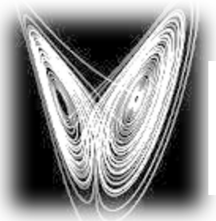
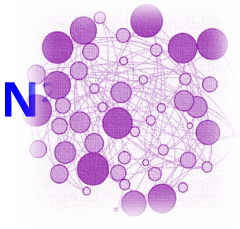
Introduction to Complex systems and networks: from epidemics to epilepsy



Lab For Networks And Nonlinear Dynamics (LN



Chittaranjan Hens

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

CCNSB

Dynamical Processes in complex networks

Requirements

- Newtonian Mechanics
- Basics of Matrices/Linear Algebra
- Matlab/Python/C++
- Graphical Plot (Must)
- Differential Equations
- Computational Complexity
- Error calculation

Module 1 Introduction

Complex systems

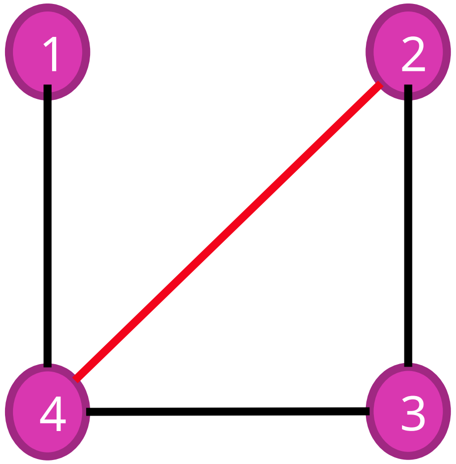
Outlines

- Introduction of graphs/networks
- Statistical physics and interacting particles
    - Complex systems: Interaction, emergence
- Why Network Science?
    - Connectivity and emergence

