Random Simplicial Complexes

Omer Bobrowski
Duke University

CAT-School 2015

Oxford

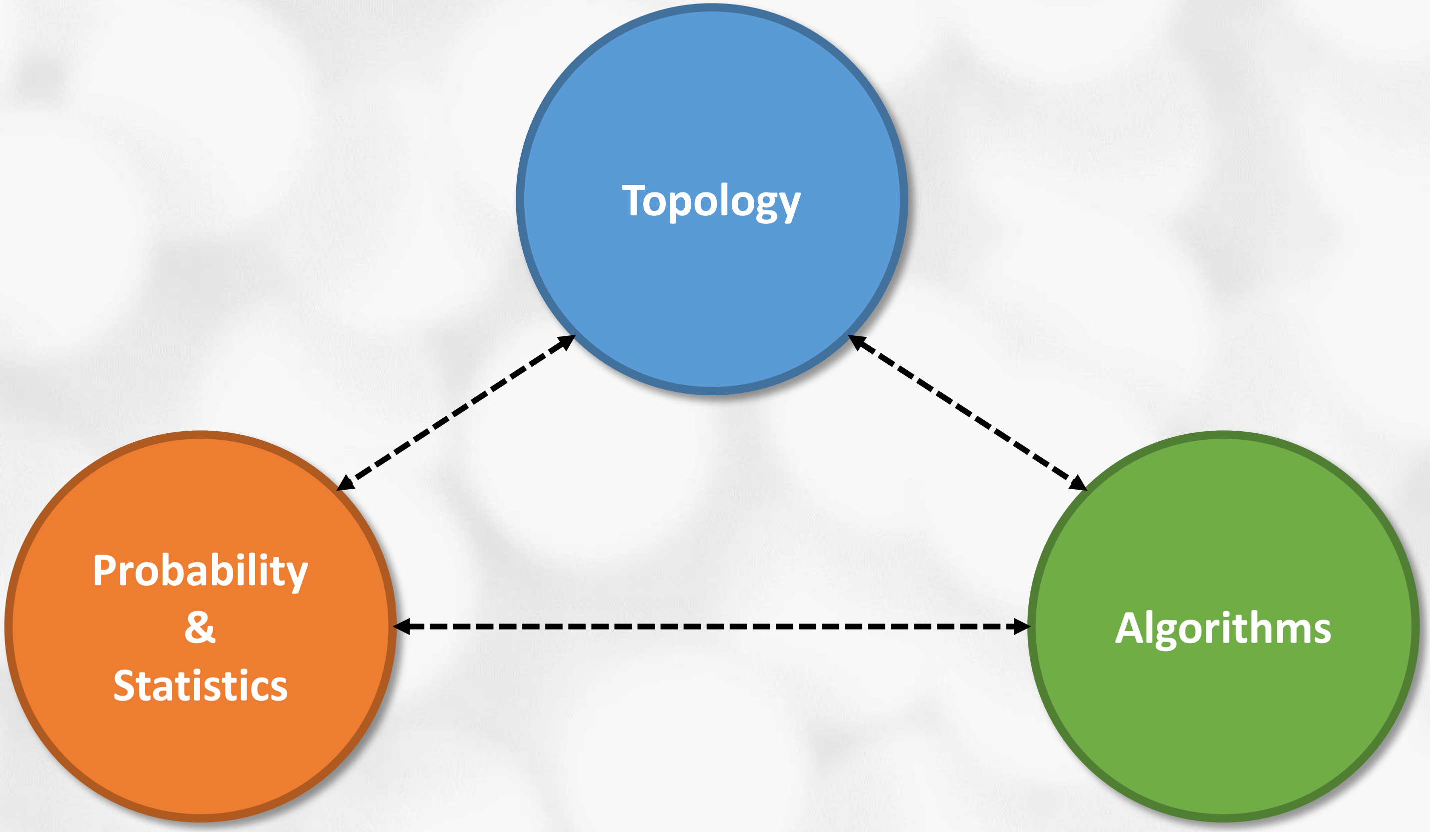
8/9/2015

Part I
Random Combinatorial Complexes

Contents

Contents

Motivation I - TDA



Topological Inference

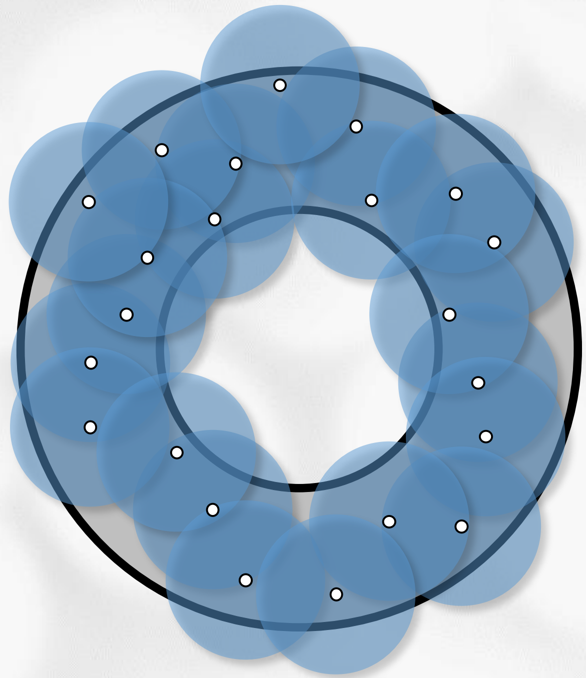
Objective: Study the topology of an unknown space from a set of samples

- Example:

$X = \text{an annulus}$

$\#components = 1$

$\#holes = 1$



$U = \bigcup B_r(x_k)$

$\#components = 12$

$\#holes = 0$

- Problems:
  - How to properly choose r ?
  - How many samples are needed?
  - How to handle noisy samples?
  - How to implement?

Homology Inference

1. How to properly choose r ?
2. How many samples are needed?
3. How to handle noisy samples?
4. How to implement?

- Solutions:
  - 4 - simplicial complex (Čech , Vietoris-Rips, Alpha, etc.)
  - 1,2,3 - Different approaches:
    - Topological:

      Don't choose r, use persistent homology instead
    - Probabilistic / Statistical:

      Given the distribution of the samples, find r,n that guarantee recovery with high probability
    - Combination of the two:

      For example - statistical models for persistence diagrams
- **Bottom line:** We should study how random complexes behave

