

introduction to *network analysis* (*ina*)

Lovro Šubelj

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

announcements *1st week*

- *homework #0* out *today*
- *homework #0* due *next Monday*
- start thinking about *course project*
- *course labs* start *next Monday*
- networks *consultations* *any time*
- networks *seminars* posted on *Piazza*
- *mobiles* at *lectures* & *laptops* at *labs*
- posts to *feedback box* get *candy* =)

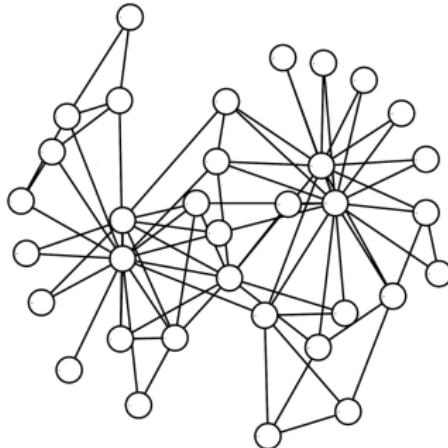
networks *motivation*

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2019/20

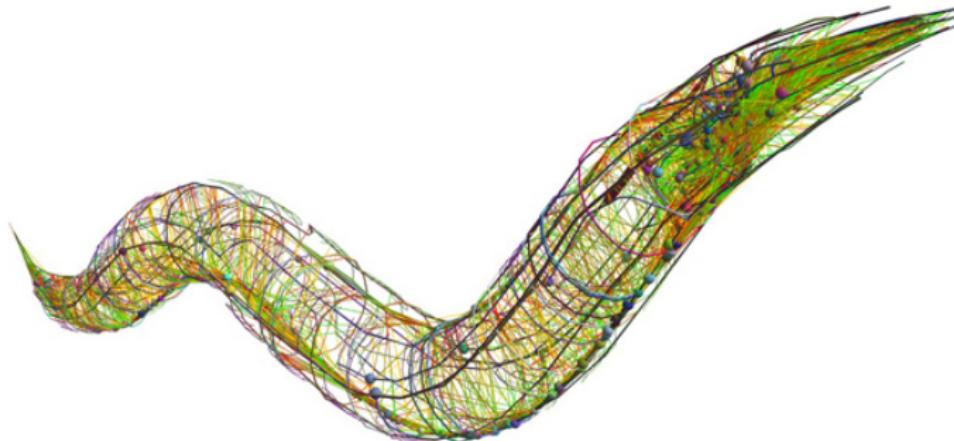
motivation *network*

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



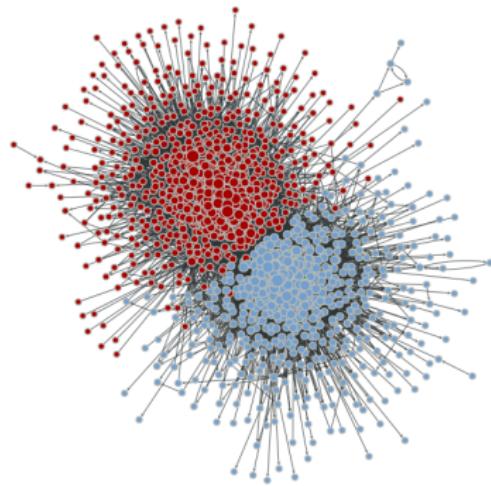
motivation *neural wiring*

- *human brain* $\approx 10^{11}$ neurons
- nodes are *C. elegans* neurons
- links are *synapses*