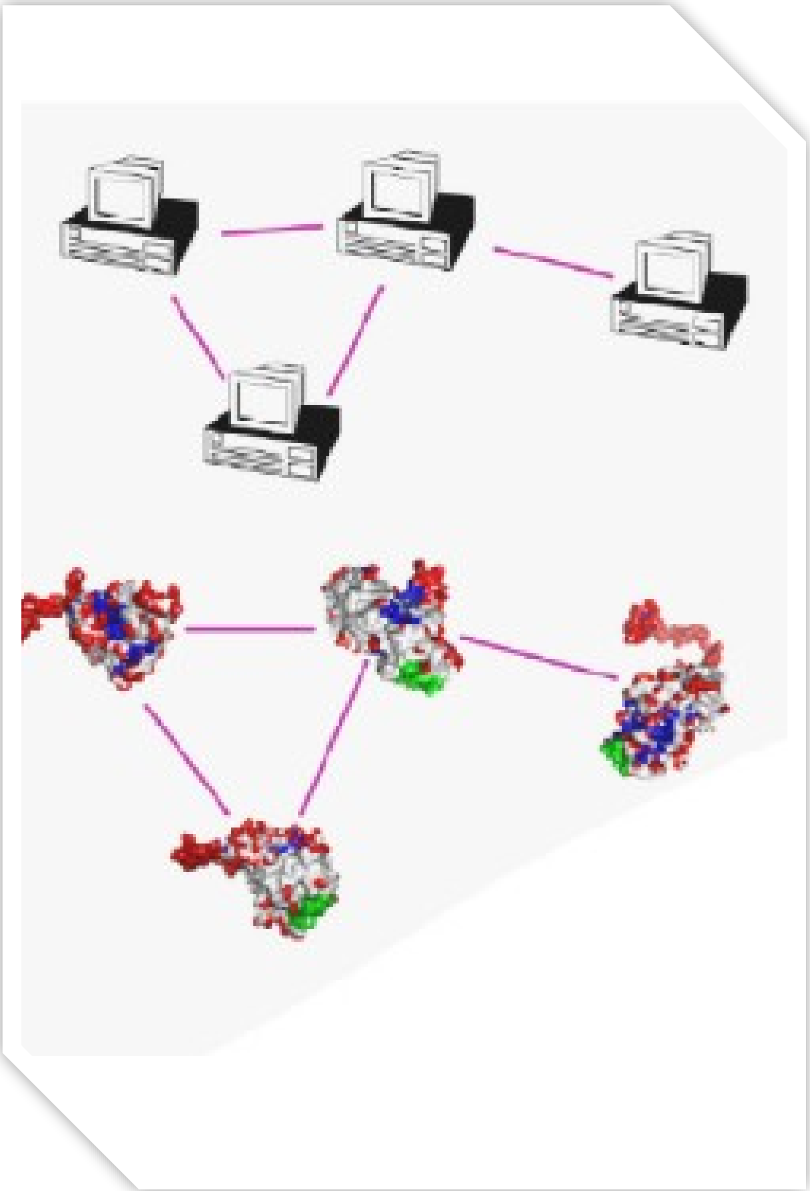
Networks/Graphs

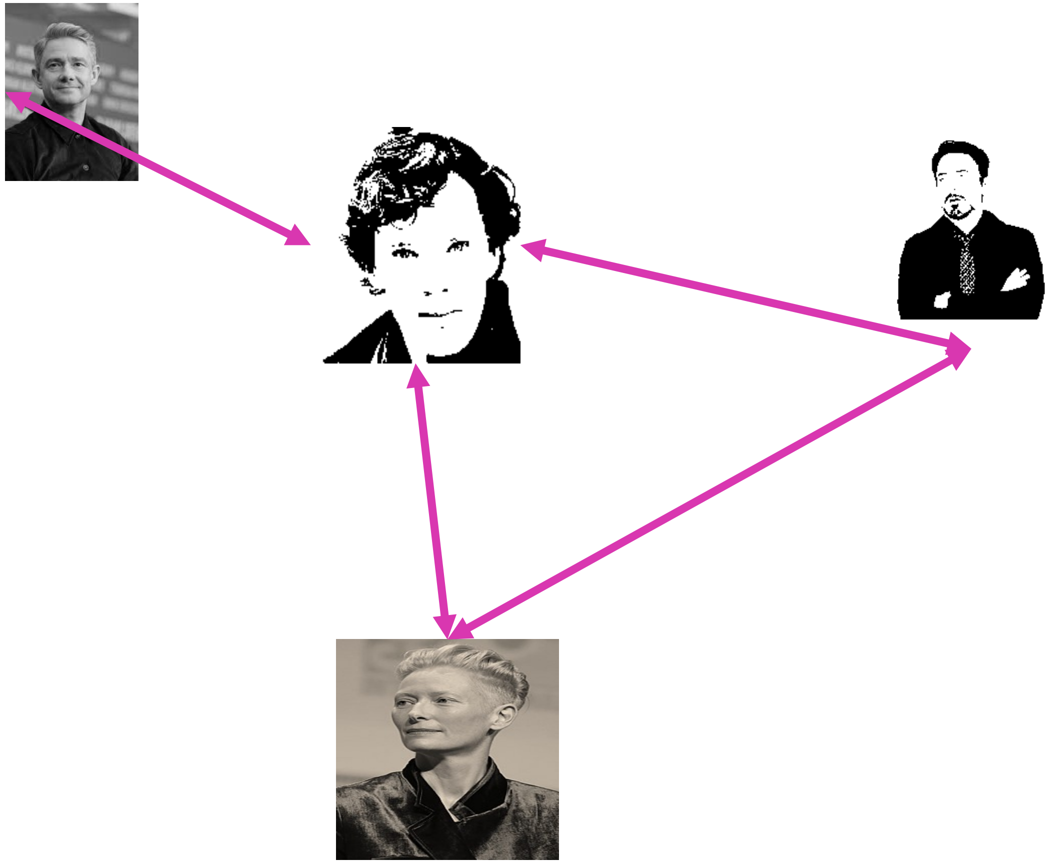
A_{ij}

1	0	0	0
1	1	0	0
1	0	1	0
0	1	1	1



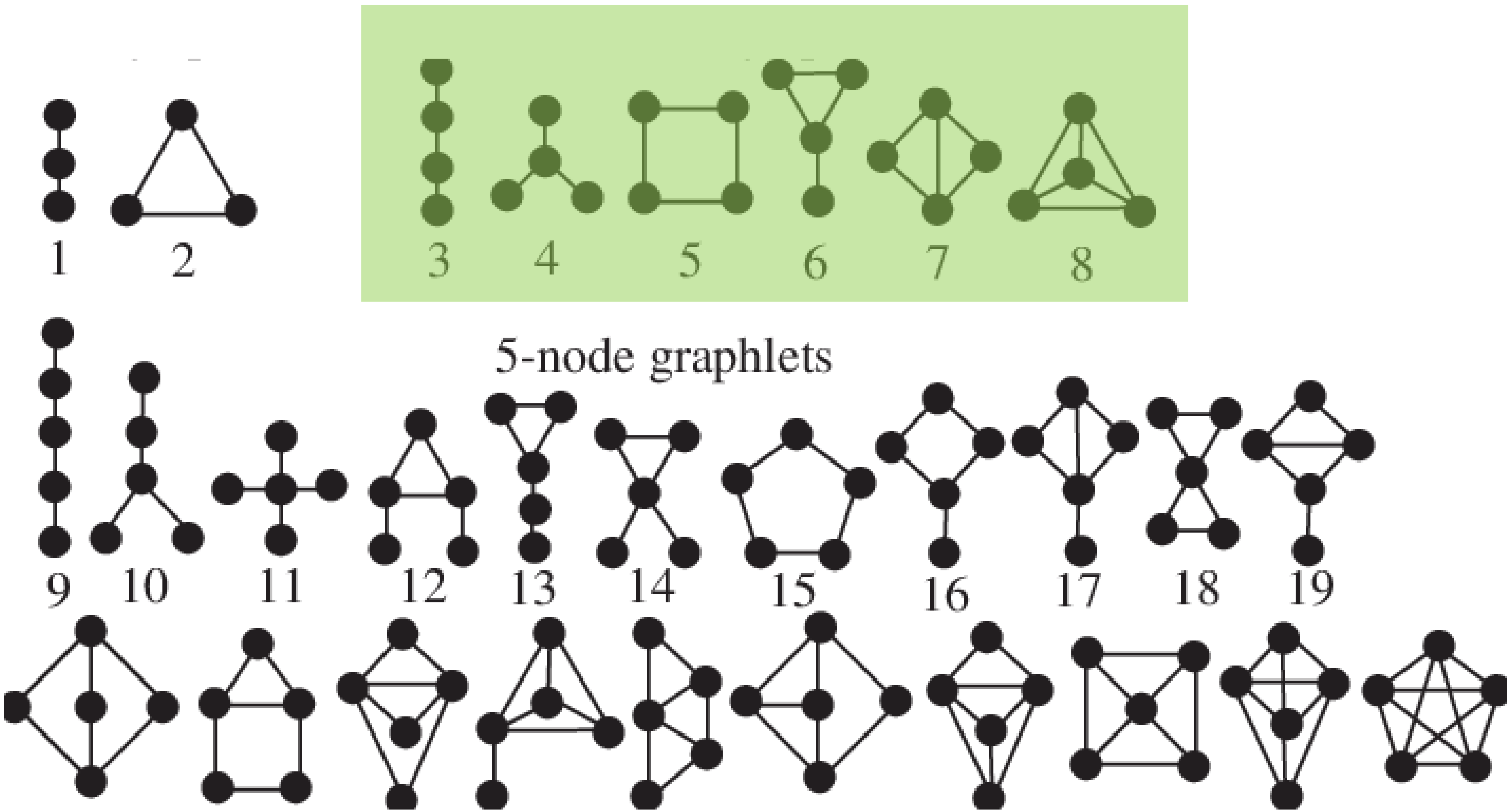
Networks/Graphs





•**Benedict Cumberbatch** *is connected to* **Robert Downey Jr.** *through their appearances in the Marvel Cinematic Universe (MCU).*
•**Benedict Cumberbatch** *and* **Martin Freeman** *are linked through their work on the TV series* **"Sherlock"**.
•**Tilda Swinton** *has connections both with* **Cumberbatch** *(from* **"Doctor Strange"***)* *and* **Downey Jr.** *(from* **"Avengers: Endgame"***).*

