Armand M. Makowski

ECE & ISR/HyNet
University of Maryland at College Park
armand@isr.umd.edu

RANDOM GRAPHS

Graph $G = (V, E)$

- A **graph** $G = (V, E)$ with set V of nodes and edge set $E \subseteq V \times V$
  - Undirected
  $$(x, y) \in E \quad \text{iff} \quad (y, x) \in E$$
  - No self-loop
  $$(x, x) \notin E$$
- Convention
  $$V = \{1, \ldots, n\} = V_n$$

An algebraic view $\boldsymbol{A} \equiv (V, E)$

- **Adjacency matrix** of $G = (V, E)$ is the $n \times n$ matrix $\boldsymbol{A} = (a_{xy})$

$$a_{xy} = \begin{cases} 1 & \text{if } (x, y) \in E \\ 0 & \text{if } (x, y) \notin E \end{cases}$$

  – Undirected – Symmetric matrix

$$a_{xy} = a_{yx}, \quad x, y = 1, \ldots, n$$

  – No self-loop – Zero diagonal elements

$$a_{xx} = 0, \quad x = 1, \ldots, n$$

Counting edges and graphs

- There are **at most**

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

possible edges, i.e., for any $G = (V_n, E)$,

$$|E| \leq \binom{n}{2}$$

- If $\mathcal{G}(V_n)$ denotes the collection of **all** graphs on V_n, then

$$|\mathcal{G}(V_n)| = 2^{\binom{n}{2}} = 2^{\frac{n(n-1)}{2}}$$

Graph properties

- A **graph property** A for graphs on V_n is simply a subset $\mathcal{A}$ of $\mathcal{G}(V_n)$, i.e.,

$$\mathcal{A} \subseteq \mathcal{G}(V_n)$$

- Example 1 – Graph connectivity

$$\mathcal{A}_{\text{Con}} := \{(V_n, E) \in \mathcal{G}(V_n) : \ (V_n, E) \text{ connected}\}$$

- Example 2 – Absence of isolated nodes

$$\begin{aligned} & \mathcal{A}_{\text{No isolated node}} \\ := \quad & \{(V_n, E) \in \mathcal{G}(V_n) : \ (V_n, E) \text{ contains no isolated node}\} \end{aligned}$$

Monotone graph properties

- A graph property A for graphs on V_n is said to be **monotone increasing** if the corresponding subset $\mathcal{A} \subset \mathcal{G}(V_n)$ has the following monotonicity property: For (V_n, E) and (V_n, E') in $\mathcal{G}(V_n)$, the conditions

$$E \subset E' \quad \text{and} \quad (V_n, E) \in \mathcal{A}$$

imply

$$(V_n, E) \in \mathcal{A}$$

- Graph connectivity and absence of isolated nodes are monotone increasing properties

Random graphs

- The finite set $\mathcal{G}(V_n)$ has a natural measurable structure, namely

$$(\mathcal{G}(V_n), \mathcal{P}(\mathcal{G}(V_n)))$$

- A **random graph** over the vertex set V_n is a **probability measure** P_n defined on this measurable space $(\mathcal{G}(V_n), \mathcal{P}(\mathcal{G}(V_n)))$ with **pmf**

$$\{P_n(G),\ G = (V_n, E) \in \mathcal{G}(V_n)\}$$

- Many different ways to generate the pmf P_n
  - Structure!

- A more concrete definition: A random graph over the vertex set V_n is a $\mathcal{G}(V_n)$-valued rv $\mathbb{G}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, i.e.,

$$\mathbb{G} : \Omega \to \mathcal{G}(V_n)$$

with

$$P_n(G) = \mathbb{P}\left[\mathbb{G} = G\right], \quad G = (V_n, E) \in \mathcal{G}(V_n).$$

- For any graph property A on V_n,

$$P_n(A) = \mathbb{P}\left[\mathbb{G} \in \mathcal{A}\right] = \sum_{G \in \mathcal{A}} \mathbb{P}\left[\mathbb{G} = G\right]$$

Examples (Non-geometric)

- Erdős-Renyi graphs
  - $\mathbb{G}(n; m)$ $(1 \leq m \leq \binom{n}{2})$
  - $\mathbb{G}(n; p)$ $(0 \leq p \leq 1)$
- Random intersection graphs
  - $\mathbb{K}(n; K, p)$ $(K = 1, 2, \ldots$ and $0 \leq p \leq 1)$

Geometry!

- A population of n nodes located at $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ in a **compact convex** region $\Omega \subset \mathbb{R}^d$
  - Unit cube $[0,1]^d$, unit ball
- Assume $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ **i.i.d.** distributed according to some non-atomic probability measure μ on Ω
  - The pm μ admits a density $f : \Omega \rightarrow \mathbb{R}_+$, so that

$$\mu(B) = \int_B f(\boldsymbol{x}) d\boldsymbol{x}, \quad B \in \mathcal{B}(\Omega)$$

- Metric $\delta : \mathbb{R}^d \rightarrow \mathbb{R}_+$
  - ℓ_p $(1 \leq p \leq \infty)$

Examples (Geometric)

- Waxman graphs
  - $\mathbb{W}(n; a)$ $(a > 0)$
- Random K-nearest neighbor graphs
  - $\mathbb{N}(n; K)$ $(K = 1, 2, \ldots)$
- Random Yao graphs
  - $\mathbb{Y}(n; \theta)$ $(0 < \theta < 2\pi)$
- Metric random graphs (a.k.a. geometric random graphs)
  - $\mathbb{G}(n; \tau)$ $(\tau > 0)$

The search for typicality

- Consider a family of random graphs

$$\{\mathbb{G}(n;\theta),\ \theta \in \Theta;\ n = 2, 3, \ldots\}$$

and for some graph property A, define

$$P_A(n;\theta) = \mathbb{P}\left[\mathbb{G}(n;\theta) \in \mathcal{A}\right]$$

- Find a scaling function $\theta : \mathbb{N}_0 \rightarrow \Theta : n \rightarrow \theta_n$ such that either

$$\lim_{n\rightarrow\infty} P_A(n;\theta_n) = 1$$

or

$$\lim_{n\rightarrow\infty} P_A(n;\theta_n) = 0$$

- Often, there exists a separation of scales via a **critical** scaling function

$$\theta^\star : \mathbb{N}_0 \to \Theta : n \to \theta_n$$

in the form of a **zero-one** law

$$\lim_{n\to\infty} P_A(n;\theta_n) = \begin{cases} 0 & \text{if } \theta_n \text{ much smaller than } \theta^\star_n \\ \\ 1 & \text{if } \theta_n \text{ much larger than } \theta^\star_n \end{cases}$$

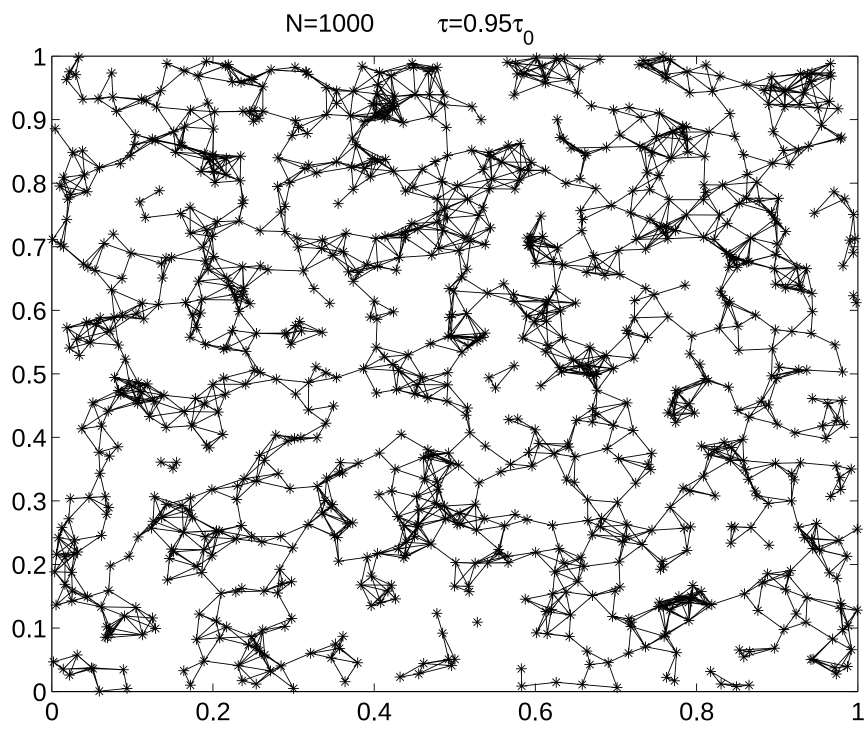
- Basic questions
  - Identify $\theta^\star$ for property A of interest
  - Give precise meaning to statements "θ_n much smaller than $\theta^\star_n$" and "θ_n much larger than $\theta^\star_n$"

