Random Simplicial Complexes

Omer Bobrowski
Duke University

CAT-School 2015

Oxford

9/9/2015

Part II
Random Geometric Complexes

Contents

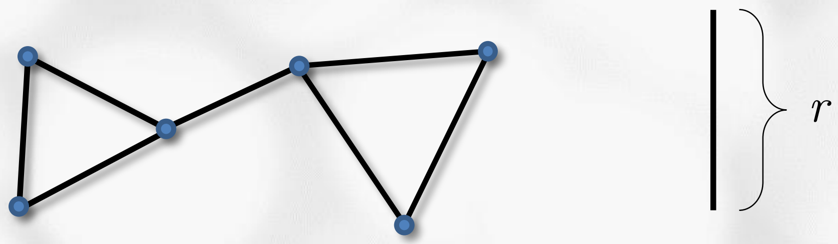
Contents

Contents

Random Graphs and Neworks

Random Geometric Graphs (RGG)

- $\mathcal{G}(\mathcal{X}, r)$ - undirected graph
  - Vertices: $\mathcal{P}$ – either the binomial ($\mathcal{X}_n$) or the Poisson ($\mathcal{P}_n$) process
  - Edges: for every $x,y \in \mathcal{P}$, $x \sim y$ iff $d(x,y) \le r$



- Triangle condition: $\mathbb{P}(x \sim z \mid x \sim y \text{ and } y \sim z) > \mathbb{P}(x \sim z)$
- Drawback – more difficult to analyze
- Common application:

  Wireless networks – transceivers with transmission range r are deployed in a region

Connectivity

- From here on:
  - $f : \mathbb{T}^d \to \mathbb{R}$- the uniform distribution ($f \equiv 1$)
  - $G(n,r) = G(\mathcal{X}_n, r)$
  - ω_d= volume of a d-dimensional unit ball

Theorem [Penrose, 97]

For $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(G(n,r) \text{ is connected}\right) = \begin{cases} 1 & r = \left(\frac{(1+\epsilon)\log n}{\omega_d n}\right)^{\frac{1}{d}} \\ 0 & r = \left(\frac{(1-\epsilon)\log n}{\omega_d n}\right)^{\frac{1}{d}} \end{cases}$$

- Connectivity threshold: $r = \left(\frac{\log n}{\omega_d n}\right)^{1/d}$
- Other distributions – compact support $\longrightarrow$ same results (up to constants)
- Compare to Erdős–Rényi : $p = \frac{\log n}{n}$

Cycles in Random Geometric Graphs

- From here on: $\Lambda = \omega_d n r^d$

Theorem [?]

$$\lim_{n\to\infty} \mathbb{P}\left(G(n,r) \text{ is acyclic}\right) = \begin{cases} 1 & \Lambda \gg \sqrt{1/n} \\ 0 & \Lambda \ll \sqrt{1/n} \end{cases}$$

- Acyclicity threshold: $\Lambda = \sqrt{1/n}$
- In Erdős–Rényi: $np = 1 \quad (p = 1/n)$

} very different

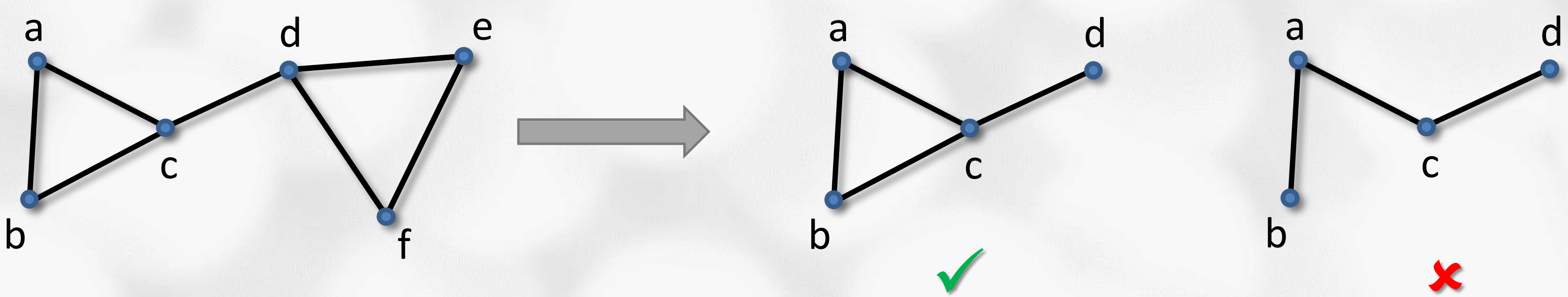
- Partial explanation:

The triangle condition $\to$ "easier" to form tringles in $G(n,r)$ than in $G(n,p)$

Induced Subgraphs

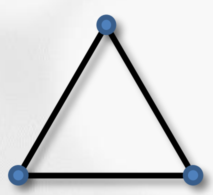
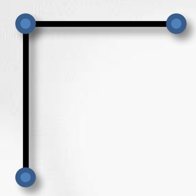
- **Induced subgraphs of $G(n,r)$:**

  All the graphs of the form $G(S,r)$ where $S \subset \mathcal{X}_n$



- Γ = a graph on k vertices

- N_Γ = the number induced subgraphs of $G(n,r)$ that are isomorphic to Γ

$\Gamma_1 =$  $\Gamma_2 =$  $\Rightarrow \quad N_{\Gamma_1} = 2, \quad N_{\Gamma_2} = 4$

- **Q:** What is $\mathbb{E}\{N_\Gamma\}$?

The Expected Number of Subgraphs

- Recall: $\Lambda = \omega_d n r^d$

Theorem

Let Γ be a graph on k vertices, and N_Γ the number of induced subgraphs of $G(n,r)$ isomorphic to Γ. Then

$$\mathbb{E}\left\{N_\Gamma\right\} \approx n\Lambda^{k-1}\mu_\Gamma,$$

where μ_Γ depends on Γ, k, d only (not on n, r).

- Example – triangles (k=3):

$$\mathbb{E}\left\{N_\Delta\right\} \approx n\Lambda^2\mu_\Delta \rightarrow \begin{cases} 0 & \text{if } \Lambda \ll 1/\sqrt{n} \\ \infty & \text{if } \Lambda \gg 1/\sqrt{n} \end{cases}$$

acyclicity

Giant Component

- Recall: L = largest connected component

Theorem [Penrose & Pisztora, 96; Penrose 03]

Let $\Lambda = \lambda$ (or $r = \left(\frac{\lambda}{\omega_d n}\right)^{1/d}$). There exists $\lambda_c > 0$ such that

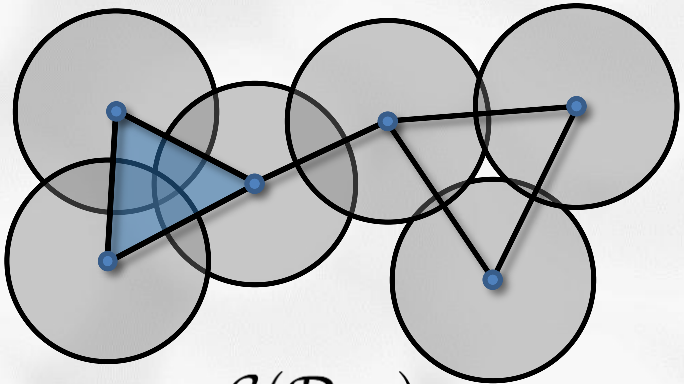
- $\lambda < \lambda_c \quad \Rightarrow \quad L \sim \log n$
- $\lambda > \lambda_c \quad \Rightarrow \quad L \sim n$

- Similar to Erdős–Rényi: $np = 1$
- Continuum percolation

Contents

The Čech Complex

- Take a set of vertices $\mathcal{P}$ in a metric space (0-simplexes)
- Draw balls with radius $r/2$
- Intersection of 2 balls → an edge (1-simplex) $\longrightarrow G(n,r)$
- Intersection of 3 balls → a triangle (2-simplex)
- Intersection of k balls → a (k-1)-simplex



