Deviations of triangle counts in the binomial random graph

Wojciech Samotij (Tel Aviv University)

joint work with Matan Harel & Frank Mousset

SCMS online seminar

Our setting:

$G_{n,p}$: the binomial random graph with vx set $\{1, \dots, n\}$ and edge prob. p

$X :=$ #triangles in $G_{n,p}$ $\qquad \mathbb{E}[X] = \binom{n}{3} p^3$

Problem. For every $\delta > 0$, determine the asymptotics of

$$\underbrace{\log \mathbb{P}\big(X \geq (1+\delta)\,\mathbb{E}[X]\big)}_{\text{logarithmic upper tail probability}}.$$

Theorem (Chatterjee / DeMarco–Kahn 2012)

If $p \gg \log n / n$, then, for every $\delta > 0$,

$$-\log \mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) = \Theta_\delta(n^2 p^2 \log(1/p)).$$

Question. What is the implicit constant?

Definition. For $\delta > 0$,

$$\psi(\delta) := \min\{e(G) : \underbrace{\mathbb{E}[X \mid G \subseteq G_{n,p}]}_{\mathbb{E}_G[X]} \geq (1+\delta)\mathbb{E}[X]\}.$$

Proposition. If $\psi(\delta) \to \infty$, then
$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \geq p^{(1+o(1))\psi(\delta)}.$$

Proof: $\mathbb{P}_G(\cdot) = \mathbb{P}(\cdot \mid G \subseteq G_{n,p})$

Pick a small $\varepsilon > 0$ and a $G \subseteq K_n$ with $\psi(\delta+\varepsilon)$ edges s.t. $\mathbb{E}_G[X] \geq (1+\delta+\varepsilon)\mathbb{E}[X]$. Note that:

(i) $\mathbb{P}(UT_\delta) \geq \mathbb{P}(G \subseteq G_{n,p}) \cdot \mathbb{P}_G(UT_\delta) = p^{\psi(\delta+\varepsilon)} \cdot \mathbb{P}_G(UT_\delta)$.

(ii) $\mathbb{E}_G[X] \leq \mathbb{P}_G(UT_\delta^c) \cdot (1+\delta)\mathbb{E}[X] + \mathbb{P}_G(UT_\delta) \cdot \binom{n}{3}$.

Finally, (ii) $\Rightarrow \mathbb{P}_G(UT_\delta) \geq \varepsilon \mathbb{E}[X] / \binom{n}{3} = \varepsilon p^3 = p^{o(\psi(\delta))}$. $\square$

Theorem (...) If $n^{-\alpha} \tilde{\ll} p \ll 1$, then

$$\mathbb{P}(X \geqslant (1+\delta)\mathbb{E}[X]) \leqslant p^{(1-o(1))\psi(\delta)}.$$

- Chatterjee–Dembo (2014) $\qquad \alpha = \frac{1}{42}$
- Eldan (2016) $\qquad \alpha = \frac{1}{18}$
- Cook–Dembo (2018) $\qquad \alpha = \frac{1}{3}$
- Augeri (2018) $\qquad \alpha = \frac{1}{2}$

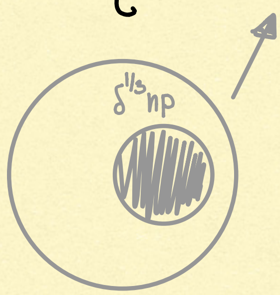
Theorem (Harel, Mousset, S. 2019) If $n^{-1}\log n \ll p \ll 1$, then, for every $\delta > 0$,

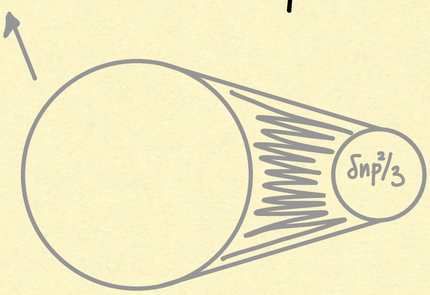
$$\mathbb{P}\big(X \geq (1+\delta)\mathbb{E}[X]\big) \leq p^{(1-o(1))\psi(\delta)}.$$

Theorem (Lubetzky, Zhao 2014) For every $\delta > 0$,

$$\psi(\delta)/n^2p^2 \to \begin{cases} \delta^{2/3}/2 & \text{if } n^{-1} \ll p \ll n^{-1/2}, \\ \min\{\delta^{2/3}/2,\ \delta/3\} & \text{if } n^{-1/2} \ll p \ll 1. \end{cases}$$

clique





hub

Proof: Fix a small $\varepsilon > 0$ and a large C.

