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

Lovro Šubelj & Jure Leskovec

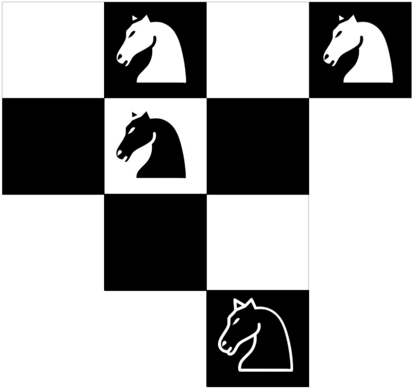
University of Ljubljana
Faculty of Computer and Information Science
spring 2019/20

announcements *F1 week*

- *four knights* out *today*
- *four knights* due *next week*
- *PhD presentations* in *two weeks*
- start thinking about *course project*

- *seminars* posted on Piazza
- keep your *reading list*!

four knights challenge



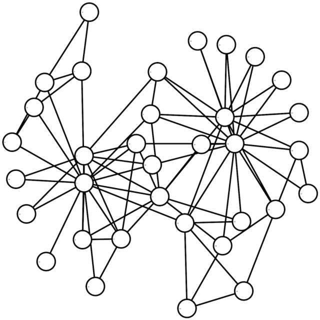
networks *motivation*

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

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

motivation *network*

- *network/graph* wiring diagram
- points called *nodes/vertices*
- connected by *links/edges*

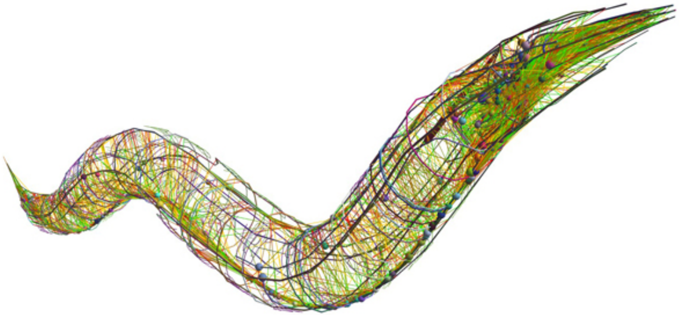


motivation *neural wiring*

— *human brain* $\approx 10^{11}$ neurons

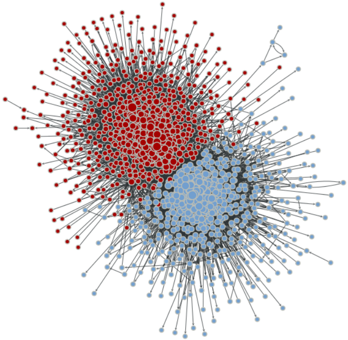
— nodes are *C. elegans neurons*

— links are *synapses*



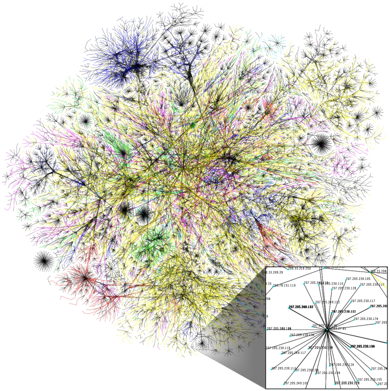
motivation *Web*

- *Web graph* $> 10^{12}$ pages
- nodes are *web pages*
- links are *hyperlinks*



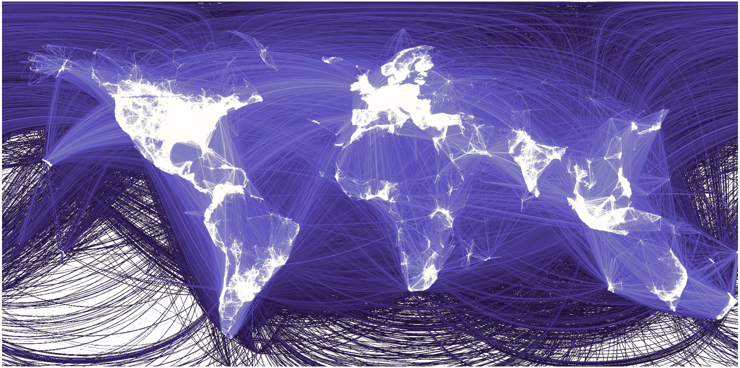
motivation *Internet*

- Internet *overlay map*
- nodes are *class C subnets*
- links are *packet routes*



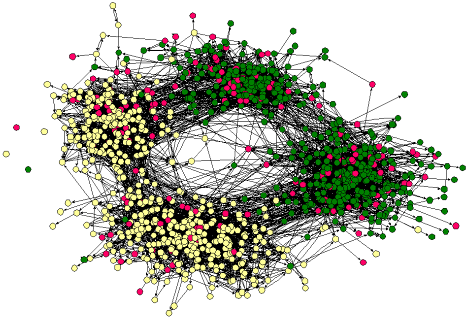
motivation *Facebook*

- *online social* network $> 10^9$ users
- nodes are *Facebook users*
- links are *social connections*