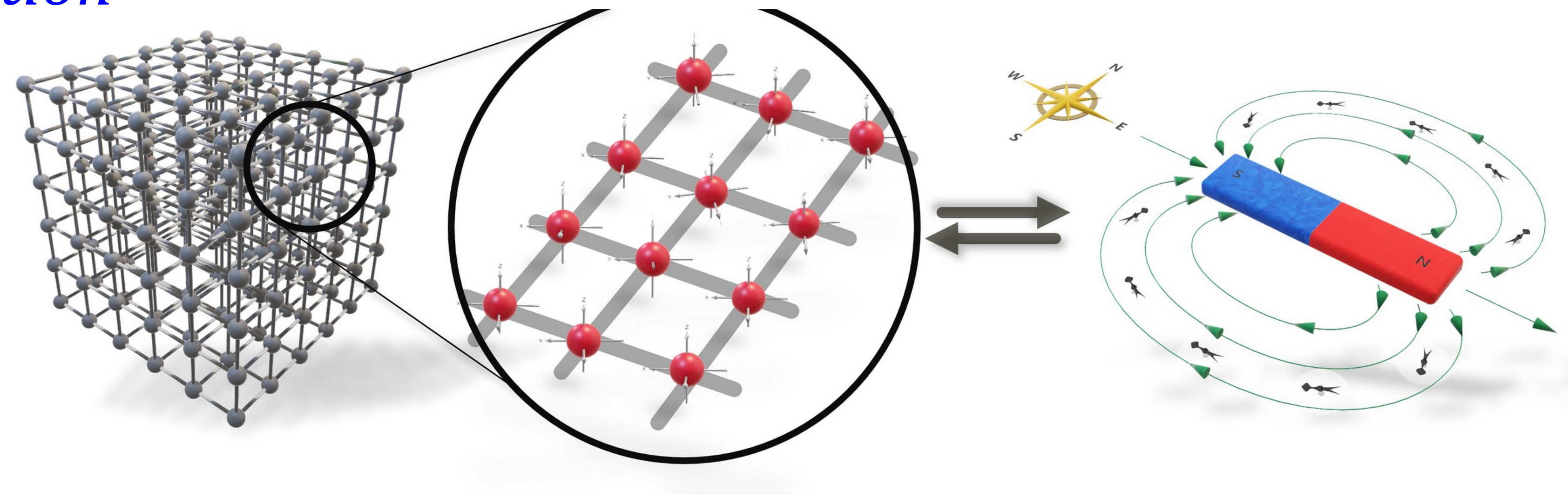
Networks/Graphs

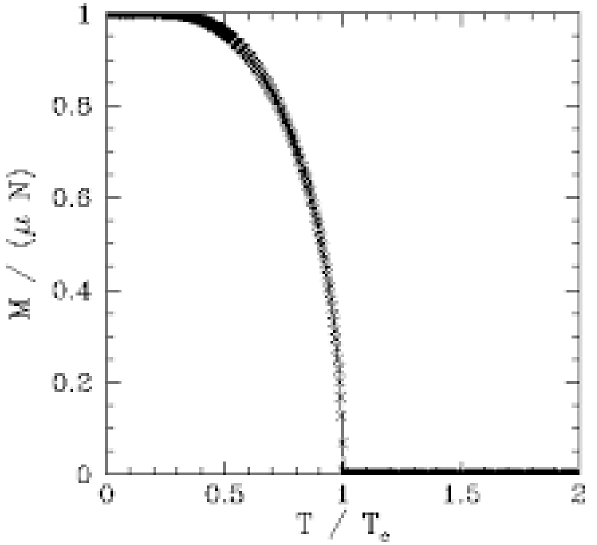


Statistical physics and interacting particles

Magnetization

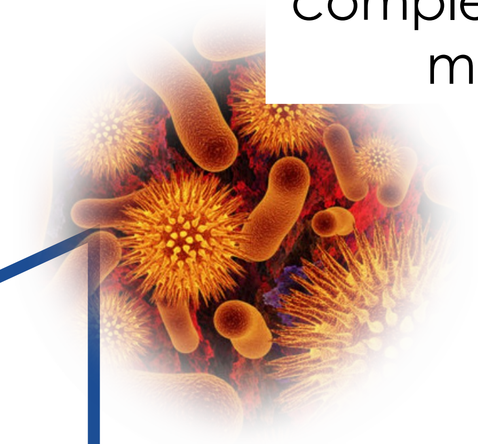
Statistical physics and interacting particles





is the atomic magnetic moment. is the net magnetization

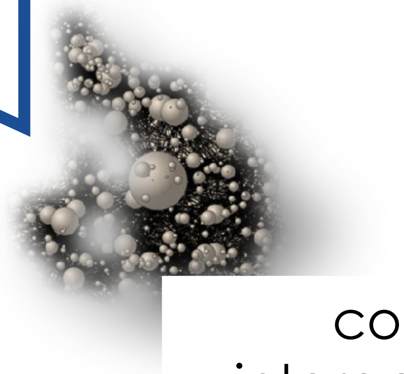
Statistical physics and interacting particles



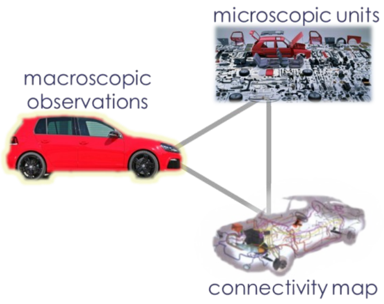
complex interaction mechanisms

macroscopic observation

?