GRG $\mathbb{G}_d(n;\tau)$ on $\Omega \subset \mathbb{R}^d$

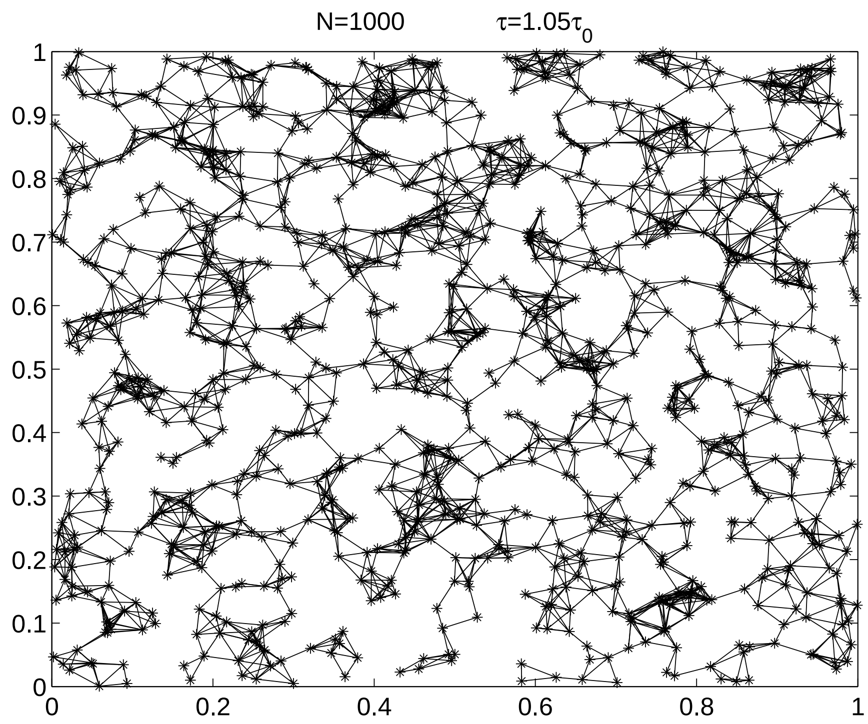
- A population of n nodes located at $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ in **compact convex** region $\Omega \subset \mathbb{R}^d$
- Nodes i and j are connected if $\|\boldsymbol{X}_i - \boldsymbol{X}_j\| \leq \tau$
- Assume $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ **i.i.d.** and **uniformly** distributed on Ω

- Applications to statistical physics, cluster analysis, hypothesis testing and wireless networks
- Appel and Russo, Penrose, Gupta and Kumar, etc.

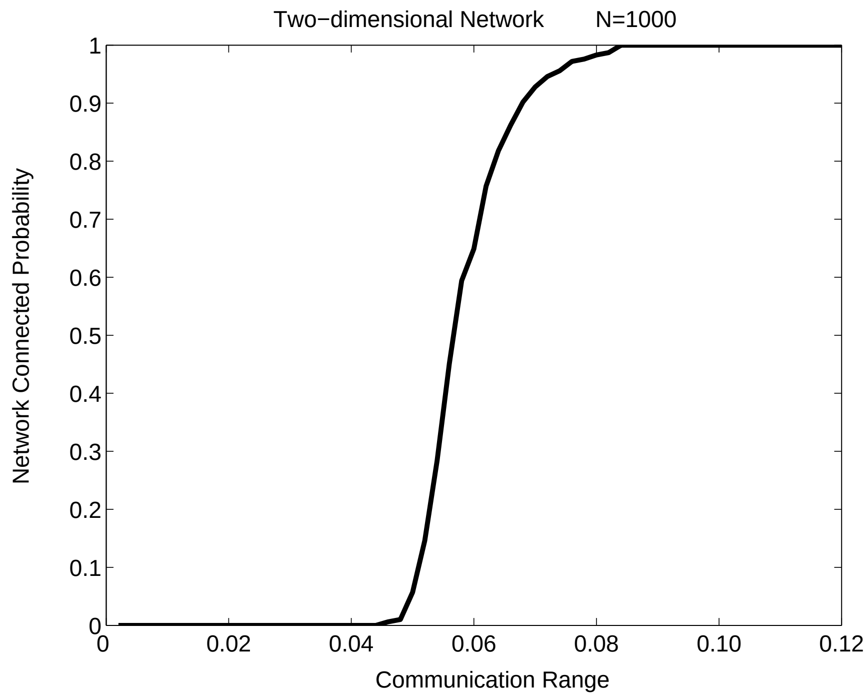
Not yet connected



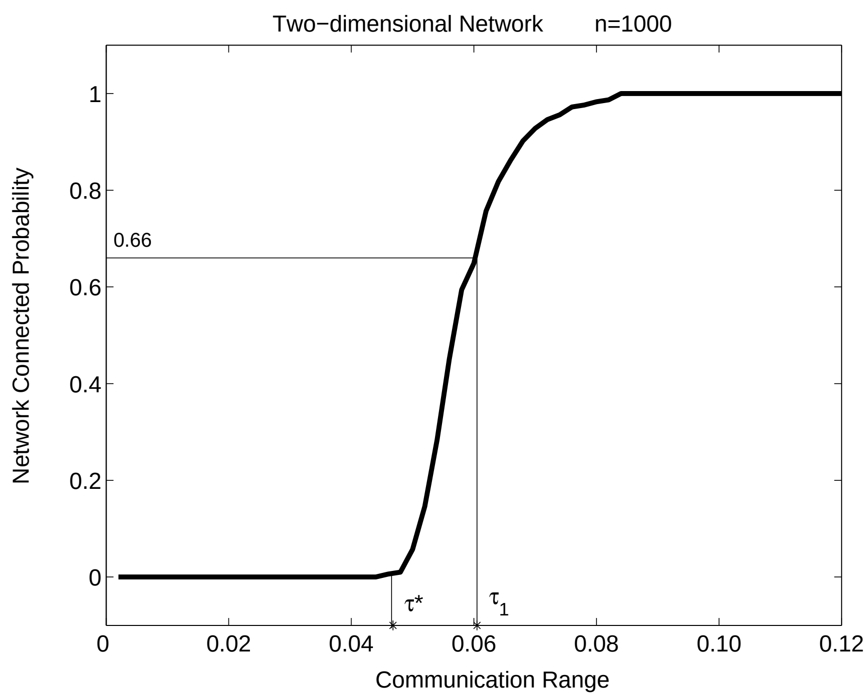
Just connected



Transitions! Transitions!



Phase transitions



CONNECTIVITY
IN THE ONE-DIMENSIONAL MODEL

GRG $\mathbb{G}(n; \tau)$ on $[0, 1]$

- A population of n nodes located at $X_1, \ldots, X_n$ in $[0, 1]$
- Nodes i and j are connected if $|X_i - X_j| \leq \tau$
- Assume $X_1, \ldots, X_n$ **i.i.d.** and **uniformly** distributed on $[0, 1]$
- E.g., Highway networks

Graph connectivity

- For each $n = 2, 3, \ldots$, write

$$P(n;\tau) := \mathbb{P}\left[\mathbb{G}(n;\tau) \text{ is connected}\right], \quad \tau \geq 0$$

- Kendall and Moran (1963), Godehardt and Jaworski (1996), Desai and Manjunath (2002)

$$P(n;\tau) = \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \left((1-k\tau)_+\right)^n$$

Order statistics

- Let $X_{n,1}, \ldots, X_{n,n}$ denote the locations of the n nodes arranged in **increasing** order, i.e.,

$$X_{n,1} \leq \ldots \leq X_{n,n}$$

with the convention $X_{n,0} = 0$ and $X_{n,n+1} = 1$.

- Also define

$$L_{n,k} := X_{n,k} - X_{n,k-1}, \quad k = 1, \ldots, n+1.$$

- For all $\tau \in (0,1)$,

$$P(n;\tau) = \mathbb{P}\left[L_{n,k} \leq \tau,\ k = 2, \ldots, n\right]$$

A useful fact

- For any subset $I \subseteq \{1, \ldots, n\}$,

$$\mathbb{P}\left[L_{n,k} > t_k,\ k \in I\right] = \left(1 - \sum_{k \in I} t_k\right)_+^n, \quad t_k \in [0,1],\ k \in I$$

with the notation

$$x_+^n = \begin{cases} x^n & \text{if } x \geq 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Leads to closed form expression for $P(n;\tau)$ by the mutual inclusion-exclusion principle

ZERO-ONE LAWS

- Does there exists a separation of scales via a **critical** scaling function

$$\tau^\star : \mathbb{N}_0 \longrightarrow \mathbb{R}_+ : n \longrightarrow \tau_n$$

in the form of a **zero-one** law

$$\lim_{n\to\infty} P(n;\tau_n) = \begin{cases} 0 & \text{if } \tau_n \text{ much smaller than } \tau_n^\star \\ \\ 1 & \text{if } \tau_n \text{ much larger than } \tau_n^\star \end{cases}$$

Range functions

No loss of generality in writing a range function

$$\tau : \mathbb{N}_0 \longrightarrow \mathbb{R}_+ : n \longrightarrow \tau_n$$

in the form

$$\tau_n = \frac{1}{n}\left(\log n + \alpha_n\right), \quad n = 1, 2, \ldots . \tag{1}$$

for some deviation function

$$\alpha : \mathbb{N}_0 \longrightarrow \mathbb{R} : n \longrightarrow \alpha_n$$

$$\alpha_n = n\tau_n - \log n, \quad n = 1, 2, \ldots$$

Zero-one law for graph connectivity

Theorem 1 *For any range function* $\tau : \mathbb{N}_0 \to \mathbb{R}_+$ *written in the form (1), we have*

$$\lim_{n\to\infty} P(n;\tau_n) = \begin{cases} 0 & \textit{if } \lim_{n\to\infty}\alpha_n = -\infty \\ \\ 1 & \textit{if } \lim_{n\to\infty}\alpha_n = +\infty. \end{cases}$$

Critical scaling

$$\tau_n^\star = \frac{\log n}{n}, \quad n = 1, 2, \ldots$$

acts as **boundary** in the space of scalings.