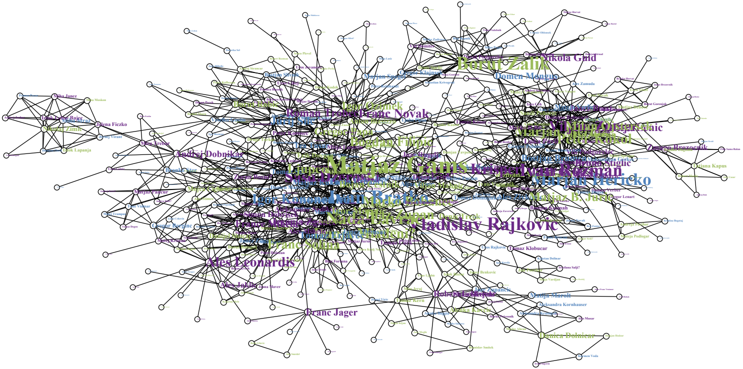
motivation *society*

- *offline social* network
- nodes are *school children*
- links are *friendship ties*



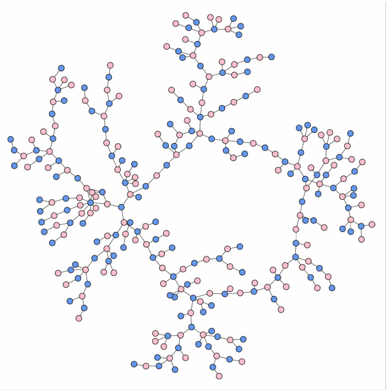
motivation *collaboration*

- *author collaboration* network
- nodes are *Slovenian computer scientists*
- links are *paper coauthorships* ≤ 2000



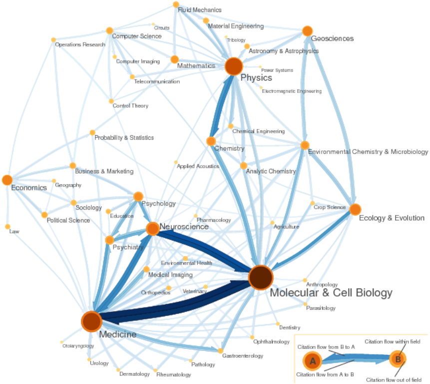
motivation *sex*

- *sexual* network
- nodes are *men/women*
- links are *sexual contacts*



motivation *science*

- *map of science* network
- nodes are *scientific fields*
- links are *journal citations*

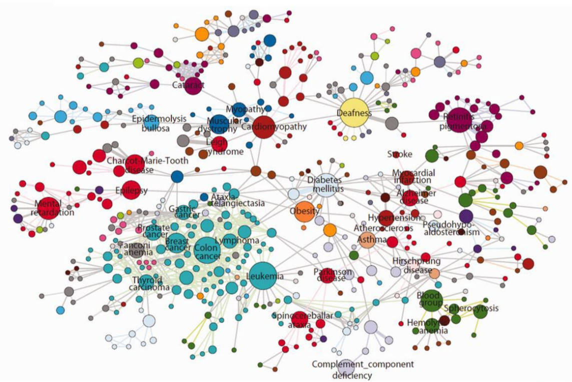


motivation *medicine*

— *human diseaseome* network

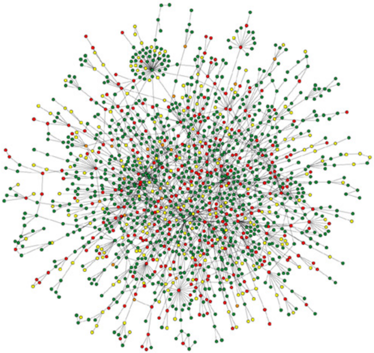
— nodes are *human diseases*

— links are *shared genes*



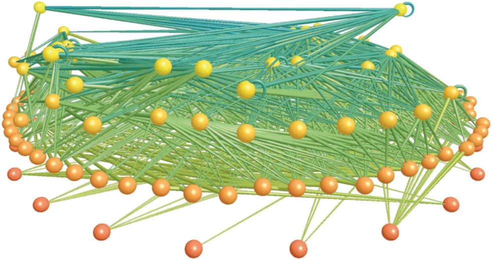
motivation *biology*

- *protein interaction* network
- nodes are *S. cerevisiae proteins*
- links are *physical interactions*



motivation *ecology*

- ecosystem *food web*
- nodes are *lake species*
- links are *predatory interactions*