Motivation II - Random Graphs and Networks

- Various random graph models:
  - Erdős–Rényi: take n vertices, flip a coin for every edge
  - Geometric: take n random points in some metric space, connect by proximity
  - Regular
  - Scale free
  - Small world
- Applications:
  - Network analysis (computer, social, biological)
  - Combinatorics (the probabilistic method)
  - Randomized algorithms
  - Statistical physics
  - Many more…

Graphs $\longrightarrow$ Simplicial Complexes

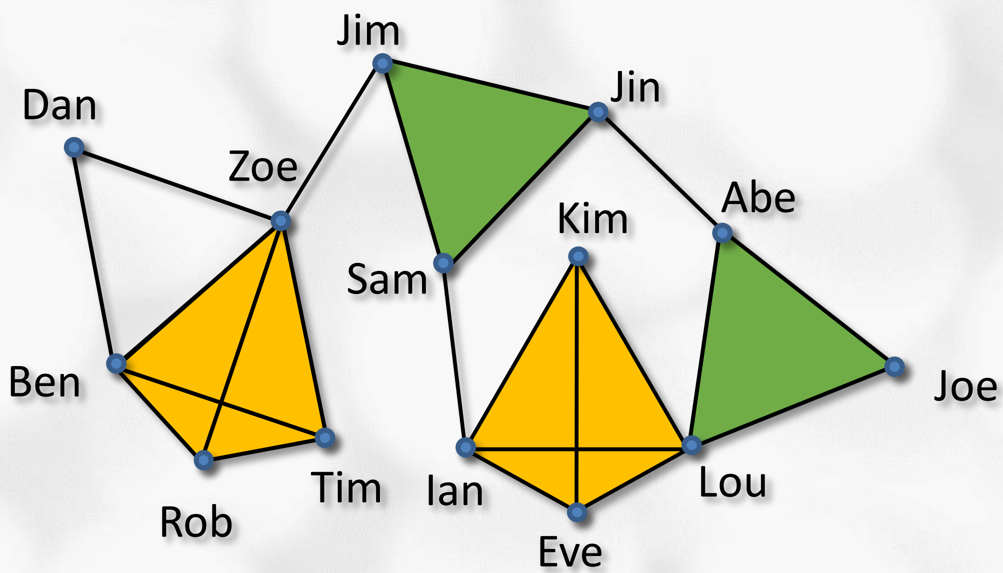
- **Graphs:**

  Modeling pairwise interaction (between computers, sensors, people, etc.)

- **Simplicial complexes:**

  Modeling higher-order interaction

- For example – social / collaboration networks:



Beyond Connectivity & Cycles- Homology

Random graph theory:

- Connectivity:
  - Is a graph connected or not?
  - How many components are there?
  - What can we say about the size of the components?

$H_0(G)$

- Cycles:
  - Is a graph acyclic or not?
  - How many cycles are there?
  - What can we say about cycle sizes?

$H_1(G)$

Random simplicial complexes:

Extend these questions to higher degrees of homology:

- Is $H_k(X)$ trivial or not?
- What is the rank of $H_k(X)$ (Betti numbers)?
- What can we say about the "size" of k-cycles?