Several proofs

- Several representations for $P(n;\tau)$
- Method of first and second moments applied to the number of breakpoint users
- An interpolation result
  - Results by P. Lévy (1939) for maximal spacings
  - Poisson convergence for the the number of breakpoint users

A proof of Theorem 1
by counting
the number of breakpoint nodes

Breakpoint nodes

- For each $i = 1, \ldots, n$, node i is said to be a **breakpoint** node in $\mathbb{G}(n;\tau)$ whenever
  - it is not the leftmost node in $[0,1]$ and
  - there is no node in the random interval $[X_i - \tau, X_i]$.
- The number $C_n(\tau)$ of breakpoint nodes in $\mathbb{G}(n;\tau)$ is given by

$$C_n(\tau) = \sum_{k=2}^{n} \chi_{n,k}(\tau)$$

with indicators

$$\chi_{n,k}(\tau) := \mathbf{1}\left[L_{n,k} > \tau\right], \quad k = 1, \ldots, n+1.$$

- For all $\tau \in (0,1)$,

$$\begin{aligned} P(n;\tau) &= \mathbb{P}\left[L_{n,k} \leq \tau,\ k = 2,\ldots,n\right] \\ &= \mathbb{P}\left[C_n(\tau) = 0\right]. \end{aligned}$$

For all $\tau \in (0,1)$,

$$C_n(\tau) + 1 = \text{Number of connected components in } \mathbb{G}(n;\tau)$$

For future reference

- For all $\tau \in (0,1)$ and all $n = 1, 2, \ldots,$

$$\mathbb{E}\left[C_n(\tau)\right] = (n-1)\,(1-\tau)^n$$

and

$$\begin{aligned}\mathbb{E}\left[C_n(\tau)^2\right] &= \mathbb{E}\left[C_n(\tau)\right] + (n-1)(n-2)\,(1-2\tau)_+^n \\ &= (n-1)\,(1-\tau)^n + (n-1)(n-2)\,(1-2\tau)_+^n\end{aligned}$$

- Observe that

$$
\begin{aligned}
C_n(\tau)^2 &= \left(\sum_{k=2}^{n} \chi_{n,k}(\tau)\right)^2 \\
&= \sum_{k=2}^{n} \chi_{n,k}(\tau) \\
&\quad + \sum_{k,\ell=2,k\neq\ell}^{n} \chi_{n,k}(\tau)\chi_{n,\ell}(\tau)
\end{aligned}
$$

- For all $k, \ell = 1, \ldots, n$, with $k \neq \ell$,

$$\mathbb{E}\left[\chi_{n,k}(\tau)\right] = \mathbb{P}\left[L_{n,k} > \tau\right] = (1-\tau)^n$$

and

$$\mathbb{E}\left[\chi_{n,k}(\tau)\chi_{n,\ell}(\tau)\right] = \mathbb{P}\left[L_{n,k} > \tau, L_{n,\ell} > \tau\right] = (1-2\tau)_+^n$$

Basic inequalities (I)

For any $\mathbb{N}$-valued rv X with $\mathbb{E}[X] < \infty$, we have

$$1 - \mathbb{E}[X] \leq \mathbb{P}[X = 0]$$

A proof

Note that

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=1}^{\infty} x\mathbb{P}[X = x] \\ &\geq \sum_{x=1}^{\infty} \mathbb{P}[X = x] \\ &= \mathbb{P}[X > 0] \end{aligned}$$

Basic inequalities (II)

For any $\mathbb{N}$-valued rv X with $0 < \mathbb{E}\left[X^2\right] < \infty$, we have

$$\mathbb{P}\left[X = 0\right] \leq 1 - \frac{\mathbb{E}\left[X\right]^2}{\mathbb{E}\left[X^2\right]} = \frac{\mathrm{Var}[X]}{\mathbb{E}\left[X^2\right]}$$

A proof

By Cauchy-Schwartz,

$$\begin{aligned} \mathbb{E}\left[X\right]^2 &= \mathbb{E}\left[\mathbf{1}\left[X \neq 0\right] X\right]^2 \\ &\leq \mathbb{E}\left[\mathbf{1}\left[X \neq 0\right]^2\right] \mathbb{E}\left[X^2\right] \end{aligned}$$

so that

$$\frac{\mathbb{E}\left[X\right]^2}{\mathbb{E}\left[X^2\right]} \leq \mathbb{P}\left[X \neq 0\right]$$

A first proof of Theorem 1

Method of **first** moment:

$$1 - \mathbb{E}\left[C_n(\tau)\right] \leq P(n;\tau)$$

for each $n = 2, 3, \ldots$ and τ in $[0, 1]$.

Method of **second** moment:

$$P(n;\tau) \leq 1 - \frac{\mathbb{E}\left[C_n(\tau)\right]^2}{\mathbb{E}\left[C_n(\tau)^2\right]}$$

for each $n = 2, 3, \ldots$ and τ in $[0, 1]$.

The zero-one law follows if for any range function $\tau : \mathbb{N}_0 \to \mathbb{R}_+$ of the form (1), we show that

$$\lim_{n\to\infty} \mathbb{E}\left[C_n(\tau_n)\right] = 0 \quad \text{if} \qquad \lim_{n\to\infty} \alpha_n = \infty$$

and

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[C_n(\tau_n)^2\right]}{\mathbb{E}\left[C_n(\tau_n)\right]^2} = 1 \quad \text{if} \qquad \lim_{n\to\infty} \alpha_n = -\infty.$$

Easily done once we note that

$$\mathbb{E}\left[C_n(\tau)\right] = (n-1)\left(1-\tau\right)_+^n$$

and

$$\frac{\mathbb{E}\left[C_n(\tau)^2\right]}{\mathbb{E}\left[C_n(\tau)\right]^2} = \frac{1}{(n-1)\left(1-\tau\right)_+^n} + \frac{(n-2)}{(n-1)}\frac{\left(1-2\tau\right)_+^n}{\left(1-\tau\right)_+^{2n}}.$$

A proof of Theorem 1 by limiting results on maximal spacings

Maximal spacing

- The **maximal spacing** associated with $X_1, \ldots, X_n$ is given by

$$M_n := \max\left(L_{n,k},\ k = 2, \ldots, n\right)$$

- For all $\tau \in (0, 1)$,

$$\begin{aligned} P(n;\tau) &= \mathbb{P}\left[L_{n,k} \leq \tau,\ k = 2, \ldots, n\right] \\ &= \mathbb{P}\left[M_n \leq \tau\right]. \end{aligned}$$

Variations on a theme by Lévy (1939)

Theorem 2 *It holds that*

$$\frac{M_n}{\tau_n^\star} \xrightarrow{P}{}_n 1$$

and

$$nM_n - \log n \Longrightarrow_n \text{Gumbel } \Lambda$$

The $\mathbb{R}$-valued rv X is Gumbel (Λ) if

$$\mathbb{P}\left[X \leq x\right] = e^{-e^{-x}}, \quad x \in \mathbb{R}$$

Relevance?

For each x in $\mathbb{R}$, consider the range function $\sigma(x) : \mathbb{N}_0 \to \mathbb{R}_+$ given by

$$\sigma_n(x) = \left(\frac{\log n + x}{n} \right)_+ , \quad n = 1, 2, \ldots$$

and

$$\sigma_n(x) = \frac{\log n + x}{n} = \tau_n^\star + \frac{x}{n}$$

for n large enough.

For n large enough,

$$\begin{aligned} P(n;\sigma_n(x)) &= \mathbb{P}\left[M_n \leq \sigma_n(x)\right] \\ &= \mathbb{P}\left[M_n \leq \frac{\log n + x}{n}\right] \\ &= \mathbb{P}\left[nM_n - \log n \leq x\right] \\ &\to_n e^{-e^{-x}} \end{aligned}$$

by Theorem 2.

Interpolating the zero-one law

Theorem 3 *For each x in $\mathbb{R}$, we have*

$$\lim_{n\to\infty} P(n;\sigma_n(x)) = e^{-e^{-x}} =: g(x)$$

- Godehardt and Jaworski (1996)
- Subsumes the zero-one law (Theorem 1)
- A natural question: Where is Theorem 3 coming from?