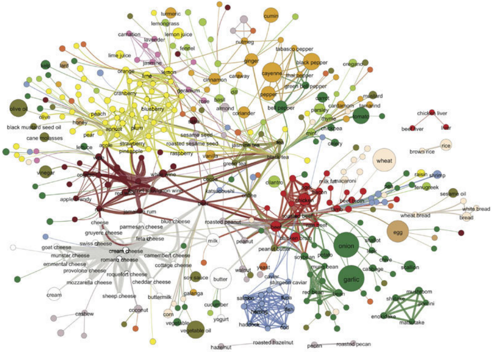


motivation *transport*

- *air transportation* network
- nodes are *world airports*
- links are *passenger flux*

motivation *gastronomy*

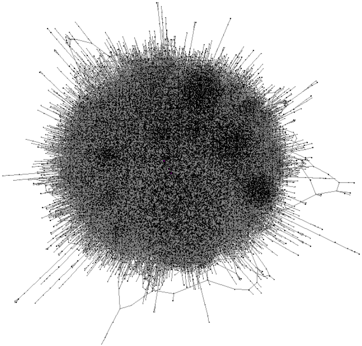
- *ingredient/flavor* network
- nodes are *recipe ingredients*
- links are *flavor compounds*



motivation *hairballs*

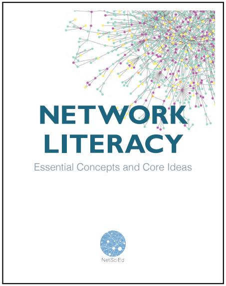
- most networks are *large/dense/complex*
- visualizations look like *ridiculograms*

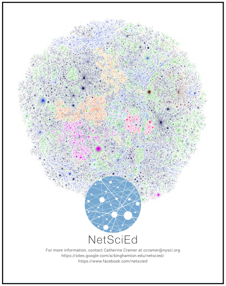
  visually stunning but scientifically worthless



motivation *networks*

- must *study networks* to *understand real systems*
- how to see what is too complex to visualize?
- through *structure*, *evolution* and *dynamics*





motivation *documentary*

connected

the power of six degrees

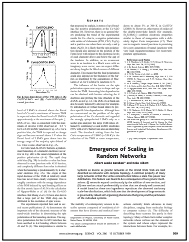
documentary on small-world and scale-free networks

letters to nature

Collective dynamics of 'small-world' networks

Duncan J. Watts & Steven H. Strogatz

[WS98]



REPORTS

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

[BA99]

letters to nature

Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong & Albert-László Barabási

[AJB00]

motivation *references*

Reka Albert, Hawoong Jeong, and Albert Laszlo Barabasi.
Error and attack tolerance of complex networks.
Nature, 406(6794):378–382, 2000.

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.

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

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

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

graph theory → *network science*

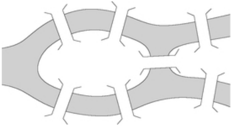
advanced topics in *network science* (*ants*)

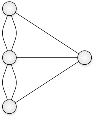
Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

history *graph theory*

1736 seven *bridges of Königsberg* [Eul36] (Leonhard Euler)

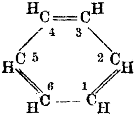
1800s *travelling salesman* problem (William Hamilton)

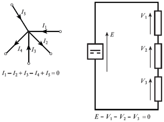




1845 *electrical circuit* laws (Gustav Kirchhoff)

1857 *chemical structure* theory (August Kekulé)





history *operations research*

1956 *shortest paths* (Edsger Dijkstra)

1956 minimum *spanning tree* (Joseph Kruskal)

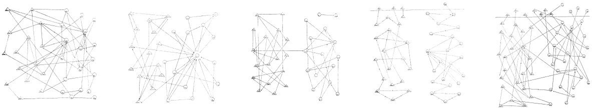
1956 maximum *flow*/minimum *cut* (Ford & Fulkerson)

1956 *signed graph* theory [CH56] (Cartwright & Harary)

1959 *random graph* theory [ER59] (Erdős & Rényi)

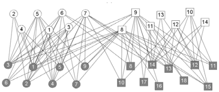
history *sociometry*

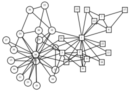
1934 children *sociograms* [Mor34] (Jacob Moreno)



1941 *Southern women* [DGG41] (Allison Davis)

1970 university *karate club* [Zac77] (Wayne Zachary)





1967 *small-world* experiment [Mil67] (Stanley Milgram)

1973 strength of *weak ties* [Gra73] (Mark Granovetter)

1977 measures of *centrality* [Fre77] (Linton Freeman)

history *bibliometrics*

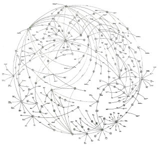
1965 scientific *paper citations* [Pri65] (Derek de Solla Price)

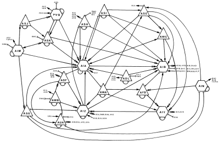
SCIENCE CITATION INDEX

1980s *political scandals* [HL03] (Mark Lombardi)

1986 *neural wirings* [WSTB86] (White et al.)