Definition. A graph $G \subseteq K_n$ is a **seed** if

(1) $e(G) \leq C n^2 p^2 \log(1/p)$ $\qquad \psi(\delta) = \Theta(n^2 p^2)$

(2) $\mathbb{E}_G[X] \geq (1+\delta-\varepsilon)\mathbb{E}[X]$

Lemma 1. $\mathbb{P}(UT) \leq (1+o(1)) \cdot \mathbb{P}(G_{n,p} \text{ contains a seed})$.

Union bound over all seeds? Too naive...

G_0: a seed with $m := \psi(\delta+\varepsilon)$ edges & $e(G_1) \leq 2m$

$\Rightarrow G_0 \cup G_1$ is a seed

$$\mathbb{E}[\#\text{seeds in } G_{n,p}] \geq \binom{n^2/3}{2m} p^{3m} \geq \left(\frac{n^2 p^{3/2}}{6m}\right)^m = \Omega\left(p^{-1/2}\right)^m \to \infty.$$

Definition. A graph G^* is a **core** if

(1) $e(G^*) \le Cn^2p^2\log(1/p)$

(2) $\mathbb{E}_{G^*}[X] \geq (1+\delta-2\varepsilon)\,\mathbb{E}[X]$

(3) $\forall\, uv \in G^* \quad \mathbb{E}_{G^*}[X] - \mathbb{E}_{G^*\setminus uv}[X] \geq \dfrac{\varepsilon\,\mathbb{E}[X]}{Cn^2p^2\log(1/p)}$

Fact. Every seed contains a core.

Proof: Iteratively remove from a seed all edges failing (3).

Lemma 2. For every m,

$$\#\text{cores with } m \text{ edges} \le p^{-\varepsilon m}.$$

Assuming both lemmas

$$\mathbb{P}(UT) \overset{L1}{\lessapprox} \mathbb{P}(G_{n,p} \text{ contains a seed})$$

$$\overset{F}{\lesssim} \mathbb{P}(G_{n,p} \text{ contains a core})$$

$$\leq \mathbb{E}[\#\text{cores in } G_{n,p}]$$

$$\leq \sum_m \mathbb{E}[\#\text{cores with } m \text{ edges in } G_{n,p}]$$

$$\overset{L2}{\leq} \sum_{m \geq \psi(\delta-2\varepsilon)} p^{-\varepsilon m} \cdot p^{m} \approx p^{(1-\varepsilon)\,\psi(\delta-2\varepsilon)}$$

$\psi(\delta-2\varepsilon) \leq$ #edges in a core

Proof of Lemma 1: (based on Janson-Oleszkiewicz-Ruciński)

Let $Z := \mathbb{1}[G_{n,p}$ does not contain a seed$]$. As $Z \in \{0,1\}$,

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq \mathbb{P}(Z=0) + \underbrace{\mathbb{P}(XZ \geq (1+\delta)\mathbb{E}[X])}_{⊛}$$

Enough to show: $⊛ \ll p^{2\psi(\delta)} \leq \mathbb{P}(X \geq (1+\delta)\mathbb{E}[X])$

Markov's inequality $\Rightarrow \forall \ell \geq 1 \quad ⊛ \leq \dfrac{\mathbb{E}[(XZ)^\ell]}{(1+\delta)^\ell \cdot \mathbb{E}[X]^\ell}$

Claim. $3\ell \leq Cn^2p^2\log(1/p) \implies \mathbb{E}[(XZ)^\ell] \leq (1+\delta-\varepsilon)^\ell \mathbb{E}[X]^\ell$.

Claim. $3\ell \leq Cn^2p^2\log(1/p) \implies \mathbb{E}[(XZ)^\ell] \leq (1+\delta-\varepsilon)^\ell \mathbb{E}[X]^\ell$.