Theorem 3 implies Theorem 1

- Pick x in $\mathbb{R}$. With $\lim_{n\to\infty} \alpha_n = \infty$, we have $x \le \alpha_n$ for $n \ge n(x)$, whence

$$\sigma_n(x) \le \tau_n, \quad n \ge n(x)$$

- Thus, by monotonicity,

$$P(n; \sigma_n(x)) \le P(n; \tau_n), \quad n \ge n(x)$$

- Letting n go to infinity, we have

$$g(x) = \lim_{n\to\infty} P(n; \sigma_n(x)) \le \liminf_{n\to\infty} P(n; \tau_n)$$

  and the one-law follows since

$$1 = \lim_{x\to\infty} g(x) \le \liminf_{n\to\infty} P(n; \tau_n)$$

- Pick x in $\mathbb{R}$. With $\lim_{n\to\infty} \alpha_n = -\infty$, we have $\alpha_n \leq x$ for $n \geq n(x)$, whence

$$\tau_n \leq \sigma_n(x), \quad n \geq n(x)$$

- Thus, by monotonicity,

$$P(n;\tau_n) \leq P(n;\sigma_n(x)), \quad n \geq n(x)$$

- Letting n go to infinity, we have

$$\limsup_{n\to\infty} P(n;\tau_n) \leq g(x) = \lim_{n\to\infty} P(n;\sigma_n(x))$$

  and the zero-law follows since

$$\limsup_{n\to\infty} P(n;\tau_n) \leq \lim_{x\to\infty} g(x) = 0$$

Strengthening Theorem 1

Theorem 4 *For any range function $\tau : \mathbb{N}_0 \to \mathbb{R}_+$ written in the form (1), we have*

$$\lim_{n\to\infty} P(n;\tau_n) = \begin{cases} 0 & \textit{iff } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \textit{iff } \lim_{n\to\infty} \alpha_n = +\infty. \end{cases}$$

Preparing the proof of Theorem 2

For each $n = 2, 3, \ldots$, write

$$\Lambda_n = nM_n - \log n$$

so that

$$\begin{aligned}\frac{M_n}{\tau_n^\star} &= \frac{1}{\tau_n^\star} \cdot \frac{1}{n} (\Lambda_n + \log n) \\ &= 1 + \frac{\Lambda_n}{\log n}\end{aligned}$$

Thus, $\Lambda_n \Longrightarrow_n \Lambda$ implies

$$\frac{\Lambda_n}{\log n} \Longrightarrow_n 0 \quad \text{whence} \quad \frac{M_n}{\tau_n^\star} \xrightarrow{P}{}_n 1.$$

$\frac{M_n}{\tau_n^\star} \xrightarrow{P}{}_n 1$ implies

Lemma 1 *The threshold function* $\tau^\star$ *is a* **weak** *threshold in the sense that*

$$\lim_{n\to\infty} P(n;\tau_n) = 0 \quad \textit{if} \quad \lim_{n\to\infty} \frac{\tau_n}{\tau_n^\star} = 0$$

while

$$\lim_{n\to\infty} P(n;\tau_n) = 1 \quad \textit{if} \quad \lim_{n\to\infty} \frac{\tau_n}{\tau_n^\star} = \infty$$

for range function $\tau : \mathbb{N}_0 \to \mathbb{R}_+$.

$\frac{M_n}{\tau^\star_n} \xrightarrow{P}{}_n 1$ implies

Lemma 2 *The threshold function $\tau^\star$ is a* **strong** *threshold in the sense that*

$$\lim_{n\to\infty} P(n; c\tau^\star_n) = \begin{cases} 0 & \textit{if } \ 0 < c < 1 \\ \\ 1 & \textit{if } \ 1 < c. \end{cases}$$

Best possible result

$$\text{Zero} - \text{one Law} \Longrightarrow \text{Strong threshold} \Longrightarrow \text{Weak threshold}$$

A very strong threshold

Theorem 5 *For any range function $\tau : \mathbb{N}_0 \to \mathbb{R}_+$ written in the form (1), we have*

$$\lim_{n\to\infty} P(n;\tau_n) = \begin{cases} 0 & \textit{iff} \;\; \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \textit{iff} \;\; \lim_{n\to\infty} \alpha_n = +\infty. \end{cases}$$

$$\tau_n = \frac{1}{n}\left(\log n + \alpha_n\right), \quad n = 1, 2, \ldots .$$

Appropriate to call the threshold function $\tau^\star$ a **very strong** threshold – Early indicator that the phase transition will be sharp

A useful representation of the spacings

Consider a sequence $\{\xi, \xi_n,\ n = 1, 2, \ldots\}$ of i.i.d. $\mathbb{R}_+$-valued rvs with $\xi > 0$ a.s. and set

$$T_n = \xi_1 + \ldots + \xi_n, \quad n = 1, 2, \ldots$$

Lemma 3 *With ξ exponentially distributed with parameter 1, we have*

$$(L_{n,1}, \ldots, L_{n,n+1}) =_{st} \left(\frac{\xi_1}{T_{n+1}}, \ldots, \frac{\xi_{n+1}}{T_{n+1}} \right)$$

A proof of Theorem 2

Fix $n = 1, 2, \ldots$. We have

$$\begin{aligned} M_n &= \max_{k=2,\ldots,n} L_{n,k} \\ &=_{st} \max_{k=2,\ldots,n} \left(\frac{\xi_k}{T_{n+1}} \right) \\ &= \frac{1}{T_{n+1}} \left(\max_{k=2,\ldots,n} \xi_k \right) \end{aligned}$$

Therefore,

$$
\begin{aligned}
nM_n - \log n \quad &=_{st} \quad \frac{n}{T_{n+1}} \left(\max_{k=2,\ldots,n} \xi_k \right) - \log n \\
&= \quad \frac{n}{T_{n+1}} \left(\max_{k=2,\ldots,n} \xi_k - \log n \right) \\
&\quad\quad + \left(\frac{n}{T_{n+1}} - 1 \right) \cdot \log n
\end{aligned}
$$

with

$$
\begin{aligned}
\left(\frac{n}{T_{n+1}} - 1 \right) \cdot \log n \quad &= \quad \frac{n}{T_{n+1}} \left(1 - \frac{T_{n+1}}{n} \right) \cdot \log n \\
&= \quad \frac{n}{T_{n+1}} \cdot \sqrt{n} \left(1 - \frac{T_{n+1}}{n} \right) \cdot \frac{\log n}{\sqrt{n}}
\end{aligned}
$$

But, by SLLNs

$$\lim_{n\to\infty} \frac{T_{n+1}}{n} = 1 \quad a.s.$$

while CLT yields

$$\sqrt{n}\left(\frac{T_{n+1}}{n} - 1\right) \Longrightarrow_n \sigma^2 U$$

with $U =_{st} N(0,1)$ and $\sigma^2 = 1$.

Therefore,

$$\left(\frac{n}{T_{n+1}} - 1\right) \cdot \log n = \frac{n}{T_{n+1}} \cdot \sqrt{n}\left(1 - \frac{T_{n+1}}{n}\right) \cdot \frac{\log n}{\sqrt{n}} \Longrightarrow_n 0$$

Finally, for each x in $\mathbb{R}$,

$$\begin{aligned}
\mathbb{P}\left[\max_{k=2,\ldots,n} \xi_k - \log n \leq x\right] &= \mathbb{P}\left[\xi_k \leq x + \log n,\ k = 2, \ldots n\right] \\
&= \prod_{k=2}^{n} \mathbb{P}\left[\xi_k \leq x + \log n\right] \\
&= \left(1 - e^{-(x+\log n)}\right)^{n-1} \\
&= \left(1 - \frac{1}{n} e^{-x}\right)^{n-1} \\
&\to_n g(x)
\end{aligned}$$

In short,

$$\max_{k=2,\ldots,n} \xi_k - \log n \Longrightarrow_n \text{Gumbel } \Lambda$$

THE WIDTH OF THE PHASE TRANSITION AND POISSON CONVERGENCE

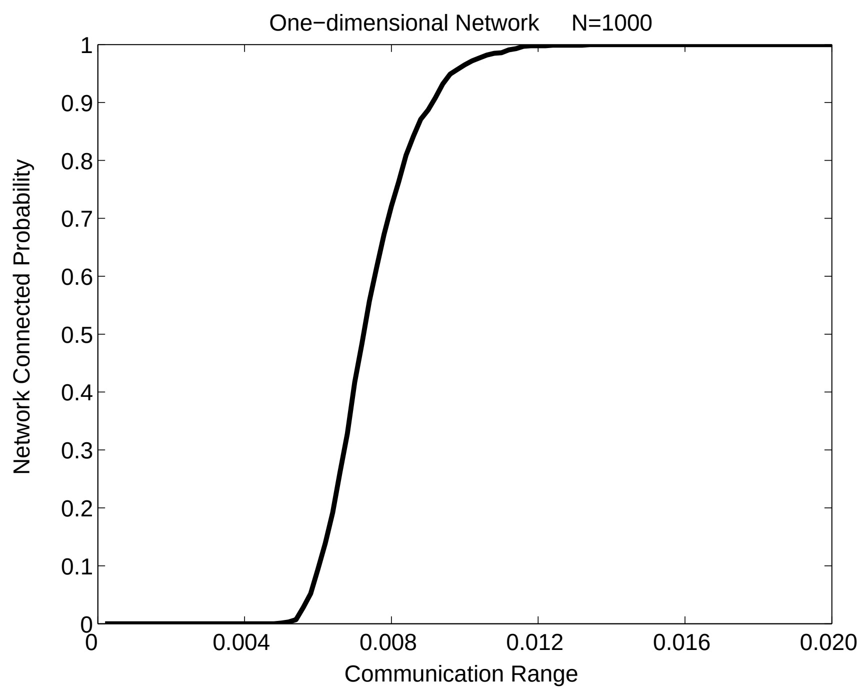
GRG $\mathbb{G}(n;\tau)$ on $[0,1]$

- A population of n nodes located at $X_1, \ldots, X_n$ in $[0,1]$
- Nodes i and j are connected if $|X_i - X_j| \le \tau$
- Assume $X_1, \ldots, X_n$ **i.i.d.** and **uniformly** distributed on $[0,1]$

For each $n = 2, 3, \ldots$, we have

$$P(n;\tau) := \mathbb{P}\left[\mathbb{G}(n;\tau) \text{ is connected}\right], \quad \tau \ge 0$$

Phase transitions



The width of the phase transition

- For $n = 2, 3, \ldots$ and $a \in (0,1)$, let $\tau_n(a)$ denote the **unique** solution to

$$P(n; \tau) = a, \quad \tau \in (0,1).$$

- Also define the transition width

$$\delta_n(a) := \tau_n(1-a) - \tau_n(a), \quad a \in (0, \frac{1}{2}).$$

Question – How does $\delta_n(a)$ vary with n large? Beyond Goel et al.