1999 transportation [Pel99] (Jon Pelletier)





networks *boom*

$<$ 2000 *small graphs* 10^2-10^3 nodes

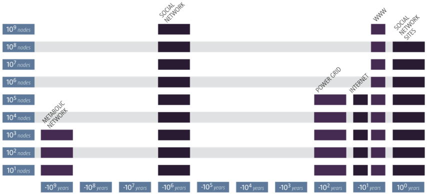
$\approx$ 2000 *communication networks* 10^5-10^8 nodes

$\approx$ 2005 *online social networks* 10^8 nodes

today *Bitcoin blockchain* 10^8-10^9 addresses

today *Facebook graph* $> 10^9$ users

today *Web graph* $> 10^{12}$ pages



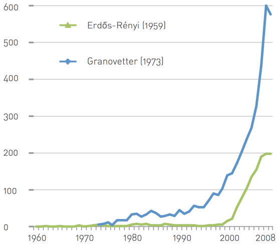
network *models*

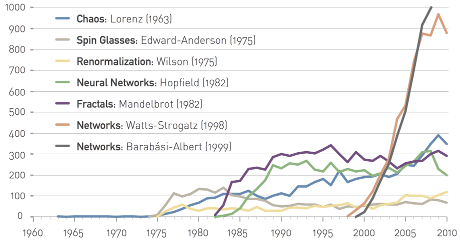
1959 *random graph* theory [ER59]

1973 *valued graphs* theory [Gra73]

1998 *small-world network* structure [WS98]

1999 *scale-free network* structure [BA99]





networks *language*

"A key discovery of network science is that the architecture of networks emerging in various domains of science, nature, and technology are similar to each other, a consequence of being governed by the same organizing principles. Consequently we can use a common set of tools to explore these systems."

Albert-László Barabási

"Networks are ideal structures to describe problems of organized complexity."

César A. Hidalgo

"I think the next century will be the century of complexity."

Stephen Hawking

networks *impact*

- *management:* internal structure of organization
- *economic:* from web search to social networking
- *epidemics:* from forecasting to halting deadly viruses
- *health:* from drug design to metabolic engineering
- *security:* fraud detection and fighting terrorism
- *neuroscience:* mapping human brain
- *many other:* your course project

network *science*

problem

understanding *real networks*

means

study of *network properties*

design of *mathematical models*

implementation of *efficient algorithms*

goals

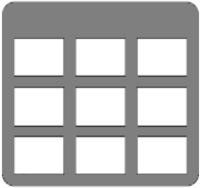
network *structure* and *evolution*

nodes, fragments, clusters, layers, network

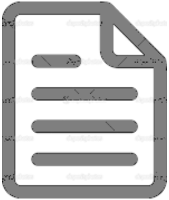
network *dynamics* and *processes*

spreading, diffusion, epidemics

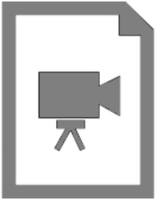
network *analysis*



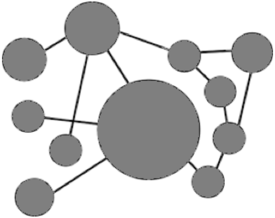
data mining



text mining



computer vision



network analysis

history *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.

Dorwin Cartwright and Frank Harary.
Structural balance: A generalization of Heider's theory.
Psychological Review, 63(5):277–293, 1956.

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.

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

Leonhard Euler.
Solutio problematis ad geometriam situs pertinentis.
Comment. Academiae Sci. I. Petropolitanae, 8:128–140, 1736.

L. Freeman.
A set of measures of centrality based on betweenness.
Sociometry, 40(1):35–41, 1977.

history *references*

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

César A. Hidalgo.
Disconnected, fragmented, or united? A trans-disciplinary review of network science.
Appl. Netw. Sci., 1:6, 2016.

Robert Hobbs and Mark Lombardi.
Mark Lombardi: Global Networks.
Independent Curators International, New York, 2003.

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

J. L. Moreno.
Who Shall Survive?
Beacon House, Beacon, 1934.

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

Jon D. Pelletier.
Self-organization and scaling relationships of evolving river networks.
Journal of Geophysical Research, 104(4):7359–7375, 1999.

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

history *references*

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

J. G. White, E. Southgate, J. N. Thomson, and S. Brenner.
The structure of the nervous system of the nematode Caenorhabditis elegans.
Phil. Trans. R. Soc. Lond. B, 314(1165):1–340, 1986.

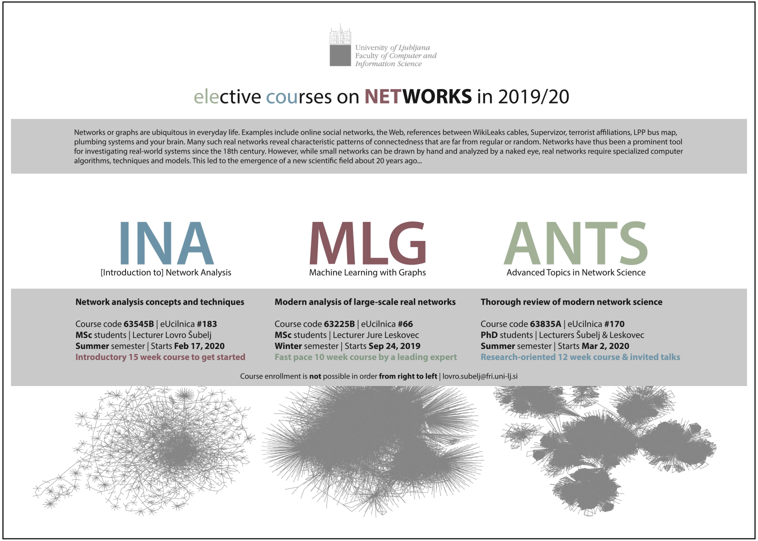
Wayne W. Zachary.
An information flow model for conflict and fission in small groups.
J. Anthropol. Res., 33(4):452–473, 1977.

