On a phase transition of the random intersection graph: supercritical region

Sang June Lee

Duksung Women's University
Seoul, Korea

2015 Combinatorics Workshop
NIMS
July 16, 2015

This is joint work with Jeong Han Kim and Joohan Na (KIAS).

Motivation

Question

What was the most issued word in Korea for the last two months?

Motivation

Question

What was the most issued word in Korea for the last two months?

Question (for foreigners)

What makes you hesitate for visiting Korea?

Motivation

Question

What was the most issued word in Korea for the last two months?

Question (for foreigners)

What makes you hesitate for visiting Korea?

Answer: Mers

Motivation

Question

What was the most issued word in Korea for the last two months?

Question (for foreigners)

What makes you hesitate for visiting Korea?

Answer: Mers

Question

What is the proper network (or graph) model explaining epidemic of Mers?

Motivation

Question

What was the most issued word in Korea for the last two months?

Question (for foreigners)

What makes you hesitate for visiting Korea?

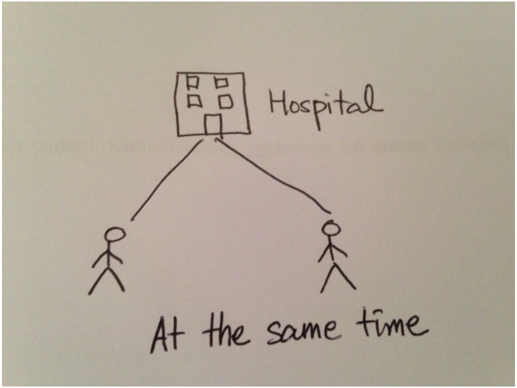
Answer: Mers

Question

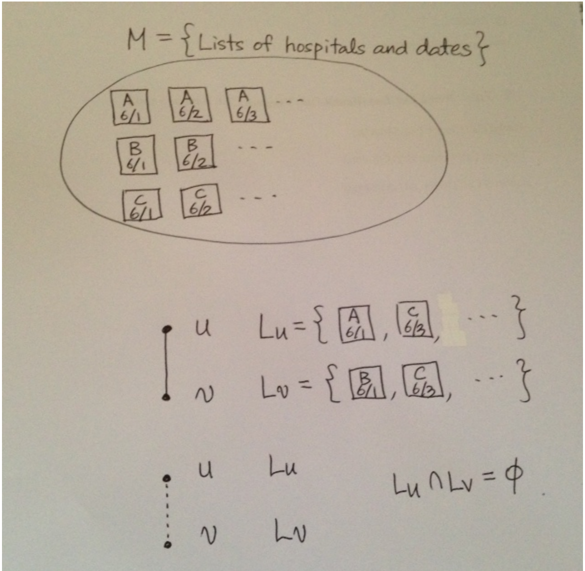
What is the proper network (or graph) model explaining epidemic of Mers?

Answer: Random intersection graph

Feature of Mers in Korea



Graph model about epidemic of Mers



Sec 1) Definition

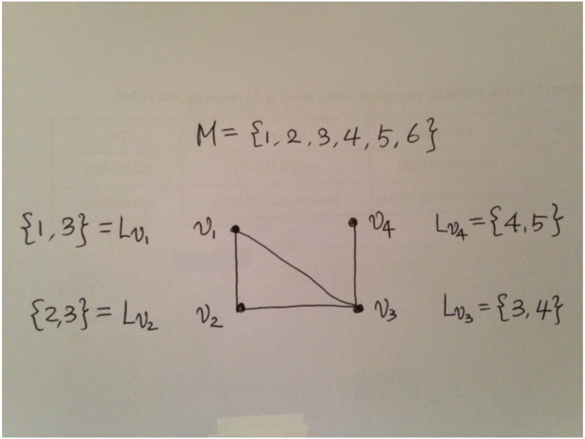
- $V := \{v_1, \ldots, v_n\}$
- $\{L_1, \ldots, L_n\}$: a collection of sets

Definition (Intersection graph)

The *intersection graph* on V generated by $\{L_1, \ldots, L_n\}$ is the graph on V in which

$$v_i \sim v_j \quad \text{if and only if} \quad L_i \cap L_j \neq \emptyset.$$

Example



Random intersection graph

Definition (Random Intersection graph $G(n, m; p)$)

- M: a set of size m.
- L_i: a random subset obtained by choosing each element in M indepedently with probabiity p.