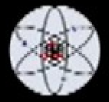
complex interaction maps



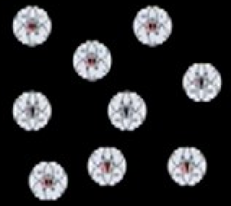
More is different

P. W. Anderson
Nature (1972)

Interaction between smaller entities leads to new physics!



Microscopic Physics: individual constituent particles



Many interacting particles

Emergence of collective order!

$$Z = \sum_{\{s\}} \exp[-\beta \mathscr{H}(s)]$$

Macroscopic properties

Complex Systems

Physics

Examples: deterministic chaos, quantum entanglement, spin glasses

Study of complex systems has a long history in Physics, dating back to Aristotle's time, and more relevant than ever in this century

"I think the next [21st] century will be the century of complexity"

— Stephen Hawking

- *Things derive their being and nature by mutual dependence and are nothing in themselves.*
  -Nagarjuna, second century Buddhist philosopher

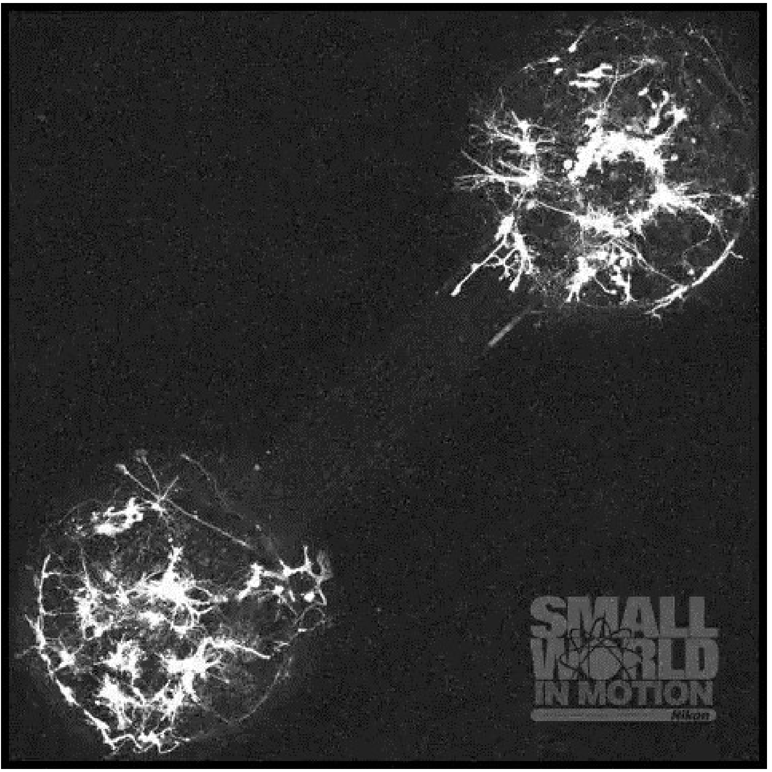
More is different,

— P. Anderson, Science (1972)

philosophy of science and emergent phenomena; limitations of reductionism and the existence of hierarchical levels of science

Module 1 Introduction

Why Network Science?

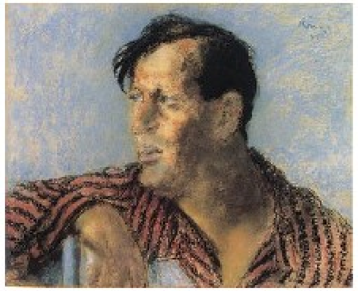


Neurons communicate with each other through specialized junctions called synapses. The formation of new synapses is crucial for learning and memory. It's a dynamic process that involves the growth and pruning of connections between neurons.

https://www.nikonsmallworld.com/galleries/2016-small-world-in-motion-competition/neurons-seeded-in-two-different-micro-compartments-extend-their-neurites

Why Network Science?

Complex networks– Construction 1

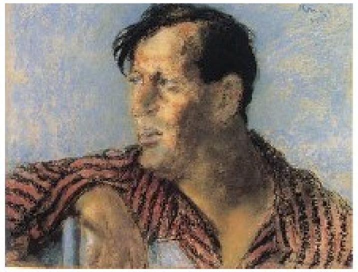


1929 Frigyes Karinthy

"If you choose a person out of the 1.5 billions of our planet, I bet that using no more than ***five*** individuals, one of them my acquaintance, I could contact the person you chose, using only the list of acquaintances of each one"

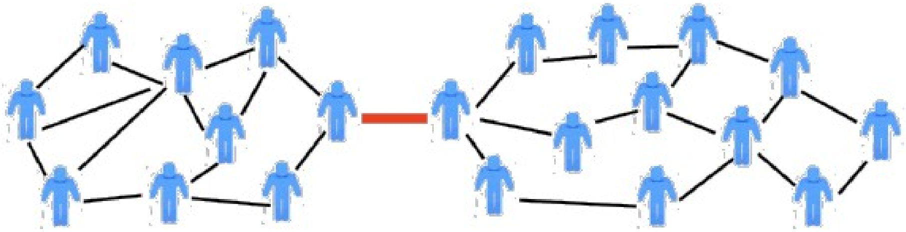
Why Network Science?

