Phase Transitions in Random Hypergraphs

Mihyun Kang

Joint work with Oliver Cooley and Christoph Koch

Emergence of Giant Component in $G(n, p)$

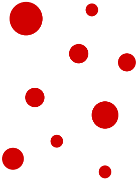
$L_1(d)$ $=$ # vertices in the largest component with $d = p \cdot (n-1)$

Emergence of Giant Component in $G(n, p)$

$L_1(d) = \#$ vertices in the largest component with $d = p \cdot (n-1)$

Theorem [Erdős–Rényi 60]

- If $d < 1$, whp $\quad L_1(d) = O(\log n)$



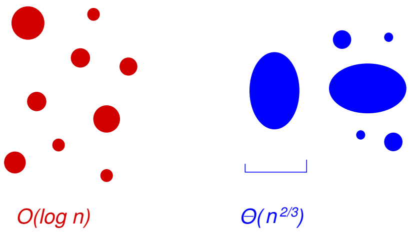
O(log n)

Emergence of Giant Component in $G(n, p)$

$L_1(d) = \#$ vertices in the largest component with $d = p \cdot (n-1)$

Theorem [ERDŐS–RÉNYI 60]

- If $d < 1$, whp $L_1(d) = O(\log n)$
- If $d = 1$, whp $L_1(d) = \Theta(n^{2/3})$

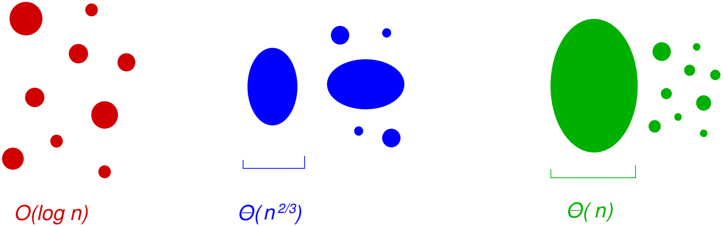


Emergence of Giant Component in $G(n, p)$

$L_1(d) = \#$ vertices in the largest component with $d = p \cdot (n-1)$

Theorem [Erdős–Rényi 60]

- If $d < 1$, whp $L_1(d) = O(\log n)$
- If $d = 1$, whp $L_1(d) = \Theta(n^{2/3})$
- If $d > 1$, whp $L_1(d) = \Theta(n)$



Critical Phenomenon in $G(n, p)$

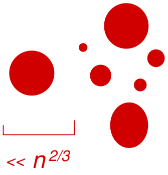
Let $d = p \cdot (n-1) = 1 + \epsilon$ for $\epsilon = o(1)$.

Critical Phenomenon in $G(n, p)$

Let $d = p \cdot (n-1) = 1 + \epsilon$ for $\epsilon = o(1)$.

Theorem [Bollobás 84; Łuczak 90; Aldous 97]

- If $\epsilon^3 n \to -\infty$, whp $L_1(d) \sim 2\epsilon^{-2} \log \epsilon^3 n \ll n^{2/3}$

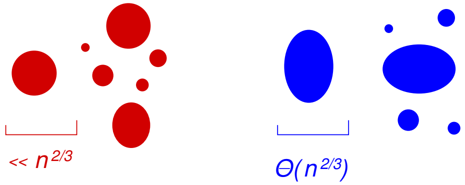


Critical Phenomenon in $G(n, p)$

Let $d = p \cdot (n-1) = 1 + \epsilon$ for $\epsilon = o(1)$.

Theorem [Bollobás 84; Łuczak 90; Aldous 97]

- If $\epsilon^3 n \to -\infty$, whp $L_1(d) \sim 2\epsilon^{-2} \log \epsilon^3 n \ll n^{2/3}$
- If $\epsilon^3 n \to \lambda \in (-\infty, \infty)$, whp $L_1(d) = \Theta(n^{2/3})$

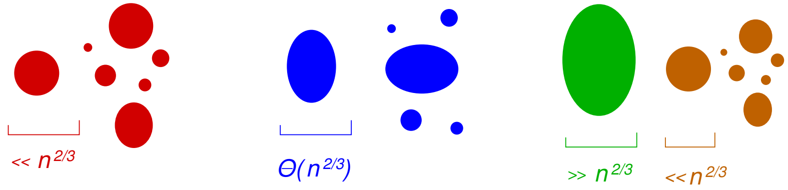


Critical Phenomenon in $G(n,p)$

Let $d = p \cdot (n-1) = 1 + \epsilon$ for $\epsilon = o(1)$.