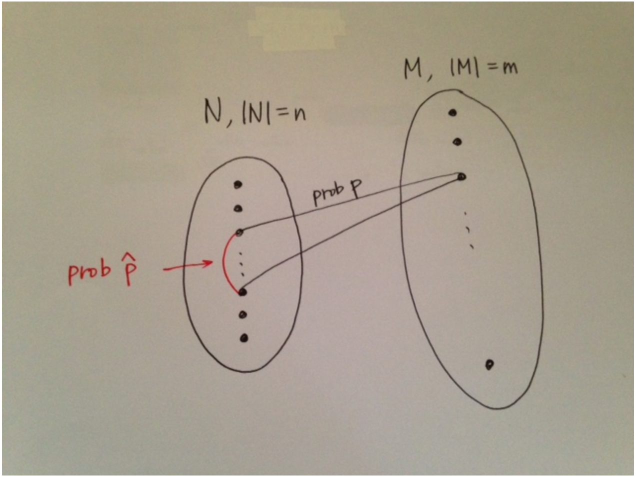
Random intersection graph

Definition (Random Intersection graph $G(n, m; p)$)

- M: a set of size m.
- L_i: a random subset obtained by choosing each element in M indepedently with probabiity p.
- The *random intersection graph* $G(n, m; p)$ is the intersection graph generated by i.i.d. L_i as above.

It was defined by Karoński, Scheinerman, and Singer-Cohen (1999).

Visualization: Random bipartite graph



Application

1. A random intersection graph has received a lot of attention because of a great diversity of applications:
   - Epidemic
   - Network user profiling
   - Analysis of complex networks.

Application

1. A random intersection graph has received a lot of attention because of a great diversity of applications:
   - Epidemic
   - Network user profiling
   - Analysis of complex networks.
2. The special case when L_i's are uniformly distributed as subsets of M of the same size has been applied to security of wireless sensor networks.

Question

When is $G(n, m; p)$ essentially the same as the binomial random graph $G(n, \hat{p})$ with the same expected number of edges?

Remark: $\hat{p} := 1 - (1 - p^2)^m$.

Question

When is $G(n, m; p)$ essentially the same as the binomial random graph $G(n, \hat{p})$ with the same expected number of edges?

Remark: $\hat{p} := 1 - (1 - p^2)^m$.

Notion

Distance between two random graphs: Total variation.

Question

When is $G(n, m; p)$ essentially the same as the binomial random graph $G(n, \hat{p})$ with the same expected number of edges?

Remark: $\hat{p} := 1 - (1 - p^2)^m$.

Notion

Distance between two random graphs: Total variation.

Definition

The total variation between two random graphs X and Y is defined by

$$\mathrm{TV}(X, Y) := \frac{1}{2} \sum_{G} \Big| \Pr[X = G] - \Pr[Y = G] \Big|,$$

where the sum is over all possible graphs G of X and Y.

Sec 2) Previous results

Observation

Let $\omega \to \infty$ as $n \to \infty$.

1. If $p \le \frac{1}{\omega n \sqrt{m}}$,
   then two random graphs are the empty graph
   with high probability.

Sec 2) Previous results

Observation

Let $\omega \to \infty$ as $n \to \infty$.

1. If $p \leq \frac{1}{\omega n \sqrt{m}}$,
   then two random graphs are the empty graph with high probability.
2. If $p \geq \sqrt{\frac{2 \ln n + \omega}{m}}$,
   then two random graphs are the complete graph with high probability.

Sec 2) Previous results

Observation

Let $\omega \to \infty$ as $n \to \infty$.

1. If $p \leq \frac{1}{\omega n \sqrt{m}}$,
   then two random graphs are the empty graph with high probability.
2. If $p \geq \sqrt{\frac{2 \ln n + \omega}{m}}$,
   then two random graphs are the complete graph with high probability.

Assumption

$$\frac{1}{\omega n \sqrt{m}} \leq p \leq \sqrt{\frac{2 \ln n + \omega}{m}}.$$

Proposition

If $m \ll n^3$, then

$$\mathrm{TV}\Big(G(n, m; p), G(n, \hat{p})\Big) \to 1.$$

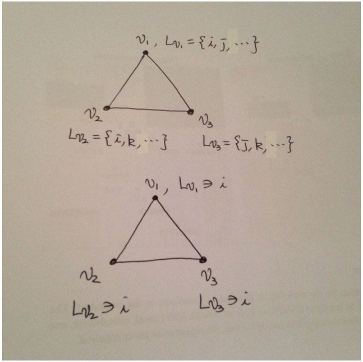
Remark: By comparing the number of triangles.

Proposition

If $m \ll n^3$, then

$$\mathrm{TV}\Big(G(n, m; p), G(n, \hat{p})\Big) \to 1.$$

Remark: By comparing the number of triangles.



Proposition

If $m \ll n^3$, then

$$\mathrm{TV}\Big(G(n, m; p), G(n, \hat{p})\Big) \to 1.$$

Remark: By comparing the number of triangles.