Complex networks– Construction 1



1929 Frigyes Karinthy

"If you choose a person out of the 1.5 billions of our planet, I bet that using no more than ***five*** individuals, one of them my acquaintance, I could contact the person you chose, using only the list of acquaintances of each one"



Why Network Science?

Complex networks– Construction 1

In 1967, psychologist Stanley Milgram coined the phrase "*six degrees of separation*" to describe the world-shrinking effects of social networks.

Why Network Science?

Complex networks– Construction 1

1969 Stanley Milgram

- People chosen at random on a US State
- Request to send a letter to a given final person in another state :
  - If you know the final person, send directly to him
  - If not, send to someone you think it is more likely to know him

An Experimental Study of the Small World Problem*

JEFFREY TRAVERS
Harvard University
AND
STANLEY MILGRAM
The City University of New York

Why Network Science?

Complex networks– Construction 1

1969 Stanley Milgram

- People chosen at random on a US State
- Request to send a letter to a given final person in another state :
  - If you know the final person, send directly to him
  - If not, send to someone you think it is more likely to know him

Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.

Why Network Science?

Complex networks– Construction 1

1969 Stanley Milgram

- People chosen at random on a US State
- Request to send a letter to a given final person in another state :
  - If you know the final person, send directly to him
  - If not, send to someone you think it is more likely to know him

An Experimental Study of the Small World Problem*

JEFFREY TRAVERS
Harvard University
AND
STANLEY MILGRAM
The City University of New York

Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.

Why Network Science?

Complex networks– Construction 1

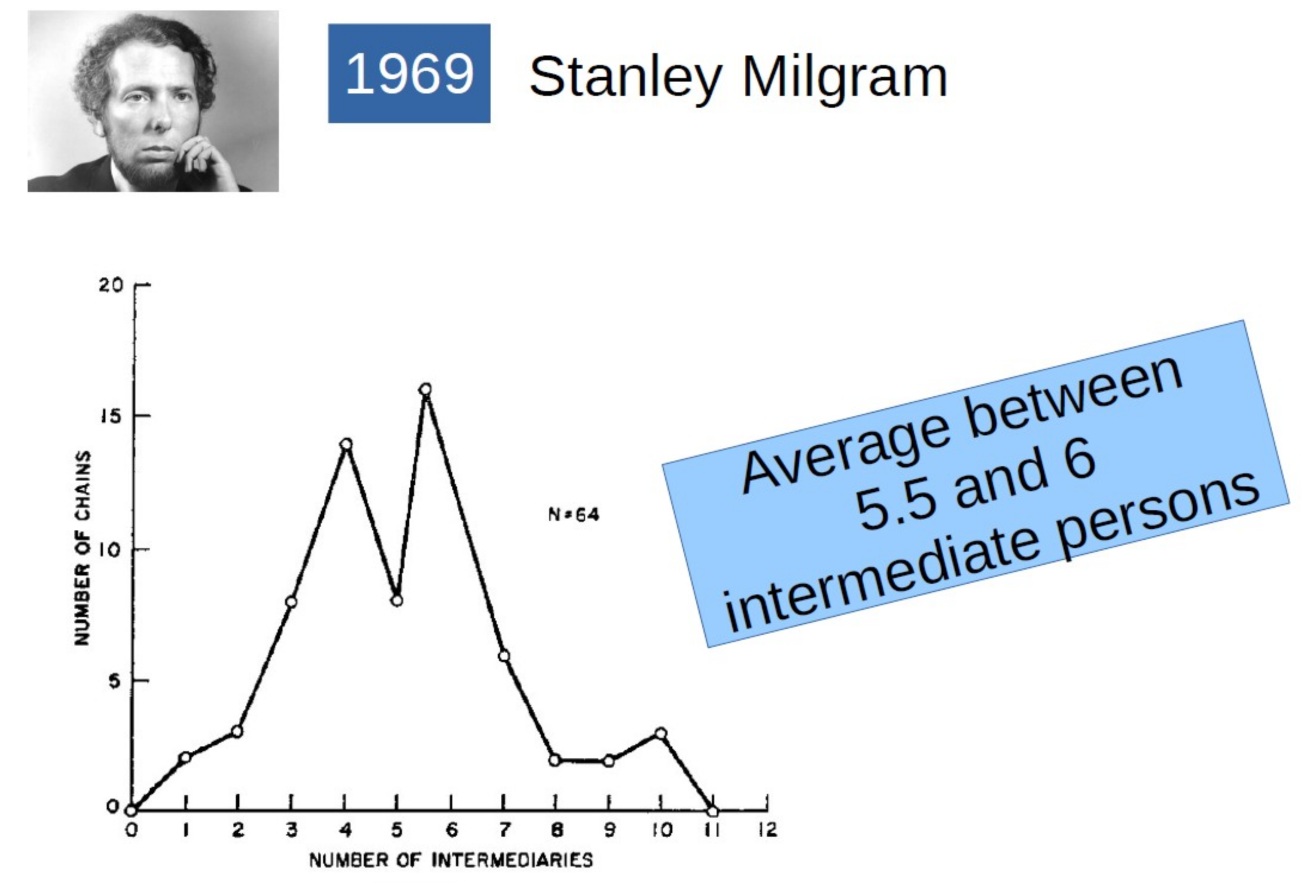


FIGURE 1

Lengths of Completed Chains

Why Network Science?

