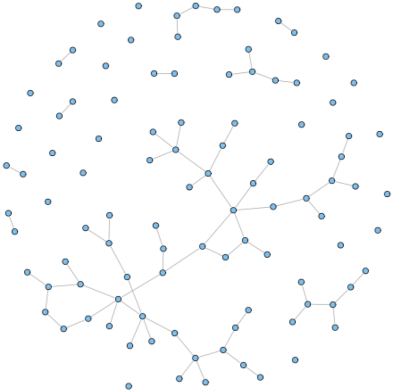
Number of isomorphic copies of a given graph in a random graph

Grzegorz Serafin

Joint work with Nicolas Privault (Singapore)

Probability and Analysis
Będlewo, 20-24.05.2019

(Binomial) Erdős-Rényi random graph $\mathbb{G}(n, p_n)$**:** A graph with n vertices, wherein any of two vertices are connected independently with probability $p_n \in (0, 1)$.



Notation:

- G a fixed graph

Notation:

- G a fixed graph

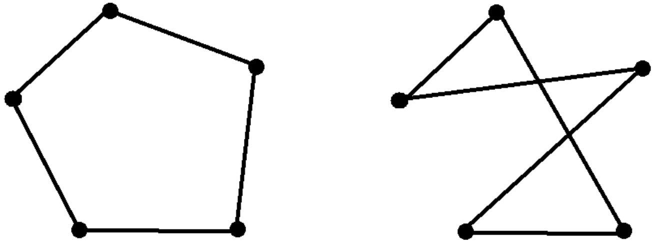
- e_G - number of edges in G
  v_G - number of vertices in G

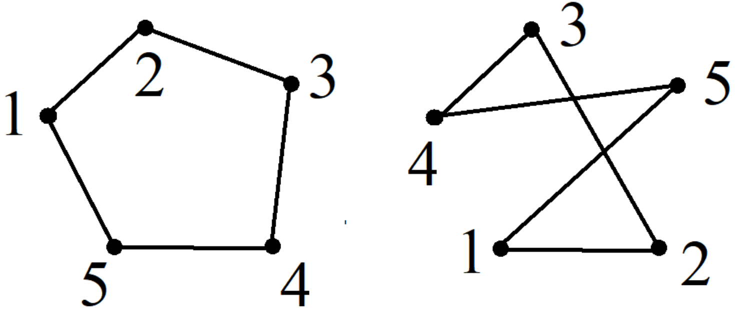
Notation:

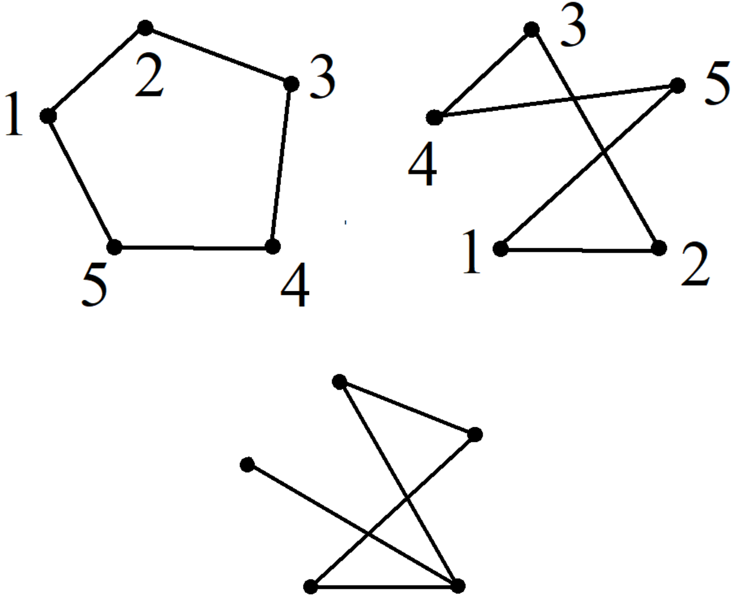
- G a fixed graph

- e_G - number of edges in G
  v_G - number of vertices in G

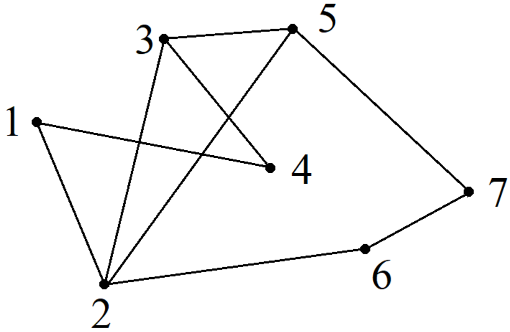
- N_n^G - number of graphs in $\mathbb{G}_n(p_n)$ that are isomorphic to a fixed graph G.



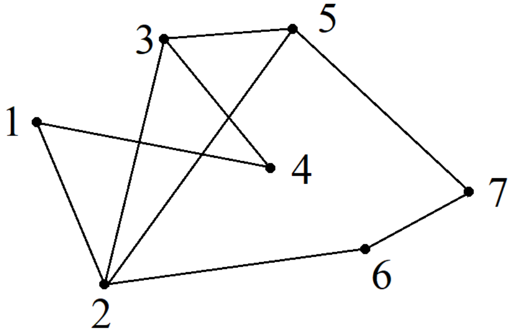

1
2
3
4
5
3
5
4
1
2


1
2
3
4
5
3
5
4
1
2

$\mathbb{G}_7(p_7)$

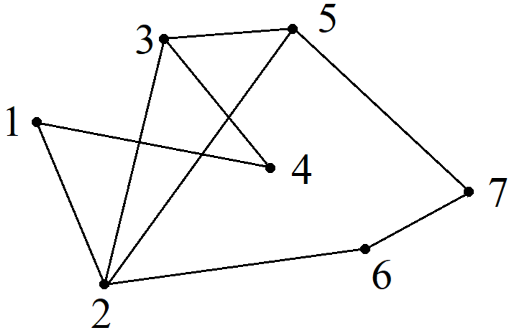


$\mathbb{G}_7(p_7)$



$$G = \triangle \quad \to \quad N_7^G = \#\{(2,3,5)\} = 1,$$

$\mathbb{G}_7(p_7)$



$$G = \triangle \quad \to \quad N_7^G = \#\{(2,3,5)\} = 1,$$

$$G = \square \quad \to \quad N_7^G = \#\{(2,6,7,5),(1,2,3,4)\} = 2.$$

- We have the equivalence:

$$\mathrm{Var}\left[N_n^G\right] \approx (1-p_n) n^{2v_G} p_n^{2e_G} \left(\min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p_n^{e_H} \right)^{-1}.$$

- We have the equivalence:

$$\mathrm{Var}\left[N_n^G\right] \approx (1-p_n)n^{2v_G}p_n^{2e_G}\left(\min_{\substack{H\subset G\\ e_H\geq 1}} n^{v_H}p_n^{e_H}\right)^{-1}.$$

- Consider the renormalization

$$\widetilde{N}_n^G := \frac{N_n^G - \mathbb{E}[N_n^G]}{\sqrt{\mathrm{Var}[N_n^G]}}.$$

- We have the equivalence:

$$\mathrm{Var}\left[N_n^G\right] \approx (1-p_n)n^{2v_G}p_n^{2e_G}\left(\min_{\substack{H\subset G\\ e_H\geq 1}} n^{v_H}p_n^{e_H}\right)^{-1}.$$

- Consider the renormalization

$$\widetilde{N}_n^G := \frac{N_n^G - \mathbb{E}[N_n^G]}{\sqrt{\mathrm{Var}[N_n^G]}}.$$

- Ruciński 1988:

$$\widetilde{N}_n^G \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \textit{iff} \quad (1-p_n)\min_{\substack{H\subset G\\ e_H\geq 1}}\{n^{v_H}p_n^{e_H}\} \to \infty,$$