$\mathcal{C}(\mathcal{P}, r)$

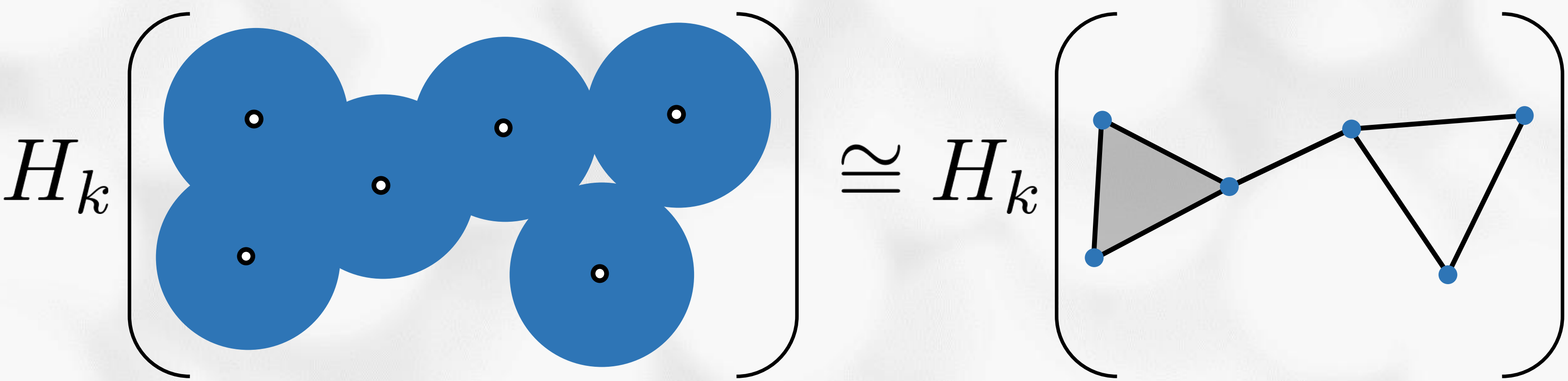
The Nerve Lemma

$$\mathcal{U}(\mathcal{P}, r) := \bigcup_{p\in\mathcal{P}} B_{r/2}(p)$$

Lemma [Borsuk, 48]

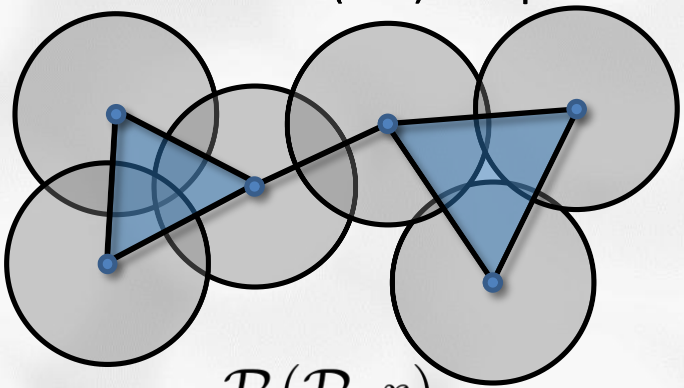
The Čech complex $\mathcal{C}(\mathcal{P}, r)$ is homotopy equivalent to $\mathcal{U}(\mathcal{P}, r)$.

In particular, $H_k\left(\mathcal{U}(\mathcal{P}, r)\right) \cong H_k\left(\mathcal{C}(\mathcal{P}, r)\right)$.



The Vietoris - Rips Complex

- Take a set of vertices $\mathcal{P}$ in a metric space (0-simplexes)
- Draw balls with radius $r/2$
- Intersection of 2 balls → an edge (1-simplex) $\longrightarrow G(n,r)$
- All pairwise intersections of 3 balls → a triangle (2-simplex)
- All pairwise intersections of k balls → a (k-1)-simplex



$\mathcal{R}(\mathcal{P},r)$

Some Useful Facts

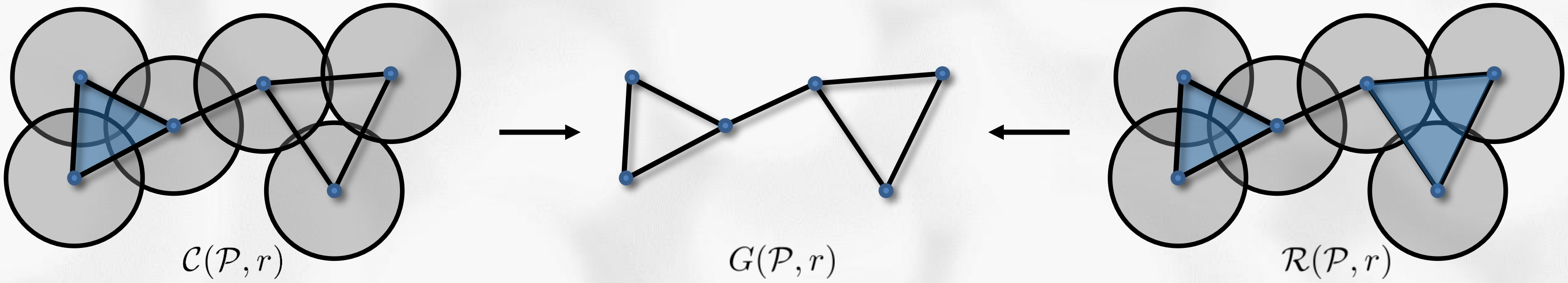
- Even if $\mathcal{P} \subset \mathbb{R}^d$ it is possible for $\mathcal{C}(\mathcal{P},r), \mathcal{R}(\mathcal{P},r)$ to have simplexes in any dimension $\geq d$