2008 Microsoft Messenger

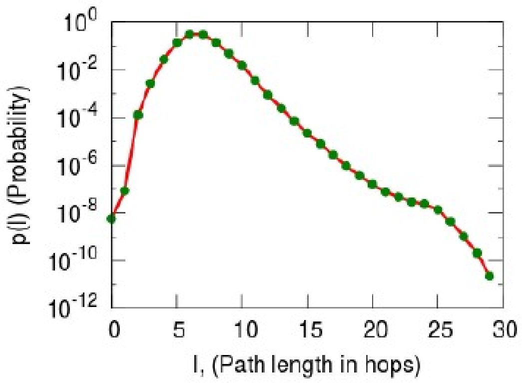
- 30 billion conversations between 240 million persons

Planetary-Scale Views on a Large Instant-Messaging Network

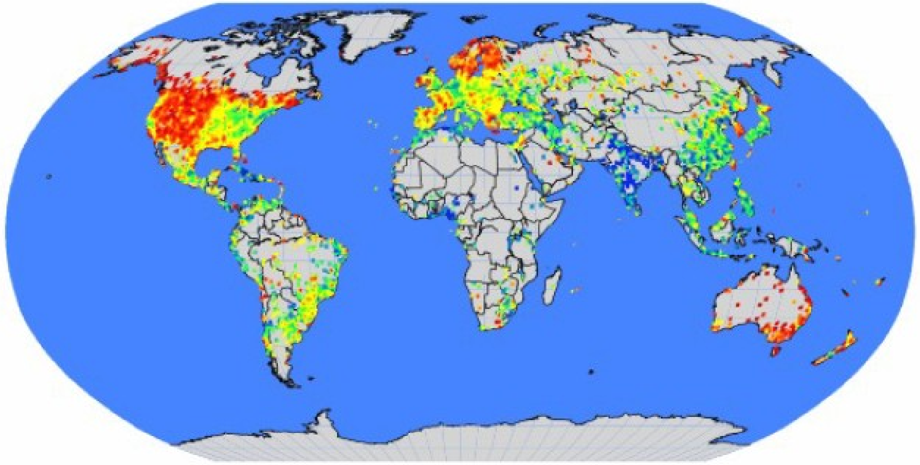
Jure Leskovec*
Carnegie Mellon University
jure@cs.cmu.edu

Eric Horvitz
Microsoft Research
horvitz@microsoft.com

Global Average: 5.6



(a) Diameter



Why Network Science?

- Imagine that a person has, on average, 100 friends
  - 0 intermediates: 100
  - 1 intermediate: $100^2 = 10.000$
  - 2 intermediates: $100^3 = 1.000.000$
  - 3 intermediates: $100^4 = 100.000.000$
  - 4 intermediates: $100^5 = 10.000.000.000$
  - 5 intermediates: $100^6 = 1.000.000.000.000$