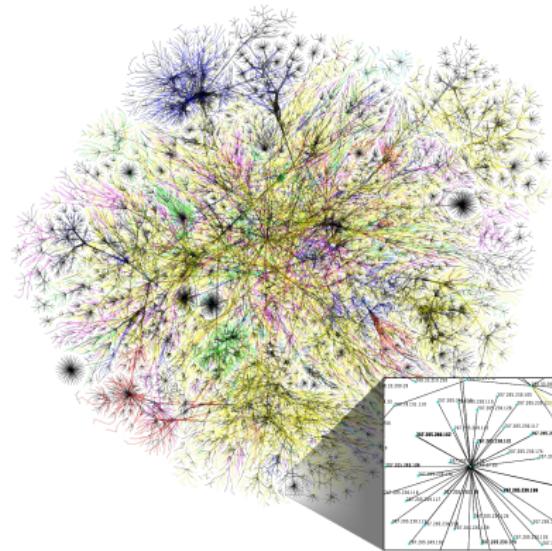
motivation *Web*

- *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*



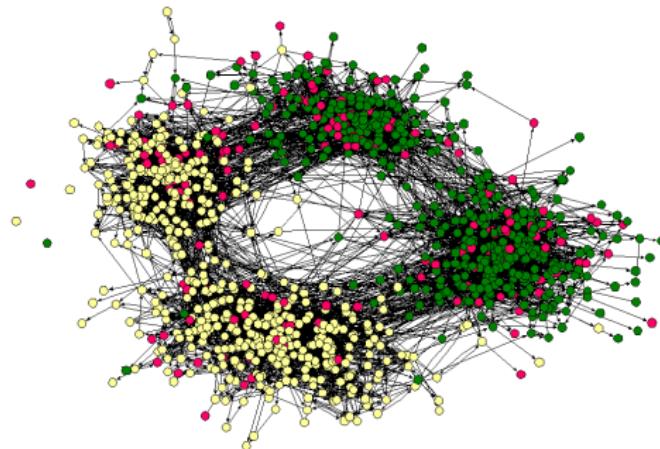
motivation *Facebook*

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



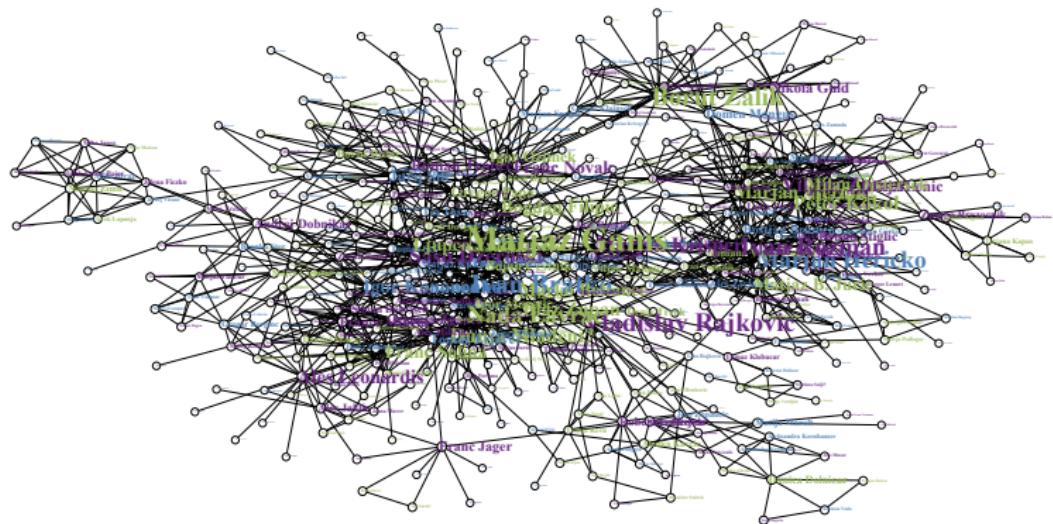
motivation *society*

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



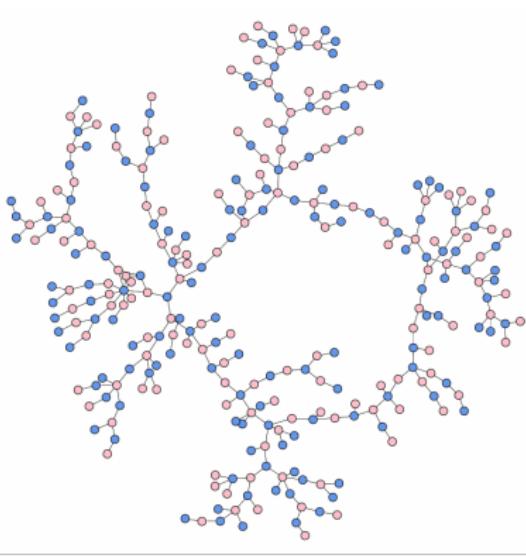
motivation *collaboration*

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



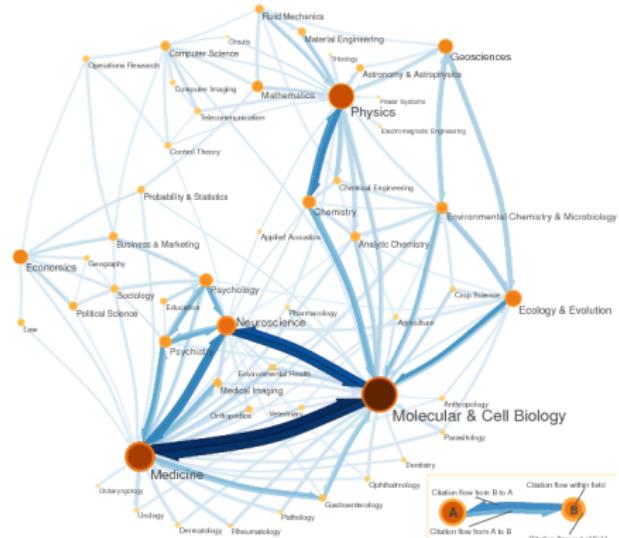
motivation *sex*

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