- Still: $H_k(\mathcal{C}(\mathcal{P},r)) = 0$ for $k \geq d$ (Nerve Lemma)

- Rips "approximates" Čech (de-Silva & Ghrist, 07):

$$\mathcal{R}(\mathcal{P},\tfrac{1}{\sqrt{2}}r) \subset \mathcal{C}(\mathcal{P},r) \subset \mathcal{R}(\mathcal{P},r)$$

- The 1-skeleton of both $\mathcal{C}(\mathcal{P},r), \mathcal{R}(\mathcal{P},r)$ is the same geometric graph



Contents

Overview

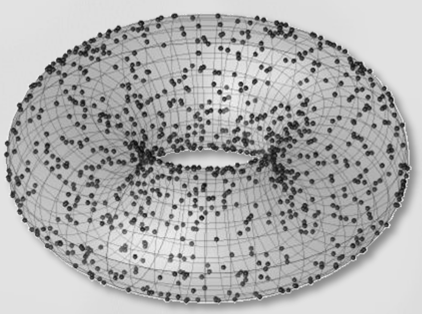
- **Goal:**

  Study the limiting behavior of $H_k(\mathcal{C}(n,r))$ and $H_k(\mathcal{R}(n,r))$ as $n \to \infty$, $r = r(n) \to 0$

- **Three main regimes:**

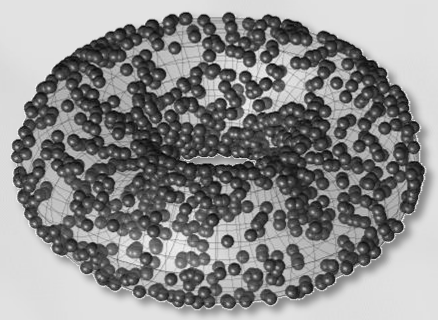
$$\Lambda := \omega_d n r^d$$ ~ vertex degree

Subcritical



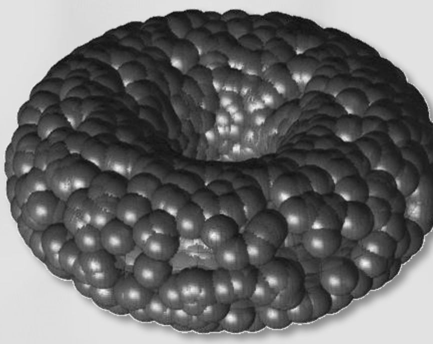
$\Lambda \to 0$

Critical



$\Lambda \to \lambda \in (0, \infty)$

Supercritical



$\Lambda \to \infty$

Connected Components (H_0)

- Random Geometric Graphs Theory (Penrose, Bollobás and others)
- Subcritical ($\Lambda \to 0$):

$$\frac{\beta_0}{n} \to 1 \qquad \text{(dust)}$$

- Critical ($\Lambda \to \lambda$):

$$\frac{\beta_0}{n} \to C_\lambda < 1 \qquad \text{(continuum percolation)}$$

- Supercritical ($\Lambda \to \infty$):

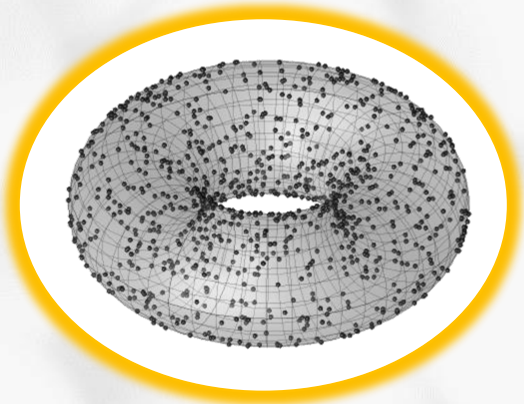
$$\beta_0 = o(n)$$

- Connectivity threshold - Λ_0:

$$\Lambda_0 = \log n \qquad r = \left(\frac{\log n}{\omega_d n}\right)^{1/d}$$

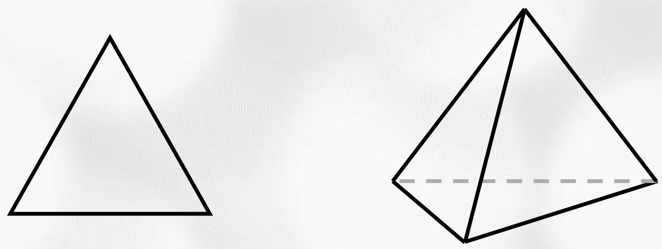
The Subcritical Regime ($\Lambda \to 0$)