Main result – Very sharp asymptotics

Theorem 6 *For every a in the interval $(0,1)$,*

$$\tau_n(a) = \frac{\log n}{n} - \log\left(\log\left(\frac{1}{a}\right)\right)\cdot\frac{1}{n} + o\left(n^{-1}\right).$$

Corollary 1 *For every a in the interval $(0,\frac{1}{2})$, we have*

$$\delta_n(a) = \log\left(\frac{\log a}{\log(1-a)}\right)\cdot\frac{1}{n} + o\left(n^{-1}\right)$$

Goel et al. ($d = 1$)

- For **every** monotone graph property A,

$$\delta_{A,n}(a) = O\left(\sqrt{\frac{-\log a}{n}}\right).$$

- There **exists** some monotone graph property, say B, such that

$$\delta_{B,n}(a) = \Omega\left(\sqrt{\frac{-\log a}{n}}\right).$$

Theorem 6 gives **sharper** (and **exact**) asymptotics in the case of graph connectivity!

The big picture (revisited)

- Guessing Theorem 6 from Theorem 3
- **Poisson convergence** (Theorem 9)
  - **Poisson approximation** by Chen-Stein method
  - Theorem 9 implies Theorem 3 which implies Theorem 6
  - Information on **rate** of convergence, hence a handle on finite node graphs!

Range functions

No loss of generality in writing a range function

$$\tau : \mathbb{N}_0 \to \mathbb{R}_+ : n \to \tau_n$$

in the form

$$\tau_n = \frac{1}{n} \left(\log n + \alpha_n \right), \quad n = 1, 2, \ldots . \tag{2}$$

for some

$$\alpha : \mathbb{N}_0 \to \mathbb{R} : n \to \alpha_n$$

$$\alpha_n = n\tau_n - \log n, \quad n = 1, 2, \ldots$$

Zero-one Law for graph connectivity

Theorem 1 *For any range function $\tau : \mathbb{N}_0 \to \mathbb{R}_+$ written in the form (2), we have*

$$\lim_{n\to\infty} P(n;\tau_n) = \begin{cases} 0 & \textit{iff } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \textit{iff } \lim_{n\to\infty} \alpha_n = +\infty. \end{cases}$$

Critical scaling

$$\tau_n^\star = \frac{\log n}{n}, \quad n = 1, 2, \ldots$$

acts as **boundary** in the space of scalings.

Solving $P(n;\tau) = a$?

- **Interpolate** between 0 and 1 through **mild fluctuations** about $\tau^\star : \mathbb{N}_0 \to \mathbb{R}_+$
- For each x in $\mathbb{R}$, consider the range function $\sigma(x) : \mathbb{N}_0 \to \mathbb{R}_+$ given by

$$\sigma_n(x) = \left(\frac{\log n + x}{n}\right)_+, \quad n = 1, 2, \ldots$$

and

$$\sigma_n(x) = \frac{\log n + x}{n} = \tau_n^\star + \frac{x}{n}$$

for n large enough.

Interpolating the zero-one law

Theorem 3 *For each* x *in* $\mathbb{R}$*, we have*

$$\lim_{n\to\infty} P(n;\sigma_n(x)) = e^{-e^{-x}} =: g(x)$$

Guessing Theorem 6 from Theorem 3

- For each x in $\mathbb{R}$, Theorem 3 yields the **approximation**

$$P(n;\sigma_n(x)) \simeq g(x)$$

  for large enough n.

- The mapping $g : \mathbb{R} \to \mathbb{R}_+ : x \to g(x)$ is **strictly monotone** and **continuous** with $\lim_{x\to-\infty} g(x) = 0$ and $\lim_{x\to\infty} g(x) = 1$.

- Thus, for each $a \in (0,1)$, there exists a **unique** scalar x_a such that $g(x_a) = a$, namely

$$x_a = -\log\left(-\log a\right).$$

- Given $a \in (0,1)$, we find

$$P(n;\sigma_n(x_a)) \simeq a$$

for large n.

- By definition,

$$P(n; \tau_n(a)) = a$$

so that

$$P(n; \sigma_n(x_a)) \simeq P(n; \tau_n(a))$$

for large n.

- This strongly suggests that **asymptotically** $\sigma_n(x_a)$ and $\tau_n(a)$ behave **in tandem**, laying the grounds for the validity of

$$\tau_n(a) = \sigma_n(x_a) + o(n^{-1})$$

or equivalently,

$$\tau_n(a) = \frac{\log n}{n} - \log\left(\log\left(\frac{1}{a}\right)\right) \cdot \frac{1}{n} + o\left(n^{-1}\right).$$

Origins of Theorem 3?

- Property of maximal spacings (Lévy 1939)
  - Makes sense only for $d = 1$

$$\lim_{n \to \infty} P(n; \sigma_n(x)) = \lim_{n \to \infty} \mathbb{P}\left[M_n \leq \sigma_n(x)\right]$$

- Poisson convergence
  - Works (in principle) for all dimensions

$$\lim_{n \to \infty} P(n; \sigma_n(x)) = \lim_{n \to \infty} \mathbb{P}\left[C_n(\sigma_n(x)) = 0\right]$$

(Classical) Poisson convergence

For each $p \in [0,1]$, let $\{B_n(p),\ n = 1,2,\ldots\}$ denote a collection of **i.i.d.** $\{0,1\}$-valued (Bernoulli) rvs with

$$\mathbb{P}\left[B_n(p) = 1\right] = 1 - \mathbb{P}\left[B_n(p) = 0\right] = p, \quad n = 1,2,\ldots$$

and define

$$S_n(p) := B_1(p) + \ldots + B_n(p), \quad n = 1,2,\ldots$$

$$S_n(p) =_{st} \mathrm{Bin}(n;p)$$

Theorem 7 *Consider a* $[0,1]$*-valued sequence* $\{p_n,\ n = 1, 2, \ldots\}$ *with*

$$\lim_{n\to\infty} np_n = \lambda$$

for some $\lambda > 0$. *Then, it holds that*

$$S_n(p_n) \Longrightarrow_n \Pi(\lambda)$$

where $\Pi(\lambda)$ *denotes a Poisson rv with parameter* λ.

For n large,

$$p_n \sim \frac{\lambda}{n} \quad \text{and} \quad S_n(p_n) \simeq_{st} \Pi(np_n)$$

The Poisson paradigm

- For each $r = 1, 2, \ldots$, let

$$\{B_{r,k}(p_{r,k}),\ k = 1, \ldots, k_r\}$$

denote a collection of $\{0, 1\}$-valued rvs, which are not necessarily independent, and write

$$S_r(p_{r,1}, \ldots, p_{r,k_r}) = B_{r,1}(p_{r,1}) + \ldots + B_{r,k_r}(p_{r,k_r})$$

- A typical result takes the following form: With $\lim_{r\to\infty} k_r = \infty$, if

$$\lim_{r\to\infty} \left(\max_{k=1,\ldots,k_r} p_{r,k} \right) = 0$$

and

$$\lim_{r\to\infty} (p_{r,1} + \ldots + p_{r,k_r}) = \lambda$$

for some $\lambda > 0$, then under additional conditions of **vanishingly weak** correlations,

$$S_r(p_{r,1}, \ldots, p_{r,k_r}) \Longrightarrow_r \Pi(\lambda)$$

Thus,

$$\mathbb{E}\left[S_r(p_{r,1}, \ldots, p_{r,k_r})\right] = p_{r,1} + \ldots + p_{r,k_r} \simeq \lambda$$

and

$$S_r(p_{r,1}, \ldots, p_{r,k_r}) \simeq_{st} \Pi(\lambda)$$

Obvious ideas

Via pmfs:

$$\lim_{r\to\infty} \mathbb{P}\left[S_r(p_{r,1},\ldots,p_{r,k_r}) = x\right] = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x \in \mathbb{N}$$

Via pgfs:

$$\lim_{r\to\infty} \mathbb{E}\left[z^{S_r(p_{r,1},\ldots,p_{r,k_r})}\right] = e^{-\lambda(1-z)}, \quad z \in \mathbb{R}$$

Via the method of moments: For each $p = 0, 1, \ldots$,

$$\lim_{r\to\infty} \mathbb{E}\left[S_r(p_{r,1}, \ldots, p_{r,k_r})^p\right] = \mathbb{E}\left[\Pi(\lambda)^p\right]$$

Via the method of factorial moments – Brun's Sieve: For each $p = 0, 1, \ldots$,

$$\lim_{r\to\infty} \mathbb{E}\left[\prod_{\ell=0}^{p} \left(S_r(p_{r,1}, \ldots, p_{r,k_r}) - \ell\right)\right] = \lambda^{p+1}$$

Total variation

For pmfs $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ on $\mathbb{N}$, with $X \sim \boldsymbol{\mu}$ and with $Y \sim \boldsymbol{\nu}$,

$$d_{TV}(\boldsymbol{\mu}; \boldsymbol{\nu}) := \frac{1}{2} \sum_{x=0}^{\infty} |\mu(x) - \nu(x)| = d_{TV}(X; Y)$$

This defines a distance on the space of all pmfs on $\mathbb{N}$!