Goals

What would we like to know?

- Probability:
  - Distribution for topological quantities (Betti numbers, Euler characteristic, etc.)
  - Phase transitions (appearance/vanishing of homology)
  - Extreme values (outliers and “big” cycles)
  - Ultimately: distributions of barcodes/persistence diagrams
- Statistical TDA:
  - Conditions for homology recovery
  - Likelihood functions, and priors (Bayesian) for barcodes/persistence diagrams
  - Confidence intervals, p-values, error estimates, null models, etc.

Plan

- **Part I (now):**

  Random Combinatorial Complexes (coin-flipping type)

- **Part II (tomorrow):**

  Random Geometric Complexes

- **Part III (Thursday):**

  Extensions & Applications (still Geometric)

Asymptotics

- **Notation:**

$$a_n \approx b_n \Leftrightarrow \frac{a_n}{b_n} \to 1$$

$$a_n \sim b_n \Leftrightarrow \frac{a_n}{b_n} \to \ell \in (0,\infty)$$

$$a_n \ll b_n \Leftrightarrow \frac{a_n}{b_n} \to 0$$

- **With high probability (w.h.p.) / Asymptotically almost surely (a.a.s.):**

E = an event that depends on n

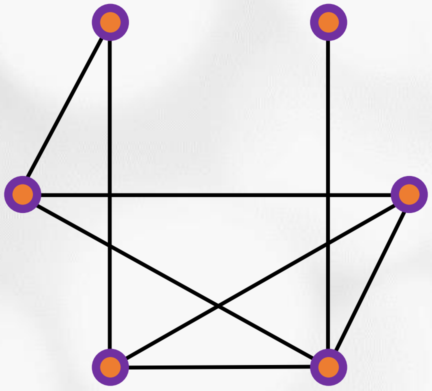
(for example: a random graph with n vertices is connected)

$$E \text{ occurs w.h.p. (or a.a.s.)} \quad \Leftrightarrow \quad \lim_{n\to\infty} \mathbb{P}(E) = 1$$

Contents

The Erdős – Rényi Random Graph

- $G(n,p)$- undirected graph
  - Vertices: $\{1,2,\ldots,n\}$
  - Edges: for every i,j flip a coin – heads $\Rightarrow$ $i \sim j$
- Example:



Goals

Study the asymptotic behavior of $G(n,p)$ as $n \to \infty,\ p = p(n) \to 0.$

More specifically:

- Connectivity
- Cycles
- Distribution of subgraphs
- “Giant” components
- Vertex degree
- Coloring
- Expanders
- More…

Applications - I

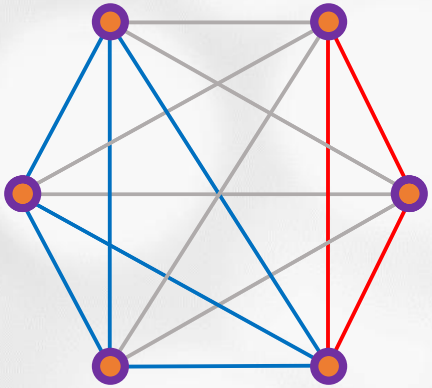
- Network modeling and analysis
- Example – epidemics:
  - n individuals, connected randomly as $G(n,p)$
  - A random individual is infected with a virus
  - α = probability of an individual to be immune
  - How will the epidemic spread?
- One can show that:
  - $\alpha < 1 - \frac{\log n}{np} \quad \Rightarrow \quad \#\text{infected} \sim (1-\alpha)n$ - almost all
  - $1 - \frac{\log n}{np} < \alpha < 1 - \frac{1}{np} \quad \Rightarrow \quad \#\text{infeceted} \sim \beta(1-\alpha)n$ - a fraction
  - $\alpha > 1 - \frac{1}{np} \quad \Rightarrow \quad \#\text{infected} \sim \log n \ll n$ - very few

Applications - II

- Combinatorics – **The Probabilistic Method (Erdős)**

Prove existence of a complicated (nonrandom) object, by showing that a random setting generates such an object with a nonzero probability

- **Example:** Consider a two-coloring of the complete graph on n vertices:



- Look for monochromatic cliques
- The Ramsey number:

$$R(k,\ell) = \min\{n : \underline{\text{every}} \text{ coloring has either a blue } k\text{-clique or a red } \ell\text{-clique}\}$$

Applications - II

$$R(k,\ell) = \min\{n : \underline{\text{every}} \text{ coloring has either a blue } k\text{-clique or a red } \ell\text{-clique}\}$$

Theorem

For $k \ge 3$: $\quad R(k,k) > 2^{k/2-1}$

Proof:

- Consider the random graph - $G(n,p)$, with $p = \frac{1}{2}$. Then:

number of edges

$$\mathbb{P}\left(G(n,1/2) \text{ has a blue or red } k\text{-clique}\right) \le 2\binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} < 2n^k 2^{-\frac{k(k-1)}{2}}$$

red or blue

choose k vertices

color the edges

- If $n = 2^{k/2-1}$ this probability is **< 1**
- $\Rightarrow$ There exists a “bad” graph on $2^{k/2-1}$ vertices $\Rightarrow \; R(k,k) > 2^{k/2-1}$

Connectivity

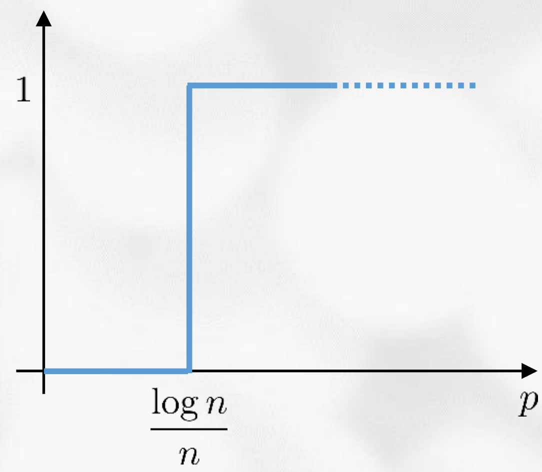
- Probably the first "random topology" result:

Theorem [Erdős & Rényi, 59]

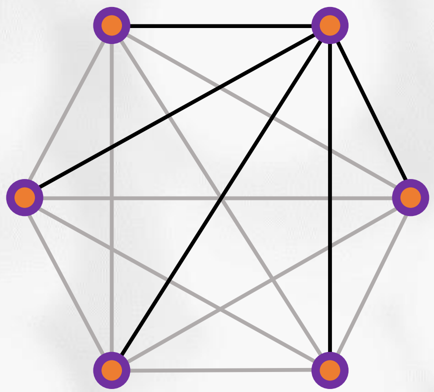
Let $G \sim G(n,p)$, then for any $w(n) \to \infty$,

$$\lim_{n\to\infty} \mathbb{P}\left(G \text{ is connected}\right) = \begin{cases} 1 & p = \frac{\log n + w(n)}{n} \\ 0 & p = \frac{\log n - w(n)}{n} \end{cases}$$

- Phase transition
- Connectivity threshold: $p = \dfrac{\log n}{n}$



Intuition for Connectivity



- The degree distribution:

$$\deg(v_i) \sim \mathrm{Binom}(n-1, p)$$

- Number of isolated vertices:

$$\mathbb{P}(v_i \text{ is isolated}) = (1-p)^{n-1} \approx e^{-np}$$

$$\mathbb{E}\{\#\text{isolated vertices}\} \approx ne^{-np} \to \begin{cases} 0 & \text{if } p = \frac{\log n + w(n)}{n} \\ \infty & \text{if } p = \frac{\log n - w(n)}{n} \end{cases}$$

- Show that the graph consists of a big component + isolated points

Acyclic Graphs

- Another topological result:

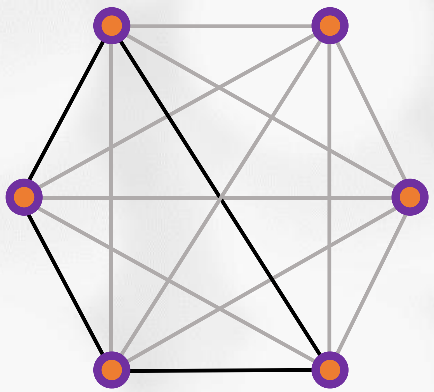
Theorem [Erdős & Rényi, 60]

Let $G \sim G(n,p)$, then

$$\lim_{n\to\infty} \mathbb{P}\left(G \text{ is acyclic}\right) = \begin{cases} 1 & p \ll \frac{1}{n} \\ 0 & p \geq \frac{1}{n} \end{cases}$$

- Threshold for acyclicity: $p = \dfrac{1}{n} \ll \dfrac{\log n}{n}$

Intuition for Acyclicity



- The probability to see a given cycle γ on k vertices:

$$\mathbb{P}\left(\gamma \in G\right) = p^k$$

- Number of all possible cycles on k vertices:

$$\binom{n}{k}\frac{(k-1)!}{2} \approx \frac{n^k}{2k}$$

- Expected number of cycles:

$$\mathbb{E}\left\{N\right\} \approx \sum_{k=3}^{n} \frac{(np)^k}{2k} \quad \begin{cases} 0 & \text{if } p \ll \frac{1}{n} \\ \infty & \text{if } p \geq \frac{1}{n} \end{cases}$$

Giant Component

- L = size of the largest connected component

Theorem [Erdős & Rényi, 60]