- We have the equivalence:

$$\mathrm{Var}\left[N_n^G\right] \approx (1-p_n) n^{2v_G} p_n^{2e_G} \left(\min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p_n^{e_H} \right)^{-1}.$$

- Consider the renormalization

$$\widetilde{N}_n^G := \frac{N_n^G - \mathbb{E}[N_n^G]}{\sqrt{\mathrm{Var}[N_n^G]}}.$$

- Ruciński 1988:

$$\widetilde{N}_n^G \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \textit{iff} \quad (1-p_n) \min_{\substack{H \subset G \\ e_H \geq 1}} \{ n^{v_H} p_n^{e_H} \} \to \infty,$$

- How to quantify the speed of convergence?

Distance between random variables

- The Wasserstein distance between the laws of X and Y:

$$d_W(X, Y) := \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|,$$

where $\mathrm{Lip}(1)$ is the class of real-valued Lipschitz functions with Lipschitz constant less than or equal to 1.

Distance between random variables

- The Wasserstein distance between the laws of X and Y:

$$d_W(X,Y) := \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|,$$

where $\mathrm{Lip}(1)$ is the class of real-valued Lipschitz functions with Lipschitz constant less than or equal to 1.

- The Kolmogorov distance between the laws of X and Y:

$$d_K(X,Y) := \sup_{x \in \mathbb{R}} |P(X \leq x) - P(Y \leq x)|,$$

Distance between random variables

- The Wasserstein distance between the laws of X and Y:

$$d_W(X,Y) := \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|,$$

  where $\mathrm{Lip}(1)$ is the class of real-valued Lipschitz functions with Lipschitz constant less than or equal to 1.

- The Kolmogorov distance between the laws of X and Y:

$$d_K(X,Y) := \sup_{x \in \mathbb{R}} |P(X \leq x) - P(Y \leq x)|,$$

If the random variable Y has Lebesgue density bounded by $C > 0$, then

$$d_K(X,Y) \leq \sqrt{2C d_W(X,Y)}.$$

- Barbour, Karoński, Ruciński (1989):

$$d_W(\tilde{N}_n^G, \mathcal{N}) \leq C\varepsilon_n,$$

where $\varepsilon_n := \left((1 - p_n) \min_{\substack{H \subset G \\ e_H \geq 1}} \{ n^{v_H} p_n^{e_H} \} \right)^{-1/2}$.

- Barbour, Karoński, Ruciński (1989):

$$d_W(\tilde{N}_n^G, \mathcal{N}) \le C\varepsilon_n,$$

where $\varepsilon_n := \left((1-p_n) \min_{\substack{H \subset G \\ e_H \ge 1}} \{ n^{v_H} p_n^{e_H} \} \right)^{-1/2}.$

Note: this implies $d_K(\tilde{N}_n^G, \mathcal{N}) \le C_G \sqrt{\varepsilon_n}$.

- Barbour, Karoński, Ruciński (1989):

$$d_W(\tilde{N}_n^G, \mathcal{N}) \leq C\varepsilon_n,$$

where $\varepsilon_n := \left((1-p_n) \min_{\substack{H \subset G \\ e_H \geq 1}} \{ n^{v_H} p_n^{e_H} \} \right)^{-1/2}.$

Note: this implies $d_K(\tilde{N}_n^G, \mathcal{N}) \leq C_G \sqrt{\varepsilon_n}$.

- Krokowski, Reichenbachs, Thaele (2015):
$G = \triangle$, $p_n \approx n^{-\alpha}$, $\alpha \in (0,1)$.

$$d_K(\tilde{N}_n^{\triangle}, \mathcal{N}) \leq C[\text{better than } \sqrt{\varepsilon_n}, \text{ worse than } \varepsilon_n].$$

- Barbour, Karoński, Ruciński (1989):

$$d_W(\tilde{N}_n^G, \mathcal{N}) \leq C\varepsilon_n,$$

where $\varepsilon_n := \left((1-p_n) \min_{\substack{H \subset G \\ e_H \geq 1}} \{ n^{v_H} p_n^{e_H} \} \right)^{-1/2}$.