For $\mathbb{N}$-valued rvs $\{X,\ X_n,\ n = 1, 2, \ldots\}$, $X_n \Longrightarrow_n X$ if and only if

$$\lim_{n \to \infty} d_{TV}(X_n; X) = 0$$

The coupling inequality

Lemma 4 *For pmfs* $\boldsymbol{\mu}$ *and* $\boldsymbol{\nu}$ *on* $\mathbb{N}$*, we have*

$$d_{TV}(\boldsymbol{\mu};\boldsymbol{\nu}) \leq \mathbb{P}\left[X \neq Y\right]$$

for any pair of $\mathbb{N}$*-valued rvs* X *and* Y*, with* $X \sim \boldsymbol{\mu}$ *and with* $Y \sim \boldsymbol{\nu}$*, which are defined on a* **common** *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*.*

A pair of $\mathbb{N}$-valued rvs X and Y, with $X \sim \boldsymbol{\mu}$ and with $Y \sim \boldsymbol{\nu}$, which are defined on the **common** probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **coupling** for the pair of pmfs $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$.

$$
\begin{aligned}
& d_{TV}(\boldsymbol{\mu};\boldsymbol{\nu}) \\
= \quad & \frac{1}{2}\sum_{x=0}^{\infty} \left|\mathbb{P}\left[X = x\right] - \mathbb{P}\left[Y = x\right]\right| \\
= \quad & \frac{1}{2}\sum_{x=0}^{\infty} \left|\mathbb{P}\left[X \neq Y, X = x\right] - \mathbb{P}\left[X \neq Y, Y = x\right]\right| \\
\leq \quad & \frac{1}{2}\sum_{x=0}^{\infty} \left(\mathbb{P}\left[X \neq Y, X = x\right] + \mathbb{P}\left[X \neq Y, Y = x\right]\right) \\
\leq \quad & \frac{1}{2}\sum_{x=0}^{\infty} \mathbb{P}\left[X \neq Y, X = x\right] + \frac{1}{2}\sum_{x=0}^{\infty} \mathbb{P}\left[X \neq Y, Y = x\right] \\
= \quad & \mathbb{P}\left[X \neq Y\right]
\end{aligned}
$$

Maximal coupling

Theorem 8 *For pmfs $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ on $\mathbb{N}$, we have*

$$d_{TV}(\boldsymbol{\mu};\boldsymbol{\nu}) = \inf\left(\mathbb{P}\left[X \neq Y\right] : \ (X, Y) \in \mathcal{C}(\boldsymbol{\mu}, \boldsymbol{\nu})\right)$$

where $\mathcal{C}(\boldsymbol{\mu}, \boldsymbol{\nu})$ denotes the collection of all couplings for the pair $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$.

Corollary 2 *For pmfs $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ on $\mathbb{N}$, there exists a coupling $(X^\star, Y^\star)$ in $\mathcal{C}(\boldsymbol{\mu}, \boldsymbol{\nu})$ such that*

$$d_{TV}(\boldsymbol{\mu};\boldsymbol{\nu}) = \mathbb{P}\left[X^\star \neq Y^\star\right]$$

Such a coupling is called a **maximal** *coupling for the pair $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$.*

An easy example

- Pick $0 < p < p' < 1$. It is easy to verify that

$$d_{TV}(B(p), B(p')) = |p - p'|$$

- The independent coupling is **not** maximal
- The maximal coupling is achieved by taking

$$B^\star(p) = \mathbf{1}\left[U \le p\right] \quad \text{and} \quad B^\star(p') = \mathbf{1}\left[U \le p'\right]$$

with U uniform on $(0, 1)$. Indeed,

$$\mathbb{P}\left[\mathbf{1}\left[U \le p\right] \neq \mathbf{1}\left[U \le p'\right]\right] = \mathbb{P}\left[p < U \le p'\right] = |p - p'|$$

A useful fact via coupling

Proposition 1 *For arbitrary pmfs* $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n$ *on* $\mathbb{N}$, *it holds*

$$d_{TV}(\boldsymbol{\mu}_1 \star \ldots \star \boldsymbol{\mu}_n; \boldsymbol{\nu}_1 \star \ldots \star \boldsymbol{\nu}_n) \leq \sum_{i=1}^{n} d_{TV}(\boldsymbol{\mu}_i; \boldsymbol{\nu}_i)$$

Proposition 2 *Consider mutually independent* $\mathbb{N}$*-valued rvs* $X_1, \ldots, X_n$ *defined on a common probability space with* $X_i \sim \boldsymbol{\mu}_i$ *for all* $i = 1, \ldots, n$*. Similarly, consider mutually independent* $\mathbb{N}$*-valued rvs* $Y_1, \ldots, Y_n$ *defined on a common (possibly different) probability space with* $Y_i \sim \boldsymbol{\nu}_i$ *for all* $i = 1, \ldots, n$*. Then, it holds*

$$d_{TV}(X_1 + \ldots + X_n; Y_1 + \ldots + Y_n) \leq \sum_{i=1}^{n} d_{TV}(X_i; Y_i)$$

A proof of Proposition 1

- For each $i = 1, \ldots, n$, consider any coupling (X_i, Y_i) in $\mathcal{C}(\boldsymbol{\mu}_i, \boldsymbol{\nu}_i)$ such that the $\mathbb{N}^2$-valued rvs $(X_1, Y_1), \ldots, (X_n, Y_n)$ are mutually independent pairs defined on a common probability space.
- By construction,

$$X_1 + \ldots + X_n \sim \boldsymbol{\mu}_1 \star \ldots \star \boldsymbol{\mu}_n$$

and

$$Y_1 + \ldots + Y_n \sim \boldsymbol{\nu}_1 \star \ldots \star \boldsymbol{\nu}_n$$

- By the coupling inequality,

$$\begin{aligned} & d_{TV}(\boldsymbol{\mu}_1 \star \ldots \star \boldsymbol{\mu}_n; \boldsymbol{\nu}_1 \star \ldots \star \boldsymbol{\nu}_n) \\ = \;& d_{TV}(X_1 + \ldots + X_n; Y_1 + \ldots + Y_n) \\ \leq \;& \mathbb{P}\left[X_1 + \ldots + X_n \neq Y_1 + \ldots + Y_n\right] \\ \leq \;& \mathbb{P}\left[\cup_{i=1}^n [X_i \neq Y_i]\right] \\ \leq \;& \sum_{i=1}^n \mathbb{P}\left[X_i \neq Y_i\right] \end{aligned}$$

- Now use the maximal coupling for each $i = 1, \ldots, n$ so that

$$d_{TV}(\boldsymbol{\mu}_i; \boldsymbol{\nu}_i) = \mathbb{P}\left[X_i^\star \neq Y_i^\star\right]$$

so that

$$d_{TV}(\boldsymbol{\mu}_1 \star \ldots \star \boldsymbol{\mu}_n; \boldsymbol{\nu}_1 \star \ldots \star \boldsymbol{\nu}_n) \leq \sum_{i=1}^n d_{TV}(\boldsymbol{\mu}_i; \boldsymbol{\nu}_i)$$

An easy Poisson approximation result

- Consider a collection $\{B_k(p_k),\ k = 1, 2, \ldots, n\}$ of **mutually independent** $\{0, 1\}$-valued (Bernoulli) rvs with

$$\mathbb{P}\left[B_k(p_k) = 1\right] = 1 - \mathbb{P}\left[B_k(p_k) = 0\right] = p_k, \quad k = 1, \ldots, n$$

  and define

$$S_n := B_1(p_1) + \ldots + B_n(p_n).$$

- Also write

$$\lambda_n = p_1 + \ldots + p_n.$$

Question – How well is S_n approximated by a Poisson rv, say with parameter λ_n? In particular, what can we say about

$$d_{TV}(S_n; \Pi(\lambda_n))?$$

Answer – With **mutually independent** Poisson rvs $\Pi(p_1), \ldots, \Pi(p_n)$, we get

$$\begin{aligned} & d_{TV}(S_n; \Pi(\lambda_n)) \\ = \quad & d_{TV}(B_1(p_1) + \ldots + B_n(p_n); \Pi(p_1) + \ldots + \Pi(p_n)) \\ \leq \quad & \sum_{i=1}^{n} d_{TV}(B_i(p_i); \Pi(p_i)). \end{aligned}$$

Computing $d_{TV}(B(p);\Pi(p))$ **(**$0 < p < 1$**)**

- The maximal coupling $(B^\star(p), \Pi^\star(p))$ is given by

$$\mathbb{P}\left[B^\star(p) = x, \Pi^\star(p) = y\right] = \begin{cases} 1 - p & \text{if } \ x = y = 0 \\ \frac{p^y}{y!} e^{-p} & \text{if } \ x = 1, y = 1, 2, \ldots \\ e^{-p} - (1 - p) & \text{if } \ x = 1, y = 0 \end{cases}$$

- It is easy to see that

$$\begin{aligned}\mathbb{P}\left[B^\star(p) \neq \Pi^\star(p)\right] &= \left(e^{-p} - (1-p)\right) + \sum_{y=2}^{\infty} \frac{p^y}{y!} e^{-p} \\ &= \left(e^{-p} - (1-p)\right) + \left(1 - e^{-p} - pe^{-p}\right) \\ &= \left(1 - e^{-p}\right) p\end{aligned}$$

Thus,

$$d_{TV}(B(p); \Pi(p)) \leq \left(1 - e^{-p}\right) p \leq p^2$$

for all $0 < p < 1$.