course *logistics*

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

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

course *design*



course *overview*

- main *course objective*
  - *graph theory* vs *network science*
  - present rich set of *network science tools*
  - *theoretical foundations* and *practical applicability*

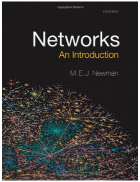
- high-level *course outline*
  1. fundamental *concepts* and *techniques*
  2. selected *advanced topics* (not in-depth)
  3. *invited talks* (physics, mathematics, sociology)

- *prerequisites* and *background*
  - clearly identified *PhD topic* and problem
  - *linear algebra*, *probability theory* and *statistics*
  - good *programming skills* (C/C++, Python, Java)

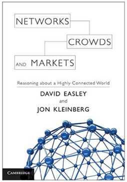
course *literature*

- *handouts/printouts* as *notes*
- *book chapters* as *background reading*
- *scientific papers* as *further readings*

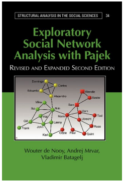
[Bar16]



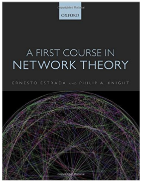
[New10]



[EK10]



[dNMB11]



[EK15]

course *schedule*

- *course syllabus* online
- weekly *course schedule*
  - *lectures* on *Thursdays at 4:00pm*
  - *consultations* on *Thursdays at 6:00pm*
  - *office hours* on *Thursdays at 3:00pm*
- course *communication*
  - classroom and consultations
  - *questions* and *comments* on Piazza
  - *announcements* on eUcilnica
  - avoid private e-mails!

course *coursework*

- *course lectures* and *book chapters*
- keep *reading list* of *scientific papers*
- *ongoing topics* within *weekly challenges*

- substantial *course project*
  - network science applied to own PhD research
  - empirical evaluation, derivation or implementation
  - *project proposal*, *milestone report* and *final paper*
  - eight-page *scientific paper submitted* (to arXiv.org)

- informal *project presentations*
  - *PhD topic*, *project proposal* and *milestone report*
  - final *project presentation* (in front of *INA students*)
  - possible *faculty talk* (within Cookie/AI seminars)

course *grading*

- course *grade breakdown*
  - 15% on *project proposal*
  - 25% on *project milestone*
  - 40% on *project paper*
  - 10% on *presentations*
  - 10% on *commitment*
  - possible *oral exam*

- course *assignments submission*
  - typed on computer in English
  - *hard-copy* in *submission box* (grading)
  - *electronic version* to eUcilnica (archive)
  - *cover sheet* with signed *honor code*
  - *late days twice* in semester

course *deadlines*

- *assignments* due on *Thursdays at 3:00pm*
- *late days* expire on *Tuesdays at 3:00pm*
- *presentations* on *Thursdays at 6:00pm*

week	lectures	presentations	assignments
1			
2			
3			
4			
5	*fundamentals of network science*	*PhD* (Mar 19th)	
6			
7		*proposal* (Apr 2nd)	
8			*proposal* (Apr 9th)
9			
10			
11	*advanced topics in network science*	*milestone* (Apr 30th)	
12			*milestone* (May 7th)
13			
14			
15	*applications of network science*	*project* (May 28th)	
16			*paper* (Jun 4th)
...			

course *references*

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

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.

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

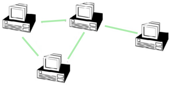
graphology & *networkology*

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

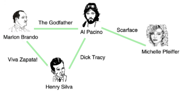
Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

terminology *graphs & networks*

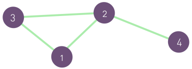
- *network science* perspective
  - *network* is some *real-world system*
  - *graph* is *representation of network*
- *graph theory* perspective
  - *graph* is formal *mathematical object*
  - *network* is *graph with real data*
- *social science* perspective
- but Web graph, Internet map



network



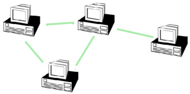
another network



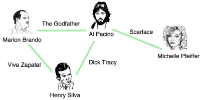
graph

terminology *nodes & links*

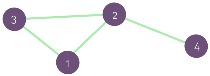
- *network science* terminology
  - *nodes* and *links*
- *graph theory* terminology
  - *vertices* and *edges/relations*
- *social science* terminology
  - *agents/brokers/units* and *ties*