Note: this implies $d_K(\tilde{N}_n^G, \mathcal{N}) \leq C_G \sqrt{\varepsilon_n}$.

- Krokowski, Reichenbachs, Thaele (2015):
$G = \triangle$, $p_n \approx n^{-\alpha}$, $\alpha \in (0,1)$.

$$d_K(\tilde{N}_n^{\triangle}, \mathcal{N}) \leq C[\text{better than } \sqrt{\varepsilon_n}, \text{ worse than } \varepsilon_n].$$

- Röllin (2017): $G = \triangle$.

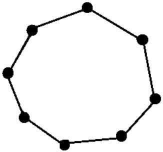
$$d_K(\tilde{N}_n^{\triangle}, \mathcal{N}) \leq C\varepsilon_n.$$

Let G be a graph without isolated vertices.

Theorem (N.P., G. Serafin (2018+))

We have

$$d_K(\tilde{N}_n^G, \mathcal{N}) \lesssim \left((1-p) \min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p^{e_H} \right)^{-1/2}.$$

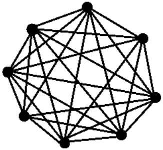
$G =$ 

Corollary

Let C_r be a cycle graph with r vertices, $r \geq 3$. We have

$$d_K\left(\widetilde{N}_n^{C_r}, \mathcal{N}\right) \lesssim \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } n^{-(r-2)/(r-1)} < p_n, \\ \dfrac{1}{(np_n)^{r/2}} & \text{if } 0 < p_n \leq n^{-(r-2)/(r-1)}. \end{cases}$$

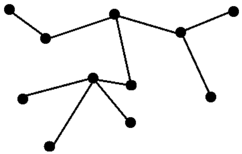
Complete graphs

$G =$ 

Corollary

Let K_r be a complete graph with $r \geq 3$ vertices, $r \geq 3$. We have

$$d_K\left(\widetilde{N}_n^G, \mathcal{N}\right) \lesssim \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } n^{-2/(r+1)} < p_n, \\[2ex] \dfrac{1}{n^{r/2} p_n^{r(r-1)/4}} & \text{if } 0 < p_n \leq n^{-2/(r+1)}. \end{cases}$$

$G =$ 

Corollary

Let T be any tree (a connected graph without cycles) with r edges, and $c \in (0,1)$. We have

$$d_K\left(\widetilde{N}_n^T, \mathcal{N}\right) \lesssim \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } \dfrac{1}{n} < p_n, \\[2ex] \dfrac{1}{n^{(r+1)/2} p_n^{r/2}} & \text{if } 0 < p_n \le \dfrac{1}{n}. \end{cases}$$

Observation:

If G is a cycle, complete graph tree trees we have

$$\min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p^{e_H} = \min\{n^2 p, n^{v_G} p^{e_G}\}$$

- We consider the class

$$\mathcal{B} := \left\{ graph\ G : v_G \geq 3 \wedge \left(\max_{H \subset G} \frac{e_H - 1}{v_H - 2} = \frac{e_G - 1}{v_G - 2} \right) \right\}$$

which is reminiscent of balanced graphs

- We consider the class

$$\mathcal{B} := \left\{ graph\ G : v_G \geq 3 \wedge \left(\max_{H \subset G} \frac{e_H - 1}{v_H - 2} = \frac{e_G - 1}{v_G - 2} \right) \right\}$$

  which is reminiscent of balanced graphs
- $\mathcal{B}$ is a subclass of balanced graphs defined by $\max_{H \subset G} \frac{e_H}{v_H} = \frac{e_G}{v_G}$.

- We consider the class

$$\mathcal{B} := \left\{ graph\ G : v_G \geq 3 \wedge \left(\max_{H \subset G} \frac{e_H - 1}{v_H - 2} = \frac{e_G - 1}{v_G - 2} \right) \right\}$$

  which is reminiscent of balanced graphs
- $\mathcal{B}$ is a subclass of balanced graphs defined by $\max_{H \subset G} \frac{e_H}{v_H} = \frac{e_G}{v_G}$.
- All complete graphs, cycles and trees with at least 3 vertices belong to the class $\mathcal{B}$.

