advanced topics in *network science* (*ants*)

Lovro Šubelj & Jure Leskovec

University of Ljubljana
Faculty of Computer and Information Science
spring 2019/20

announcements *F2 week*

- *random selection* out *today*
- *random selection* due *next week*
- *PhD presentations* due *next week*

- think about *course project*
- keep your *reading list*!

random *selection* challenge

Erdős-Rényi *random graph*

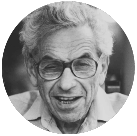
advanced topics in *network science* (*ants*)

Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

graph *models*

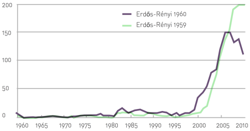
- *graph model* is *ensemble* of random graphs
- *algorithm* for random graphs of given parameters
  - *baseline* for *network structure* statistics
  - for *reasoning* about *network evolution*
  - for *generating* large *random graphs*
- *random graph* refers to *Erdős-Rényi model* [ER59]

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*

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

input parameters n, m
output graph G

$G \leftarrow n$ isolated nodes
while not G has m links **do**
    add link for random node pair
end while
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 $\quad$ $m, \langle k \rangle$ *unknown*

input parameters n, p
output graph G

```
1: G ← n isolated nodes
2: for all (n choose 2) node pairs in G do
3:     add link with probability p
4: end for
5: 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 *binomial distribution* $\mathrm{B}(n-1, p)$

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

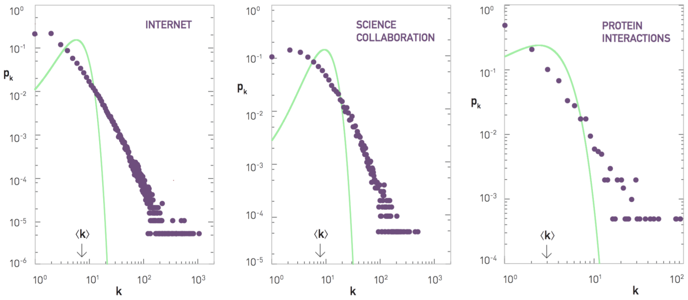
$x \sim \mathrm{Pois}(\lambda)$ then $p_x = \frac{\lambda^x e^{-\lambda}}{x!}$ and $\langle x \rangle = \lambda$

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

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

network *degree distribution*

- *scale-free* p_k of real networks [Bar16]
- real networks are *not Poisson graphs*
- random graphs *lack hubs* with $k \gg \langle k \rangle$

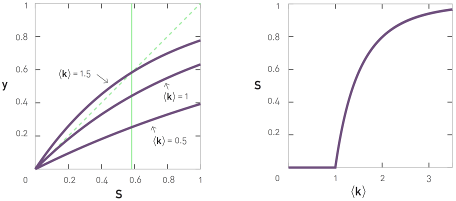


graph *connectivity*

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

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

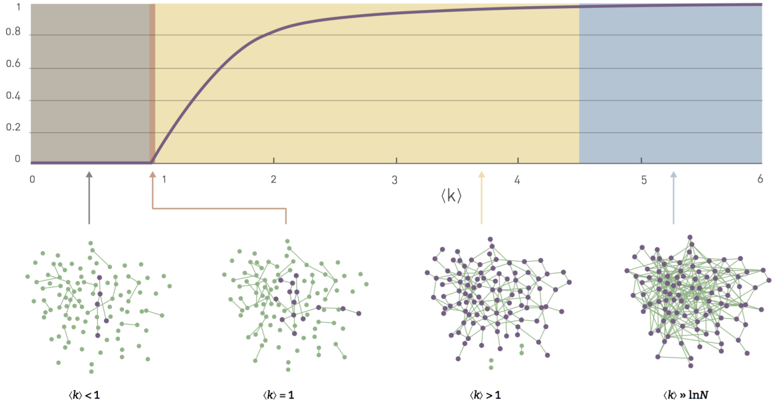
$$1 - S = (1 - p + p(1-S))^{n-1} \qquad S = 1 - e^{-\langle k \rangle S}$$



— *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* $n_S \sim n^{2/3}$ *supercritical* $n_S \sim n^{\frac{\langle k \rangle - 1}{n-1}}$ *fully connected* $n_S \approx n$

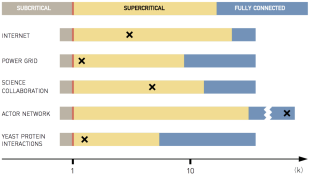
see random graph evolution NetLogo demo

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* with $S > 0.997$

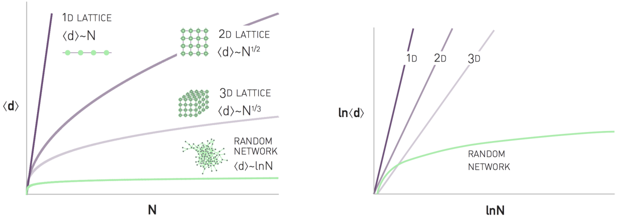
graph *diameter & distance*

- *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 *small-world* opposed to *lattices*



network *diameter & distance*

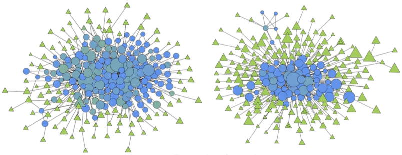
- *diameter* d_{max} and *distance* $\langle d \rangle$ of real networks [Bar16]
- $\langle d \rangle$ *well estimated* by $\frac{\ln n}{\ln \langle k \rangle}$ whereas $d_{max} \gg \frac{\ln n}{\ln \langle k \rangle}$

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]

$$C = \langle C \rangle = \langle C_i \rangle = \frac{2\langle t_i \rangle}{k_i(k_i-1)} = \frac{2p\binom{k_i}{2}}{k_i(k_i-1)} = p$$

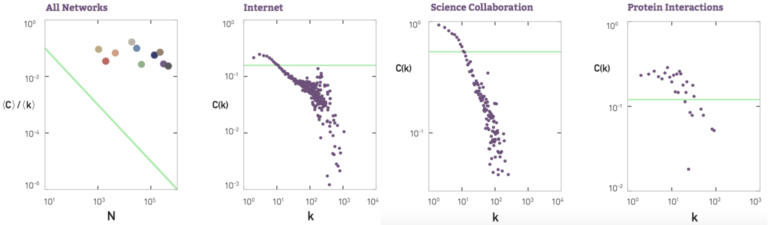


— $\langle C \rangle = 0.61$ for *Facebook* social circles [NL12] while $\rho < 10^{-6}$

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

network *clustering*

- *clustering* $\langle C \rangle$ and $C_i(k)$ of real networks [Bar16]
- C_i *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.

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: An Introduction.
Oxford University Press, Oxford, 2010.

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

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.

graph *references*

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.

configuration graph model

advanced topics in *network science* (*ants*)

Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

configuration *model*

- random graphs *Poisson distribution* $p_k \simeq \frac{\langle k\rangle^k e^{-\langle k\rangle}}{k!}$ [ER59]
- real networks *power-law degree distribution* $p_k \sim k^{-\gamma}$ [BA99]
- *configuration model* random graph for *arbitrary* $\{k\}$ [NSW01]

assume *undirected* G from now on

Mark Newman

Steven Strogatz

Duncan Watts

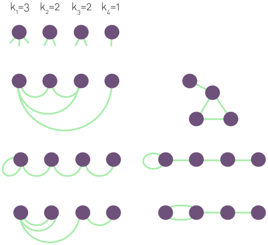
configuration $G(\{k\})$ *model*

- $G(\{k\})$ *configuration model* [NSW01]
- *randomly link* m *stub pairs* between n nodes
- *computationally convenient* and *analytically tractable*

$$\textit{graphical } k_1, k_2 \ldots k_n \qquad m = \tfrac{1}{2} \textstyle\sum_i k_i$$

input sequence $\{k\}$
output graph G