Proof:

$$\mathbb{E}[(XZ)^\ell] = \sum_{T_1,\dots,T_\ell} \mathbb{P}(T_1 \cup \dots \cup T_\ell \subseteq G_{n,p} \wedge Z = 1)$$

$$\leq \sum_{\substack{T_1,\dots,T_\ell \\ T_1\cup\dots\cup T_\ell \text{ not a seed}}} \mathbb{P}(T_1 \cup \dots \cup T_\ell \subseteq G_{n,p})$$

$e(T_1\cup\dots\cup T_\ell) \leq 3\ell \leq Cn^2p^2\log(1/p)$

$$\leq \sum_{\substack{T_1,\dots,T_{\ell-1} \\ T_1\cup\dots\cup T_{\ell-1} \text{ not a seed}}} \mathbb{P}(T_1 \cup \dots \cup T_{\ell-1} \subseteq G_{n,p}) \cdot \overbrace{\sum_{T_\ell} \mathbb{P}(T_\ell \subseteq G_{n,p} \mid T_1\cup\dots\cup T_{\ell-1} \subseteq G_{n,p})}^{\mathbb{E}_{T_1\cup\dots\cup T_{\ell-1}}[X] < (1+\delta-\varepsilon)\mathbb{E}[X]}$$

$$\leq \dots \leq \big((1+\delta-\varepsilon)\mathbb{E}[X]\big)^\ell$$

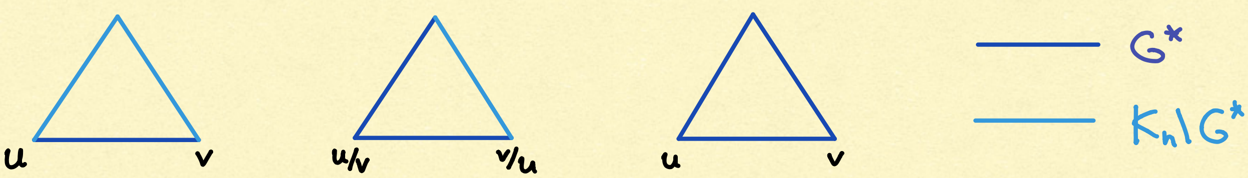
$\square$

Proof of Lemma 2:

Suppose G^*: core graph with m edges

$$G^* \ni uv \mapsto \partial_{uv} := \mathbb{E}_{G^*}[X] - \mathbb{E}_{G^* \setminus uv}[X]$$

- $\partial_{uv} \geq \dfrac{\varepsilon \mathbb{E}[X]}{C n^2 p^2 \log(1/p)} \geq \dfrac{\gamma n p}{\log(1/p)}$
- $\partial_{uv} \leq n(p^2 - p^3) + (d_u + d_v)(p - p^2) + d_{u,v}(1-p)$



$\Longrightarrow$ either $d_u + d_v \geq \gamma' n / \log(1/p)$ or $d_{u,v} \geq \gamma' np / \log(1/p)$.

Cor. $\forall uv \in G^*$ $d_u + d_v \geq \gamma' n / \log(1/p)$ or $d_{u,v} \geq \gamma' np / \log(1/p)$.

Define $A := \{ w \in V(G^*) : d_w \geq \gamma' n / (2 \log(1/p)) \}$,

$B := \{ w \in V(G^*) : d_w \geq \gamma' np / \log(1/p) \}$.

Obs. $2m = 2e(G^*) = \sum_{w \in V(G^*)} d_w \geq \max \left\{ |A| \cdot \frac{\gamma' n}{2 \log(1/p)} , |B| \cdot \frac{\gamma' np}{\log(1/p)} \right\}$.

$$\Rightarrow \quad |A| \leq a_{max} := \frac{4m \log(1/p)}{\gamma' n} \quad \& \quad |B| \leq b_{max} := \frac{2m \log(1/p)}{\gamma' np}$$

every edge has 2 endpts in B or ≥1 endpt in A

$$\#\text{cores w/ } m \text{ edges} \leq \underbrace{\binom{n}{a_{max}}}_{\#A} \cdot \underbrace{\binom{n}{b_{max}}}_{\#B} \cdot \binom{b_{max}^2 + n \cdot a_{max}}{m} \leq \ldots \leq p^{-\varepsilon m}.$$

$\square$

Problem. Calculate

$$-\log \mathbb{P}\big(X_H \geq (1+\delta)\mathbb{E}[X_H]\big),$$

where $X_H :=$ #copies of H in $G_{n,p}$.

State-of-the-art:

- $H = K_r$ Harel-Mousset-S.
- H: connected & regular HMS / Basak-Basu
- H: irregular Cook-Dembo