Lemma

Let G be a graph with $v_H \geq 3$ *and* $e_H \geq 1$*. Then we have*

$$G \in \mathcal{B} \Leftrightarrow \left(\forall_{p \in (0,1)} \min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p^{e_H} = \min\{n^2 p, n^{v_G} p^{e_G}\} \right).$$

Lemma

Let G be a graph with $v_H \geq 3$ and $e_H \geq 1$. Then we have

$$G \in \mathcal{B} \Leftrightarrow \left(\forall_{p \in (0,1)} \min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p^{e_H} = \min\{n^2 p, n^{v_G} p^{e_G}\} \right).$$

Corollary

For $G \in \mathcal{B}$ we have

$$d_K(\tilde{N}_G, \mathcal{N}) \leq C \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } \; n^{-(v_G-2)/(e_G-1)} < p_n, \\ \dfrac{1}{n^{v_G/2} p^{e_G/2}} & \text{if } \; 0 < p_n \leq n^{-(v_G-2)/(e_G-1)}. \end{cases}$$

Rademacher functionals

- Consider $\{X_i\}$ an i.i.d. sequence of Bernoulli random variables with $P(X_i = 1) = p$, $P(X_i = -1) = 1 - p$.

Rademacher functionals

- Consider $\{X_i\}$ an i.i.d. sequence of Bernoulli random variables with $P(X_i = 1) = p$, $P(X_i = -1) = 1 - p$.
- Consider the normalization

$$Y_i = \frac{X_i - p + q}{2\sqrt{pq}}$$

Rademacher functionals

- Consider $\{X_i\}$ an i.i.d. sequence of Bernoulli random variables with $P(X_i = 1) = p$, $P(X_i = -1) = 1 - p$.
- Consider the normalization

$$Y_i = \frac{X_i - p + q}{2\sqrt{pq}}$$

- Define the multiple stochastic integrals

$$J_k(f_k) := \sum_{\substack{i_1,\dots,i_k \\ i_r \neq i_s \text{ if } r \neq s}} f_k(i_1, \dots, i_k) Y_{i_1} \dots Y_{i_k},$$

where $f_k : \mathbb{N}^k \to \mathbb{R}$.
- For suitable f_k we have (Privault, S.)

$$N_n^G = \sum_{k=0}^{e_G} I_k(f_k).$$

General results

Most remarkable previous results:

General results

Most remarkable previous results:

- L.H.Y. Chen and Q.M. Shao (2007):

  Normal approximation for U-statistics

General results

Most remarkable previous results:

- L.H.Y. Chen and Q.M. Shao (2007):

  Normal approximation for U-statistics

  Note: no sums, no weighted U-statistics

General results

Most remarkable previous results:

- L.H.Y. Chen and Q.M. Shao (2007):

  Normal approximation for U-statistics

  Note: no sums, no weighted U-statistics

- K. Krokowski, A. Reichenbachs, and C. Thaele (2016):

  Normal approximation for $J_k(f_k)$ in symmetric case

General results

Most remarkable previous results:

- L.H.Y. Chen and Q.M. Shao (2007):

  Normal approximation for U-statistics

  Note: no sums, no weighted U-statistics

- K. Krokowski, A. Reichenbachs, and C. Thaele (2016):

  Normal approximation for $J_k(f_k)$ in symmetric case

  Note: no sums, only $p = 1/2$.

