation retractions do. The answer does become Yes when we assume that X and Y are both simply connected. In particular, this tells that a simply connected space X is contractible if and only if X is collapsible.

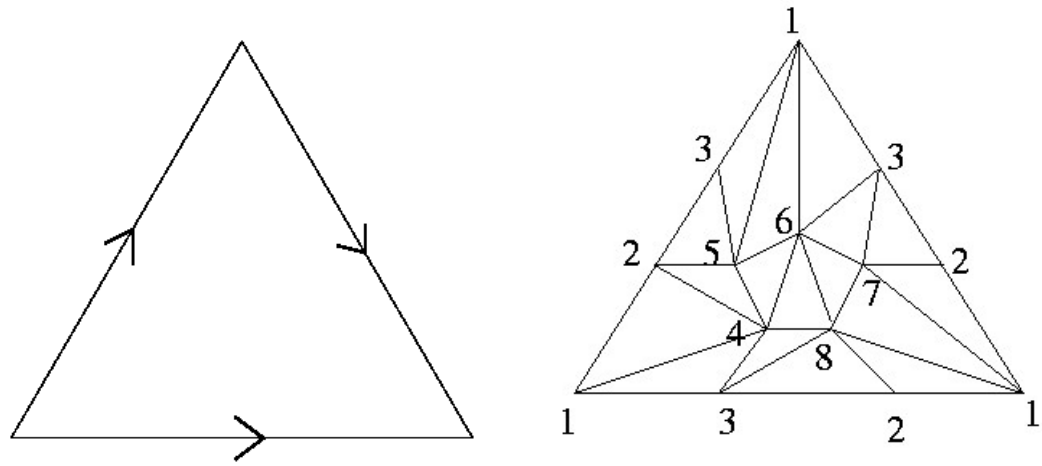


FIGURE 14. The Dunce Hat with a triangulation.

A Nerve Theorem. Suppose that X is a simplicial complex, and $C \subseteq 2^X$ is some collection of subcomplexes of X.

Definition 37. The set of subcomplexes C is called a cover or covering of X when, for every simplex $\sigma \in X$, we have some subcomplex $Y \in C$ such that $Y \ni \sigma$. The nerve of the cover C is the simplicial complex

$$N(C) = \left\{ D \subseteq C \;\middle|\; \bigcap_{Y \in D} Y \neq \varnothing \right\}.$$

Theorem 23. *If for every $D \subseteq C$, the intersection $\bigcap_{Y \in D} Y$ is either empty or contractible, then the nerve $N(C)$ is homotopy equivalent to the original complex X.*

This theorem is the real reason we want our collection C to cover the complex X. You could apply the above definitions and theorem to any set of subcomplexes and their union. We may think of this theorem as saying that the nerve is what you get when you contract the intersections among the pieces of a cover, giving a simpler complex to work with.

Example. Let's calculate the nerve of a cover C of the dodecagon X (12 vertices linked by 1-dimensional edges in a cycle). We cover X with three 5-vertex path graphs, as shown in Figure 15. To calculate the nerve complex $N(C)$, we start with one vertex for each subcomplex making up our cover. Since each pair of subcomplexes intersects, we get an edge in $N(C)$ between each pair of vertices. Since we have empty intersection across all three subcomplexes, we get no 2-dimensional face, so our nerve complex is K_3. Notice that our nerve theorem applies: each pair of subcomplexes intersects in a point, and all three have empty intersection, and notice that $N(C) = K_3 \simeq S^1$ is homotopy equivalent (homeomorphic, in fact) to $X = C_{12} \simeq S^1$.

Example. Let's try a larger example. Let X be the surface of the icosahedron, and let our cover C consist of four pieces as shown in Figure 16. Take each piece to include precisely the highlighted faces, as well as their boundaries. Since each pair and triple of elements of our cover has nonempty (and contractible!) intersection, the nerve C has all 1- and 2-dimensional