```
G ← n nodes with {k} stubs
while G has node stubs do
    link random node stub pair
end while
return G
```



configuration *probability*

- *probability of self-loop* p_i on i

$$p_i = m\frac{\binom{k_i}{2}}{\binom{2m}{2}} \approx \frac{k_i(k_i-1)}{4m}$$

- *probability of link* p_{ij} between i and j

$$p_{ij} = m\frac{k_i k_j}{\binom{2m}{2}} = k_i\frac{k_j}{2m-1} \approx \frac{k_i k_j}{2m}$$

- thus *number of multilinks* and *self-loops* is

$$\left[\frac{\langle k^2\rangle - \langle k\rangle}{\sqrt{2}\langle k\rangle}\right]^2 \qquad \sum_i p_i = \sum_i \frac{k_i(k_i-1)}{2n\langle k\rangle} = \frac{\langle k^2\rangle - \langle k\rangle}{2\langle k\rangle}$$

configuration *neighbors*

- *neighbor degree distribution* p_k is *not* p_k
  - n_k is *number of degree-k* nodes thus $n_k = np_k$

$$\{neighbor\ p_k\} = n_k \frac{k}{2m-1} \approx \frac{kp_k}{\langle k \rangle}$$

- *average neighbor degree* $\langle k \rangle$ is *not* $\langle k \rangle$

$$\tfrac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \tfrac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle} = \tfrac{\sigma_k^2}{\langle k \rangle} > 0$$

$$\langle neighbor\ k \rangle \approx \sum_k k \frac{kp_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle$$

- $\frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle^2 + \langle k \rangle}{\langle k \rangle} = \langle k \rangle + 1$ even for *Poisson graph* [ER59]

network *neighbors*

— *friendship paradox* $\langle neighbor\ k \rangle > \langle k \rangle$ [Fel91] in real networks

— $\langle neighbor\ k \rangle$ *well estimated* by $\frac{\langle k^2 \rangle}{\langle k \rangle}$ whereas $\langle k \rangle \ll \frac{\langle k^2 \rangle}{\langle k \rangle}$

network	n	$\langle k \rangle \ll$	$\langle$**neighbor** $k\rangle$	$\approx \frac{\langle k^2 \rangle}{\langle k \rangle}$
Southern women [DGG41]	32	5.56	7.57	7.02
Karate club [Zac77]	34	4.59	9.61	7.77
American football [GN02]	115	10.71	10.78	10.79
Java dependencies [ŠB11]	1368	16.20	207.52	140.53
Facebook circles [ML12]	4039	43.69	105.55	106.57
Physics collaboration [New01]	36 458	9.42	21.65	27.88
Enron e-mails [LLDM09]	36 692	20.04	472.86	280.16
Internet map [HJJ+03]	75 885	9.42	1853.73	1461.54
Actors collaboration [BA99]	382 219	78.69	282.72	417.69
Physics citation [ŠFB14]	438 943	21.56	78.38	77.72
Patent citation [HJT01]	3 774 768	8.75	17.15	21.33
Facebook snowball [Fer12]	8 217 272	3.06	308.52	157.06

configuration *clustering*

— (*neighbor*) *excess degree distribution* q_k defined as

excess degree is remaining neighbor degree or neighbor degree-1

$$q_k = \frac{(k+1)p_{k+1}}{\langle k\rangle}$$

— then *network clustering coefficient* C [NSW01] is

$$\sum_{k_i,k_j} q_{k_i} q_{k_j} \frac{k_i k_j}{2m} = \frac{1}{2m}\left[\sum_k k q_k\right]^2 = \frac{1}{2m\langle k\rangle^2}\left[\sum_k k(k+1)p_{k+1}\right]^2 = \frac{1}{n\langle k\rangle^3}\left[\sum_k (k-1)kp_k\right]^2$$

$$C = \sum_{k_i,k_j} q_{k_i} q_{k_j} p_{ij} \approx \frac{\left[\langle k^2\rangle - \langle k\rangle\right]^2}{n\langle k\rangle^3}$$

network *clustering*

— *average clustering coefficient* $\langle C \rangle$ [WS98] of real networks

— *neither* $G(n,p)$ [ER59] *nor* $G(\{k\})$ [NSW01] *explain* $\langle C \rangle \gg 0$

network	n	$\langle C \rangle$	$\gg \frac{[\langle k^2 \rangle - \langle k \rangle]^2}{n \langle k \rangle^3}$	$\gg \frac{\langle k \rangle}{n-1}$
Southern women [DGG41]	32	0.000	0.204	0.179
Karate club [Zac77]	34	0.571	0.294	0.139
American football [GN02]	115	0.403	0.078	0.094
Java dependencies [ŠB11]	1368	0.497	0.879	0.012
Facebook circles [ML12]	4039	0.606	0.063	0.011
Physics collaboration [New01]	36 458	0.657	0.002	0.000