Let $p = \frac{c}{n}$. Then:

- $c < 1 \quad \Rightarrow \quad L \sim \log n$
- $c > 1 \quad \Rightarrow \quad L \sim n$

- Threshold for "giant component" emergence: $p = \dfrac{1}{n}$
- Not exactly topological, but still relevant
- (animation)

Modeling Networks

- In many network models we would like to have a "triangle condition":

$$\mathbb{P}\left(i \sim k \mid i \sim j \text{ and } j \sim k\right) > \mathbb{P}\left(i \sim k\right)$$

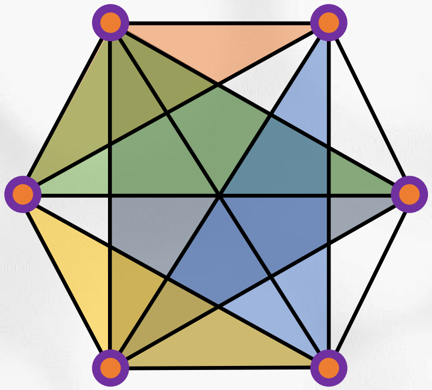
(i and j are friends, j and k are friends $\Rightarrow$ more likely that i and k are friends)

- Not true for $G(n,p)$ - everything is independent
- Degree distribution ~ Poisson, light tail (no "hubs")
- There are more realistic network models…

Contents

The Linial – Meshulam Complex

- Start with the complete graph on n vertices
- For each triangle - flip a coin
- $Y_2(n,p)$ - a random 2-complex
- Example:



- Linial & Meshulam, *Homological connectivity of random 2-complexes*, 2006

Homological Connectivity

Random graphs:

- No edges $\longrightarrow$ all possible components (0-cycles)
- Adding edges $\longrightarrow$ terminating components (0-cycles)
- Connectivity $\Leftrightarrow H_0 = \mathbb{Z} \Leftrightarrow \tilde{H}_0 = 0$

(reduced homology)

Random 2-complexes:

- No triangles $\longrightarrow$ all possible holes (1-cycles)
- Adding triangles $\longrightarrow$ terminating holes (1-cycles)
- Homological connectivity $\Leftrightarrow H_1 = 0$

Homological Connectivity

Theorem [Linial & Meshulam, 06]

Let $Y \sim Y_2(n,p)$, then

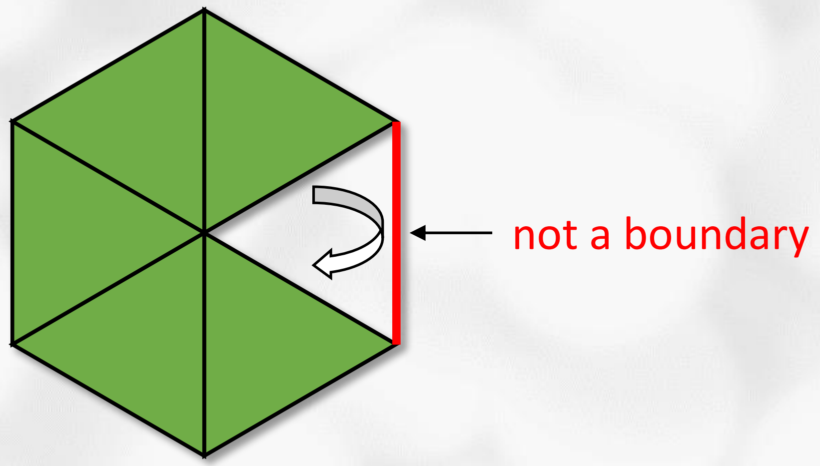
$$\lim_{n\to\infty} \mathbb{P}\left(H_1(Y) = 0\right) = \begin{cases} 1 & p = \frac{2\log n + w(n)}{n} \\ 0 & p = \frac{2\log n - w(n)}{n} \end{cases},$$

where $w(n) \to \infty$.

- Twice the threshold needed for graph connectivity

Intuition for Homological Connectivity

- **Recall:** Graph connectivity $\Leftrightarrow$ isolated points
- **An isolated edge:** Not on the boundary of any 2-simplex
- **Random 2-complexes:** Homological connectivity $\Leftrightarrow$ isolated edges
- **Why?**



- **Alternatively -**
  - For an edge e_0 define the co-chain $g(e) = \begin{cases} 1 & e = e_0 \\ 0 & e \neq e_0 \end{cases}$
  - Then g is a nontrivial co-cycle ($\delta g \equiv 0$)

Random d-Complexes

Construction:

- Start with the <u>full</u> (d–1)-dimensional skeleton on n vertices
- For each d-dimensional simplex - flip a coin
- $Y_d(n,p)$ - the random d-complex

Homological connectivity:

- No d-faces $\rightarrow$ all possible (d–1)-cycles
- Adding d-faces $\rightarrow$ terminating (d–1)-cycles
- Homological connectivity $\Leftrightarrow H_{d-1} = 0$
- $d{=}1 \rightarrow$ Erdős–Rényi graph

Homological Connectivity

Theorem [Meshulam & Wallach, 09]

Let $Y \sim Y_d(n,p)$, then

$$\lim_{n\to\infty} \mathbb{P}\left(H_{d-1}(Y)=0\right) = \begin{cases} 1 & p = \frac{d\log n + w(n)}{n} \\ 0 & p = \frac{d\log n - w(n)}{n} \end{cases},$$

where $w(n) \to \infty$.

- Note: $d{=}1$ – the Erdős–Rényi connectivity result
- A reasonable extension to the notion of “connectivity”
- Homological connectivity ⇔ isolated (d-1)-simplexes

"Acyclic" d-Complex

Random graphs:

- No edges (1-faces) $\longrightarrow$ no cycles
- Adding edges (1-faces) $\longrightarrow$ might create cycles
- Acyclic graph = no cycles $\Leftrightarrow$ $H_1 = 0$

Random d-complexes:

- No d-faces $\longrightarrow$ no d-cycles
- Adding d-faces $\longrightarrow$ might create d-cycles
- "Acyclic d-complex" = no d-cycles $\Leftrightarrow$ $H_d = 0$

Another Phase Transition

Theorem [Kozlov, 10 ; Aronshtam et al. , 2013]

Let $Y \sim Y_d(n,p)$, then

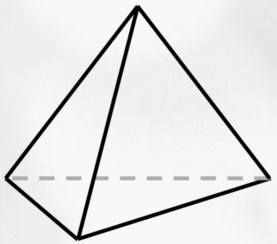
$$\lim_{n\to\infty} \mathbb{P}\left(H_d(Y) \neq 0\right) = \begin{cases} 1 & p \geq \frac{c}{n} \\ 0 & p \ll \frac{1}{n}, \end{cases}$$

where $c > c_d$ (known constant).

- The threshold for acyclicity - $p \sim \dfrac{1}{n}$ (same scale as graphs)

Intuition

$d=2$



$4=d+2$ vertices

- A (d+1)-empty simplex is a nontrivial d-cycle
- How many do we have?

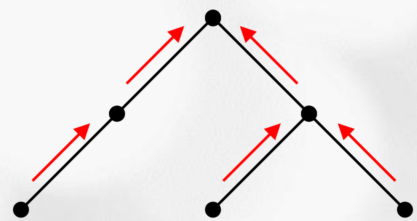
$$\mathbb{E}\left\{N_\Delta\right\} = \binom{n}{d+2} p^{d+2} \sim (np)^{d+2} \to \begin{cases} 0 & \text{if } p \ll \frac{1}{n} \\ \infty & \text{if } p \gg \frac{1}{n} \end{cases}$$

- Need to consider a whole bunch of other structures

Collapsibility

In a graph:

- Pick a free vertex (degree 1), remove it and its edge



- Collapsible = the end result has no edges $\Leftrightarrow$ acyclic $\Leftrightarrow$ $H_1 = 0$

In a k-dimensional simplicial complex:

- Pick a free (k-1)-simplex, remove it and its k-coface

- **Collapsible** = the end result is (k-1)-dimensional $\underset{\not\Leftarrow}{\Rightarrow} H_k = 0$

Collapsibility Threshold

Theorem [Aronshtam & Linial 2014]

Let $Y \sim Y_d(n,p)$, then

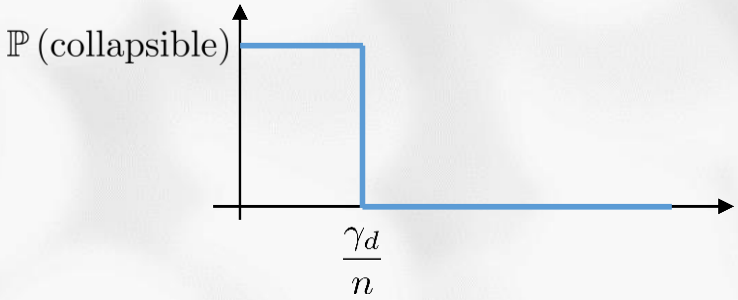
$$\lim_{n\to\infty} \mathbb{P}\left(Y \text{ is not collapsible}\right) = \begin{cases} 1 & p \ge \frac{c}{n} \\ 0 & p \ll \frac{1}{n}, \end{cases}$$

where $c > \gamma_d$ (known constant).