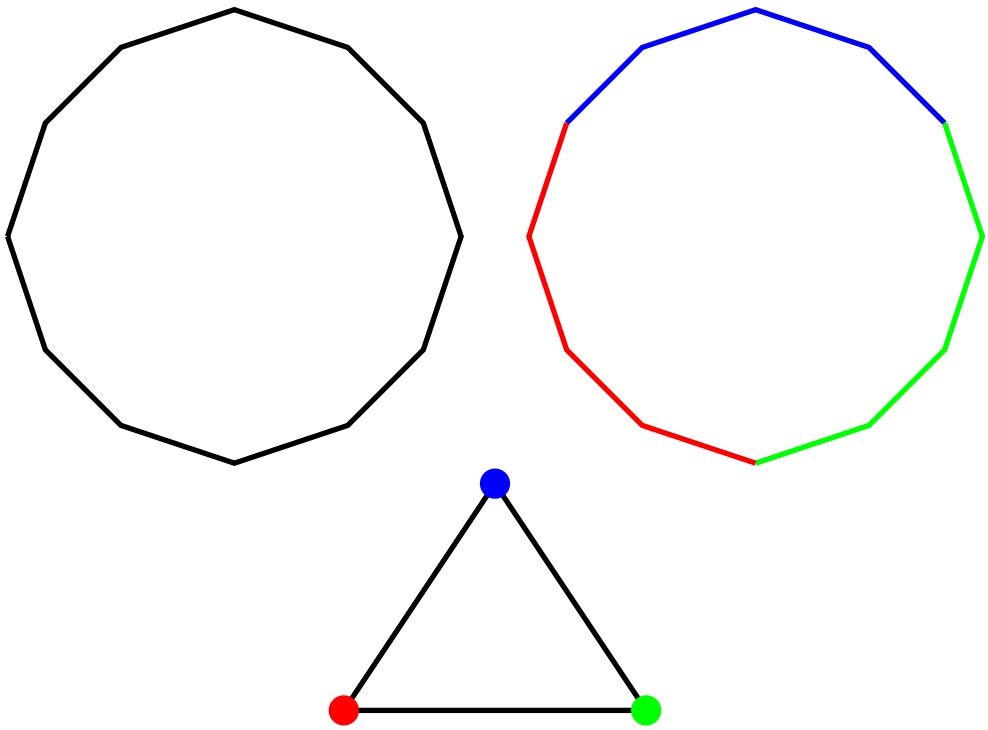


FIGURE 15. The dodecagon, a three-element cover, and the resulting nerve.

cells between its four vertices, but not the three dimensional cell, since the 4-way intersection is empty. Thus $N(C)$ is the surface of the tetrahedron. Comparing with the nerve theorem, we again see that $X \simeq S^2$ and $N(C) \simeq S^2$ are homotopy equivalent (homeomorphic, in fact).

Discrete Morse Theory. Raoul Bott is quoted as saying, "Every mathematician has a secret weapon. Mine is Morse Theory." Hopefully, discrete Morse theory has the same power, without the need for full knowledge of classical Morse theory.

Definition 38. Let S be a finite simplicial complex. A <u>discrete vector field</u> V is a collection of ordered pairs of faces $[\alpha_i, \beta_i]$ such that

- for every i, $\alpha_i \subset \beta_i$ with $\dim \alpha_i = \dim \beta_i - 1$ and
- no face is in more than one pair.

For a discrete vector field V, a <u>closed V walk</u> is a sequence of distinct faces $\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_k, \beta_k$ such that

- for each i, $[\alpha_i, \beta_i] \in V$,
- for $i = 1, 2, \ldots, k-1$, α_{i+1} is a codimension-1 face of β_i, and
- α_1 is a face of β_k.

We call a discrete vector field V <u>gradient</u> when there are no closed V walks. Given a discrete gradient vector field, we call a face $\sigma \in S$ <u>critical</u> when it is not in any pair $[\alpha_i, \beta_i]$.

Example. Let's check out an example. We will triangulate the real projective plane, then put a discrete gradient vector field V on it, as shown in Figure 17. We express the pairs $[\alpha_i, \beta_i]$ by drawing an arrow from α_i to β_i, in the hopes that this makes it seem more like

