Erdős-Rényi *random graph*

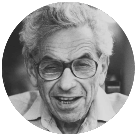
introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2024/25

graph *models*

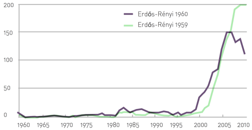
- *graph model* is *ensemble* of random graphs
- *algorithm* for random graphs of given parameters
  - *baseline* for *network structure* statistics
  - for *reasoning* about *network evolution*
  - for *generating* new *large graphs*
- *random graph* refers to *Erdős-Rényi model* [ER59]

assume *undirected* G from now on



Pál Erdős

Alfréd Rényi



Erdős-Rényi model

graph $G(n, m)$ *model*

— *$G(n, m)$ random graph* model [ER59]

— *randomly place m links* between $\binom{n}{2}$ node pairs

— *computationally convenient* but *analytically hard*

n, m given $\qquad \langle k \rangle = 2m/n$

input parameters n, m
output graph G
1: $G \leftarrow n$ isolated nodes
2: **while not** G has m links **do**
3: add link btw random node pair
4: **return** G

graph $G(n, p)$ *model*

- $G(n, p)$ *random graph* model [SR51]
- *place links* between $\binom{n}{2}$ node pairs *with probability* p
- *computationally hard* but *analytically convenient*

n, p given $\qquad$ $m, \langle k \rangle$ *unknown*

input parameters n, p
output graph G

$G \leftarrow n$ isolated nodes
for all $\binom{n}{2}$ node pairs in G **do**
    add link with probability p
return G

graph *density & degree*

— *number of links* m follows *binomial distribution* $\mathrm{B}(\binom{n}{2}, p)$

$$x \sim \mathrm{B}(n,p) \text{ then } p_x = \binom{n}{x} p^x (1-p)^{n-x} \text{ and } \langle x \rangle = np$$

$$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m\mathrm{P}(m) = \sum_{m=0}^{\binom{n}{2}} m \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} = \binom{n}{2} p$$

— then *density* $\rho = p$ and *average degree* $\langle k \rangle = (n-1)p$

graph *degree distribution*

— *degree distribution* p_k is also *binomial distribution* $\mathrm{B}(n-1,p)$

$$x \sim \mathrm{B}(n,p) \text{ then } p_x = \binom{n}{x} p^x (1-p)^{n-x} \text{ and } \langle x \rangle = np$$

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

— p_k approximately *Poisson distribution* $\mathrm{Pois}(\langle k \rangle)$ for $n \gg \langle k \rangle$

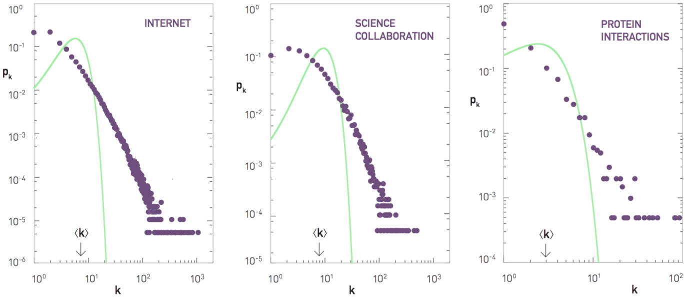
$$x \sim \mathrm{Pois}(\lambda) \text{ then } p_x = \frac{\lambda^x e^{-\lambda}}{x!} \text{ and } \langle x \rangle = \lambda$$

$$\ln\left[(1-p)^{n-1-k}\right] = (n-1-k)\ln\left(1-\frac{\langle k \rangle}{n-1}\right) \simeq -(n-1-k)\frac{\langle k \rangle}{n-1} \simeq -\langle k \rangle$$

$$p_k \simeq \frac{(n-1)^k}{k!}\left(\frac{\langle k \rangle}{n-1}\right)^k e^{-\langle k \rangle} = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$$

network *degree distribution*

- *scale-free* $p_k \sim k^{-\gamma}$ of real networks [Bar16]
- real networks are *not random graphs* [ER59]
- random graphs *lack hubs* with $k \gg \langle k \rangle$