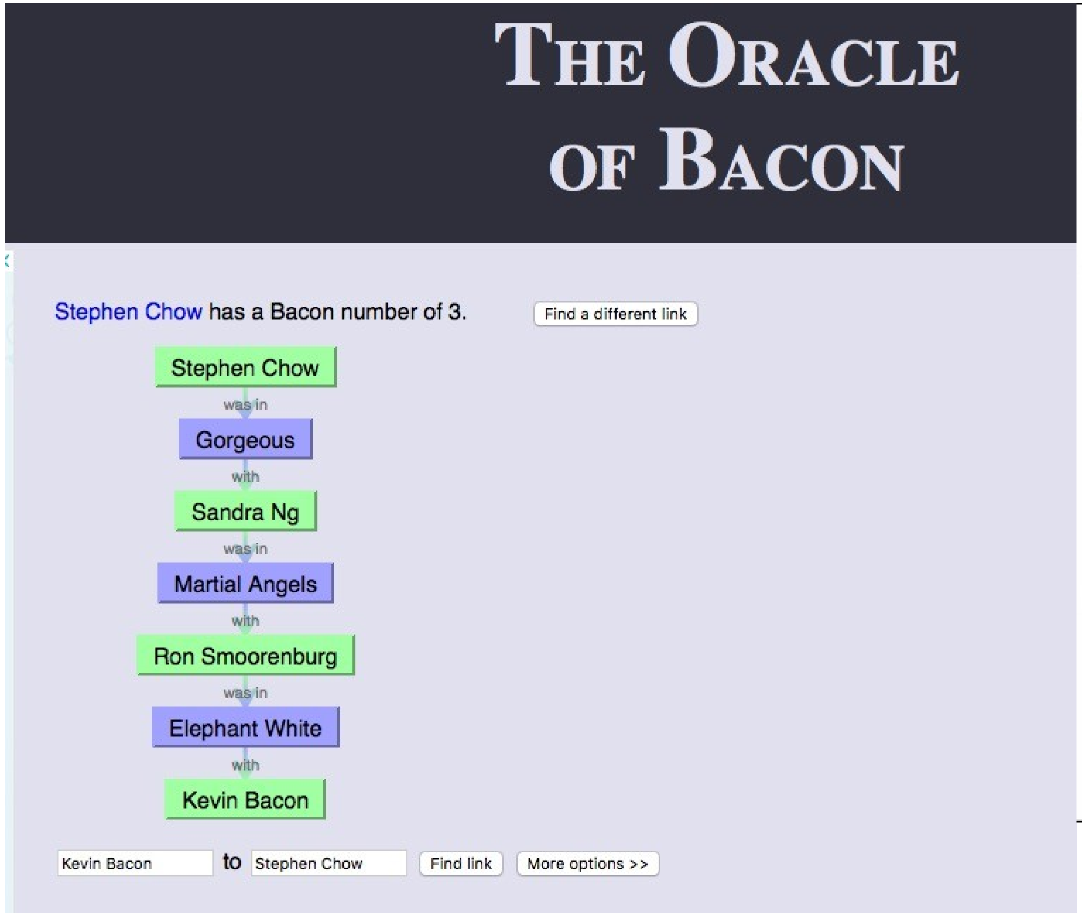
The power of exponentiation

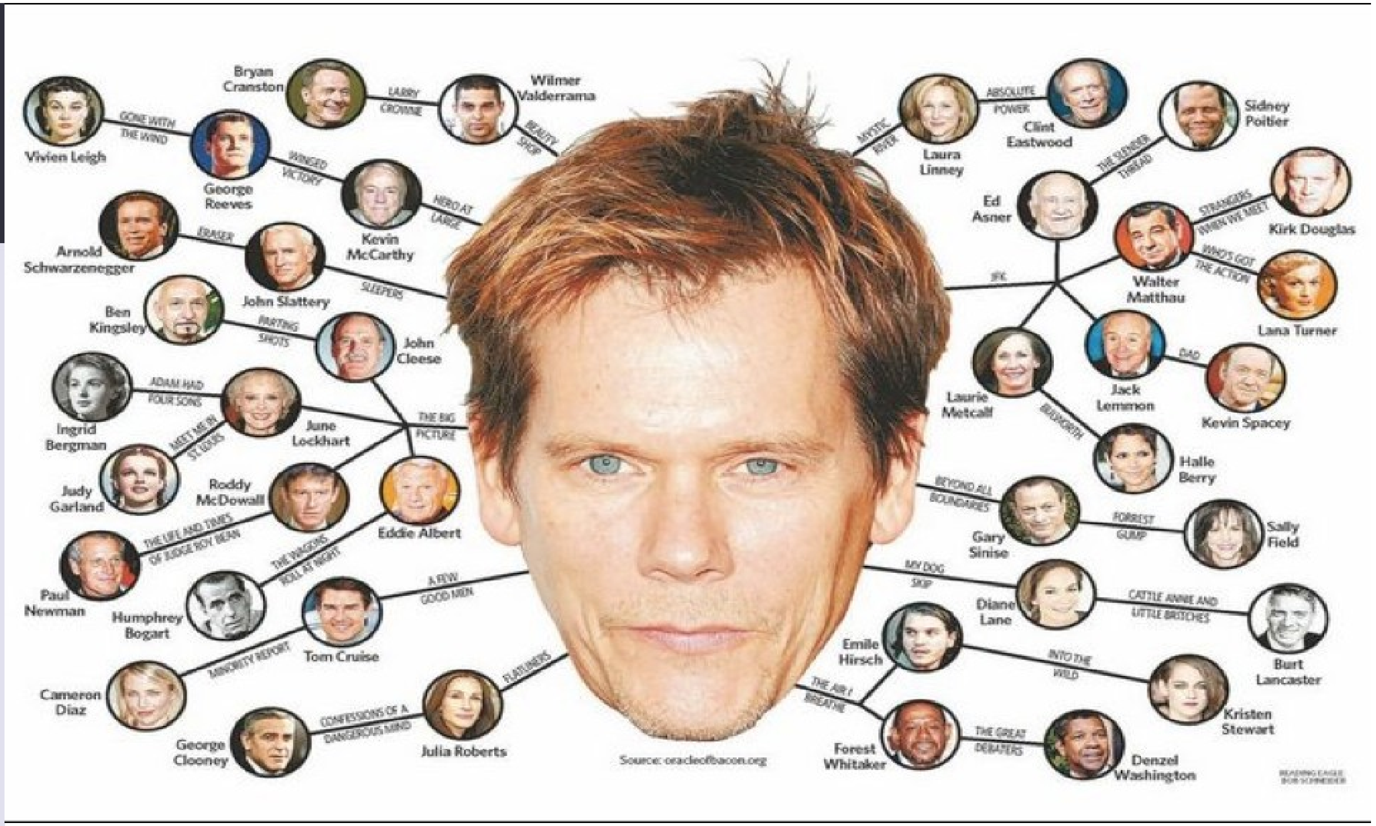
- In practice, not all friends are new, but still there is a very fast growth

Why Network Science?

Complex networks– Construction 2

+ *Kevin Bacon Game http://oracleofbacon.org/movielinks.php*

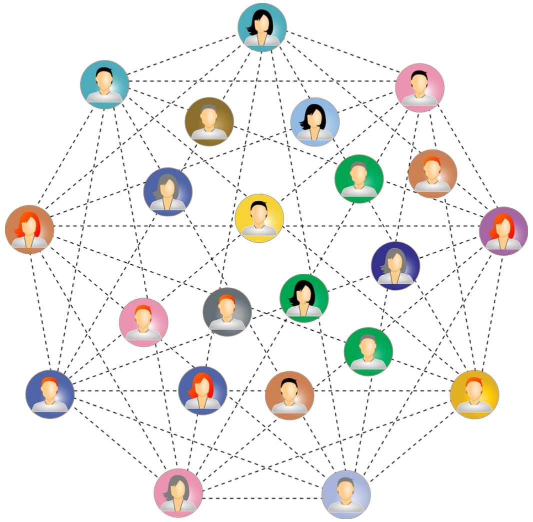


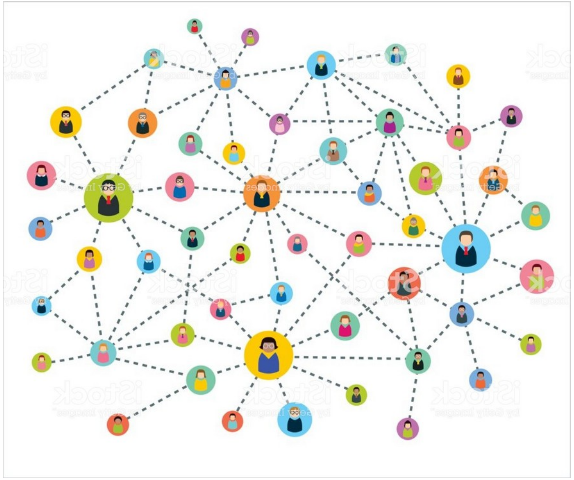


Why Network Science?