Theorem [Bollobás 84; Łuczak 90; Aldous 97]

- If $\epsilon^3 n \to -\infty$, whp $L_1(d) \sim 2\epsilon^{-2} \log \epsilon^3 n \ll n^{2/3}$
- If $\epsilon^3 n \to \lambda \in (-\infty, \infty)$, whp $L_1(d) = \Theta(n^{2/3})$
- If $\epsilon^3 n \to +\infty$, whp $L_1(d) \sim 2\epsilon n \gg n^{2/3}$



Asymptotic Normality of Giant Component

Assume $d = p \cdot (n-1) > 1$ and $0 < \rho < 1$ satisfies $1 - \rho = e^{-d \cdot \rho}$.

Let $\mu := \rho \cdot n$ and $\sigma^2 := \frac{\rho(1-\rho)}{(1-d(1-\rho))^2} \cdot n$

Asymptotic Normality of Giant Component

Assume $d = p \cdot (n-1) > 1$ and $0 < \rho < 1$ satisfies $1 - \rho = e^{-d \cdot \rho}$.

Let $\mu := \rho \cdot n$ and $\sigma^2 := \frac{\rho(1-\rho)}{(1-d(1-\rho))^2} \cdot n$

Central limit theorem

Let $N(0,1)$ denote the standard normal distribution. Then

$$\frac{L_1(d) - \mu}{\sigma} \xrightarrow{\mathrm{d}} N(0,1)$$

for d constant [STEPANOV 70; BEHRISCH–COJA-OGHLAN–K. 09]

for $(d-1)^3 n \to \infty$ [PITTEL–WORMALD 05; BOLLOBÁS–RIORDAN 12]

Proof techniques

- Counting connected graphs inside-out [PW 05]
- Stein's method [BC-OK 09]
- Random walk [BR 12]

Local Limit Theorem for Giant Component

Assume $d \; = \; p \cdot (n-1) \; > \; 1$ and $0 < \rho < 1$ satisfies $1 - \rho = e^{-d \cdot \rho}$.

Let $\mu := \rho \cdot n$ and $\sigma^2 := \frac{\rho(1-\rho)}{(1-d(1-\rho))^2} \cdot n$

Theorem [Stepanov 70; Pittel–Wormald 05; Behrisch–Coja-Oghlan–K. 09]

Let $d > 1$ be constant and $I \subset \mathbb{R}$ compact. For any $k \in \mathbb{N}$ with $\sigma^{-1}(k - \mu) \in I$

$$\mathbb{P}[\, L_1(d) = k \,] \; \sim \; \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right)$$

Local Limit Theorem for Giant Component

Assume $d = p \cdot (n-1) > 1$ and $0 < \rho < 1$ satisfies $1 - \rho = e^{-d \cdot \rho}$.

Let $\mu := \rho \cdot n$ and $\sigma^2 := \frac{\rho(1-\rho)}{(1-d(1-\rho))^2} \cdot n$

Theorem [STEPANOV 70; PITTEL–WORMALD 05; BEHRISCH–COJA-OGHLAN–K. 09]

Let $d > 1$ be constant and $I \subset \mathbb{R}$ compact. For any $k \in \mathbb{N}$ with $\sigma^{-1}(k-\mu) \in I$

$$\mathbb{P}[\, L_1(d) = k \,] \sim \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right)$$

LLT for joint distribution of # vertices and # edges

- Recurrence formulas for # connected graphs [S 70]
- Counting connected graphs inside-out [PW 05]
- Two round exposure and smoothing (for $L_1(d)$) [BC-OK 09]
- Fourier analysis (for joint distribution) [BC-OK 14]

Part II

Random k-uniform Hypergraph $H_k(n,p),\ k \geq 2$

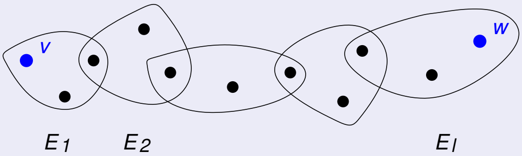
Standard Notion of Components

Vertex connectivity

- A vertex v is said to be reachable from a vertex w

  if there is a sequence $E_1, \ldots, E_\ell$ of hyperedges such that

  $v \in E_1$, $w \in E_\ell$ and $|E_i \cap E_{i+1}| \geq 1$ for each $i = 1, \ldots, \ell - 1$.