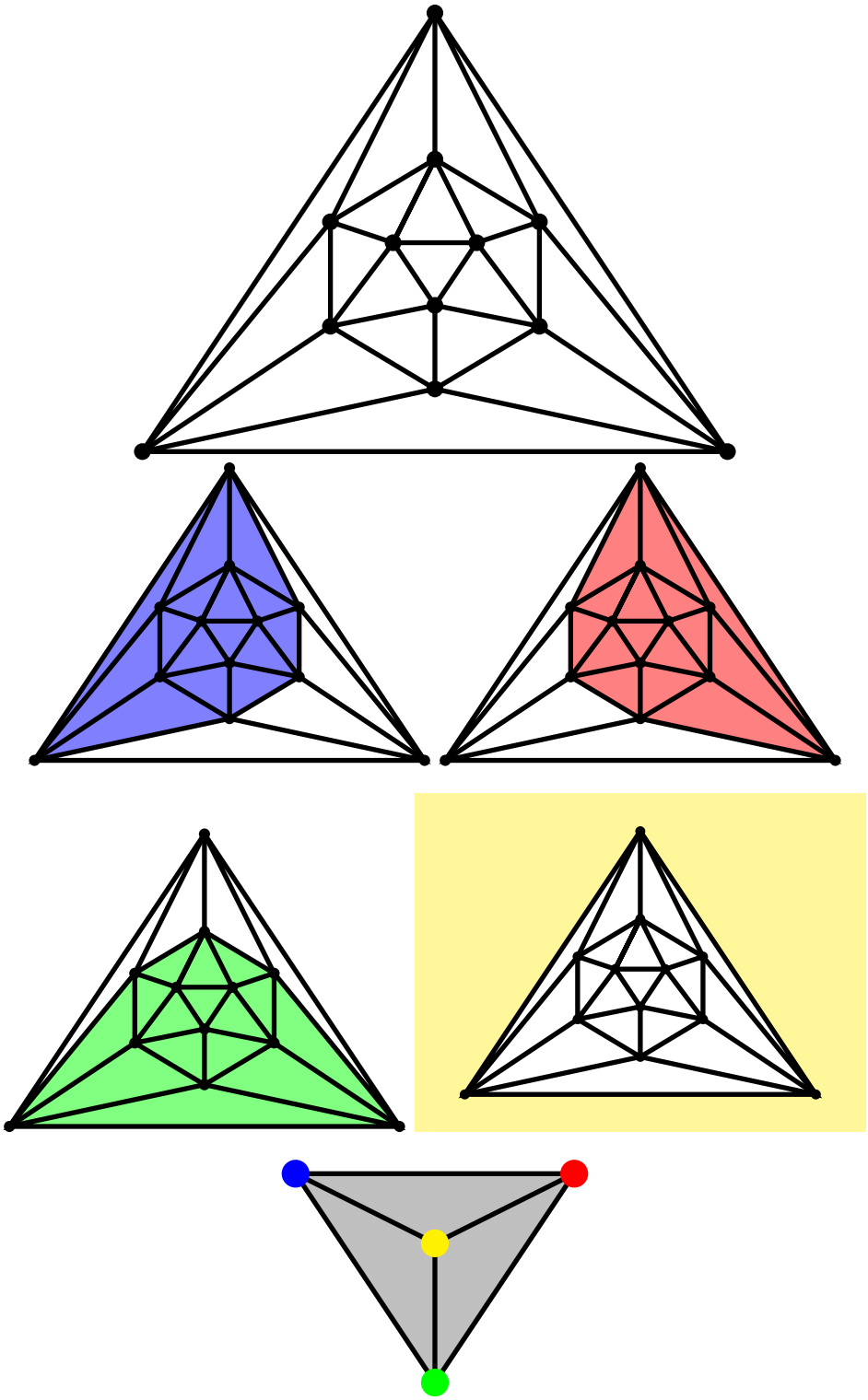
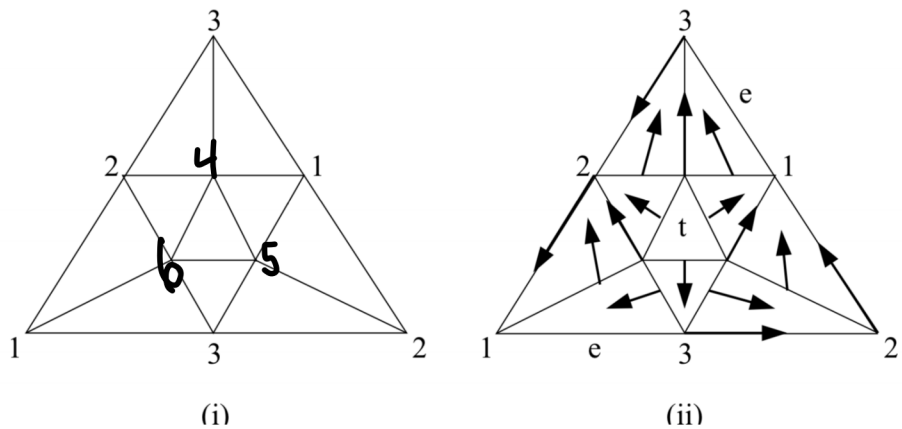


FIGURE 16. The icosahedron, a four-element cover, and the resulting nerve.



(i) A triangulation of the real projective plane. (ii) A discrete gradient vector field on $\mathbb{P}^2$.