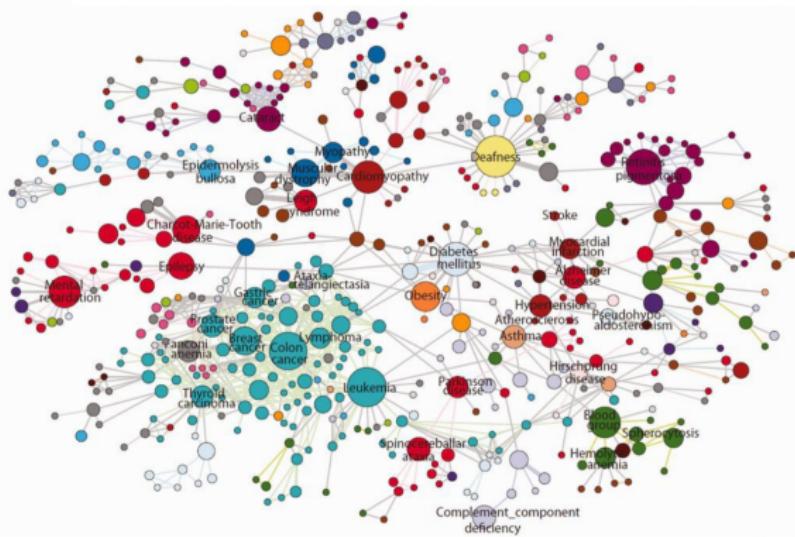
motivation *science*

- *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*



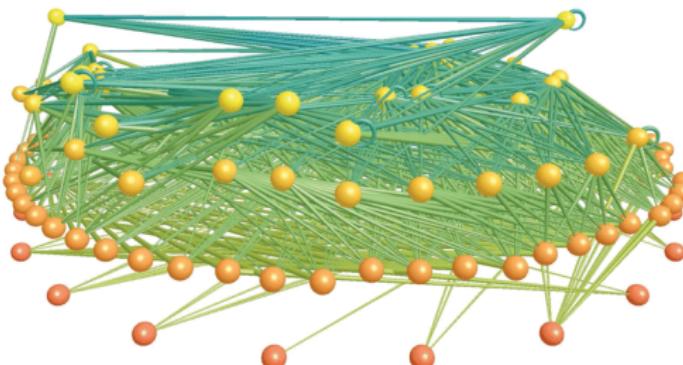
motivation *biology*

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



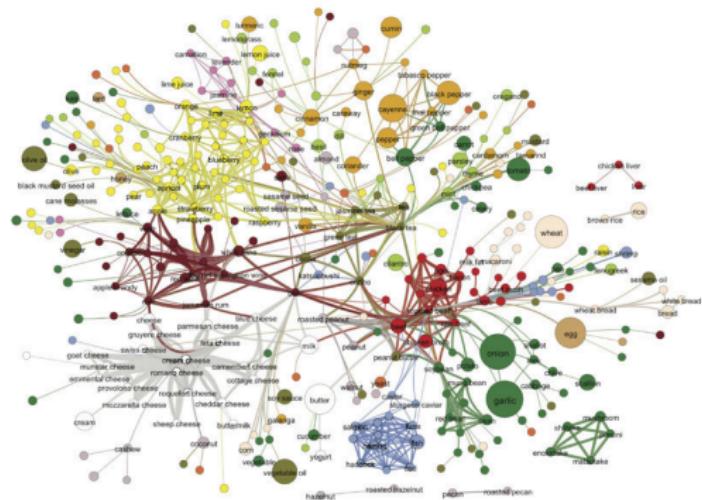
motivation *ecology*

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



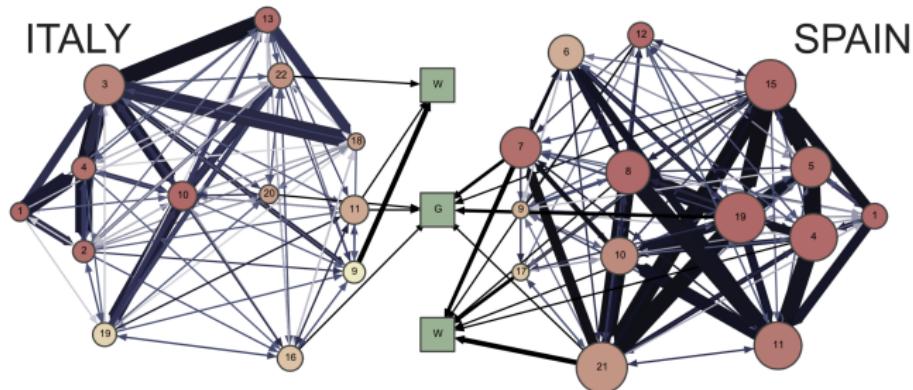
motivation *gastronomy*

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



motivation *sports*

- *football player* network
- nodes are *football players*
- links are *player passes*