- The reachability is an equivalence relation, and

  the equivalence classes are called components

Phase Transition in $H_k(n, p)$

$L_1(d) = \#$ vertices in the largest component, where $d = p \cdot (k-1) \cdot \binom{n-1}{k-1}$

Emergence of giant component [Schmidt-Pruzan–Shamir 85]

- If $d < 1$, whp $L_1(d) = O(\log n)$
- If $d > 1$, whp $L_1(d) = \Theta(n)$

Phase Transition in $H_k(n, p)$

$L_1(d) = \#$ vertices in the largest component, where $d = p \cdot (k-1) \cdot \binom{n-1}{k-1}$

Emergence of giant component

[Schmidt-Pruzan–Shamir 85]

- If $d < 1$, whp $\quad L_1(d) = O(\log n)$
- If $d > 1$, whp $\quad L_1(d) = \Theta(n)$

Local limit theorem for (# vertices, # edges) in the giant component

- $(d-1)^3 n \to \infty, \quad (d-1)^3 n = o(\frac{\log n}{\log \log n})$ [Karoński–Łuczak 02]
- $d > 1$ constant [Behrisch–Coja-Oghlan–K. 14]
- $(d-1)^3 n \to \infty$, $d - 1 \to 0$ [Bollobás–Riordan 14+]

Counting Connected k-uniform Hypergraphs

... with n vertices and m edges

- $m - \frac{n}{k-1} \ll \frac{\log n}{\log \log n}$ [Karoński–Łuczak 02]
- $m - \frac{n}{k-1} = \Theta(n)$ [Behrisch–Coja-Oghlan–K. 14]
- $m - \frac{n}{k-1} = o(n)$ [Bollobás–Riordan 14+]
- $n^{1/3}\log^2 n \ll m - \frac{n}{2} \ll n$ for $k = 3$ [Sato–Wormald 14+]

Proof techniques

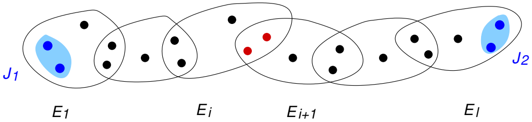
- Combinatorial enumeration [KŁ02]
- Local limit theorem for the giant in $H_k(n,p)$ [BC-OK 14; BR 14+]
- Counting connected graphs inside-out (cores and kernels) [SW 14+]

Higher Order Connectivity

[Bollobás–Riordan 12]

Let $1 \leq j \leq k - 1$.

- A j-element subset J_1 is said to be reachable from another j-set J_2 if there is a sequence $E_1, \ldots, E_\ell$ of hyperedges such that $J_1 \subseteq E_1$, $J_2 \subseteq E_\ell$ and $|E_i \cap E_{i+1}| \geq j$ for each $i = 1, \ldots, \ell - 1$.



- The reachability is an equivalence relation on j-sets, and the equivalence classes are called j-connected component.

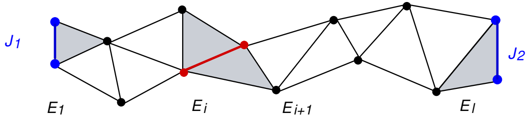
Higher Order Connectivity

[Bollobás–Riordan 12]

Let $1 \leq j \leq k-1$.

- A j-element subset J_1 is said to be reachable from another j-set J_2 if there is a sequence $E_1, \ldots, E_\ell$ of hyperedges such that $J_1 \subseteq E_1$, $J_2 \subseteq E_\ell$ and $|E_i \cap E_{i+1}| \geq j$ for each $i = 1, \ldots, \ell - 1$.

e.g. $k = 3,\ j = 2$



- The reachability is an equivalence relation on j-sets, and the equivalence classes are called j-connected component.