Theorem (Fill, Scheinerman and Singer-Cohen (2000))

If $m = n^\alpha$ *and* $\alpha > 6$, *then*

$$\mathrm{TV}\Big(G(n, m; p), G(n, \hat{p})\Big) \to 0.$$

Proposition

If $m \ll n^3$, then

$$\mathrm{TV}\Big(G(n,m;p), G(n,\hat{p})\Big) \to 1.$$

Remark: By comparing the number of triangles.

Theorem (Fill, Scheinerman and Singer-Cohen (2000))

If $m = n^\alpha$ *and* $\alpha > 6$, *then*

$$\mathrm{TV}\Big(G(n,m;p), G(n,\hat{p})\Big) \to 0.$$

Theorem (Rybarczyk (2011))

If $m = n^\alpha$ *and* $3 \le \alpha \le 6$, *for any monotone property* $\mathcal{P}$, $\Pr[G(n,m;p) \in \mathcal{P}]$ *is similar to* $\Pr[G(n,\hat{p}) \in \mathcal{P}]$ *(with a technical statement).*

Question

What is the smallest constant α such that for $m = n^{\alpha}$ and any $p = p(n, m)$,

$$\mathrm{TV}\Big(G(n, m; p), G(n, \hat{p})\Big) = o(1)?$$

Previous result

$$3 \leq \alpha \leq 6.$$

Question

What is the smallest constant α such that for $m = n^{\alpha}$ and any $p = p(n, m)$,

$$\mathrm{TV}\Big(G(n, m; p), G(n, \hat{p})\Big) = o(1)?$$

Previous result

$$3 \leq \alpha \leq 6.$$

Main Theorem (Kim, Lee, Na (2015+))

For $m \gg n^4$ and $0 \leq p \leq 1$,

$$\mathrm{TV}\Big(G(n, m; p), G(n, \hat{p})\Big) = o(1).$$

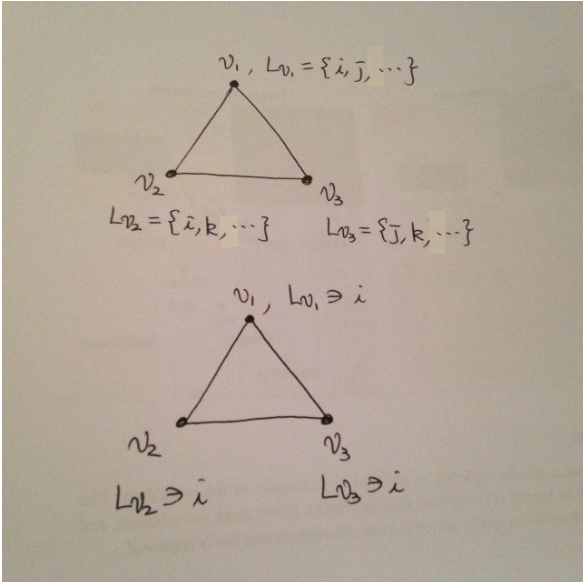
* We believe that 4 in the exponent is tight.

Artifact triangles

Definition

An artifact triangle is a triangle formed by the same element in M.

1. Fill, Scheinerman and Singer-Cohen (2000):
   the case when there is no artifact triangle.
2. Kim, Lee and Na (2015+):
   the case when there are not so many artifact triangles.

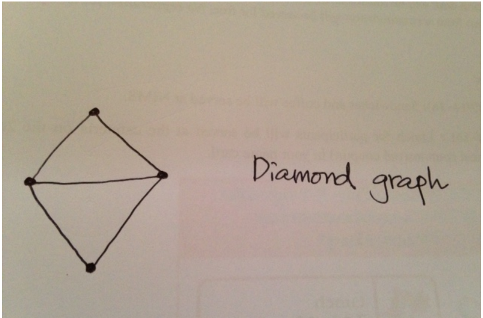

v_1, $L_{v_1} = \{i, j, \cdots\}$
v_2
v_3
$L_{v_2} = \{i, k, \cdots\}$
$L_{v_3} = \{j, k, \cdots\}$
v_1, $L_{v_1} \ni i$
v_2
v_3
$L_{v_2} \ni i$
$L_{v_3} \ni i$

Key object

Key object: Diamond graph

- A diamond graph $= K_4$ minus one edge.
- The number of diamond graphs with two artifact triangles in $G(n,m;p)$ is small iff

$$G(n,m;p) \sim G(n,\hat{p}).$$



Sec 3) Outline of Proof of Main Theorem

Recall: Main Theorem (Kim, Lee, Na (2015+))

For $m \gg n^4$ and $0 \leq p \leq 1$,

$$\mathrm{TV}\Big(G(n,m;p), G(n,\hat{p})\Big) = o(1).$$

Remark (essentially by Rybarczyk)

$G(n,m;p)$ is approximated by a random graph $G(n,(p_2,p_3,p_4))$.

Definition of $G(n,(p_2,p_3,p_4))$

1. $p_k := 1 - e^{-mp^k(1-p)^{n-k}}$.