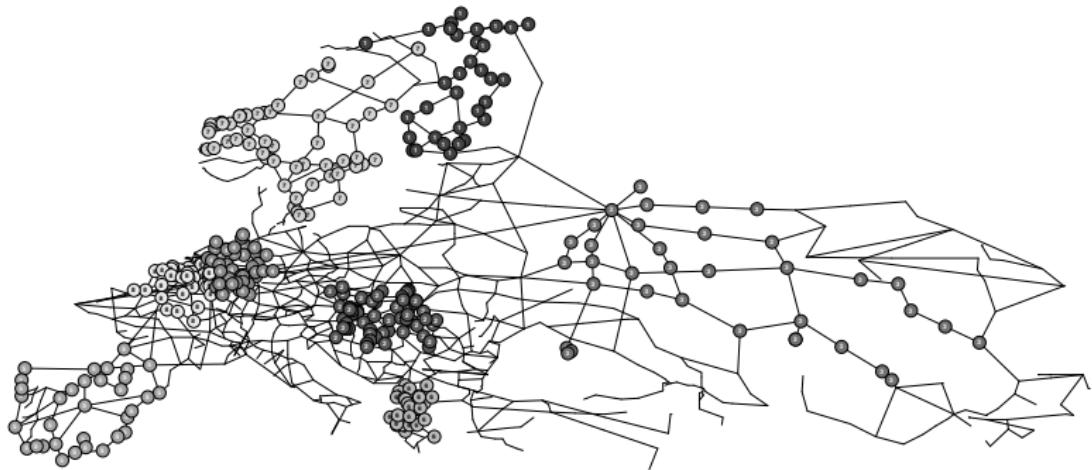
motivation *transport*

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



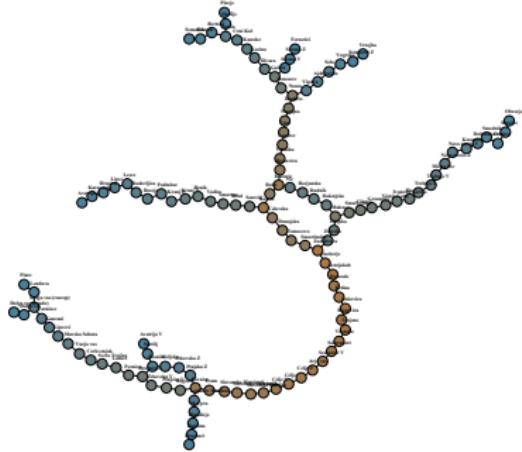
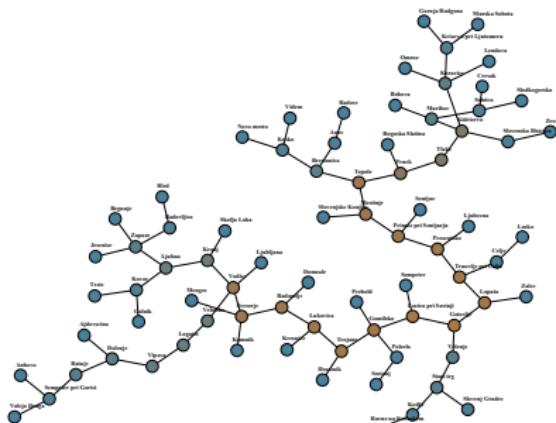
motivation *Europe*

- *road infrastructure* network
- nodes are *European cities*
- links are *highways*



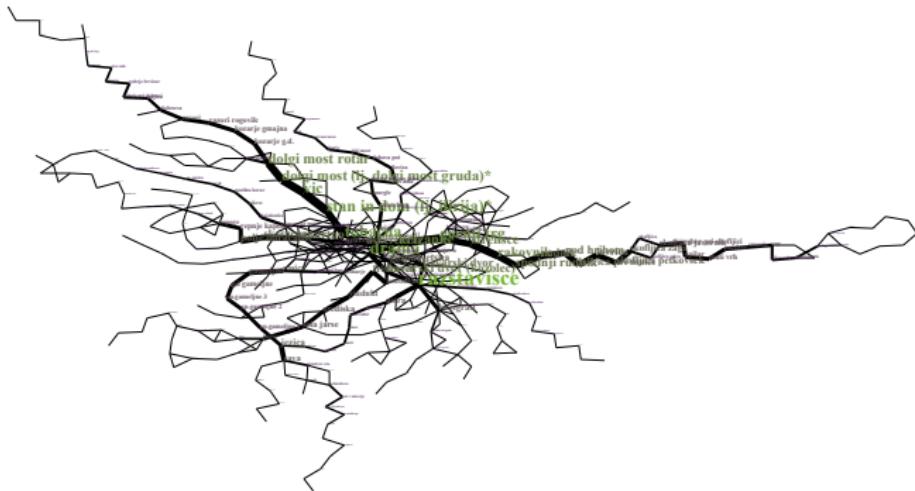
motivation *Slovenia*