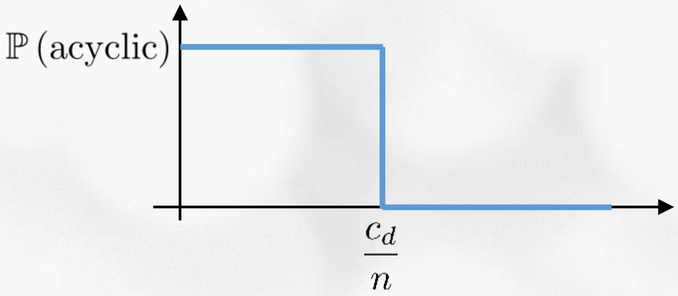
- The threshold for collapsibility - $p \sim \dfrac{1}{n}$
- Same as acyclicity, but a bit earlier - $\gamma_d < c_d$

Random d-Complexes - Conclusion

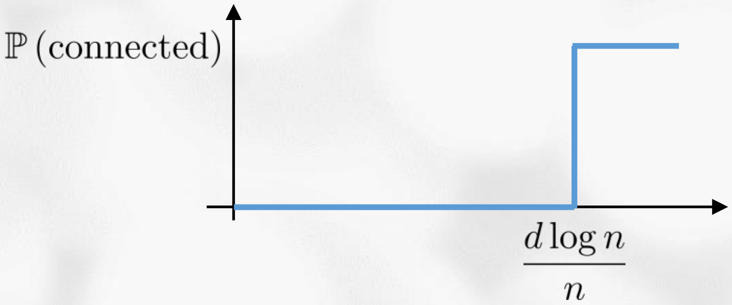
- A simplicial complex where d-faces are added at random
- **Collapsibility:**
  - collapsible $\longrightarrow$ not collapsible
  - Threshold: $p = \frac{\gamma_d}{n}$



- **Acyclicity:**
  - $H_d = 0 \longrightarrow H_d \neq 0$
  - Threshold: $p = \frac{c_d}{n}$



- **Connectivity:**
  - $H_{d-1} \neq 0 \longrightarrow H_{d-1} = 0$
  - Threshold: $p = \frac{d \log n}{n}$



More

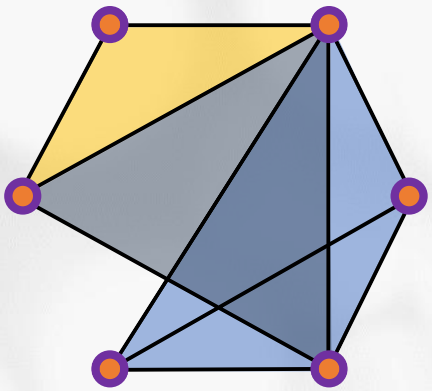
Some other topics that have been studied:

- Expander complexes
- The fundamental group
- The Betti numbers distribution
- Analogue for the giant component emergence (“shadows”)

Contents

Random Clique Complexes

- Start with the Erdős–Rényi graph $G(n,p)$
- Add a k-simplex for every ($k{+}1$)-clique
- $X(n,p)$ - the random clique (flag) complex
- Example:



- Can have simplexes of any dimension

Main Differences

The Linial-Meshulam d-complex:

- Adding d-simplexes at random
- Including all simplexes in dimension $0,\ldots,d\text{–}1$, nothing in dimension $> d$
- The only nontrivial homology is in degrees $d\text{–}1$ and d
- Monotone behavior -

$$\text{(a) } H_{d-1} \neq 0 \longrightarrow H_{d-1} = 0 \qquad \text{(b) } H_d = 0 \longrightarrow H_d \neq 0$$

The random clique complex:

- Adding edges at random
- May have simplexes in any dimension
- Can have nontrivial homology in any degree
- Non-monotone behavior -

$$H_k = 0 \longrightarrow H_k \neq 0 \longrightarrow H_k = 0$$

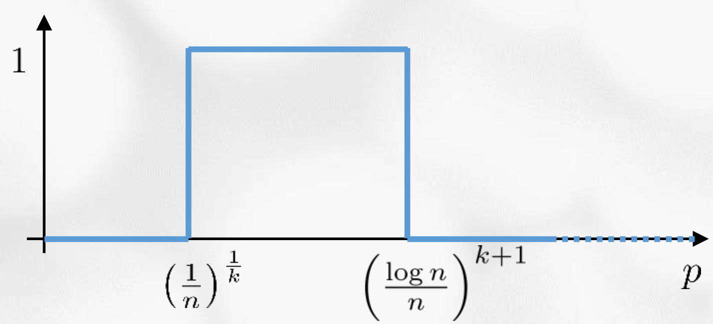
not enough k-faces   good   too many (k+1)-faces