nodes & links



agents & ties



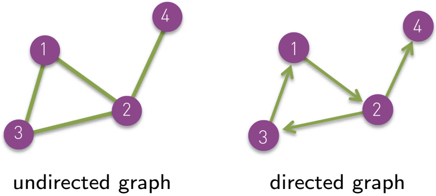
vertices & edges

terminology *classes*

- *social* networks
  - nodes are *people or animals*, links are relations or interactions
  - Facebook, offline, online, affiliation, author/actor collaboration
- *information* networks
  - nodes are information sources, links resemble *information flow*
  - Web, Twitter, citation, communication, peer-to-peer
- *technological* networks
  - human-made infrastructure with *technological constraints*
  - Internet, telephone, transportation, power grid, software
- *biological* networks
  - interaction between genes, cells, neurons in *living beings*
  - gene regulatory, metabolic, protein interaction, neural
- *ecological*, *lexical*, *financial*, *sports* etc. networks

graphology *graphs & digraphs*

- *graph* G is defined by
  - set of nodes $N = \{1, 2 \dots n\}$
  - set of links L where $m = |L|$
- if G is *undirected* then $L \subseteq \{\{i, j\} |\ i, j \in N\}$
- if G is *directed* then $L \subseteq \{(i, j) |\ i, j \in N\}$



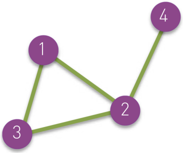
graphology *adjacency*

- *adjacency matrix* A is $n \times n$ matrix defined as
  - $A_{ij} = 1$ if there is link *from* j *to* i
  - $A_{ij} = 0$ if $i = j$ or *otherwise*
- if G is *undirected* then $A_{ij} = A_{ji}$ and $\sum_{ij} A_{ij} = 2m$
- if G is *directed* then $A_{ij} \neq A_{ji}$ and $\sum_{ij} A_{ij} = m$

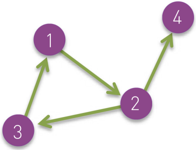
$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

undirected graph   directed graph

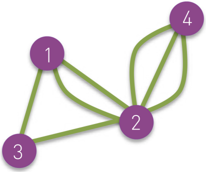
graphology *multigraphs*



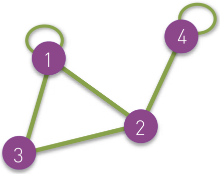
simple undirected
$A_{ij} = A_{ji} \in \{0, 1\}$



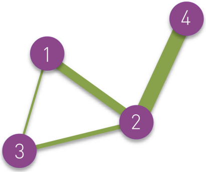
simple directed
$A_{ij} \neq A_{ji} \in \{0, 1\}$



multigraph $A_{ij} \in \mathbb{N}_0$



self-loops $A_{ii} = 2$

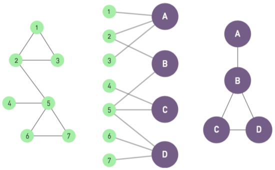


weighted $W_{ij} \in \mathbb{R}_{\geq 0}$

graphology *multipartite*

- undirected *bipartite graph* G_B is defined by
  - sets of nodes $N_1 = \{1, 2 \ldots n_1\}$ and $N_2 = \{1, 2 \ldots n_2\}$
  - set of m links $\subseteq N_1 \times N_2$
- *incidence matrix* B is $n_2 \times n_1$ matrix defined as
  - $B_{ij} = 1$ if there is link *between j and i*
  - $B_{ij} = 0$ *otherwise*
- (*one-mode*) *projections* are multigraphs with

$$A = B^T B - D_1 \qquad A = BB^T - D_2$$

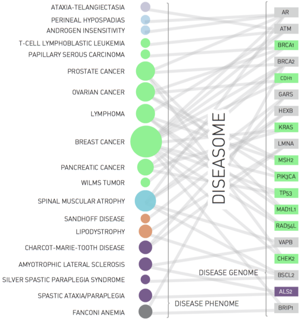


bipartite graph & projections

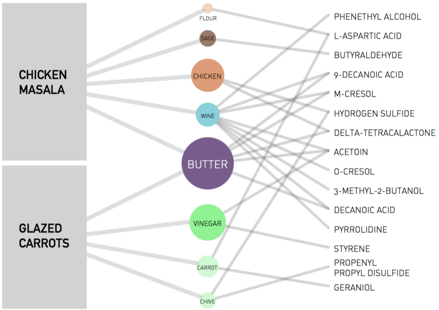
$$B = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

incidence matrix

networkology *multi-mode*



bipartite or *two-mode network*



tripartite graph or *three-mode network*

— for *undirected* G *degree* k_i of i is number of *incident links*

$$k_i = \sum_j A_{ij} = \sum_j A_{ji}$$

— for *directed* G *degree* $k_i = k_i^{in} + k_i^{out}$

- *in-degree* k_i^{in} of i is number of *incoming links*

$$k_i^{in} = \sum_j A_{ij}$$

- *out-degree* k_i^{out} of i is number of *outgoing links*

$$k_i^{out} = \sum_j A_{ji}$$

— thus (*network*) *average degrees* $\langle k \rangle$ and $\langle k^{\cdot} \rangle$ are

$$\langle k \rangle = 2m/n \qquad \langle k^{\cdot} \rangle = m/n$$

networkology *degrees*

- *average degrees* $\langle k \rangle$ of real networks [Bar16]
- mostly $\langle k \rangle \leq 10$ despite *very different* n