FIGURE 17. This discrete gradient vector field has three critical cells, marked 1, e, and t.

we are talking about about vectors. Try as we might, we can't for a closed V walk; we always seem to get pushed away from the center triangle and are unable to return to our starting point. We may also notice that V misses three faces, a 0-dimensional face labeled 1, a 1-dimensional face labelled e, and a 2-dimensional face labeled t.

One might notice, in the previous example, that the real projective plane has a CW complex structure consisting of one cell in each of dimensions 0, 1, and 2, and the discrete gradient vector field given has one critical cell in each of these dimensions. This sort of observation may be partially realized as a theorem.

Theorem 24. *If V is a discrete gradient vector field on a simplicial complex S, then there is a homomtopy equivalent CW complex X with one k-dimensional cell for every critical k-dimensional face of S.*

Corollary 9. *Suppose a simplicial complex S admits a discrete gradient vector field whose only critical cells are:*

- *one* 0*-dimensional cell and*
- *some number of d-dimensional cells.*

Then S is homotopy equivalent to a bouquet of d-dimensional spheres.

In A User's Guide to Discrete Morse Theory, Robin Forman has made a comment in this vein, saying, "Many simplicial complexes that arise in combinatorics are homotopy equivalent to a wedge of spheres. I have often wondered if perhaps there is some deeper explanation of this."

LECTURE 19: RANDOM 2-DIMENSIONAL SIMPLICIAL COMPLEXES

""I predict a new subject of statistical topology. Rather than count the number of holes, Betti numbers, etc., one will be more interested in the distribution of such objects on non-compact manifolds as one goes out to infinity." –Isadore Singer, 2004

Today we will define the Linial-Meshulam model and survey some of the results of the past 15 years.

Definition 39. The random 2-complex Y $Y(n,p)$ is the random simplicial complex with n vertices, a complete 1-skeleton, and each 2-dimensional face appearing independently with probability p.

We can see that this is a higher dimensional generalization of the Erdos-Renyi model. In that case, there were n vertices and each 1-dimensional face had probability p of appearing. More generally, there is a d-dimensional version of this model: start with the complete $(d-1)$-skeleton of a simplex on n vertices. Each d-dimensional faces appears independently with probability p.

One can think about $Y(n,p)$ as a stochastic process, in which each 2-dimensional face is added randomly, one at a time.

Note. Many topological properties are monotone for $Y(n,p)$. For example, $H_1(Y)$, $\pi_1(Y)$, and 1-collapsibility are monotone decreasing, while $H_2(Y)$ is increasing.

There is are two important chains of implications for the homology of $Y(n,p)$. We have

$$\pi_1(Y) = 0 \Rightarrow H_1(Y) = 0 \Rightarrow H_1(Y, \mathbb{Z}/2\mathbb{Z}) = 0 \Rightarrow H_1(Y, \mathbb{Q}) = 0 \Leftrightarrow H_1(Y, \mathbb{R}) = 0.$$

We also have Y is 1-collapsible $\Rightarrow H_2(Y, \mathbb{R}) = 0 \Rightarrow$ there do not exist any tetrahedron boundaries.

If we set $p = c/n$ for some fixed $c > 0$, we can calculate the expected number of tetrahedron boundaries in $Y(n,p)$ as $\mathbb{E}[X] \to c^4/24$. In fact, $X \to \mathrm{Pois}(c^4/24)$, so $\mathbb{P}[X = 0] \to \exp(-c^4/24)$.

There is much to say about the topology of $Y(n,p)$ for different values of c.

$c = 0$ 1-collapsible. Cohen, D.; Costa, A.; Farber, M.; Kappeler, T. Topology of random 2-complexes. Discrete Comput. Geom. 47 (2012), no. 1, 117–149.

$c = .5$ Giant subcomplex. Cooley, Oliver; Kang, Mihyun; Person, Yury Largest components in random hypergraphs. Combin. Probab. Comput. 27 (2018), no. 5, 741–762.

$c = 2.433\ldots$ Not 1-collapsible. This threshold is at $p = c_2/n$, where c_2 is the infimum of $\eta > 0$ such that $\exp(-\eta(1-x)^2) - x = 0$ has a root $0 < x < 1$. Aronshtam, Lior; Linial, Nathan; Luczak, Tomasz; Meshulam, Roy Collapsibility and vanishing of top homology in random simplicial complexes. Discrete Comput. Geom. 49 (2013), no. 2, 317–334.