Phase Transitions

- We expect to find <u>two</u> thresholds

Theorem [Kahle, 2009, 2014]

$$\lim_{n\to\infty} \mathbb{P}\left(H_k(X) \neq 0\right) = \begin{cases} 0 & p \ll n^{-1/k}, \\ 1 & n^{-1/k} \ll p \ll \left(\frac{\log n}{n}\right)^{\frac{1}{k+1}}, \\ 0 & p \gg \left(\frac{\log n}{n}\right)^{\frac{1}{k+1}} \end{cases}$$



- Connectivity = vanishing of $H_0 \longrightarrow p = \frac{\log n}{n}$ $(k=0)$ - Erdős–Rényi
- Completely different scales than the Linial-Meshulam thresholds

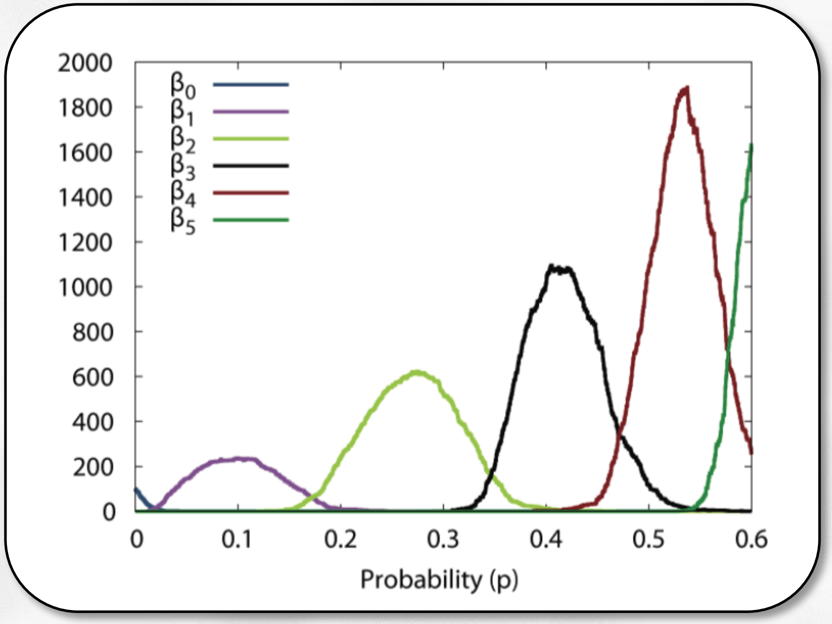
Clique Complexes - Summary

*ignoring the log-scale $(n^{-1-\epsilon} \ll \frac{\log n}{n} \ll n^{-1+\epsilon})$

$H_0 \neq 0$ $H_1 \neq 0$ $H_2 \neq 0$ $H_3 \neq 0$ $H_4 \neq 0$ ...

0 n^{-1} $n^{-\frac{1}{2}}$ $n^{-\frac{1}{3}}$ $n^{-\frac{1}{4}}$ $n^{-\frac{1}{5}}$

Different degrees are almost separated



Topology of random geometric complexes: a survey, M. Kahle, 2013

The Expected Degree

- For a k-simplex σ in a simplicial complex:

$$\deg(\sigma) = \#(k+1) - \text{simplexes containing } \sigma$$

- Expected degree in the Linial-Meshulam k-complex:

$$\mathbb{E}\left\{\deg(\sigma)\right\} = (n-k-1)p \approx np$$

- Expected degree of a k-simplex in the Clique complex:

$$\mathbb{E}\left\{\deg(\sigma)\right\} = (n-k-1)p^{k+1} \approx np^{k+1}$$

- Vanishing of k-cycles:
  - Linial-Meshulam: $p \sim \frac{\log n}{n}$
  - Clique: $p \sim \left(\frac{\log n}{n}\right)^{1/k+1}$

$$\left.\begin{array}{l} p \sim \frac{\log n}{n} \\ p \sim \left(\frac{\log n}{n}\right)^{1/k+1} \end{array}\right\} \mathbb{E}\left\{\deg\right\} \sim \log n$$

The Betti Numbers

- Recall:

  If $n^{-\frac{1}{k}} \ll p \ll \left(\frac{\log n}{n}\right)^{\frac{1}{k+1}}$ then $H_k(X(n,p)) \neq 0$

- Question:

  How many cycles do we expect to see?

Theorem [Kahle, 2009 ; Kahle & Meckes, 2013]

Let $\beta_k = \mathrm{rank}(H_k(X(n,p)))$.

- Let F_k be the number of k-faces in $X(n,p)$, then

$$\mathbb{E}\{\beta_k\} \approx \mathbb{E}\{F_k\} = \binom{n}{k+1} p^{\binom{k+1}{2}}$$

- Central limit theorem (CLT):

$$\frac{\beta_k - \mathbb{E}\{\beta_k\}}{\sqrt{\mathrm{Var}(\beta_k)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Next:

Random Geometric Complexes