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

- $\langle k \rangle = 190.5$ for *Facebook* friendships [BBR+12]

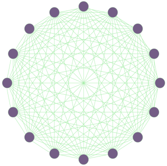
graphology *density*

— for *undirected* G *density* ρ is defined as

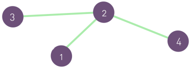
$$\rho = \frac{2m}{n(n-1)} = \frac{\langle k \rangle}{n-1}$$

— for *directed* G *density* $\rho^{\cdot}$ is defined as

$$\rho^{\cdot} = \frac{m}{n(n-1)} = \frac{\langle k^{\cdot} \rangle}{n-1}$$



complete $m = \binom{n}{2}$



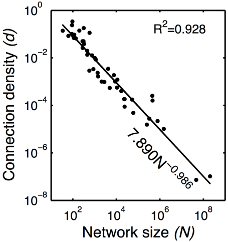
tree $m = n - 1$

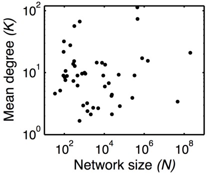
— G is *dense* if $\rho \to \mathrm{const.}$ as $n \to \infty$ thus $\langle k \rangle = \mathcal{O}(n)$

— G is *sparse* if $\rho \to 0$ as $n \to \infty$ thus $\langle k \rangle \neq \mathcal{O}(n)$

— *density* ρ and *degree* $\langle k \rangle$ of real networks [LJT+11]

— real networks are *sparse* $\rho \approx \mathcal{O}(n^{-1})$ and $\langle k \rangle \ll n$





— $\rho \approx \frac{138 \cdot 10^9}{721^2 \cdot 10^{12}} < 10^{-6}$ for *Facebook* friendships [BBR+12]

— A of real networks is *almost all zeros* $m \approx \mathcal{O}(n)$

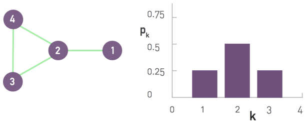
graphology *degree distribution*

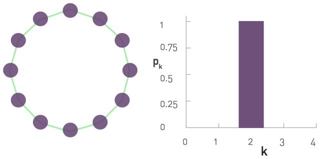
- for *undirected* G *degree distribution* p_k is defined as
  - n_k is number of *degree-k* nodes

$$p_k = n_k / n \qquad \sum_k p_k = 1 \qquad \langle k \rangle = \sum_k k p_k$$

- for *directed* G *in-/out-degree distributions* p_k^{in} and p_k^{out}
  - n_k^{in} and n_k^{out} is number of *in-/out-degree-k* nodes

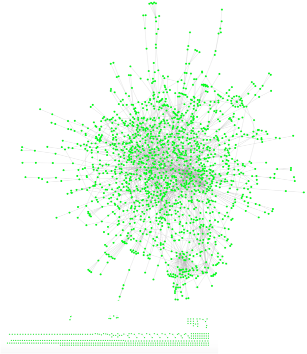
$$p_k^{in} = n_k^{in} / n \qquad p_k^{out} = n_k^{out} / n \qquad \langle k^{\cdot} \rangle = \sum_k k p_k^{\cdot}$$

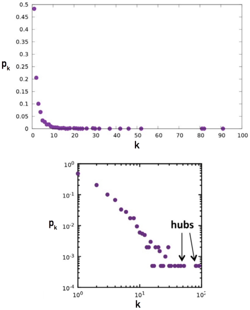




networkology *degree distribution*

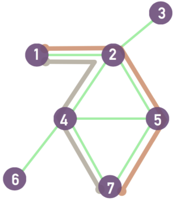
- *heavy-tail distribution* p_k of protein network [Bar16]
- nodes with *very high* $k \gg \langle k \rangle$ are called *hubs*



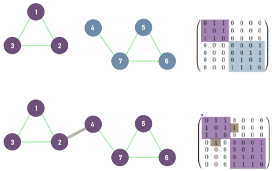


pathology *connectivity*

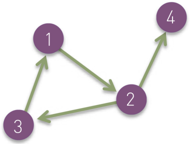
- for *undirected* G *path* P_{ij} is seq. of *links between* i *and* j
  - (*connected*) *component* is *maximal subset* thus $\forall i, j : \exists P_{ij}$
  - *giant component* contains *nontrivial fraction* of nodes
  - *connected* G has *exactly one* connected component
- for *directed* G *path* $\overrightarrow{P_{ij}}$ is seq. of *directed links from* i *to* j
  - *weak*/*strong connectivity* defined through P and $\overrightarrow{P}$