Enron e-mails [LLDM09]	36 692	0.497	0.106	0.001
Internet map [HJJ+03]	75 885	0.160	2.985	0.000
Actors collaboration [BA99]	382 219	0.780	0.006	0.000
Physics citation [ŠFB14]	438 943	0.227	0.001	0.000
Patent citation [HJT01]	3 774 768	0.076	0.000	0.000
Facebook snowball [Fer12]	8 217 272	0.019	0.001	0.000

configuration *references*

A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286(5439):509–512, 1999.

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

A. Davis, B. B. Gardner, and M. R. Gardner.
Deep South.
Chicago University Press, Chicago, 1941.

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.

Scott. L. Feld.
Why your friends have more friends than you do.
Am. J. Sociol., 96(6):1464–1477, 1991.

Stefano Ferretti.
On the degree distribution of faulty peer-to-peer overlays.
ICST Transactions on Complex Systems, 2012.

configuration *references*

M. Girvan and M. E. J Newman.
Community structure in social and biological networks.
P. Natl. Acad. Sci. USA, 99(12):7821–7826, 2002.

M Hoerdt, M Jaeger, A James, D Magoni, J Maillard, D Malka, and P Merindol.
Internet {IP}v4 overlay map produced by network cartographer (nec), 2003.

B. H. Hall, A. B. Jaffe, and M. Tratjenberg.
The NBER patent citation data file: Lessons, insights and methodological tools.
Technical report, National Bureau of Economic Research, 2001.

Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney.
Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters.
Internet Math., 6(1):29–123, 2009.

Seth A. Myers and Jure Leskovec.
Clash of the contagions: Cooperation and competition in information diffusion.
In *Proceedings of the IEEE International Conference on Data Mining*, 2012.

M. E. J. Newman.
The structure of scientific collaboration networks.
P. Natl. Acad. Sci. USA, 98(2):404–409, 2001.

Mark E. J. Newman.
Networks: An Introduction.
Oxford University Press, Oxford, 2010.

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.

configuration *references*

Lovro Šubelj and Marko Bajec.
Community structure of complex software systems: Analysis and applications.
Physica A, 390(16):2968–2975, 2011.

Lovro Šubelj, Dalibor Fiala, and Marko Bajec.
Network-based statistical comparison of citation topology of bibliographic databases.
Sci. Rep., 4:6496, 2014.

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

Wayne W. Zachary.
An information flow model for conflict and fission in small groups.
J. Anthropol. Res., 33(4):452–473, 1977.

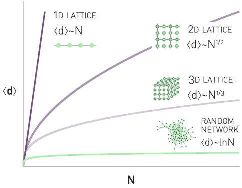
small-world networks

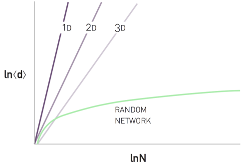
advanced topics in *network science* (*ants*)

Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

small-world *phenomenon*

- for *regular lattices*
  - *high clustering* $\langle C\rangle \gg 0$
  - long distances $\langle d\rangle \simeq n^{1/D}$
- in *random graphs* [ER59]
  - low clustering $\langle C\rangle = \frac{\langle k\rangle}{n-1}$
  - *short distances* $\langle d\rangle \simeq \frac{\ln n}{\ln\langle k\rangle}$
- real *small-world networks* [WS98]
  - *high clustering* $\langle C\rangle \gg 0$
  - *short distances* $\langle d\rangle \simeq \frac{\ln n}{\ln\langle k\rangle}$

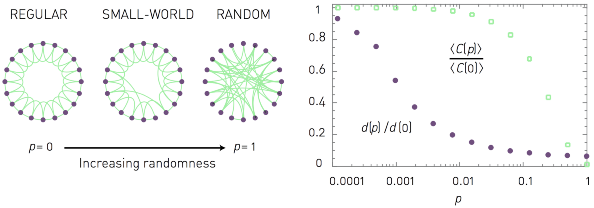




- $\langle d\rangle = 4.74$ for *Facebook* friendships [BBR+12] while $\frac{\ln n}{\ln\langle k\rangle} = 3.98$
- $\langle C\rangle = 0.61$ for *Facebook* social circles [NL12] while $\rho < 10^{-6}$

small-world *model*

- $G(n,k,p)$ *small-world* model [WS98]
- *randomly rewire* $pnk/2$ *links* of *regular lattice*
- *conceptually interesting* but *practically inapplicable*
  - for *some* p *small-world* with $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$ and $\langle C \rangle \gg 0$
