

# *configuration graph* model

introduction to *network analysis* (*ina*)

Lovro Šubelj  
University of Ljubljana  
spring 2024/25

# 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

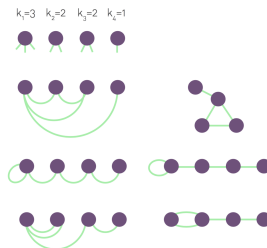
- $G(\{k\})$  configuration model [NSW01]
- randomly link  $m$  stub pairs between  $n$  nodes
- computationally convenient and analytically tractable

$$\text{graphical } k_1, k_2 \dots k_n \qquad m = \frac{1}{2} \sum_i k_i$$

input sequence  $\{k\}$

output graph  $G$

- 1:  $G \leftarrow n$  nodes with  $\{k\}$  stubs
- 2: while  $G$  has node stubs do
- 3:     link random node stub pair
- 4: 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

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

## configuration *neighbors*

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

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

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

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

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

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

# network *neighbors*

- *friendship paradox*  $\langle \text{neighbor } k \rangle > \langle k \rangle$  [Fel91] in real networks
- $\langle \text{neighbor } k \rangle$  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	$n$	$\langle k \rangle \ll$	$\langle \text{neighbor } k \rangle$	$\approx \frac{\langle k^2 \rangle}{\langle k \rangle}$
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

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

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

# 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.



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.



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

# configuration *references*



Stefano Ferretti.

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



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 Hoerdtd, 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*.  
Oxford University Press, Oxford, 2nd edition, 2018.

# configuration *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.



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.