Aronshtam, L.; Linial, N. The threshold for d-collapsibility in random complexes. Random Structures Algorithms 48 (2016), no. 2, 260–269.

$c = 2.754\ldots$ $H_2(Y, \mathbb{R}) \neq 0$. Aronshtam, Lior; Linial, Nathan When does the top homology of a random simplicial complex vanish? Random Structures Algorithms 46 (2015), no. 1, 26–35.

Linial, Nathan; Peled, Yuval On the phase transition in random simplicial complexes. Ann. of Math. (2) 184 (2016), no. 3, 745–773.

$c = 2.7538$ Some computational torsion experiments suggest that there is a burst of torsion in $H_1(Y(n,p))$ at $c = 2.7538$. Moreover, at this threshold it is conjectured that the fundamental group goes from free to not free, and $Y(n,p)$ is not embeddable in $\mathbb{R}^4$. Kahle, Matthew; Lutz, Frank H.; Newman, Andrew; Parsons, Kyle Cohen-Lenstra heuristics for torsion in homology of random complexes. Exp. Math. 29 (2020), no. 3, 347–359.

It has been shown by Uli Wagner in 2011 that there exists a constant $c > 0$ so that if $p > c/n$ then w.h.p. $Y(n,p)$ is not embeddable in $\mathbb{R}^4$.

$c = 2\log(n)$ Here, there are no isolated edges, and H_1 vanishes for $\mathbb{Z}/\ell\mathbb{Z}, \mathbb{R}$, and $\mathbb{Z}$.

Linial, Nathan; Meshulam, Roy. Homological connectivity of random 2-complexes. Combinatorica 26 (2006), no. 4, 475–487. Meshulam, R.; Wallach, N. Homological connectivity of random k-dimensional complexes. Random Structures Algorithms 34 (2009), no. 3, 408–417.

Hoffman,,Christopher. Kahle, Matthew, Paquette, Elliot. Spectral graps of random graphs and applications. IMRN. (2019), rnz077.]

Hoffman, Christopher; Kahle, Matthew; Paquette, Elliot The threshold for integer homology in random d-complexes. Discrete Comput. Geom. 57 (2017), no. 4, 810–823. Łuczak, Tomasz; Peled, Yuval Integral homology of random simplicial complexes. Discrete Comput. Geom. 59 (2018), no. 1, 131–142. Newman, Andrew and Paquette, Elliot The integer homology threshold in Yd(n, p). arXiv: 1808.10647

We now turn our attention to results about the fundamental group of $Y(n,p)$. Again suppose $p = c/n$.

$c = 2.433\ldots$ $\pi_1(Y)$ is free. Newman, Andrew Freeness of the random fundamental group. J. Topol. Anal. 12 (2020), no. 1, 29–35.

Aronshtam, L.; Linial, N. The threshold for d-collapsibility in random complexes. Random Structures Algorithms 48 (2016), no. 2, 260–269.

$c = 2.754\ldots$ $\pi_1(Y)$ is not free. Newman, Andrew Freeness of the random fundamental group. J. Topol. Anal. 12 (2020), no. 1, 29–35.

Costa, A. E.; Farber, M. Geometry and topology of random 2- complexes. Israel J. Math. 209 (2015), no. 2, 883–927.

$c = 2\log(n)$ $\pi_1(Y)$ is Kazhdan. Hoffman, Christopher. Kahle, Matthew, Paquette, Elliot Spectral graps of random graphs and applications. IMRN. (2019), rnz077.]

$c = \sqrt{n}$ $\pi_1(Y) = 0$. Babson, Eric; Hoffman, Christopher; Kahle, Matthew The fundamental group of random 2-complexes. J. Amer. Math. Soc. 24 (2011), no. 1, 1–28.

What happens between $c = 2\log(n)$ and $c = \sqrt{n}$? Are there any nontrivial finite quotients of $\pi_1(Y)$? In 2018, Luczak and Peled showed that there are no maps to any abelian groups. Łuczak, Tomasz; Peled, Yuval Integral homology of random simplicial complexes. Discrete Comput. Geom. 59 (2018), no. 1, 131–142.

Meshulam has a preprint that says there are no maps to any fixed finite targets. Meshulam, Roy. Bounded quotients of the fundamental group of a random 2-complex. arXiv:1308.3769

In 2019 Hoffman, Kahle, and Paquette showed that for Kazhdan's Property (T)—not many unitary representations. Hoffman, Christopher. Kahle, Matthew, Paquette, Elliot Spectral graps of random graphs and applications. IMRN. (2019), rnz077.]