Emergence of Giant j-Component

$L_j(d) = \#$ j-sets in the largest j-component, where $d = p \cdot \left(\binom{k}{j} - 1 \right) \cdot \binom{n-j}{k-j}$

Theorem [Cooley–Person–K. 13+]

- If $d < 1$, whp $\qquad L_j(d) = O(\log n)$
- If $d > 1$, whp $\qquad L_j(d) = \Theta(n^j)$

Emergence of Giant j-Component

$L_j(d) = \#$ j-sets in the largest j-component, where $d = p \cdot \left(\binom{k}{j} - 1 \right) \cdot \binom{n-j}{k-j}$

Theorem [Cooley–Person–K. 13+]

- If $d < 1$, whp $\quad L_j(d) = O(\log n)$
- If $d > 1$, whp $\quad L_j(d) = \Theta(n^j)$

Remarks

- Short alternative proof of [Schmidt-Pruzan–Shamir 85]
- Extension of Depth-First Search approach of [Krivelevich–Sudakov 13]
- When $d = 1 + \epsilon$ for $\epsilon \in (0, 1)$,

  whp $\exists$ a loose path of length $\Omega(\epsilon^2 n)$

Critical Phase in $H_k(n, p)$

$L_j(d) = \#$ j-sets in the largest j-component

where $d = p \cdot \left(\binom{k}{j} - 1\right) \cdot \binom{n-j}{k-j}$ and $d = 1 + \epsilon$ for $\epsilon = o(1)$

Theorem [Cooley–K.–Koch 14+]

- If $\epsilon^3 n \to -\infty$, whp $\quad L_j(d) = O(\epsilon^{-2} \log n)$
- If $\epsilon^3 n \to +\infty$, whp $\quad L_j(d) \sim 2\epsilon \frac{1}{\binom{k}{j}-1} \binom{n}{j}$

Critical Phase in $H_k(n, p)$

$L_j(d) = \#$ j-sets in the largest j-component

where $d = p \cdot \left(\binom{k}{j} - 1\right) \cdot \binom{n-j}{k-j}$ and $d = 1 + \epsilon$ for $\epsilon = o(1)$

Theorem [Cooley–K.–Koch 14+]

- If $\epsilon^3 n \to -\infty$, whp $L_j(d) = O(\epsilon^{-2} \log n)$
- If $\epsilon^3 n \to +\infty$, whp $L_j(d) \sim 2\epsilon \frac{1}{\binom{k}{j}-1} \binom{n}{j}$

Proof techniques

- Extension of Breadth-First Search, Galton-Watson branching process and second moment approach of [Bollobás–Riordan 12+]
- Smooth boundary lemma

Part III

Proof Ideas for Supercritical Regime in $H_k(n, p)$

$k \geq 2,\ j \geq 1$

Heuristic for Threshold

- Breadth-First Search process & Galton-Watson branching process

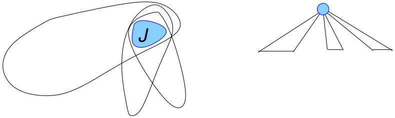
Heuristic for Threshold

- Breadth-First Search process & Galton-Watson branching process

  – Begin with a j-set J

Heuristic for Threshold

- Breadth-First Search process & Galton-Watson branching process



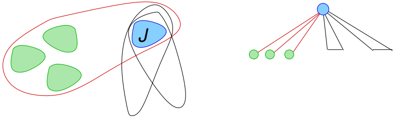
– Begin with a j-set J

– Discover all edges that contain the j-set J

$\exists$ $\binom{n-j}{k-j}$ k-sets containing J, each of which is an edge with prob. p

Heuristic for Threshold

- Breadth-First Search process & Galton-Watson branching process



– Begin with a j-set J

– Discover all edges that contain the j-set J

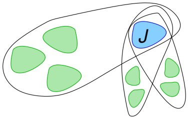
$\exists$ $\binom{n-j}{k-j}$ k-sets containing J, each of which is an edge with prob. p

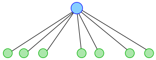
– For each edge E, discover $\binom{k}{j} - 1$ new j-sets contained in E

(It could be fewer if some of these j-sets were discovered earlier)

Heuristic for Threshold

- Breadth-First Search process & Galton-Watson branching process





– Begin with a j-set J

– Discover all edges that contain the j-set J

$\exists$ $\binom{n-j}{k-j}$ k-sets containing J, each of which is an edge with prob. p

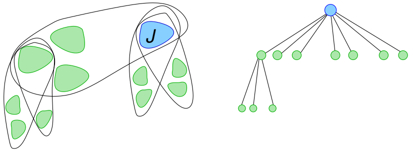
– For each edge E, discover $\binom{k}{j} - 1$ new j-sets contained in E

(It could be fewer if some of these j-sets were discovered earlier)

So, $\mathbb{E}(\#\ j\text{-sets discovered from } J \text{ in one generation}) = p \cdot \left(\binom{k}{j} - 1\right) \cdot \binom{n-j}{k-j}$