- technological/*infrastructure* networks
 - nodes are *geographical locations*
 - links are *pipes/highways*



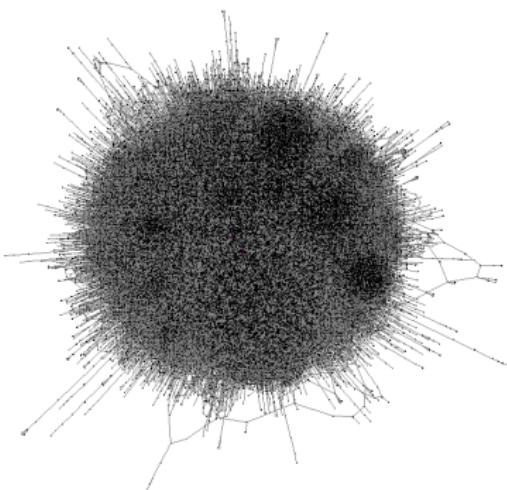
motivation *Ljubljana*

- LPP *bus map*
 - nodes are *bus stops*
 - links are *connections*



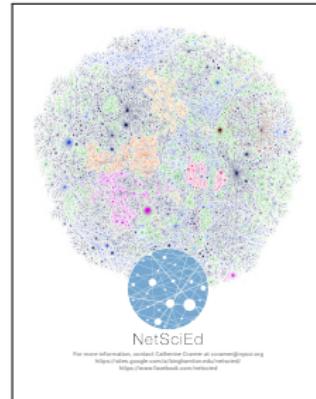
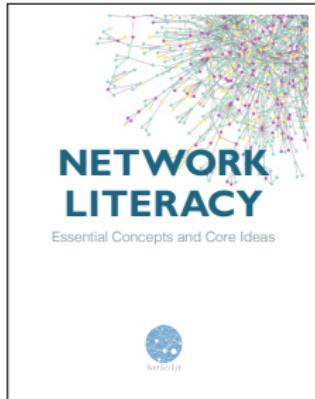
motivation *hairballs*