  - for $p=1$ *random graph* with $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$
  - for $p=0$ *regular lattice* with $C = \frac{3(k-2)}{4(k-1)}$



see small-world model NetLogo demo

small-world *networks*

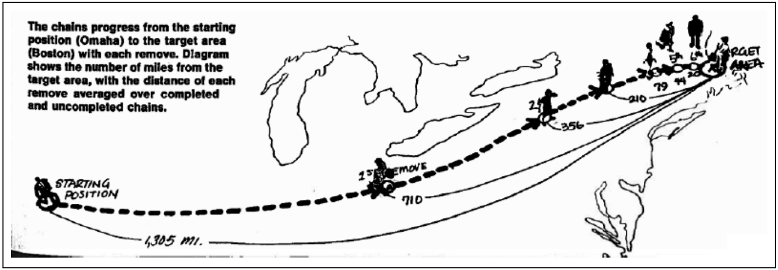
— *clustering coefficient* $\langle C \rangle$ in *real small-world* networks

— *average distance* $\langle d \rangle$ in *real small-world* networks

network	n	$\langle C \rangle$	$\gg \frac{\langle k \rangle}{n-1}$	$\langle d \rangle$	$\approx \frac{\ln n}{\ln \langle k \rangle}$
southern women [DGG41]	32	0.000	0.179	2.31	2.02
karate club [Zac77]	34	0.571	0.139	2.41	2.31
American football [GN02]	115	0.403	0.094	2.51	2.00
Java dependencies [ŠB11]	1368	0.497	0.012	2.21	2.59
Facebook circles [ML12]	4039	0.606	0.011	3.69	2.20
physics collaboration [New01]	36 458	0.657	0.000	5.50	4.68
Enron e-mails [LLDM09]	36 692	0.497	0.001	3.39	3.51
Internet map [HJJ+03]	75 885	0.160	0.000	5.83	5.01
actors collaboration [BA99]	382 219	0.780	0.000	≈ 3.6	2.94
physics citation [ŠFB14]	438 943	0.227	0.000	≈ 5.0	4.23
patent citation [HJT01]	3 774 768	0.076	0.000	≈ 8.1	6.98
Facebook snowball [Fer12]	8 217 272	0.019	0.000	≈ 6.8	14.23

small-world *experiments*

- 6 *degrees of separation* in *letter* passing as $\langle d \rangle = 6.2$ [Mil67]
- 4/7 *degrees of separation* in *e-mail* communication [DMW03]
- 4 *degrees of separation* on *Facebook* as $\langle d \rangle = 4.74$ [BBR$^+$12]



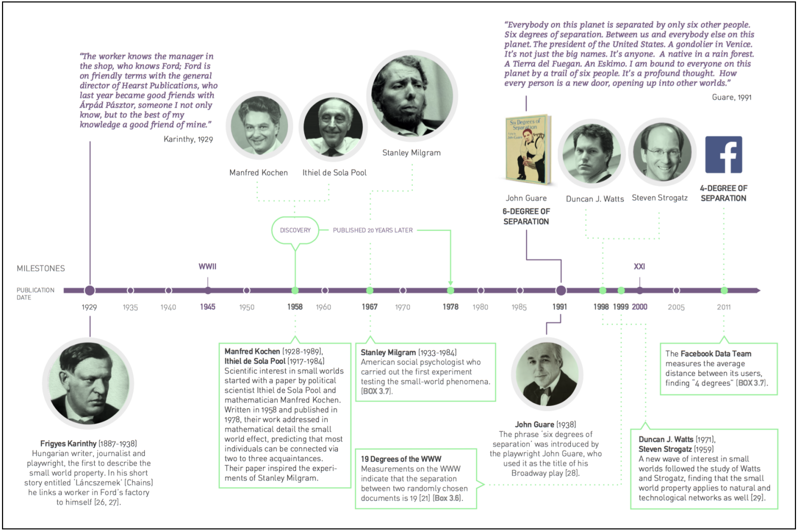
- the *strength* (*weakness*) of *weak* (*strong*) *ties* [Gra73]

small-world *navigation*

does existence of short paths imply

navigable *small-world* by *decentralized* *search*? [Kle00]

small-world *history*



small-world *references*

A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286(5439):509–512, 1999.

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.

A. Davis, B. B. Gardner, and M. R. Gardner.
Deep South.
Chicago University Press, Chicago, 1941.

Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts.
An experimental study of search in clobal social networks.
Science, 301(5634):827–829, 2003.

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

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

Stefano Ferretti.
On the degree distribution of faulty peer-to-peer overlays.
ICST Transactions on Complex Systems, 2012.

small-world *references*

M. Girvan and M. E. J Newman.
Community structure in social and biological networks.
P. Natl. Acad. Sci. USA, 99(12):7821–7826, 2002.

Mark S. Granovetter.
The strength of weak ties.
Am. J. Sociol., 78(6):1360–1380, 1973.

M Hoerdt, M Jaeger, A James, D Magoni, J Maillard, D Malka, and P Merindol.
Internet {IP}v4 overlay map produced by network cartographer (nec), 2003.

B. H. Hall, A. B. Jaffe, and M. Tratjenberg.
The NBER patent citation data file: Lessons, insights and methodological tools.
Technical report, National Bureau of Economic Research, 2001.

Jon M. Kleinberg.
Navigation in a small world.
Nature, 406(6798):845, 2000.

Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney.
Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters.
Internet Math., 6(1):29–123, 2009.

Stanley Milgram.
The small world problem.
Psychol. Today, 1(1):60–67, 1967.

Seth A. Myers and Jure Leskovec.
Clash of the contagions: Cooperation and competition in information diffusion.
In *Proceedings of the IEEE International Conference on Data Mining*, 2012.

small-world *references*

M. E. J. Newman.
The structure of scientific collaboration networks.
P. Natl. Acad. Sci. USA, 98(2):404–409, 2001.

Mark E. J. Newman.
Networks: An Introduction.