Heuristic for Threshold

- Breadth-First Search process & Galton-Watson branching process



– Begin with a j-set J

– Discover all edges that contain the j-set J

$\exists$ $\binom{n-j}{k-j}$ k-sets containing J, each of which is an edge with prob. p

– For each edge E, discover $\binom{k}{j} - 1$ new j-sets contained in E

(It could be fewer if some of these j-sets were discovered earlier)

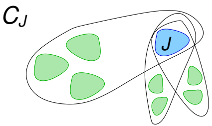
So, $\mathbb{E}(\#\ j\text{-sets discovered from } J \text{ in one generation}) = p \cdot \left(\binom{k}{j} - 1\right) \cdot \binom{n-j}{k-j}$

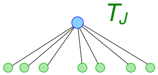
Proof Sketch

(1) Breadth-First Search

Proof Sketch

(1) Breadth-First Search





Given j-set J

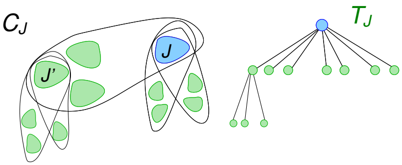
construct spanning tree T_J

of j-component C_J

consisting of j-sets as vertices

Proof Sketch

(1) Breadth-First Search



Given j-set J

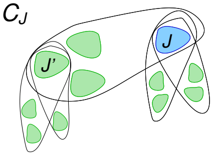
construct spanning tree T_J

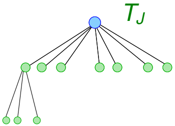
of j-component C_J

consisting of j-sets as vertices

Proof Sketch

(1) Breadth-First Search





Given j-set J

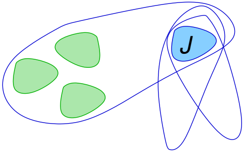
construct spanning tree T_J

of j-component C_J

consisting of j-sets as vertices

(2) Coupling T_J from above with Galton-Watson branching process

with offspring distribution $(\binom{k}{j} - 1)Bi(\binom{n-j}{k-j}, p)$

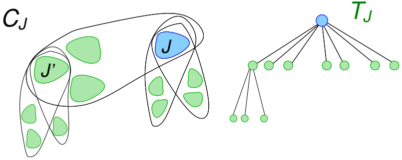


$\varrho := \mathbb{P}\,(\textit{process survives})$

$1 - \varrho = \sum_\ell \mathbb{P}\,(Bi(\binom{n-j}{k-j}, p) = \ell) \cdot (1 - \varrho)^{\ell(\binom{k}{j}-1)}$

Proof Sketch

(1) Breadth-First Search



Given j-set J

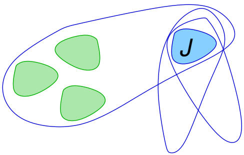
construct spanning tree T_J

of j-component C_J

consisting of j-sets as vertices

(2) Coupling T_J from above with Galton-Watson branching process

with offspring distribution $(\binom{k}{j} - 1) Bi(\binom{n-j}{k-j}, p)$



$\varrho := \mathbb{P}\,(\textit{process survives})$

$1 - \varrho = \sum_\ell \mathbb{P}\,(Bi(\binom{n-j}{k-j}, p) = \ell) \cdot (1 - \varrho)^{\ell(\binom{k}{j}-1)}$

$\longrightarrow \varrho \sim \dfrac{2\epsilon}{\binom{k}{j}-1}$

Proof Sketch – cont.

(3) First moment argument

- Let $N :=$ $\#$ j-sets in 'large' j-components with $\geq L := \epsilon\, n^j$ many j-sets
- Using upper and lower couplings with Galton-Watson branching process,

$$\mathbb{E}(N) \sim \frac{2\epsilon}{\binom{k}{j} - 1} \binom{n}{j}$$

Proof Sketch – cont.

(3) First moment argument

- Let $N :=$ # j-sets in 'large' j-components with $\geq L := \epsilon\, n^j$ many j-sets
- Using upper and lower couplings with Galton-Watson branching process,

$$\mathbb{E}(N) \sim \frac{2\epsilon}{\binom{k}{j} - 1} \binom{n}{j}$$

(4) Second moment argument

IF we could show

$$\mathbb{E}(N^2) \sim (\mathbb{E}(N))^2,$$

THEN

$$N \sim \frac{2\epsilon}{\binom{k}{j} - 1} \binom{n}{j}$$

Proof Sketch – cont.