paths P_{17}



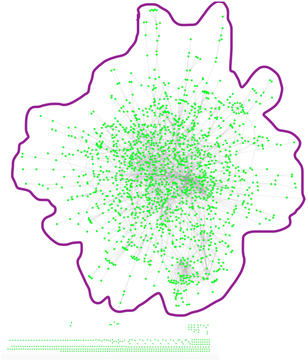
(dis)connected with *bridge*



WCC with two *SCC*

networkology *connectivity*

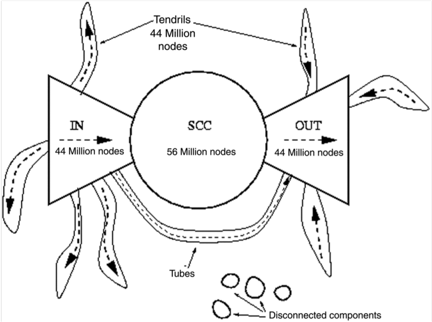
- *giant/largest component* of protein network [Bar16]



- *giant* $> 99,7\%$ for *Facebook* friendships [BBR$^+$12]
- could real network have *two giant components*?

networkology *diconnectivity*

- any *directed* G is *DAG* on its *SCC*-s
- *SCC* of i is *intersection of IN* & *OUT* of i
- *bow-tie structure* of Web graph [BKM+00]

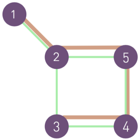


pathology *distances*

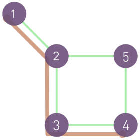
- *length* of path P or $\overrightarrow{P}$ is number of *links*/*hops*
- *geodesic path* G_{ij} or $\overrightarrow{G_{ij}}$ is any *shortest* P_{ij} or $\overrightarrow{P_{ij}}$
- *distance* d_{ij} between i and j is *length* of G_{ij} or $\overrightarrow{G_{ij}}$

- (*network*) *diameter* d_{max} or D is *maximum* d_{ij}
- (*network*) *average distance* $\langle d \rangle$ or ℓ^{-1} is defined as
  - $d_{ij} = 0$ or $d_{ij} = \infty$ for i and j in different components

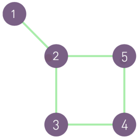
$$\langle d \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \qquad \ell^{-1} = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$$



$P_{13} \neq G_{13}$



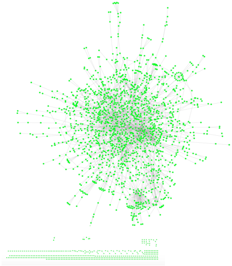
$d_{14} = 3$

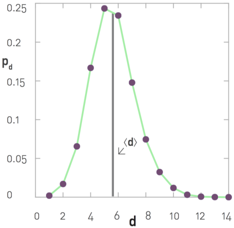


$\langle d \rangle = 1.6$

networkology *distances*

- *distance distribution* p_d of protein network [Bar16]
- most nodes are on *similar distances* $d \approx \langle d \rangle$





- $\langle d \rangle = 4.74$ for *Facebook* friendships [BBR+12]
- real networks have *surprisingly small* $\langle d \rangle \ll n$

graphology *clustering*

- for *undirected* G *node clustering coefficient* C_i of i is
  - t_i is number of *linked neighbors* or *triangles* of i

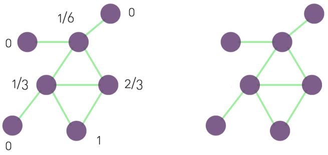
$$C_i = \frac{2t_i}{k_i(k_i-1)} \qquad C_i = 0 \text{ for } k_i \leq 1$$

- *average clustering coefficient* $\langle C\rangle$ [WS98] is defined as

$$\langle C\rangle = \frac{1}{n}\sum_i C_i$$

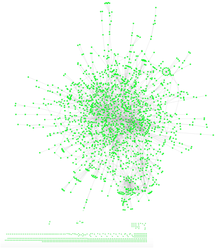
- *network clustering coefficient* C [NSW01] is defined as

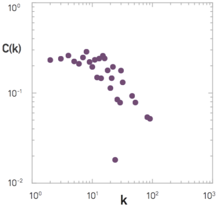
$$C = \frac{3 \text{ x number of } \textit{triangles} \text{ or } \textit{closed triads}}{\text{number of } \textit{triples} \text{ of nodes or } \textit{connected triads}}$$



$\langle C\rangle = \frac{13}{7\cdot 6} = 0.31$ $\qquad$ $C = \frac{3\cdot 2}{16} = 0.38$

- *clustering* $C_i(k)$ of protein network [Bar16]
- hubs *much lower* C_i than nodes with $k \approx \langle k \rangle$





- $\langle C \rangle = 0.61$ for *Facebook* social circles [ML12]
- real (social) networks have *significant* $\langle C \rangle \gg 0$

networkology *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.

Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener.
Graph structure in the Web.
Comput. Netw., 33(1-6):309–320, 2000.

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.

Paul J. Laurienti, Karen E. Joyce, Qawi K. Telesford, Jonathan H. Burdette, and Satoru Hayasaka.
Universal fractal scaling of self-organized networks.
Physica A, 390(20):3608–3613, 2011.

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.

networkology *references*

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.

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