Oxford University Press, Oxford, 2010.

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

Lovro Šubelj and Marko Bajec.
Community structure of complex software systems: Analysis and applications.
Physica A, 390(16):2968–2975, 2011.

Lovro Šubelj, Dalibor Fiala, and Marko Bajec.
Network-based statistical comparison of citation topology of bibliographic databases.
Sci. Rep., 4:6496, 2014.

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

Wayne W. Zachary.
An information flow model for conflict and fission in small groups.
J. Anthropol. Res., 33(4):452–473, 1977.

scale-free networks

advanced topics in *network science* (*ants*)

Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

scale-free *property*

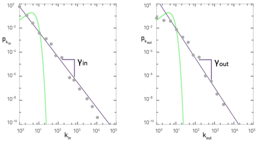
- *random graphs Poisson degree distribution* p_k [ER59]
- *real networks* contain *highly linked hubs* [Pri65, FFF99]
- *scale-free networks power-law degree distribution* p_k [BA99]

scale-free *power-law*

— *power-law degree distribution* p_k with *exponent* $\gamma > 1$

$$p_k \sim k^{-\gamma}$$

$$\log p_k \sim -\gamma \log k$$



— *theoretically correct discrete power-law* p_k for $k \geq 1$

$$\sum_{k=1}^{\infty} p_k = C \sum_{k=1}^{\infty} k^{-\gamma} = C\zeta(\gamma) = 1$$

$$p_k = Ck^{-\gamma} = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

— *analytically convenient continuos power-law* $p(k)$ for $k \geq k_{min}$

$$\int_{k_{min}}^{\infty} p(k)\,\mathrm{d}k = C \int_{k_{min}}^{\infty} k^{-\gamma}\,\mathrm{d}k = C \left.\frac{k^{-\gamma+1}}{-\gamma+1}\right|_{k_{min}}^{\infty} = C\frac{k_{min}^{-\gamma+1}}{\gamma-1} = 1$$

$$p(k) = Ck^{-\gamma} = (\gamma-1)k_{min}^{\gamma-1}k^{-\gamma}$$

scale-free *hubs*

— for *small* $k \ll \langle k \rangle$ *power-law above Poisson*

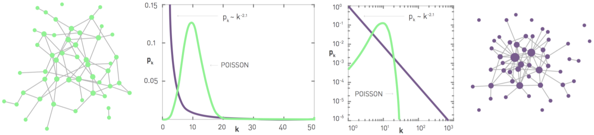
many *small degree nodes* in *scale-free network*

— for *average* $k \approx \langle k \rangle$ *power-law below Poisson*

most *nodes similar degree* in *random graph*

— for *large* $k \gg \langle k \rangle$ *power-law above Poisson*

existence of hubs in *scale-free network*



— *random graph* with $n \approx 10^{12}$ and $\langle k \rangle = 4.6$ then $n_{k \geq 100} \approx 10^{-82}$

— *scale-free network* with $n \approx 10^{12}$ and $\gamma = 2.1$ then $n_{k \geq 100} \approx 4 \cdot 10^9$

scale-free *cutoff*

- *maximum degree* k_{max} by *upper natural cutoff* of $p(k)$
- for *random graph* with *exponential* $p(k) = \lambda e^{\lambda k_{min}} e^{-\lambda k}$

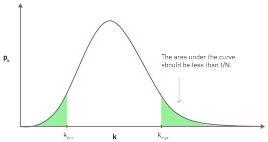
$$\int_{k_{max}}^{\infty} p(k)\,\mathrm{d}k = \lambda e^{\lambda k_{min}} \left.\frac{e^{-\lambda k}}{-\lambda}\right|_{k_{max}}^{\infty} = e^{\lambda k_{min}} e^{-\lambda k_{max}} = n^{-1}$$

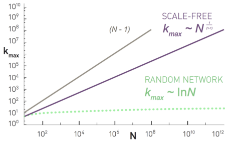
$$k_{max} = k_{min} + \frac{\ln n}{\lambda}$$

- for *scale-free network* with *power-law* $p(k) = (\gamma - 1)k_{min}^{\gamma-1} k^{-\gamma}$

$$\int_{k_{max}}^{\infty} p(k)\,\mathrm{d}k = (\gamma-1)k_{min}^{\gamma-1} \left.\frac{k^{-\gamma+1}}{-\gamma+1}\right|_{k_{max}}^{\infty} = k_{min}^{\gamma-1} k_{max}^{-\gamma+1} = n^{-1}$$

$$k_{max} = k_{min} n^{\frac{1}{\gamma-1}}$$





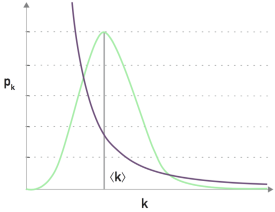
- *random graph* with $n \approx 3 \cdot 10^5$ and $\lambda = 1$ then $k_{max} \approx 14$
- *scale-free network* with $n \approx 3 \cdot 10^5$ and $\gamma = 2.1$ then $k_{max} \approx 10^5$

scale-free *moments*

- *x-th moment* $\langle k^x \rangle$ of *power-law* $p_k \sim k^{-\gamma}$
  - $\langle k^2 \rangle = \sigma_k^2 + \langle k \rangle^2$ determines *spread* and $\langle k^3 \rangle$ is *skewness*