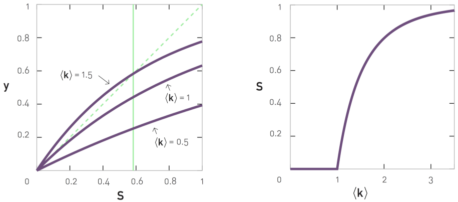


graph *connectivity*

— *fraction of nodes* in *giant component* S for $n \gg \langle k \rangle$

$$\ln(1-S) = (n-1)\ln(1-pS) \simeq -(n-1)pS = -(n-1)\frac{\langle k \rangle}{n-1}S = -\langle k \rangle S$$

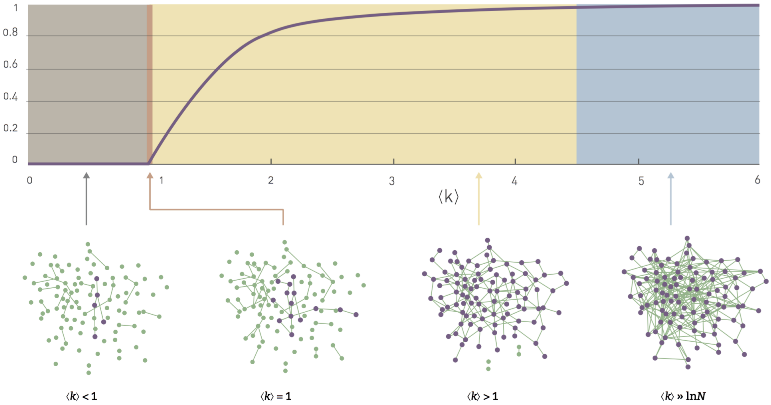
$$1-S = (1-p+p(1-S))^{n-1} \qquad S = 1 - e^{-\langle k \rangle S}$$



— *emergence* of *giant component* or *phase transition* at $\langle k \rangle = 1$

$$\frac{\mathrm{d}}{\mathrm{d}S}(1-e^{-\langle k \rangle S})\Big|_{S=0} = \langle k \rangle e^{-\langle k \rangle S}\Big|_{S=0} = \langle k \rangle > 1$$

graph *evolution*



subcritical $n_S \sim \ln n$ *critical point* *supercritical regime* *fully connected* $n_S \approx n$

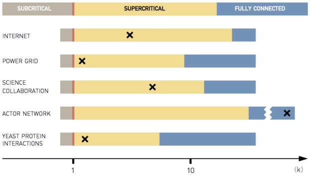
see random graph evolution NetLogo demo

network *connectivity*

— *connectivity* of real networks [Bar16]

— networks *supercritical* with $1 < \langle k \rangle < \ln n$

NETWORK	N	L	⟨k⟩	lnN
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,439	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	2.90	7.61



— *Facebook* friendships [BBR+12] *connected* $S > 0.997$

graph *diameter & distance*

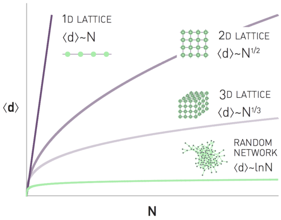
— *diameter* d_{max} and *average distance* $\langle d \rangle$ for $n \gg \langle k \rangle$

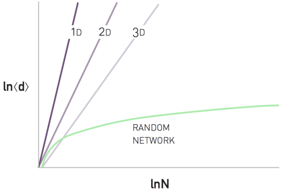
$$1 + \langle k \rangle + \langle k \rangle^2 + \cdots + \langle k \rangle^{d_{max}} = \frac{\langle k \rangle^{d_{max}+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^{d_{max}} \simeq n$$

$$d_{max} \simeq \frac{\ln n}{\ln \langle k \rangle} \qquad \langle d \rangle \approx \frac{\ln n}{\ln \langle k \rangle}$$