A Poisson approximation is born!

Thus,

$$
\begin{aligned}
d_{TV}(S_n; \Pi(\lambda_n)) &\le \sum_{i=1}^{n} d_{TV}(B_i(p_i); \Pi(p_i)) \\
&\le \sum_{i=1}^{n} p_i^2
\end{aligned}
$$

With $\boldsymbol{\mu} = \Pi(\mu)$ and $\boldsymbol{\lambda} = \Pi(\lambda)$,

$$d_{TV}(\Pi(\mu); \Pi(\lambda)) \le |\mu - \lambda|$$

Order statistics

- Let $X_{n,1}, \ldots, X_{n,n}$ denote the locations of the n nodes arranged in **increasing** order, i.e.,

$$X_{n,1} \leq \ldots \leq X_{n,n}$$

with the convention $X_{n,0} = 0$ and $X_{n,n+1} = 1$.

- Also define

$$L_{n,k} := X_{n,k} - X_{n,k-1}, \quad k = 1, \ldots, n+1.$$

- For all $\tau \in (0,1)$,

$$P(n;\tau) = \mathbb{P}\left[L_{n,k} \leq \tau,\ k = 2, \ldots, n\right]$$

A useful fact

- For any subset $I \subseteq \{1, \ldots, n\}$,

$$\mathbb{P}\left[L_{n,k} > t_k,\ k \in I\right] = \left(1 - \sum_{k \in I} t_k\right)_+^n, \quad t_k \in [0,1],\ k \in I$$

with the notation

$$x_+^n = \begin{cases} x^n & \text{if } x \geq 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Leads to closed form expression for $P(n; \tau)$

Breakpoint nodes

- For each $i = 1, \ldots, n$, node i is said to be a **breakpoint** node in $\mathbb{G}(n;\tau)$ whenever
  - it is not the leftmost node in $[0,1]$ and
  - there is no node in the random interval $[X_i - \tau, X_i]$.
- The number $C_n(\tau)$ of breakpoint nodes in $\mathbb{G}(n;\tau)$ is given by

$$C_n(\tau) = \sum_{k=2}^{n} \chi_{n,k}(\tau)$$

with indicators

$$\chi_{n,k}(\tau) := \mathbf{1}\left[L_{n,k} > \tau\right], \quad k = 1, \ldots, n+1.$$

- For all $\tau \in (0,1)$,

$$\begin{aligned} P(n;\tau) &= \mathbb{P}\left[L_{n,k} \le \tau,\ k = 2,\ldots,n\right] \\ &= \mathbb{P}\left[C_n(\tau) = 0\right]. \end{aligned}$$

For all $\tau \in (0,1)$,

$$C_n(\tau) + 1 = \text{Number of connected components in } \mathbb{G}(n;\tau)$$

For future reference

- For all $\tau \in (0,1)$ and all $n = 1, 2, \ldots$,

$$\mathbb{E}\left[C_n(\tau)\right] = (n-1)\left(1-\tau\right)^n$$

and

$$\begin{aligned} \mathbb{E}\left[C_n(\tau)^2\right] &= \mathbb{E}\left[C_n(\tau)\right] + (n-1)(n-2)\left(1-2\tau\right)_+^n \\ &= (n-1)\left(1-\tau\right)^n + (n-1)(n-2)\left(1-2\tau\right)_+^n \end{aligned}$$

Poisson convergence

Theorem 9 *For each x in $\mathbb{R}$,*

$$C_n(\sigma_n(x)) \Longrightarrow_n \Pi(e^{-x})$$

where $\Pi(\mu)$ denotes a Poisson rv with parameter μ, so that

$$\lim_{n\to\infty} P(n;\sigma_n(x)) = e^{-e^{-x}}$$

Godehardt and Jaworski (1996)

Poisson approximation (Han and Makowski 2006) – Finite node population

Poisson approximation

Theorem 10 *For each* $n = 2, 3, \ldots$ *and* τ *in the interval* $(0, 1)$, *it holds that*

$$d_{TV}(C_n(\tau); \Pi(\lambda_n(\tau))) \leq B_n(\tau)$$

with

$$\lambda_n(\tau) = \mathbb{E}\left[C_n(\tau)\right] = (n-1)\left(1-\tau\right)^n$$

and

$$B_n(\tau) = (n-1)\left(1-\tau\right)^n - (n-2)\frac{\left(1-2\tau\right)_+^n}{\left(1-\tau\right)^n}$$

Theorem 10 implies Theorem 9

The **triangular inequality** yields

$$\begin{aligned} & d_{TV}(C_n(\tau); \Pi(e^{-x})) \\ \leq \quad & d_{TV}(C_n(\tau); \Pi(\lambda_n(\tau))) + d_{TV}(\Pi(\lambda_n(\tau)); \Pi(e^{-x})) \end{aligned}$$

with

$$\lambda_n(\tau) = \mathbb{E}\left[C_n(\tau)\right] = (n-1)\,(1-\tau)^n$$

But we have

$$d_{TV}(\Pi(\lambda_n(\tau)); \Pi(e^{-x})) \leq |\lambda_n(\tau) - e^{-x}|$$

and

$$d_{TV}(C_n(\tau); \Pi(\lambda_n(\tau))) \leq B_n(\tau)$$

Substitute

$$\tau \leftarrow \sigma_n(x)$$

and check that

$$B_n(\tau) \to_n 0$$

and

$$\lambda_n(\tau) - e^{-x} \to_n 0$$

Corollary 3 *For each* $n = 2, 3, \ldots$ *and* τ *in the interval* $(0, 1)$*, it holds that*

$$d_{TV}(C_n(\tau); \Pi(e^{-x})) \leq B_n(\tau) + |\lambda_n(\tau) - e^{-x}|$$

Finite node approximations

- For each x in $\mathbb{R}$, Corollary 3 yields

$$|\mathbb{P}\left[C_n(\tau)=0\right]-e^{-e^{-x}}| \leq 2B_n(\tau)+2|\lambda_n(\tau)-e^{-x}|$$

for each $n=2,3,\ldots$ and τ in the interval $(0,1)$

- Pick a in the interval $(0,1)$ and select x_a as the unique solution to $g(x)=a$, namely

$$x_a=-\log\left(-\log a\right)$$

- Obviously,

$$e^{-x_a}=-\log a$$

- Hence,

$$|\mathbb{P}\left[C_n(\tau) = 0\right] - a| \leq 2B_n(\tau) + 2|\lambda_n(\tau) + \log a|$$

for each $n = 2, 3, \ldots$ and τ in the interval $(0, 1)$

Given $\varepsilon \in (0, 1)$ and the number n of nodes, select $\tau \in (0, 1)$ so that

$$2B_n(\tau) + 2|\lambda_n(\tau) + \log a| \leq \varepsilon$$

Given $\varepsilon \in (0, 1)$ and $\tau \in (0, 1)$, select the number n of nodes so that

$$2B_n(\tau) + 2|\lambda_n(\tau) + \log a| \leq \varepsilon$$

A proof of Theorem 10 via the Chen-Stein method

The rvs $\chi_{n,1}(\tau), \ldots, \chi_{n,n+1}(\tau)$ are **negatively related** as seen from the **coupling**

$$\begin{aligned} & [(\chi_{n,1}(\tau), \ldots, \chi_{n,n+1}(\tau))_{-i} | \chi_{n,i}(\tau) = 1] \\ =_{st} & \left(\chi_{n,1}\left(\frac{\tau}{1-\tau}\right), \ldots, \chi_{n,n+1}\left(\frac{\tau}{1-\tau}\right)\right)_{-i} \end{aligned}$$

for all $i = 1, \ldots, n+1$ with

$$\chi_{n,k}\left(\frac{\tau}{1-\tau}\right) \leq \chi_{n,k}(\tau), \quad k = 1, \ldots, n+1$$

Basic Chen-Stein inequality becomes

$$\begin{aligned} d_{TV}(C_n(\tau); \Pi(\lambda_n(\tau))) &\leq \frac{1 - e^{-\lambda_n(\tau)}}{\lambda_n(\tau)} (\lambda_n(\tau) - \mathrm{Var}[C_n(\tau)]) \\ &\leq \frac{\lambda_n(\tau) - \mathrm{Var}[C_n(\tau)]}{\lambda_n(\tau)} \end{aligned}$$

where

$$\lambda_n(\tau) = \mathbb{E}\left[C_n(\tau)\right] = (n-1)\left(1-\tau\right)^n n$$

and

$$\frac{\lambda_n(\tau) - \mathrm{Var}[C_n(\tau)]}{\lambda_n(\tau)} = B_n(\tau)$$

by direct inspection!