$$\langle k^x \rangle = \sum_{k=1}^{\infty} k^x p_k \approx \int_{k_{min}}^{k_{max}} k^x p(k)\, \mathrm{d}k \sim \frac{k_{max}^{x-\gamma+1} - k_{min}^{x-\gamma+1}}{x-\gamma+1}$$

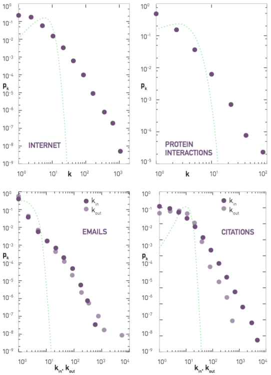
- *moments* $x \leq \gamma - 1$ *finite* whereas *moments* $x > \gamma - 1$ *diverge*

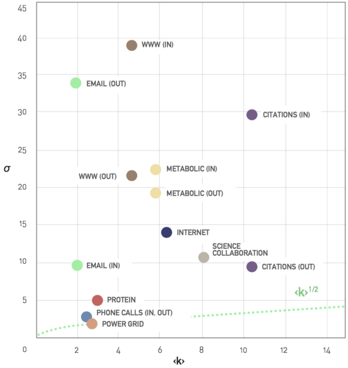


- *scale-free network* $\gamma < 3$ *lacks scale* with $k = \langle k \rangle \pm \infty$
- *random graph has scale* with $k = \langle k \rangle \pm \sqrt{\langle k \rangle}$

scale-free *networks*

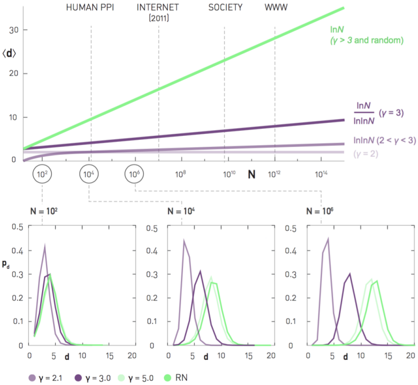
- *heavy-tail* p_k of real networks [Bar16]
- *spread* $\sigma_k = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$ in real networks



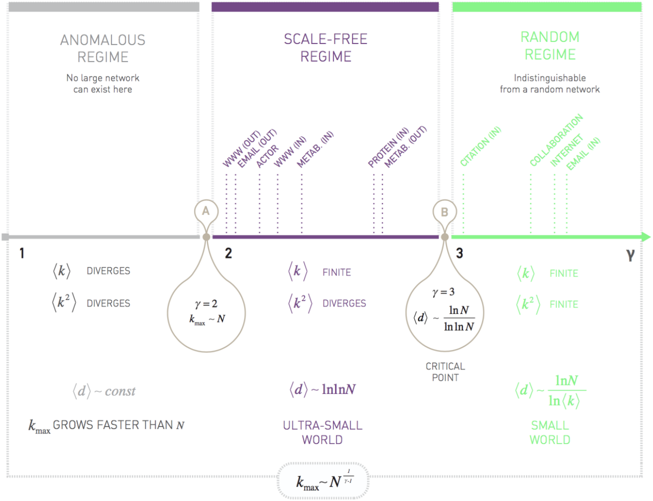


scale-free *"small-world"*

- *random graph* is *"small-world"* with $\langle d \rangle \simeq \frac{\ln n}{\ln \langle k \rangle}$
- *scale-free network* $\gamma > 3$ is *"small-world"* with $\langle d \rangle \sim \ln n$
- *scale-free network* $\gamma < 3$ *"ultrasmall-world"* with $\langle d \rangle \sim \ln \ln n$



scale-free *exponent*

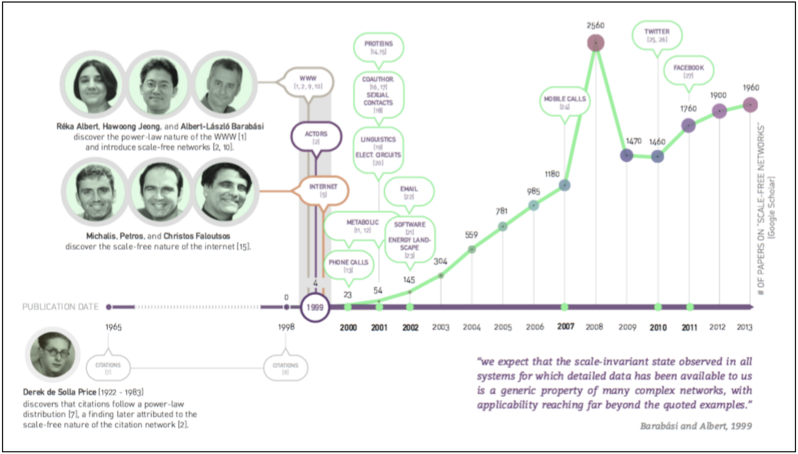


no graphical $\{k\}$ for $\gamma < 2$ $\qquad$ $n = (k_{max}/k_{min})^{\gamma-1}$ *nonexistent* for $\gamma \gg 3$

scale-free *distributions*

NAME	$p_x/p(x)$	$\langle x\rangle$	$\langle x^2\rangle$
Poisson (discrete)	$e^{-\mu}\mu^x/x!$	μ	$\mu(1+\mu)$
Exponential (discrete)	$(1-e^{-\lambda})e^{-\lambda x}$	$1/(e^\lambda-1)$	$(e^\lambda+1)/(e^\lambda-1)^2$
Exponential (continuous)	$\lambda e^{-\lambda x}$	$1/\lambda$	$2/\lambda^2$
Power law (discrete)	$x^{-\alpha}/\zeta(\alpha)$	$\begin{cases}\zeta(\alpha-2)/\zeta(\alpha), & \text{if } \alpha>2\\ \infty, & \text{if } \alpha\leq 1\end{cases}$	$\begin{cases}\zeta(\alpha-1)/\zeta(\alpha), & \text{if } \alpha>1\\ \infty, & \text{if } \alpha\leq 2\end{cases}$
Power law (continuous)	$\alpha x^{-\alpha}$	$\begin{cases}\alpha/(\alpha-1), & \text{if } \alpha>2\\ \infty, & \text{if } \alpha\leq 1\end{cases}$	$\begin{cases}\alpha/(\alpha-2), & \text{if } \alpha>1\\ \infty, & \text{if } \alpha\leq 2\end{cases}$
Power law with cutoff (continuous)	$\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)}x^{-\alpha}e^{-\lambda x}$	$\lambda^{-1}\frac{\Gamma(2-\alpha)}{\Gamma(1-\alpha)}$	$\lambda^{-2}\frac{\Gamma(3-\alpha)}{\Gamma(1-\alpha)}$
Stretched exponential (continuous)	$\beta\lambda^\beta x^{\beta-1}e^{-(\lambda x)^\beta}$	$\lambda^{-1}\Gamma(1+\beta^{-1})$	$\lambda^{-2}\Gamma(1+2\beta^{-1})$
Log-normal (continuous)	$\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-(\ln x-\mu)^2/(2\sigma^2)}$	$e^{\mu+\sigma^2/2}$	$e^{2(\mu+\sigma^2)}$
Normal (continuous)	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	$\mu^2+\sigma^2$

scale-free *history*



scale-free models

advanced topics in *network science* (*ants*)

Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

scale-free *models*

- *graph models* are *ensembles* of random graphs
- *generative models* reason about *network evolution*
- *cumulative advantage* process of *Price model* [Pri76]
- *preferential attachment* or *Barabási-Albert model* [BA99]

Pólya process Yule process Zipf's law Matthew effect
rich-get-richer proportional growth cumulative advantage

see preferential attachment model NetLogo demo

Derek de Solla Price

Albert-László Barabási

Réka Albert

scale-free $G(n, c, a)$ *model*

— $G(n, c, a)$ *cumulative advantage* model [Pri76]

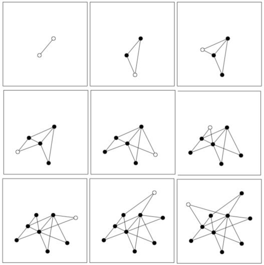
— each new node i *forms* $k_i^{out} = c > 0$ *directed links*

— node j *receives link with probability* $\sim k_j^{in} + a = q_j + a > 0$

n, c, a given $\quad\quad p_q$ *unknown*

input parameters n, c, a
output *directed* graph G