(3) First moment argument

- Let $N :=$ # j-sets in 'large' j-components with $\geq L := \epsilon\, n^j$ many j-sets
- Using upper and lower couplings with Galton-Watson branching process,

$$\mathbb{E}(N) \sim \frac{2\epsilon}{\binom{k}{j} - 1} \binom{n}{j}$$

(4) Second moment argument

IF we could show

$$\mathbb{E}(N^2) \sim (\mathbb{E}(N))^2,$$

THEN

$$N \sim \frac{2\epsilon}{\binom{k}{j} - 1} \binom{n}{j}$$

(5) Two round exposure

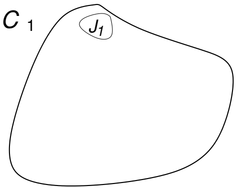
Almost all j-sets in 'large' j-components are in a single j-component

More on Second Moment Argument

Need to consider $\#$ pairs of *j*-sets in 'large' *j*-components

More on Second Moment Argument

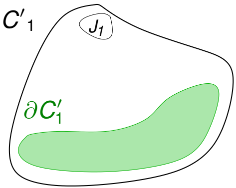
Need to consider $\#$ pairs of j-sets in 'large' j-components



- Fix j-set J_1 and grow its j-component C_1

More on Second Moment Argument

Need to consider $\#$ pairs of j-sets in 'large' j-components

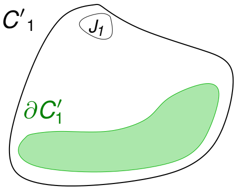


- Fix j-set J_1 and grow its j-component C'_1 until hit stopping conditions

$$S_1 = \{ |C'_1| \geq L \quad or \quad |\partial C'_1| \geq \epsilon L \}$$

More on Second Moment Argument

Need to consider $\#$ pairs of j-sets in 'large' j-components



- Fix j-set J_1 and grow its j-component C'_1 until hit stopping conditions

$$S_1 = \{ \, |C'_1| \geq L \quad or \quad |\partial C'_1| \geq \epsilon L \, \}$$