Definition of $G(n,(p_2,p_3,p_4))$

1. $p_k := 1 - e^{-mp^k(1-p)^{n-k}}$.
2. $\mathcal{H}_k(n, p_k)$:= the random k-uniform hypergraph in which each k-subset belongs independently with probability p_k.

Definition of $G(n,(p_2,p_3,p_4))$

1. $p_k := 1 - e^{-mp^k(1-p)^{n-k}}$.
2. $\mathcal{H}_k(n,p_k) :=$ the random k-uniform hypergraph in which each k-subset belongs independently with probability p_k.
3. $K(\mathcal{H}_k(n,p_k)) :=$ the graph by changing each hyperedge of $\mathcal{H}_k(n,p_k)$ into a complete subgraph.

Definition of $G(n, (p_2, p_3, p_4))$

1. $p_k := 1 - e^{-mp^k(1-p)^{n-k}}$.
2. $\mathcal{H}_k(n, p_k)$:= the random k-uniform hypergraph in which each k-subset belongs independently with probability p_k.
3. $K(\mathcal{H}_k(n, p_k))$:= the graph by changing each hyperedge of $\mathcal{H}_k(n, p_k)$ into a complete subgraph.
4. $G(n, (p_2, p_3, p_4)) := \bigcup_{2 \leq k \leq 4} K(\mathcal{H}_k(n, p_k))$.

Definition of $G(n,(p_2,p_3,p_4))$

1. $p_k := 1 - e^{-mp^k(1-p)^{n-k}}$.
2. $\mathcal{H}_k(n,p_k)$:= the random k-uniform hypergraph in which each k-subset belongs independently with probability p_k.
3. $K(\mathcal{H}_k(n,p_k))$:= the graph by changing each hyperedge of $\mathcal{H}_k(n,p_k)$ into a complete subgraph.
4. $G(n,(p_2,p_3,p_4)) := \bigcup_{2\le k\le 4} K(\mathcal{H}_k(n,p_k))$.

Remark: Why the p_k?

- For $a \in M$, $V_a := \{v : a \in L_v\}$.
- For a fixed k-subset $U \subset V$,

$$\Pr\left[\exists a \in M \text{ s.t. } V_a = U\right] = 1 - (1 - p^k(1-p)^{n-k})^m.$$

Recall: Main Theorem (Kim, Lee, Na (2015+))

For $m \gg n^4$ and $0 \le p \le 1$,

$$\mathrm{TV}\Big(G(n,m;p), G(n,\hat{p})\Big) = o(1).$$

Key Lemma (Kim, Lee, Na (2015+))

For $m \gg n^4$ and $0 \le p \le (\frac{3\log n}{m})^{1/2}$,

$$\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)), G(n,p_2)\Big) = o(1).$$

Sec 4) Proof of Lemma

$$\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)),\,G(n,p_2)\Big)$$
$$:= \frac{1}{2}\sum_{G}\Big|\Pr[X=G]-\Pr[Y=G]\Big|$$
$$= \sum_{G\in\mathcal{G}}\Big(\Pr[G(n,p_2)=G]-\min\Big\{\Pr[G(n,(p_2,p_3,p_4))=G],\Pr[G(n,p_2)=G]\Big\}\Big).$$

Sec 4) Proof of Lemma

$$\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)),G(n,p_2)\Big)$$
$$:=\frac{1}{2}\sum_G\Big|\Pr[X=G]-\Pr[Y=G]\Big|$$
$$=\sum_{G\in\mathcal{G}}\Big(\Pr[G(n,p_2)=G]-\min\Big\{\Pr[G(n,(p_2,p_3,p_4))=G],\Pr[G(n,p_2)=G]\Big\}\Big).$$

In order to show that $\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)),G(n,p_2)\Big)$ is small, it suffices to show that

$$\Pr[G(n,(p_2,p_3,p_4))=G]\geq(1-O(\epsilon))\Pr[G(n,p_2)=G].$$

$$\Pr[G(n,(p_2,p_3,p_4)) = G]$$
$$= \sum_{\substack{Q\subseteq\mathcal{H}_4(G)\\ T\subseteq\mathcal{H}_3(G)}} \Pr[\mathcal{H}_4(n,p_4) = Q, \mathcal{H}_3(n,p_3) = T, G(n,p_2) = G]$$
$$= \sum_{\substack{Q\subseteq\mathcal{H}_4(G)\\ T\subseteq\mathcal{H}_3(G)}} p_4^{|Q|}(1-p_4)^{\binom{n}{4}-|Q|} p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|} p_2^{|G|-|K(Q)\cup K(T)|}(1-p_2)^{\binom{n}{2}-|G|}$$
$$= \Pr[G(n,p_2) = G] \sum_{\substack{Q\subseteq\mathcal{H}_4(G)\\ T\subseteq\mathcal{H}_3(G)}} p_4^{|Q|}(1-p_4)^{\binom{n}{4}-|Q|} p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|} p_2^{-|K(Q)\cup K(T)|}.$$