- Complex is very sparse
- $\Rightarrow$ Homology is dominated by minimal & isolated cycles



- Čech:

$k=1$ $k=2$

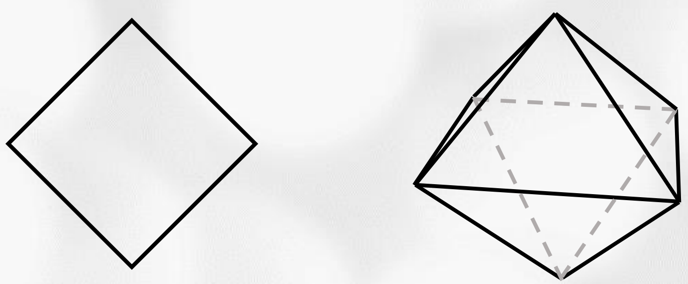


3 vertices 4 vertices

k-cycle $\Rightarrow$ empty $(k+1)$-simplex $\Rightarrow$ $k+2$ vertices

- Rips:

$k=1$ $k=2$



4 vertices 6 vertices

k-cycle $\Rightarrow$ empty cross-polytope $\Rightarrow$ $2k+2$ vertices

Betti Numbers Approximation

- Consider Čech the complex
- Define:

$$S_k = \#\text{isolated empty } (k+1)\text{-simplexs } (k+2 \text{ vertices})$$

$$F_k = \#k\text{-simplexes in components with at least } k+3 \text{ vertices}$$

- Claim:

$$S_k \leq \beta_k \leq S_k + F_k$$

- We can show:

$$F_k \ll S_k$$ (sparse regime)

The Limiting Mean

- Similarly to subgraph counting:

Theorem [Kahle, 2011]

If $\Lambda \to 0$, then

- For the Čech complex $\mathcal{C}(n,r)$:

$$\mathbb{E}\{\beta_k\} \approx n\Lambda^{k+1}\mu_k^C, \qquad 1 \le k \le d-1$$

- For the Rips complex $\mathcal{R}(n,r)$:

$$\mathbb{E}\{\beta_k\} \approx n\Lambda^{2k+1}\mu_k^R, \qquad k \ge 1$$

- Where:

$$\mu_k^C = \frac{1}{(k+2)!}\int h_1^C(0,\mathbf{y})d\mathbf{y} \qquad \mu_k^R = \frac{1}{(2k+2)!}\int h_1^R(0,\mathbf{y})d\mathbf{y}$$

$$h_r^C(S) = \mathbb{1}\{\mathcal{C}(S,r) \text{ is an empty simplex}\} \qquad h_r^R(S) = \mathbb{1}\{\mathcal{R}(S,r) \text{ is an empty cross-polytope}\}$$

Limit Theorems

Poisson Approximation

Let $Z_1, Z_2, \ldots$ - be i.i.d., $Z_i \in \{0,1\}$, with $\mathbb{P}(Z_i = 1) = p$.

Let $W = \sum_{i=1}^n Z_i$. If $p = \frac{c}{n}$ then

$$W \xrightarrow{\mathcal{L}} \text{Poisson}(c)$$

The Central Limit Theorem (CLT)

Let $Z_1, Z_2, \ldots$ - be i.i.d., with $\mathbb{E}\{Z_i\} = \mu$, and $\text{Var}(Z_i) = \sigma^2 < \infty$.