- 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



[WS98]



[BA99]



[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.

course *logistics*

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2019/20

course *design*



elective courses on NETWORKS in 2019/20

Networks or graphs are ubiquitous in everyday life. Examples include online social networks, the Web, references between WikiLeaks cables, Supervisor, terrorist affiliations, LPP bus map, plumbing systems and your brain. Many such real networks reveal characteristic patterns of connectedness that are far from regular or random. Networks have thus been a prominent tool for investigating real-world systems since the 18th century. However, while small networks can be drawn by hand and analyzed by a naked eye, real networks require specialized computer algorithms, techniques and models. This led to the emergence of a new scientific field about 20 years ago...

INA

[Introduction to] Network Analysis

Network analysis concepts and techniques

Course code **63545B** | eUcilnica #**183**
MSc students | Lecturer Lovro Šubelj
Summer semester | Starts **Feb 17, 2020**
Introductory 15 week course to get started

MLG

Machine Learning with Graphs

Modern analysis of large-scale real networks

Course code **63225B** | eUcilnica #**66**
MSc students | Lecturer Jure Leskovec
Winter semester | Starts **Sep 24, 2019**
Fast pace 10 week course by a leading expert

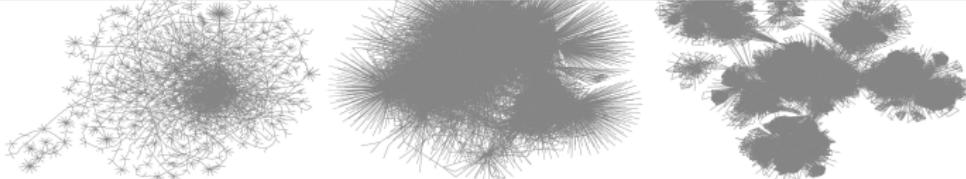
ANTS

Advanced Topics in Network Science

Thorough review of modern network science

Course code **63835A** | eUcilnica #**170**
PhD students | Lecturers Šubelj & Leskovec
Summer semester | Starts **Mar 2, 2020**
Research-oriented 12 week course & invited talks

Course enrollment is not possible in order from right to left | lovro.sobelj@fri.uni-lj.si



course *overview*

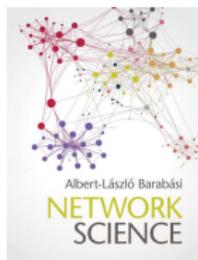
- main *course objective*
 - *graph theory* vs *network science*
 - fundamental *concepts* and *techniques*
 - selected practical *applications*
- high-level *course outline*
 - theoretical *foundations* in *lectures*
 - practical *applicability* in *labs/coursework*
- *prerequisites* and *background*
 - good *programming skills* (Python, Java, C/C++)
 - *linear algebra*, *probability theory* and *statistics* (basics)

course *syllabus*

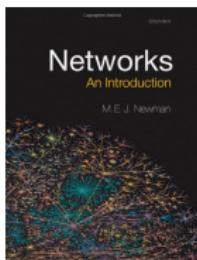
- from *graph theory* to *network science*
- *random graphs* and *real networks*
- node *position* and measures of *centrality*
- link *importance* and measures of *bridgeness*
- node *similarity* and structural *equivalence*
- graph *partitioning* and *community* detection
- network *clustering* and frequent *subgraphs*
- network structures, algorithms, libraries
- network collection, sampling, comparison
- network *structure* and *modeling*
- *scale-free* and *small-world* networks
- node *layout* and network *visualization*
- node *classification* and link *prediction*
- *applications* of network analysis

course *literature*

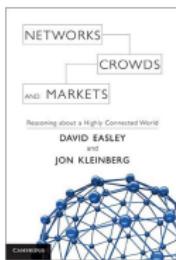
- *lecture/lab handouts as notes*
- *book chapters as background reading*
- *scientific papers as further readings*