Then $\mathbb{P}(S_1) \lesssim \frac{2\epsilon}{\binom{k}{j}-1}$

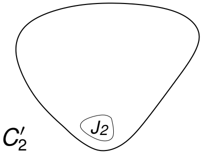
More on Second Moment Argument

Need to consider $\#$ pairs of j-sets in 'large' j-components

- Fix j-set J_1 and grow its j-component C'_1 until hit stopping conditions

$$S_1 = \{ |C'_1| \geq L \quad or \quad |\partial C'_1| \geq \epsilon L \}$$

Then $\mathbb{P}(S_1) \lesssim \frac{2\epsilon}{\binom{k}{j}-1}$

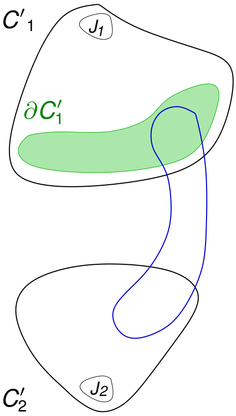


- Delete all the vertices in C'_1

& fix a j-set J_2, grow component C'_2

More on Second Moment Argument

Need to consider $\#$ pairs of j-sets in 'large' j-components



- Fix j-set J_1 and grow its j-component C'_1 until hit stopping conditions

$$S_1 = \{ |C'_1| \geq L \quad or \quad |\partial C'_1| \geq \epsilon L \}$$

Then $\mathbb{P}(S_1) \lesssim \frac{2\epsilon}{\binom{k}{j} - 1}$

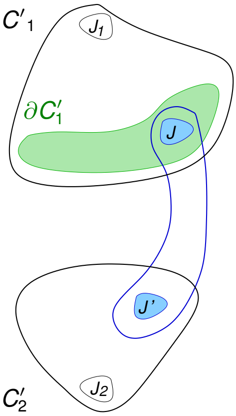
- Delete all the vertices in C'_1

& fix a j-set J_2, grow component C'_2

Need to show $\mathbb{P}(\, e(\partial C'_1,\ C'_2) \geq 1)$ is small

More on Second Moment Argument

Need to consider $\#$ pairs of j-sets in 'large' j-components



- Fix j-set J_1 and grow its j-component C'_1 until hit stopping conditions

$$S_1 = \{\, |C'_1| \geq L \quad or \quad |\partial C'_1| \geq \epsilon L \,\}$$

Then $\mathbb{P}(S_1) \lesssim \frac{2\epsilon}{\binom{k}{j}-1}$

- Delete all the vertices in C'_1

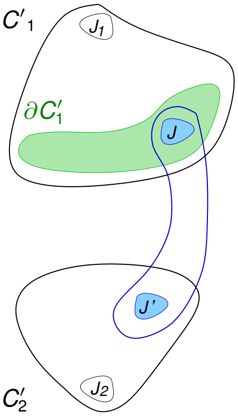
& fix a j-set J_2, grow component C'_2

Need to show $\mathbb{P}(\, e(\partial C'_1,\ C'_2) \geq 1)$ is small

$$\mathbb{P}(\, e(\partial C'_1,\ C'_2) \geq 1) \leq p \cdot |\partial C'_1| \cdot |C'_2|$$

More on Second Moment Argument

Need to consider $\#$ pairs of j-sets in 'large' j-components



- Fix j-set J_1 and grow its j-component C'_1 until hit stopping conditions

$$S_1 = \{ |C'_1| \geq L \quad or \quad |\partial C'_1| \geq \epsilon L \}$$

Then $\mathbb{P}(S_1) \lesssim \frac{2\epsilon}{\binom{k}{j}-1}$

- Delete all the vertices in C'_1

& fix a j-set J_2, grow component C'_2

Need to show $\mathbb{P}(\, e(\partial C'_1,\ C'_2) \geq 1)$ is small

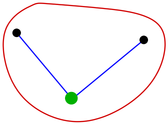
However,

$p \cdot |\partial C'_1| \cdot |C'_2|$ is not the right thing to do

More on Second Moment Argument – cont.

Instead we need

- for $k = 3$, $j = 2$,



$$\mathbb{P}\left(e(\partial C_1', C_2') \geq 1 \right)$$

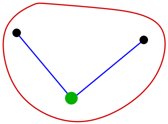
$\leq \mathbb{E}(\#$ 3-sets containing

a pair of 2-sets intersecting at a vertex)

More on Second Moment Argument – cont.

Instead we need

- for $k = 3$, $j = 2$,

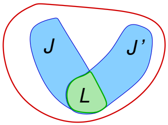


$\mathbb{P}(\ e(\partial C_1',\ C_2') \geq 1)$

$\leq\ \mathbb{E}(\#$ 3-sets containing

a pair of 2-sets intersecting at a vertex)

- for $k \geq 3$, $j \geq 2$,



$\mathbb{P}(\ e(\partial C_1',\ C_2') \geq 1)$

$\leq\ \mathbb{E}(\#$ k-sets containing

a pair of j-sets, J, J', intersecting at an ℓ-set L

for some $0 \leq \ell \leq j - 1$)

Boundary Is Smooth

Key lemma [Cooley–K.–Koch 14+]

For every $0 \le \ell \le j-1$, every ℓ-set L,

$$\#\ j\text{-sets in } \partial C'_1 \quad \text{containing } L \quad \sim \quad \frac{|\partial C'_1|}{\binom{n}{j}} \binom{n-\ell}{j-\ell}$$

'Reasonably Large' Boundary Is Smooth

Key lemma [COOLEY–K.–KOCH 14+]

Let $\partial C_1'(t)$ denote the collection of j-sets in $\partial C'_1$ after t generations of BFS.

With probability at least $1 - \exp(-\Theta(n^{1/11}))$ the following is true.

For every $0 \le \ell \le j-1$, every ℓ-set L, and every $s_\ell \le t \le s_\ell + O(\log n)$,

$$\#\ j\text{-sets in } \partial C'_1(t) \text{ containing } L \quad \sim \quad \frac{|\partial C_1'(t)|}{\binom{n}{j}} \binom{n-\ell}{j-\ell}$$

where $s_\ell := \min\{\ d\ :\ |\partial C_1'(t)| \ge n^{\ell+1/10}\}$.

Open Problems

(1) What about the number of j-set in the largest j-component at the criticality, i.e. when $d = 1$?

(2) Is the width of critical window, $(d-1)^3 n = O(1)$, best possible? Perhaps $(d-1)^j n = O(1)$?

(3) What about the number of j-set in the 2nd largest j-component in the supercritical regime?

(4) What is the actual distribution of $\#$ j-sets in the largest j-component? Central limit theorem? Local limit theorem?