```
G ← ≥ c isolated nodes
while not G has n nodes do
    add node i to G
    for c times do
        add link (i, j) with ~ q_j + a
    end for
end while
return G
```



scale-free $G(n,c,a)$ *equation*

- *master equation* for *in-degree distribution* $p_q(n)$
  - $p_q(n)$ is in-degree distribution p_q *at time* n

$$\frac{q_i+a}{\sum_i q_i+a} = \frac{q_i+a}{n(c+a)} \qquad cnp_q(n)\frac{q+a}{n(c+a)} = \frac{c(q+a)}{c+a}p_q(n)$$

$$(n+1)p_q(n+1) = np_q(n) + \frac{c(q-1+a)}{c+a}p_{q-1}(n) - \frac{c(q+a)}{c+a}p_q(n)$$

$$(n+1)p_0(n+1) = np_0(n) + 1 - \frac{ca}{c+a}p_0(n)$$

- *power-law in-degree distribution* $p_q \sim q^{-\gamma}$ with $\gamma > 2$
  - p_q is in-degree distribution *in limit* $n \to \infty$

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t \qquad B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \simeq x^{-y}\Gamma(y)$$

$$p_q = \frac{q+a-1}{q+a+1+a/c}p_{q-1} = \cdots = \frac{B(q+a,2+a/c)}{B(a,1+a/c)} \sim q^{-2-a/c}$$

$$p_0 = \frac{1+a/c}{a+1+a/c}$$

scale-free $G(n, c)$ *model*

— $G(n, c)$ *preferential attachment* model [BA99]

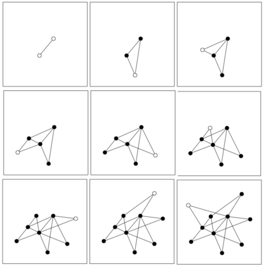
— each new node i *forms* $c > 0$ *undirected links*

— node j *receives links with probability* $\sim k_i$

n, c given $\quad$ p_k *unknown*

input parameters n, c
output *undirected* graph G