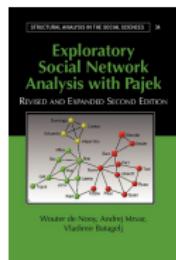
[Bar16]



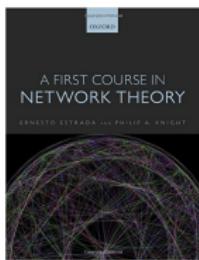
[New10]



[EK10]



[dNMB11]



[EK15]

course *schedule*

- weekly *course schedule*
 - *lectures* on *Mondays at 11:00am*
 - *office hours* on *Tuesdays at 1:00pm*
 - *labs* on *Mondays at 2:00pm* and *Tuesdays/Fridays at 11:00am*
- course *communication* in English/Slovene
 - classroom and office hours
 - *questions* and *comments* on *Piazza*
 - *announcements* on *eUcilenica*
 - avoid private e-mails!

coursework *overview*

- attendance at *lectures/labs*
- book *chapters* and course *readings*
- *three homeworks* (10-20 hours each)
- *substantial course project* (60-90 hours)
- *thorough oral exam* (theoretical & practical)
- course *participation* and *commitment*

coursework *homeworks*

- *homework #0* to get you started
- lengthy *homeworks #1, #2, #3*
 - *analytical derivations* of concepts and techniques
 - *programming* network analysis algorithms
 - *practical applications* to real networks
- students encouraged to form *study groups*
- start working on *homework #0 today*

coursework *project*

- substantial *course project*
 - empirical *evaluation* on interesting network
 - analytical *derivation* of theoretical results
 - efficient *implementation* for large networks
 - *contribution* to community/this course?
- *project proposal*, *milestone* and *report*
 - up to eight-page *technical report* (to arXiv.org)
 - *faculty talk* (within *Cookie/AI seminars*)?
 - project made *publicly available*
- informal *proposal* and (*report*) *presentations*
- students encouraged to work in *pairs/triplets*
- start thinking about *project soon/today*

coursework *submissions*

- written/typed in English/Slovene
- *hard-copy* in *submission box* (grading)
- *electronic version* to *eUcilmica* (plagiarism)
- *cover sheet* with signed *honor code*
- honor code violation *not tolerated!*
- *late days twice* in semester

coursework *deadlines*

- *assignments* due on *Mondays at 2:00pm*
- *late days* expire on *Fridays at 1:00pm*
- project *presentations* during *lectures/labs*

week	out	due	challenge	comment
1	<i>homework #0</i>			
2		<i>homework #0</i>	<i>4 knights</i>	
3				
4	<i>homework #1</i>		<i>grand graph</i>	
5				
6		<i>homework #1</i>	<i>%-centrality</i>	<i>COSTNET '20</i>
7	<i>homework #2</i>			
8			<i>5 networks</i>	
9		<i>homework #2</i>		<i>Easter days</i>
10	<i>homework #3</i>		<i>PageRank</i>	
11				<i>Labour days</i>
12		<i>homework #3</i>		
13		<i>proposal</i>		
14				
15		<i>milestone</i>		
16				<i>Semester ends</i>
17		<i>project</i>		<i>Exam period</i>

coursework *grading*

- coursework *grade* breakdown
 - 37.5% based on *homeworks*
 - 12.5% on each homework
 - homework #0 not graded
 - 47.5% based on *course project*
 - 10% on project proposal
 - 15% on project milestone
 - 22.5% on final project
 - 15% based on *commitment*
 - quizzes and challenges
 - lectures, labs, seminars
- final *grade* in *oral exam*

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.