Generl results

Theorem

For any finite sum $F = \sum_{k=1}^{m} I_k(f_k)$ *with* $\mathbb{E}[f^2] = 1$ *we have*

$$d_K(F, \mathcal{N}) \leq C_m \Bigg[\sum_{0 \leq l < i \leq m} (pq)^{l-i} \left\| f_i \star_i^l f_i \right\|_{\ell^2(\mathbb{N})^{\otimes(i-l)}} + \sum_{1 \leq l < i \leq m} \left(\left\| f_l \star_l^l f_i \right\|_{\ell^2(\mathbb{N})^{\otimes(i-l)}} + \left\| f_i \star_l^l f_i \right\|_{\ell^2(\mathbb{N})^{\otimes 2(i-l)}} \right) \Bigg],$$

where C_m *depends only on* m *and*

$$f_i \star_k^l f_j(a, b_1, b_2) = \sum_{d \in \mathbb{N}^l} f_i(a, b_1, d) f_j(a, b_2, d).$$

Random weights

- Let $X \geq 0$ such that $\mathbb{E}[X^4] < \infty$

Random weights

- Let $X \geq 0$ such that $\mathbb{E}[X^4] < \infty$
- We assign an independent weight X to each edge of the Erdős-Rényi graph $\mathbb{G}_n(p_n)$.

Random weights

- Let $X \geq 0$ such that $\mathbb{E}[X^4] < \infty$
- We assign an independent weight X to each edge of the Erdős-Rényi graph $\mathbb{G}_n(p_n)$.
- Define the weight of the graph G as the sum of its edge weights.

Random weights

- Let $X \geq 0$ such that $\mathbb{E}[X^4] < \infty$
- We assign an independent weight X to each edge of the Erdős-Rényi graph $\mathbb{G}_n(p_n)$.
- Define the weight of the graph G as the sum of its edge weights.
- Let W_n^G denote the combined weight of graphs in $\mathbb{G}_n(p_n)$ that are isomorphic to a fixed graph G.

Random weights

- Let $X \geq 0$ such that $\mathbb{E}[X^4] < \infty$
- We assign an independent weight X to each edge of the Erdős-Rényi graph $\mathbb{G}_n(p_n)$.
- Define the weight of the graph G as the sum of its edge weights.
- Let W_n^G denote the combined weight of graphs in $\mathbb{G}_n(p_n)$ that are isomorphic to a fixed graph G.
- Define the renormalized random variable

$$\widetilde{W}_n^G := \frac{W_n^G - \mathbb{E}[W_n^G]}{\sqrt{\mathrm{Var}[W_n^G]}}.$$

Random weights

Theorem (N.P., G. Serafin (2018))

Let G be a graph without isolated vertices. We have

$$d_W(\tilde{W}_G, \mathcal{N}) \lesssim \frac{\sqrt{\mathbb{E}[(X-\mathbb{E}[X])^4]} + (1-p_n)(\mathbb{E}[X])^2}{Var[X] + (1-p_n)(\mathbb{E}[X])^2} \times \left((1-p_n) \min_{\substack{H\subset G \\ e_K \geq 1}} \{ n^{v_H} p_n^{e_H} \} \right)^{-1/2}.$$

Random weights

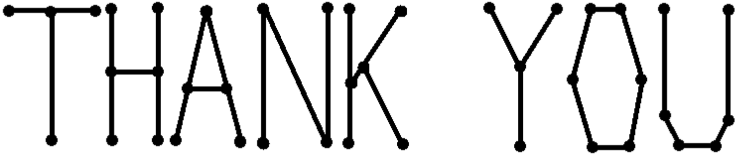
Theorem (N.P., G. Serafin (2018))

Let G be a graph without isolated vertices. We have

$$d_W(\tilde{W}_G, \mathcal{N}) \lesssim \frac{\sqrt{\mathbb{E}[(X-\mathbb{E}[X])^4]} + (1-p_n)(\mathbb{E}[X])^2}{Var[X] + (1-p_n)(\mathbb{E}[X])^2} \times \left((1-p_n) \min_{\substack{H \subset G \\ e_K \geq 1}} \{ n^{v_H} p_n^{e_H} \} \right)^{-1/2}.$$

Therefore we have the implication

$$(1-p) \min_{\substack{H \subset G \\ e_H \geq 1}} \{ n^{v_H} p_n^{e_H} \} \to \infty \quad \Rightarrow \quad \tilde{W}_G \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$


THANK YOU