EXTENSIONS AND VARIATIONS

- Arbitrary intervals
- Intermittently active nodes
- Non-uniform node placement
- Higher dimensions

Arbitrary intervals

The GRG $\mathbb{G}(n;\tau,d)$

- A population of n nodes located at $X_1,\ldots,X_n$ in $[0,d]$ with $d>0$
- Nodes i and j are connected if $|X_i - X_j| \leq \tau$
- Assume $X_1,\ldots,X_n$ **i.i.d.** and **uniformly** distributed on $[0,d]$

For each $n=2,3,\ldots$, write

$$P(n;\tau,d) = \mathbb{P}\left[\mathbb{G}(n;\tau,d) \text{ connected}\right]$$

for all $\tau>0$ and $d>0$.

Obviously,

$$P(n;\tau,d) = P(n;\frac{\tau}{d})$$

since

$$(X_1,\ldots,X_n) =_{st} d(U_1,\ldots,U_n)$$

where the rvs $U_1,\ldots,U_n$ are **i.i.d.** and **uniformly** distributed on $[0,1]$

Here, no loss of generality in taking scaling functions

$$\tau:\mathbb{N}_0 \to \mathbb{R}_+ : n \to \tau_n \quad \text{and} \quad d:\mathbb{N}_0 \to \mathbb{R}_+ : n \to d_n$$

in the form

$$\frac{\tau_n}{d_n} = \frac{\log n + \alpha_n}{n}, \quad n = 1,2,\ldots \tag{3}$$

for some $\alpha:\mathbb{N}_0 \to \mathbb{R}$

Zero-one law for graph connectivity

Theorem 11 *For scaling functions* $\tau, d : \mathbb{N}_0 \to \mathbb{R}_+$ *written in the form (3), we have*

$$\lim_{n\to\infty} P(n;\tau_n,d_n) = \begin{cases} 0 & \textit{iff } \lim_{n\to\infty}\alpha_n = -\infty \\ \\ 1 & \textit{iff } \lim_{n\to\infty}\alpha_n = +\infty. \end{cases}$$

The critical scaling $\tau^\star : \mathbb{N}_0 \to \mathbb{R}_+$ is given by

$$\tau^\star_n = d_n \frac{\log n}{n}, \quad n = 1, 2, \ldots$$

Intermittently active nodes

The GRG $\mathbb{G}(n;\tau,p)$

- A population of n nodes located at $X_1,\ldots,X_n$ in $[0,1]$
- Nodes i and j are connected if $|X_i - X_j| \le \tau$
- Assume $X_1,\ldots,X_n$ **i.i.d.** and **uniformly** distributed on $[0,1]$
- For each $p \in [0,1]$, let $B_1(p),\ldots,B_n(p)$ denote a collection of **i.i.d.** $\{0,1\}$-valued with the interpretation that for each $i = 1,\ldots,n$,

  Node i active (resp. inactive) if $B_i(p) = 1$ (resp. $B_i(p) = 0$)

- **Mutual independence** of the rvs $\{X_1,\ldots,X_n\}$ and $\{B_1(p),\ldots,B_n(p)\}$

Non-uniform node placement

The GRG $\mathbb{G}_f(n;\tau)$

- A population of n nodes located at $X_1, \ldots, X_n$ in $[0,1]$
- Nodes i and j are connected if $|X_i - X_j| \leq \tau$
- Assume $X_1, \ldots, X_n$ **i.i.d.** and distributed on $[0,1]$ according to some probability distribution function F on $[0,1]$ with probability density function (pdf) f

For each $n = 2, 3, \ldots$, write

$$P_f(n;\tau) = \mathbb{P}\left[\mathbb{G}_f(n;\tau) \text{ connected}\right]$$

for all $\tau > 0$.

Assumptions

- The pdf $f : [0,1] \to \mathbb{R}_+$ is **continuous**
- The pdf $f : [0,1] \to \mathbb{R}_+$ has an **isolated minimum** at $x = \xi$ in $(0,1)$ with

$$c = \min_{x \in [0,1]} f(x) = f(\xi) > 0$$

- There exists an integer $k = 1, 2, \ldots$ such that the pdf $f : [0,1] \to \mathbb{R}_+$ admits $2k+1$ derivatives on $(0,1)$ with

$$f^{(\ell)}(\xi) = 0, \ \ell = 1, \ldots, 2k \quad \text{and} \quad f^{(2k+1)}(\xi) > 0$$

Range functions

No loss of generality in writing a range function

$$\tau : \mathbb{N}_0 \to \mathbb{R}_+ : n \to \tau_n$$

in the form

$$\tau_n = \frac{\log n - \frac{1}{2k} \log\log n + \alpha_n}{cn}, \quad n = 1, 2, \ldots \tag{4}$$

for some $\alpha : \mathbb{N}_0 \to \mathbb{R}$

Zero-one law for graph connectivity

Theorem 12 *For any range function* $\tau : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ *written in the form (4), we have*

$$\lim_{n\to\infty} P_f(n;\tau_n) = \begin{cases} 0 & \textit{if } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \textit{if } \lim_{n\to\infty} \alpha_n = +\infty. \end{cases}$$

The **critical** scaling $\tau^{\star\star} : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ is given by

$$\tau_n^{\star\star} = \frac{\log n - \frac{1}{2k} \log\log n}{cn}, \quad n = 1, 2, \ldots$$

Open questions

For each x in $\mathbb{R}$, consider the range function $\sigma(x) : \mathbb{N}_0 \to \mathbb{R}_+$ given by

$$\sigma_n(x) = \frac{\log n - \frac{1}{2k}\log\log n + x}{cn} = \tau_n^{\star\star} + \frac{x}{cn}$$

for n large enough. What is the limit

$$C_n(\sigma_n(x)) \Longrightarrow_n \ ?$$

What are the exact asymptotics of the transition width

$$\delta_n(a), \quad a \in (0, \frac{1}{2})$$

Two-dimensional case ($d = 2$)

The GRG $\mathbb{G}_2(n;\tau)$

- A population of n nodes located at $X_1, \ldots, X_n$ in a compact convex subset $\Omega \subset \mathbb{R}^2$
- Nodes i and j are connected if $\|X_i - X_j\| \leq \tau$
- Assume $X_1, \ldots, X_n$ **i.i.d.** and **uniformly** distributed on Ω

For each $n = 2, 3, \ldots$, write

$$P_2(n;\tau) = \mathbb{P}\left[\mathbb{G}_2(n;\tau) \text{ connected}\right]$$

for all $\tau > 0$.

Critical scaling

Critical scaling (for the disk model) is the range function $\tau^\star : \mathbb{N}_0 \to \mathbb{R}_+$ given by

$$\pi \left(\tau_n^\star\right)^2 = \frac{\log n}{n}, \quad n = 1, 2, \ldots$$

Gupta and Kumar (1998), Kunniyur and Venkatesh (2006)

Perturbation $\sigma(x) : \mathbb{N}_0 \to \mathbb{R}_+$ given by

$$\sigma_n(x) = \sqrt{\left(\frac{\log n + x}{\pi n}\right)_+}, \quad n = 1, 2, \ldots$$

Poisson convergence

Poisson convergence for the number of isolated nodes, namely

$$I_n(\sigma_n(x)) \Longrightarrow_n \Pi(e^{-x})$$

so that

$$\lim_{n\to\infty} P_2(n;\sigma_n(x)) = e^{-e^{-x}}$$

by asymptotic equivalence of connectivity and absence of isolated nodes.

Poisson approximation not known

Transition width

Poisson convergence implies

$$\delta_n(a) = \frac{C(a)}{2}\sqrt{\frac{1}{\pi n \log n}}\,(1 + o(1))\,,$$

as compared to the result by Goel et al., namely

$$\delta_{A,n}(a) = O\left(\frac{(\log n)^{\frac{3}{4}}}{\sqrt{n}}\right)$$

Conclusions/Extensions

- Poisson convergence is ubiquitous in random graphs (e.g., Erdős-Renyi graphs)
  - Other properties (e.g., existence of isolated nodes)
  - Higher dimensions (e.g., $d = 2$ by Kunniyur and Venkatesh (2006))
- Poisson convergence $\equiv$ phase transition? – Chen-Stein method shows that

$$P(n;\tau) = \mathbb{P}\left[C_n(\tau) = 0\right] \simeq e^{-(n-1)\lambda}$$

  - **Small** change in τ yields a **moderate** change in λ, which in turn leads to a **significant** variation in the probability of graph connectivity