$$\Pr[G(n,(p_2,p_3,p_4))=G]$$
$$=\sum_{\substack{Q\subseteq\mathcal{H}_4(G)\\T\subseteq\mathcal{H}_3(G)}}\Pr[\mathcal{H}_4(n,p_4)=Q,\mathcal{H}_3(n,p_3)=T,G(n,p_2)=G]$$
$$=\sum_{\substack{Q\subseteq\mathcal{H}_4(G)\\T\subseteq\mathcal{H}_3(G)}} p_4^{|Q|}(1-p_4)^{\binom{n}{4}-|Q|}p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|}p_2^{|G|-|K(Q)\cup K(T)|}(1-p_2)^{\binom{n}{2}-|G|}$$
$$=\Pr[G(n,p_2)=G]\sum_{\substack{Q\subseteq\mathcal{H}_4(G)\\T\subseteq\mathcal{H}_3(G)}} p_4^{|Q|}(1-p_4)^{\binom{n}{4}-|Q|}p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|}p_2^{-|K(Q)\cup K(T)|}.$$

$$\frac{\Pr[G(n,(p_2,p_3,p_4))=G]}{\Pr[G(n,p_2)=G]}\geq\sum_{Q\subseteq\mathcal{H}_4(G)} p_4^{|Q|}(1-p_4)^{\binom{n}{4}-|Q|}p_2^{-|K(Q)|}$$
$$\times\sum_{T\subseteq\mathcal{H}_3(G\setminus K(Q))} p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|}p_2^{-|K(T)|}.$$

Three cases

Three cases

- Case I: no artifact triangles
- Case II: $\exists$ artifact triangles and no artifact quadruples
- Case III: $\exists$ artifact quadruples

Case I: no artifact triangles

In this case, the expected number of artifact triangles is small, that is,

$$p \le \frac{\varepsilon}{nm^{1/3}}.$$

Case I: no artifact triangles

In this case, the expected number of artifact triangles is small, that is,

$$p \le \frac{\varepsilon}{nm^{1/3}}.$$

- By taking $T, Q = \emptyset$, we have

$$\begin{aligned}\Pr[G(n,(p_2,p_3,p_4)) = G] &\ge \Pr[G(n,p_2) = G](1-p_4)^{\binom{n}{4}}(1-p_3)^{\binom{n}{3}} \\ &= \Pr[G(n,p_2) = G](1-O(\varepsilon)).\end{aligned}$$

Case I: no artifact triangles

In this case, the expected number of artifact triangles is small, that is,

$$p \leq \frac{\varepsilon}{nm^{1/3}}.$$

- By taking $T, Q = \emptyset$, we have

$$\begin{aligned}\Pr[G(n,(p_2,p_3,p_4)) = G] &\geq \Pr[G(n,p_2) = G](1-p_4)^{\binom{n}{4}}(1-p_3)^{\binom{n}{3}} \\ &= \Pr[G(n,p_2) = G](1-O(\varepsilon)).\end{aligned}$$

- Hence,

$$\begin{aligned}&\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)), G(n,p_2)\Big) \\ &= \sum_{G\in\mathcal{G}}\Big(\Pr[G(n,p_2)=G]-\min\big\{\Pr[G(n,(p_2,p_3,p_4))=G],\Pr[G(n,p_2)=G]\big\}\Big) \\ &\leq O(\varepsilon).\end{aligned}$$

Remark: It gives Fill–Scheinerman–Singer-Cohen (2000).

Case II: $\exists$ artifact triangles and no artifact quadruples

In this case, the expected number of artifact triangles is not small, but the expected number of artifact quadruples is small, that is,

$$\frac{\varepsilon}{nm^{1/3}} < p \leq \frac{\varepsilon}{n^{2/3}m^{1/3}}.$$

Case II: ∃ artifact triangles and no artifact quadruples

In this case, the expected number of artifact triangles is not small, but the expected number of artifact quadruples is small, that is,

$$\frac{\varepsilon}{nm^{1/3}} < p \leq \frac{\varepsilon}{n^{2/3}m^{1/3}}.$$

Remark: It is not possible for an arbitrary G to show

$$\Pr[G(n,(p_2,p_3,p_4)) = G] \geq \Pr[G(n,p_2) = G](1 - O(\varepsilon)).$$

Case II: ∃ artifact triangles and no artifact quadruples

In this case, the expected number of artifact triangles is not small, but the expected number of artifact quadruples is small, that is,

$$\frac{\varepsilon}{nm^{1/3}} < p \leq \frac{\varepsilon}{n^{2/3}m^{1/3}}.$$

Remark: It is not possible for an arbitrary G to show

$$\Pr[G(n,(p_2,p_3,p_4)) = G] \geq \Pr[G(n,p_2) = G](1 - O(\varepsilon)).$$

Idea: We consider properties of typical $G \in G(n,\hat{p})$.