Let $W = \frac{\sum_{i=1}^n (Z_i - \mu)}{\sqrt{n\sigma^2}}$. Then

$$W \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

- The Betti numbers:

Let $\mathbf{i} = (i_1, \ldots, i_{k+1})$, and $Z_{\mathbf{i}} = \mathbb{1}\{X_{i_1}, \ldots, X_{i_{k+1}} \text{ form a min. isolated cycle}\}$

$$\beta_k \approx \sum_{\mathbf{i}} Z_{\mathbf{i}}$$

- not independent

Limiting Distribution

Čech: $\mathbb{E}\{\beta_k\} \approx n\Lambda^{k+1}\mu_k^C$

Theorem [Kahle & Meckes, 2013]

If $\Lambda \to 0$, and $k \geq 1$,

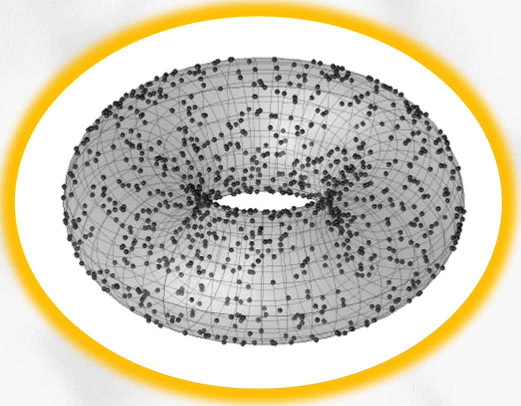
- $\mathrm{Var}(N_k) \approx n\Lambda^{k+1}\mu_k^C \approx \mathbb{E}\{\beta_k\}$
- $n\Lambda^{k+1} \to 0 \quad \Rightarrow \quad \beta_k \xrightarrow{L^2} 0$
- $n\Lambda^{k+1} \to a \in (0,\infty) \quad \Rightarrow \quad \beta_k \xrightarrow{\mathcal{L}} \mathrm{Poisson}\left(a\mu_k^C\right)$
- $n\Lambda^{k+1} \to \infty \quad \Rightarrow \quad \dfrac{\beta_k - \mathbb{E}\{\beta_k\}}{\sqrt{n\Lambda^{k+1}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\mu_k^C)$

- Proofs use Stein's method (limits of sums of "weakly dependent" variables)

The Subcritical Regime ($\Lambda \to 0$)

- Exact limit values are known, as well as limit distributions
- For example:
  - Čech: $\beta_k \sim n\Lambda^{k+1}, \quad k = 1,\ldots,d-1$
  - Rips: $\beta_k \sim n\Lambda^{2k+1}, \quad k = 1,2,\ldots$

$$\Rightarrow \quad n \approx \beta_0 \gg \beta_1 \gg \beta_2 \gg \cdots$$



- Phase transition - appearance:

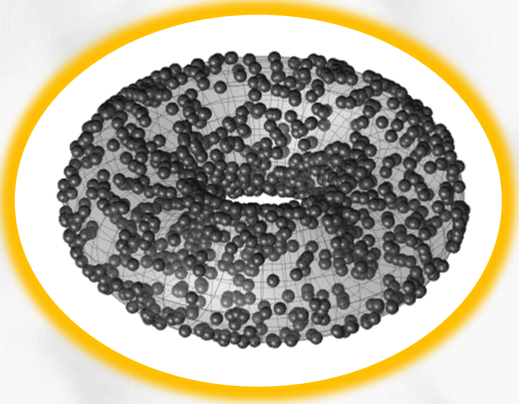
Čech: $\Lambda_k^+ = n^{-\frac{1}{k+1}}$ Rips: $\Lambda_k^+ = n^{-\frac{1}{2k+1}}$

$$\lim_{n\to\infty} \mathbb{P}\left(H_k \neq 0\right) = \begin{cases} 0 & \Lambda \ll \Lambda_k^+ \\ 1 & \Lambda \gg \Lambda_k^+. \end{cases}$$

- Behavior is independent of f (constants might be different)

Kahle, 2011
Kahle & Meckes, 2013
B & Mukherjee, 2015

The Critical Regime ($\Lambda \to \lambda$)



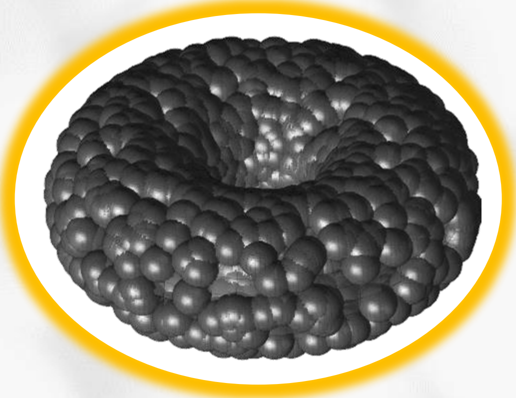
- Cycles are neither minimal nor isolated
- Inequality: $S_k \le \beta_k \le S_k + F_k$