graph theory → *network science*

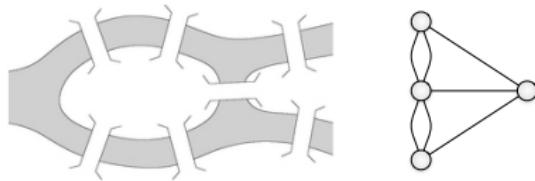
introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2019/20

history *graph theory*

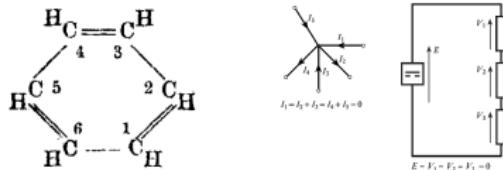
1736 seven *bridges of Königsberg* [Eul36] (Leonhard Euler)

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)

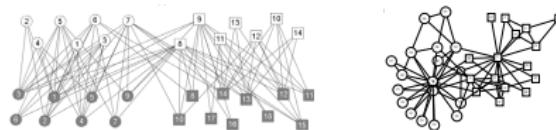
history *sociometry*

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*

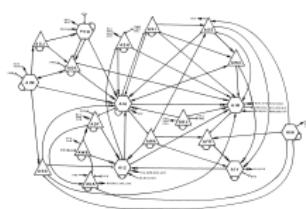
1965 scientific *paper citations* [Pri65] (Derek de Solla Price)



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

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

1999 *transportation* [Pel99] (Jon Pelletier)



networks *boom*

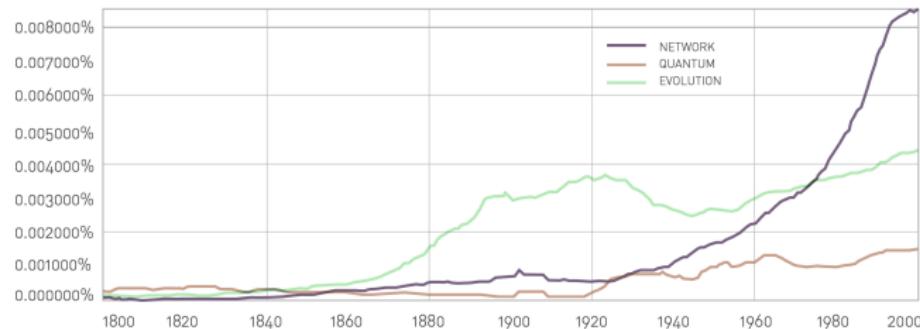
< 2000 *small graphs* $10^2\text{-}10^3$ nodes

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

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

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

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



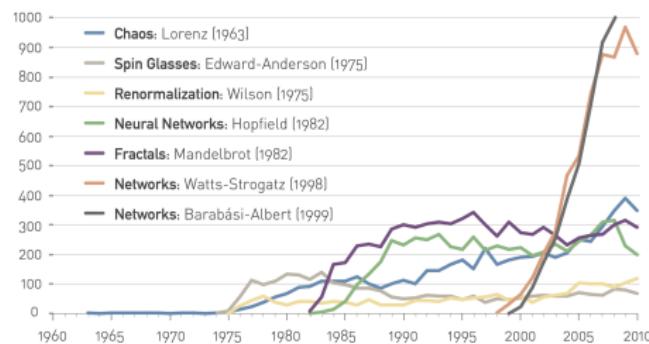
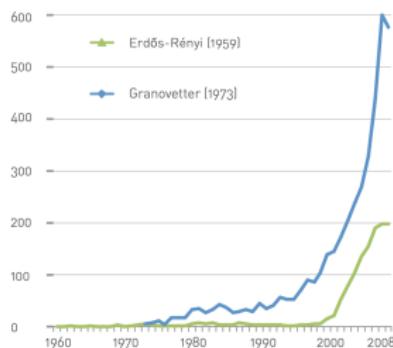
network *models*

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

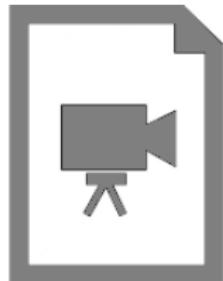
network *analysis*



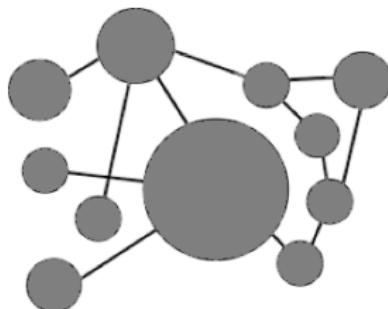
data mining



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.