For any family $\mathcal{G}_3$ of typical graphs on V,

$$\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)),G(n,p_2)\Big)$$
$$\leq \Pr[G(n,p_2)\notin\mathcal{G}_3]$$
$$+\sum_{G\in\mathcal{G}_3}\Big(\Pr[G(n,p_2)=G]-\min\Big\{\Pr[G(n,(p_2,p_3,p_4))=G],\Pr[G(n,p_2)=G]\Big\}\Big)$$
$$\leq O(\varepsilon)+\sum_{G\in\mathcal{G}_3}\Big(\Pr[G(n,p_2)=G]-\min\Big\{\Pr[G(n,(p_2,p_3,p_4))=G],\Pr[G(n,p_2)=G]\Big\}\Big).$$

For any family $\mathcal{G}_3$ of typical graphs on V,

$$\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)), G(n,p_2)\Big)$$
$$\leq \Pr[G(n,p_2)\notin\mathcal{G}_3]$$
$$+\sum_{G\in\mathcal{G}_3}\Big(\Pr[G(n,p_2)=G]-\min\big\{\Pr[G(n,(p_2,p_3,p_4))=G],\Pr[G(n,p_2)=G]\big\}\Big)$$
$$\leq O(\varepsilon)+\sum_{G\in\mathcal{G}_3}\Big(\Pr[G(n,p_2)=G]-\min\big\{\Pr[G(n,(p_2,p_3,p_4))=G],\Pr[G(n,p_2)=G]\big\}\Big).$$

Goal

For any $G\in\mathcal{G}_3$,

$$\Pr[G(n,(p_2,p_3,p_4))=G]\geq(1-O(\varepsilon))\Pr[G(n,p_2)=G].$$

- $|\mathcal{H}_3(G)|$: the number of triangles in G.
- $I(G)$: the number of diamond graphs, i.e., K_4 minus one edge.

- $|\mathcal{H}_3(G)|$: the number of triangles in G.
- $I(G)$: the number of diamond graphs, i.e., K_4 minus one edge.

Lemma

Let $\mathcal{G}_3$ be the set of all graphs G on V satisfying

$$|\mathcal{H}_3(G)| \geq (1-\delta)\binom{n}{3}p_2^3 \text{ and } I(G) \leq n^4 p_2^5/\varepsilon.$$

Then, for $\frac{\varepsilon}{nm^{1/3}} < p \leq \left(\frac{3\log n}{m}\right)^{1/2}$,

$$\Pr[G(n,p_2) \in \mathcal{G}_3] = 1 - O(\varepsilon).$$

Taking $Q = \emptyset$, we have

$$\frac{\Pr[G(n,(p_2,p_3,p_4)) = G]}{\Pr[G(n,p_2) = G]} \geq (1-p_4)^{\binom{n}{4}} \sum_{T \subseteq \mathcal{H}_3(G)} p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|} p_2^{-|K(T)|}$$

$$\geq (1-O(\varepsilon)) \sum_{t=0}^{t_0} \sum_{\substack{T \subseteq \mathcal{H}_3(G) \\ |T|=t, |K(T)|=3t}} p_3^{t}(1-p_3)^{\binom{n}{3}-t} p_2^{-3t},$$

where

$$t_0 := \frac{n^3 m p^3}{\varepsilon} = \Theta\left(\frac{n^3 p_3}{\varepsilon}\right).$$

$$\frac{\Pr[G(n,(p_2,p_3,p_4))=G]}{\Pr[G(n,p_2)=G]}$$

$$\geq (1-O(\varepsilon))\sum_{t=0}^{t_0}\sum_{\substack{T\subseteq \mathcal{H}_3(G)\\ |T|=t,|K(T)|=3t}} p_3^t(1-p_3)^{\binom{n}{3}-t}p_2^{-3t}$$

$$=(1-O(\varepsilon))\sum_{t=0}^{t_0}(1-O(\varepsilon))\binom{\binom{n}{3}}{t}p_2^{3t}p_3^t(1-p_3)^{\binom{n}{3}-t}p_2^{-3t}$$

$$\geq (1-O(\varepsilon))\sum_{t=0}^{t_0}\binom{\binom{n}{3}}{t}p_3^t(1-p_3)^{\binom{n}{3}-t}$$

$$=(1-O(\varepsilon))\left(1-\Pr\left[\mathrm{Bin}\left(\binom{n}{3},p_3\right)>t_0\right]\right)=1-O(\varepsilon).$$