```
G ← c connected nodes
while not G has n nodes do
    add node i to G
    for c times do
        add link {i, j} with ~ k_j
    end for
end while
return G
```



scale-free $G(n, c)$ *equation*

- *undirected* $G(n, c)$ is *directed* $G(n, c, c)$ for $k_i = q_i + c$
- *same master equation* for *in-degree distribution* p_q
  - p_q is in-degree distribution *in limit* $n \to \infty$

$$p_q = \frac{B(q+c, 2+c/c)}{B(c, 1+c/c)} = \frac{B(q+c, 3)}{B(c, 2)} \sim q^{-3}$$

- *power-law degree distribution* $p_k \sim k^{-3}$
  - p_k is degree distribution *in limit* $n \to \infty$

$$p_k = \frac{B(k, 3)}{B(c, 2)} = \cdots = \frac{2c(c+1)}{k(k+1)(k+2)} \sim k^{-3}$$

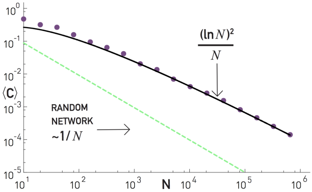
preferential $\neg$*small-world*

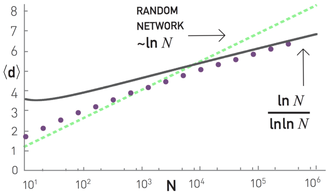
— *random graphs* are *"small-world"* as $\langle d\rangle \simeq \frac{\ln n}{\ln\langle k\rangle}$

— *random graphs* are *not small-world* as $\langle C\rangle = \frac{\langle k\rangle}{n-1}$

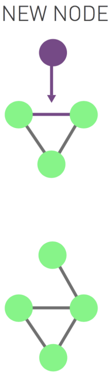
— *scale-free networks* $\gamma = 3$ are *"small-world"* as $\langle d\rangle \sim \frac{\ln n}{\ln\ln n}$

— $G(n,c)$ *scale-free model* is *not small-world* as $\langle C\rangle \simeq \frac{(\ln n)^2}{n}$

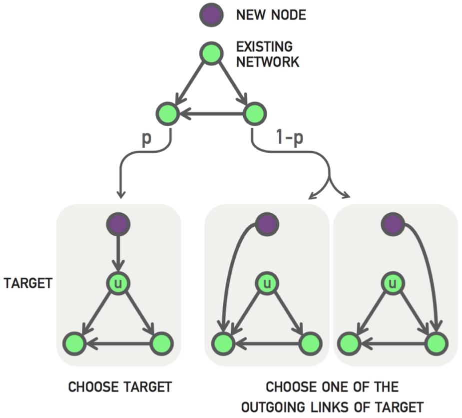




preferential *models*

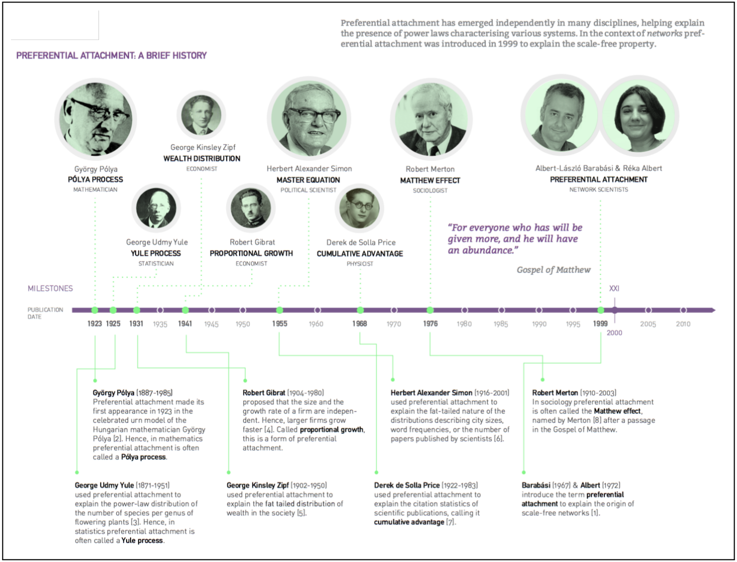


link selection [DM02]



random *link copying* model [KKR+99]

scale-free *history*



scale-free *references*

A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286(5439):509–512, 1999.

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

S. N. Dorogovtsev and J. F. F. Mendes.
Evolution of networks.
Adv. Phys., 51(4):1079–1187, 2002.

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

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

Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos.
On power-law relationships of the Internet topology.
Comput. Commun. Rev., 29(4):251–262, 1999.

Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins.
The web as a graph: Measurements, models, and methods.
In *Proceedings of the International Conference on Computing and Combinatorics*, pages 1–17, Tokyo, Japan, 1999.

scale-free *references*

Mark E. J. Newman.
Networks: An Introduction.
Oxford University Press, Oxford, 2010.

D. J. de Solla Price.
Networks of scientific papers.
Science, 149:510–515, 1965.

Derek De Solla Price.
A general theory of bibliometric and other cumulative advantage processes.
J. Am. Soc. Inf. Sci., 27(5):292–306, 1976.