— $\langle d \rangle = 4.74$ for *Facebook* [BBR+12] while $\frac{\ln n}{\ln \langle k \rangle} = 3.98$

— random graphs *short distances* opposed to *lattices*





network *diameter & distance*

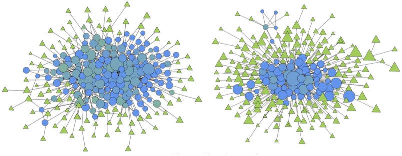
- *diameter* d_{max} and *distance* $\langle d \rangle$ of real networks [Bar16]
- $\langle d \rangle$ *well estimated* by $\frac{\ln n}{\ln \langle k \rangle}$ whereas $d_{max} \gg \frac{\ln n}{\ln \langle k \rangle}$

NETWORK	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\ln N}{\ln \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,439	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83,71	3,91	14	3,04
Citation Network	449,673	4,707,958	10.43	11,21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

graph *clustering*

— *clustering coefficients* $\langle C\rangle$ [WS98] and C [NSW01]

$$C = \langle C\rangle = \langle C_i\rangle = \frac{\langle t_i\rangle}{\binom{k_i}{2}} = \frac{p\binom{k_i}{2}}{\binom{k_i}{2}} = p$$

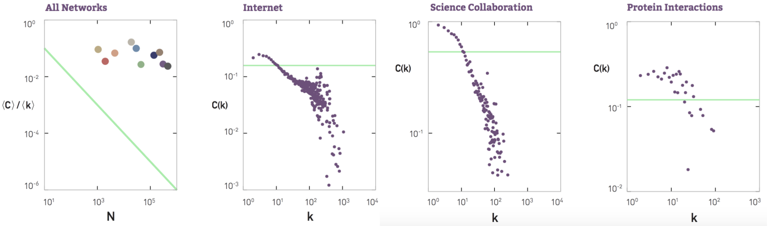


— $\langle C\rangle = 0.61$ for *Facebook* social circles [NL12] while $\rho < 10^{-6}$

— random graphs *lack clustering* for $n \gg \langle k\rangle$ opposed to *lattices*

network *clustering*

- *clustering* $\langle C \rangle$ and $C(k)$ of real networks [Bar16]
- C is *under-*/*overestimated* for *low-*/*high-k* nodes
- random graphs *substantially underestimate* $\langle C \rangle$



graph *references*

A.-L. Barabási.
Network Science.
Cambridge University Press, Cambridge, 2016.

Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna.
Four degrees of separation.
In *Proceedings of the ACM International Conference on Web Science*, pages 45–54, Evanston, IL, USA, 2012.

Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.
Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition.
Cambridge University Press, Cambridge, 2011.

David Easley and Jon Kleinberg.
Networks, Crowds, and Markets: Reasoning About a Highly Connected World.
Cambridge University Press, Cambridge, 2010.

Ernesto Estrada and Philip A. Knight.
A First Course in Network Theory.
Oxford University Press, 2015.

P. Erdős and A. Rényi.
On random graphs I.
Publ. Math. Debrecen, 6:290–297, 1959.

Mark E. J. Newman.
Networks.
Oxford University Press, Oxford, 2nd edition, 2018.

Azree Nazri and Pietro Lio.
Investigating meta-approaches for reconstructing gene networks in a mammalian cellular context.
PLoS ONE, 7(1):e28713, 2012.

graph *references*

M. E. J. Newman, S. H. Strogatz, and D. J. Watts.
Random graphs with arbitrary degree distributions and their applications.
Phys. Rev. E, 64(2):026118, 2001.

Ray Solomonoff and Anatol Rapoport.
Connectivity of random nets.
Bulletin of Mathematical Biophysics, 13(2):107–117, 1951.

D. J. Watts and S. H. Strogatz.
Collective dynamics of 'small-world' networks.
Nature, 393(6684):440–442, 1998.