It implies that

$$\mathrm{TV}\Big(G(n,(p_2,p_3,p_4)),G(n,p_2)\Big)=O(\varepsilon).$$

Case III: ∃ artifact triangles and quadruples

In this case, the expected number of artifact quadruples is not small, that is,

$$\frac{\varepsilon}{n^{2/3}m^{1/3}} < p \leq \left(\frac{3\log n}{m}\right)^{1/2}.$$

Case III: ∃ artifact triangles and quadruples

In this case, the expected number of artifact quadruples is not small, that is,

$$\frac{\varepsilon}{n^{2/3}m^{1/3}} < p \leq \left(\frac{3\log n}{m}\right)^{1/2}.$$

- $|\mathcal{H}_4(G)|$: the number of quadruples in G.

Lemma

Let $\mathcal{G}_4 \subset \mathcal{G}_3$ be the set of all graphs G on V satisfying

$$|\mathcal{H}_4(G)| \geq \left(1 - \frac{1}{\varepsilon n}\right)\binom{n}{4}p_2^6.$$

Then, for $\frac{\varepsilon}{n^{2/3}m^{1/3}} < p \leq \left(\frac{3\log n}{m}\right)^{1/2}$,

$$\Pr[G(n, p_2) \in \mathcal{G}_4] = 1 - O(\varepsilon).$$

Goal

For any $G \in \mathcal{G}_4$,

$$\Pr[G(n, (p_2, p_3, p_4)) = G] \geq (1 - O(\varepsilon)) \Pr[G(n, p_2) = G].$$

Goal

For any $G \in \mathcal{G}_4$,

$$\Pr[G(n,(p_2,p_3,p_4)) = G] \geq (1 - O(\varepsilon)) \Pr[G(n,p_2) = G].$$

$$\frac{\Pr[G(n,(p_2,p_3,p_4)) = G]}{\Pr[G(n,p_2) = G]}$$
$$\geq \sum_{Q \subseteq \mathcal{H}_4(G)} p_4^{|Q|}(1-p_4)^{\binom{n}{4}-|Q|} p_2^{-|K(Q)|} \cdot \sum_{T \subseteq \mathcal{H}_3(G \setminus K(Q))} p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|} p_2^{-|K(T)|}$$
$$\geq (1 - O(\varepsilon)) \cdot \min_{\substack{Q \subseteq \mathcal{H}_4(G) \\ |Q| \leq q_0}} \sum_{T \subseteq \mathcal{H}_3(G \setminus K(Q))} p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|} p_2^{-|K(T)|},$$

where $q_0 := \frac{n^4 m p^4}{\varepsilon} = \Theta\left(\frac{n^4 p_4}{\varepsilon}\right)$.

Let $t_0 := \frac{n^3 m p^3}{\varepsilon} = \Theta\left(\frac{n^3 p_3}{\varepsilon}\right)$ and $r := \frac{n^4 m^2 p^6}{\varepsilon^3} = \Theta\left(\frac{n^4 p_3^2}{\varepsilon^3}\right)$.

We have that

$$\sum_{T \subseteq \mathcal{H}_3(G \setminus K(Q))} p_3^{|T|}(1-p_3)^{\binom{n}{3}-|T|} p_2^{-|K(T)|}$$

$$\geq \sum_{t=0}^{t_0} \sum_{\substack{T \subseteq \mathcal{H}_3(G \setminus Q) \\ |T|=t, I(T) \leq r}} p_3^t (1-p_3)^{\binom{n}{3}-t} p_2^{-6t+r}$$

$$\geq \sum_{t=0}^{t_0} (1-O(\varepsilon)) \binom{\binom{n}{3}}{t} p_2^{6t} \cdot p_3^t (1-p_3)^{\binom{n}{3}-t} p_2^{-6t+r}$$

$$\geq (1-O(\varepsilon)) p_2^r \cdot \sum_{t=0}^{t_0} p_3^t (1-p_3)^{\binom{n}{3}-t} \geq (1-O(\varepsilon)),$$

since $p_2^r = (1-e^{-mp^2(1-p)^{n-2}})^r \geq 1 - O(re^{-mp^2}) = 1 - O(\varepsilon)$.

Problem

Problem

Fix $3 < \alpha < 6$, and let $m = n^{\alpha}$.
Find a probability $p^* = p^*(n, m)$ such that

- If $p \ll p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) = o(1)$.
- If $p \gg p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) \geq c$,
  for some positive constant $c > 0$.

Problem

Problem

Fix $3 < \alpha < 6$, and let $m = n^{\alpha}$.
Find a probability $p^* = p^*(n, m)$ such that

- If $p \ll p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) = o(1)$.
- If $p \gg p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) \geq c$, for some positive constant $c > 0$.

Question

When $G(n, m; p) \not\approx G(n, \hat{p})$,
what are interesting properties and structures of $G(n, m; p)$?