  No longer true that: $F_k \ll S_k$
- Scale is known: $\beta_k \sim n, \quad k \ge 0$
- Law of large number & central limit theorems are proved
- Limiting constants are <u>unknown</u>
- Euler characteristic (later)
- Behavior is independent of f and $\mathrm{supp}(f)$

Kahle, 2011
B. & Adler, 2014
Yogeshwaran, Subag & Adler, 2014
B. & Mukherjee, 2015

The Subcritical Regime ($\Lambda \to \infty$)

- Highly connected, almost everything is covered
- β_k decays from $\sim n$ (critical) to ~ 1 (coverage)
- Recall: $\mathrm{supp}(f) = d$-dimensional torus
- Phase transition – "vanishing":



Threshold: $\Lambda^- = 2^d \log n$ (connectivity threshold = $\log n$)

$$\lim_{n\to\infty} \mathbb{P}\left(H_k(\mathcal{C}(n,r)) \cong H_k(\mathbb{T}^d)\right) = \begin{cases} 1 & \Lambda = (1+\epsilon)\Lambda^- \\ 0 & \Lambda = (1-\epsilon)\Lambda^- \end{cases}$$

$k \geq 1$

NSW, 2008
Kahle, 2011
B. & Mukherjee, 2015
B. & Weinberger, 2015

The Expected Betti Numbers

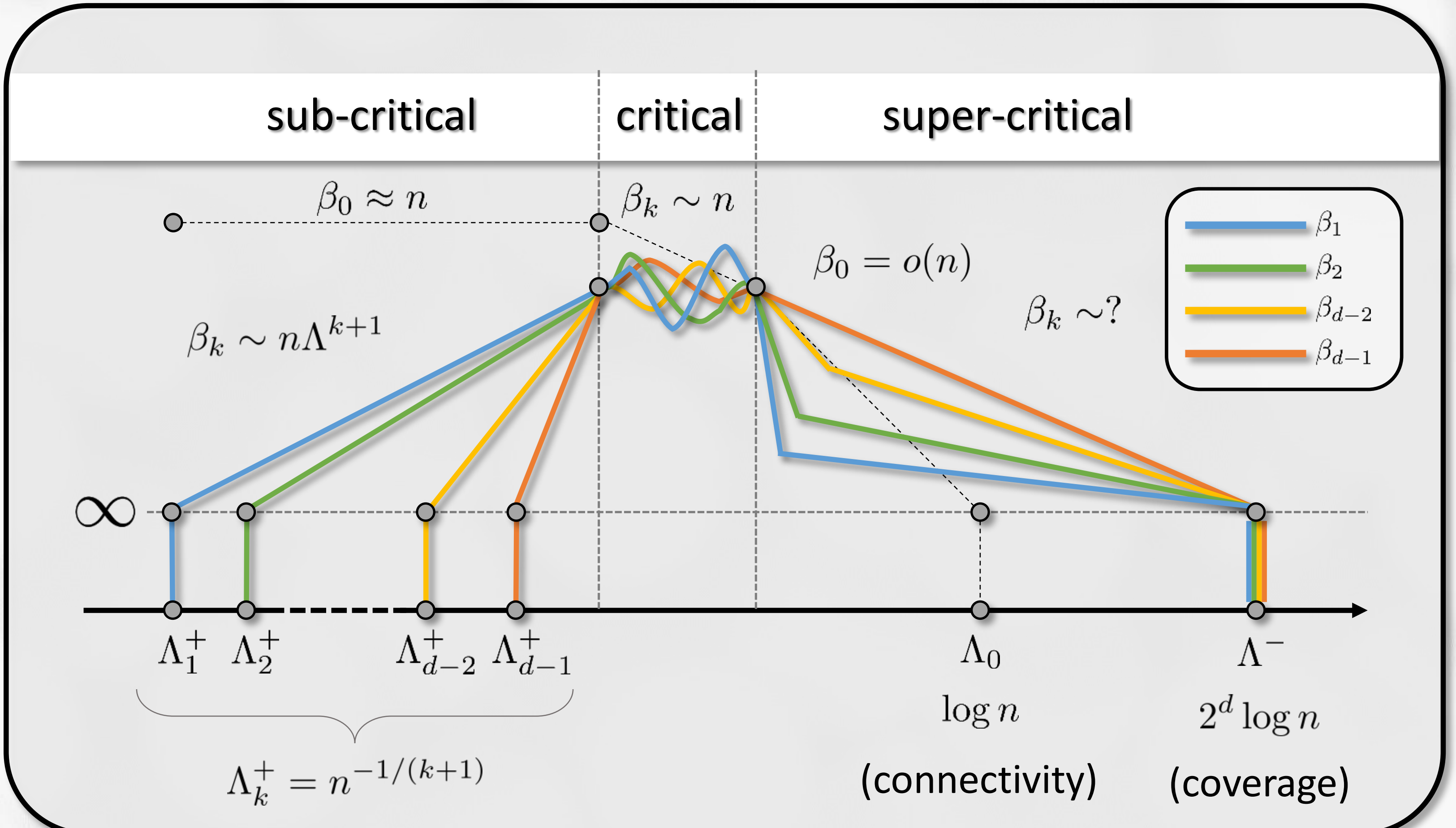
- Using Morse Theory, we can show in addition that:

Theorem [B. & Weinberger, 15]

For $1 \le k \le d$, if $\Lambda \to \infty$ then

$$\mathbb{E}\left\{\beta_k(r)\right\} = \beta_k(\mathbb{T}^d) + O(n\Lambda^k e^{-\Lambda/2^d})$$

Summary



B & Kahle – *Topology of Random Geometric Complexes: a survey*

Topological Inference

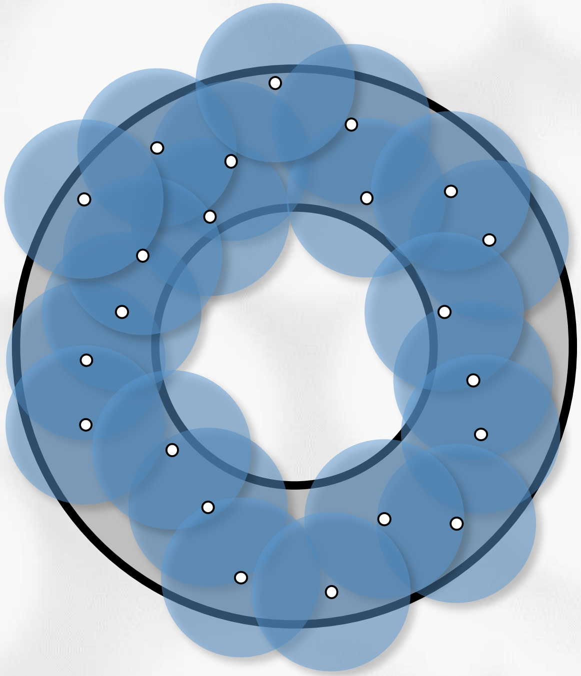
- **Objective:** Study the homology of an unknown space from a set of samples

- **Example:**

$X = \text{an annulus}$

$\beta_0(X) = 1$

$\beta_1(X) = 1$



$U = \bigcup B_r(x_k)$

$\beta_0(U) = 1$

$\beta_1(U) = 1$

- **Problem:** How should we choose r ?

Larger sample $n \to \infty$ → Smaller radius $r \to 0$

More General Manifolds

- $\mathcal{M} \subset \mathbb{R}^D$ - closed manifold, with $\dim(\mathcal{M}) = d$
- $f : \mathcal{M} \to \mathbb{R}$ - a probability density function, with $f_{\min} = \inf_{x \in \mathcal{M}} f(x) > 0$
- $X_1, X_2, \ldots, X_n$ - iid random samples, generated from f
- $n \to \infty, \quad r = r(n) \to 0$

Theorem [B & Mukherjee]

If $\Lambda \geq \frac{1+\epsilon}{f_{\min}} \log n$, then

$$\lim_{n \to \infty} \mathbb{P}\left(H_*(\mathcal{C}(n,r)) \cong H_*(\mathcal{M})\right) = 1$$

- Similar to the torus $\mathbb{T}^d$, but here – showing coverage is not enough
- Morse Theory helps (later)

Geometric vs. Clique Complexes

- Clique complexes - $H_k \neq 0$ if :

$$n^{-\frac{1}{k}} \ll p \ll \left(\frac{\log n}{n}\right)^{\frac{1}{k+1}}$$

$$\Downarrow$$

Degree: $$n^{\frac{k-1}{k}} \ll np \ll (n^k \log n)^{\frac{1}{k+1}}$$

∞ (past connectivity)

Different degrees are almost separated

- Geometric Complexes - $H_k \neq 0$ if:

$$n^{-\alpha^k} \ll \Lambda \ll \log n$$

0 (way before connectivity)

Different degrees coexist

Next:
Extensions and Applications