Thank you for your attention!

Thank you for your attention!

Proof of Lemma 1

Lemma

1. $\mathrm{TV}\left(G(n,m;p), G(n,(p_k))\right) = o(1)$. (essentially by Rybarczyk)
2. $\mathrm{TV}(G(n,p_2,p_3,p_4), G(n,p_2)) = o(1)$. (Main part)
3. $\mathrm{TV}(G(n,p_2), G(n,\hat{p})) = o(1)$. (Not hard)

Proof of Lemma 1

Lemma

1. $\mathrm{TV}\left(G(n, m; p), G(n, (p_k))\right) = o(1)$. (essentially by Rybarczyk)
2. $\mathrm{TV}(G(n, p_2, p_3, p_4), G(n, p_2)) = o(1)$. (Main part)
3. $\mathrm{TV}(G(n, p_2), G(n, \hat{p})) = o(1)$. (Not hard)

Idea of Proof

1. Coupling argument
2. Property of Poisson distribution

Coupling argument

Definition

For two random variables X and Y, the coupling (X', Y') of X and Y is a random variable on the product of the sample spaces of X and Y such that the marginal distributions of X' and Y' are the distributions of X and Y, respectively.

Coupling argument

Definition

For two random variables X and Y, the coupling (X', Y') of X and Y is a random variable on the product of the sample spaces of X and Y such that the marginal distributions of X' and Y' are the distributions of X and Y, respectively.

Lemma

X, Y : random variables.

1. *Any coupling (X', Y') of X and Y satisfies*

$$\mathrm{TV}(X, Y) \leq \Pr[X' \neq Y'].$$

2. *There exists a coupling such that*

$$\mathrm{TV}(X, Y) = \Pr[X' \neq Y'].$$

Proof of Lemma (1)

- X := the number of columns of the matrix $R(n, m; p)$ with two or more 1's.
- $\Pr[|V_a| = k] = \binom{n}{k} p^k (1-p)^{n-k} =: r_k$
- $X = Binom(m, q_2)$ where $q_2 := \sum_{k \geq 2} r_k$.

$G(n, m; p)$ can be constructed as follows:

1. $K^{(1)}, \ldots, K^{(h)}, \ldots$: i.i.d. random complete graphs on subsets of V
   - the number of vertices in $K^{(1)}$ is $k(\geq 2)$ with probability r_k / q_2
   - Then, once the number is given to be k, every k-subset of V is equally likely to be the vertex set of $K^{(1)}$.
   - In other words, for a k-subset U of V with $k \geq 2$, the probability of U being the vertex set of $K^{(1)}$ is $\frac{r_k}{q_2}\binom{n}{k}^{-1}$.
2. $G(n, m; p)$ is the edge union of X random complete graphs $K^{(1)}, \ldots, K^{(X)}$.

Definition (G_Y)

- $Y := Poisson(mq_2)$ that is coupled with X so that
$$\Pr[X \neq Y] = \mathrm{TV}(X, Y).$$
- Let G_Y be the graph whose edge set is the (edge) union of $K^{(1)}, \ldots, K^{(Y)}$.

Property

1. G_Y has the same distribution as $G(n, (p_k))$.
2. $$\begin{aligned}\mathrm{TV}(G(n, m; p), G_Y) &\leq \Pr[G(n, m; p) \neq G_Y] \\ &\leq \Pr[X \neq Y] = \mathrm{TV}(X, Y).\end{aligned}$$

Lemma (Barbour and Holst (1989))

Let $X := Binom(m, q_2)$ and $Y := Poisson(mq_2)$. Then

$$\mathrm{TV}(X, Y) \le q_2.$$

$$\mathrm{TV}(X, Y) \le q_2 = \sum_{k\ge 2} \binom{n}{k} p^k (1-p)^{n-k} \le \sum_{k\ge 2} n^k p^k = O\left(n^2 p^2\right) = O\left(\frac{n^2 \log n}{m}\right) = o(1).$$

Problem

Problem

Fix $3 < \alpha < 6$, and let $m = n^{\alpha}$.
Find a probability $p^* = p^*(n, m)$ such that

- If $p \ll p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) = o(1)$.
- If $p \gg p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) \geq c$, for some positive constant $c > 0$.

Problem

Problem

Fix $3 < \alpha < 6$, and let $m = n^{\alpha}$.
Find a probability $p^* = p^*(n, m)$ such that

- If $p \ll p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) = o(1)$.
- If $p \gg p^*$, then $\mathrm{TV}\left(G(n, m; p), G(n, \hat{p})\right) \geq c$,
  for some positive constant $c > 0$.

Question

When $G(n, m; p) \not\approx G(n, \hat{p})$,
what are interesting properties and structures of $G(n, m; p)$?