Complex networks– Construction 3

+ *Friendship network*



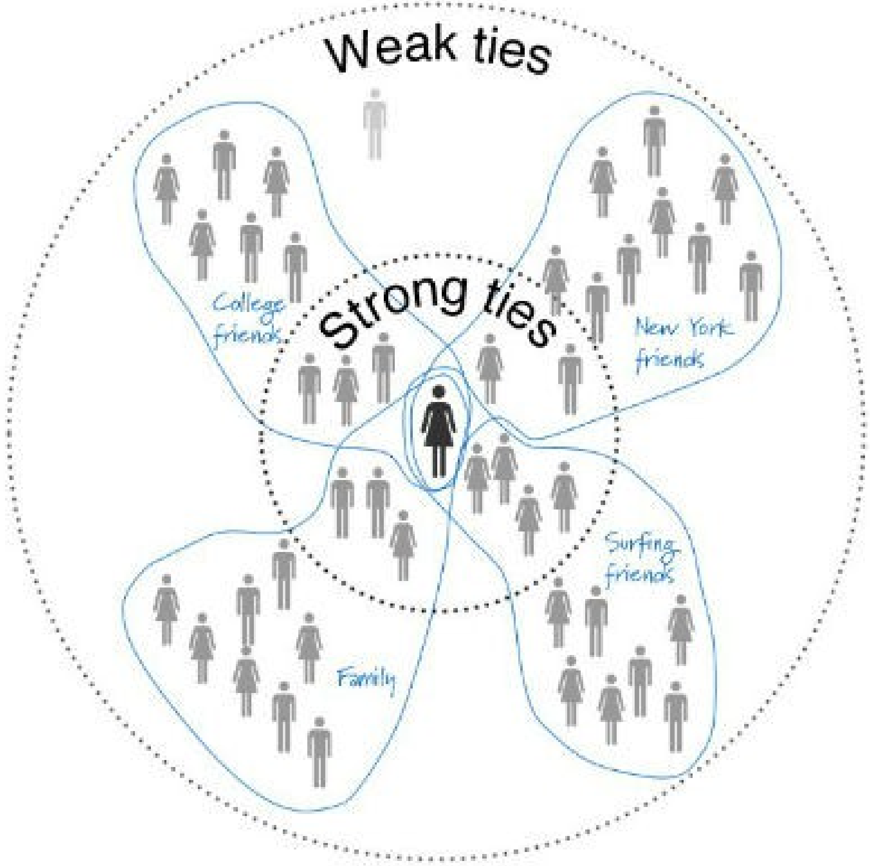


Time evolution

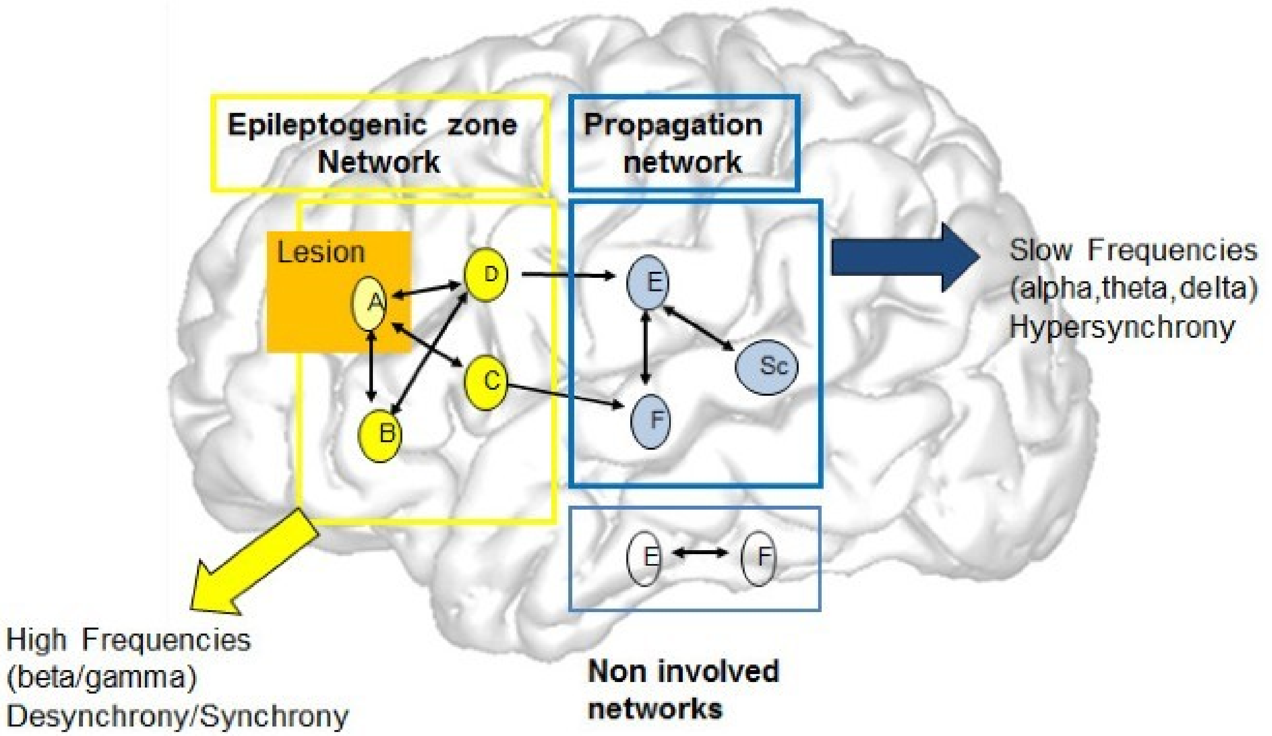
+ *Limits of friendship*

Complex networks– Functionality 1

- *Weak ties*



Friendship



Brain network of networks, Integration

Complex networks– Functionality 2

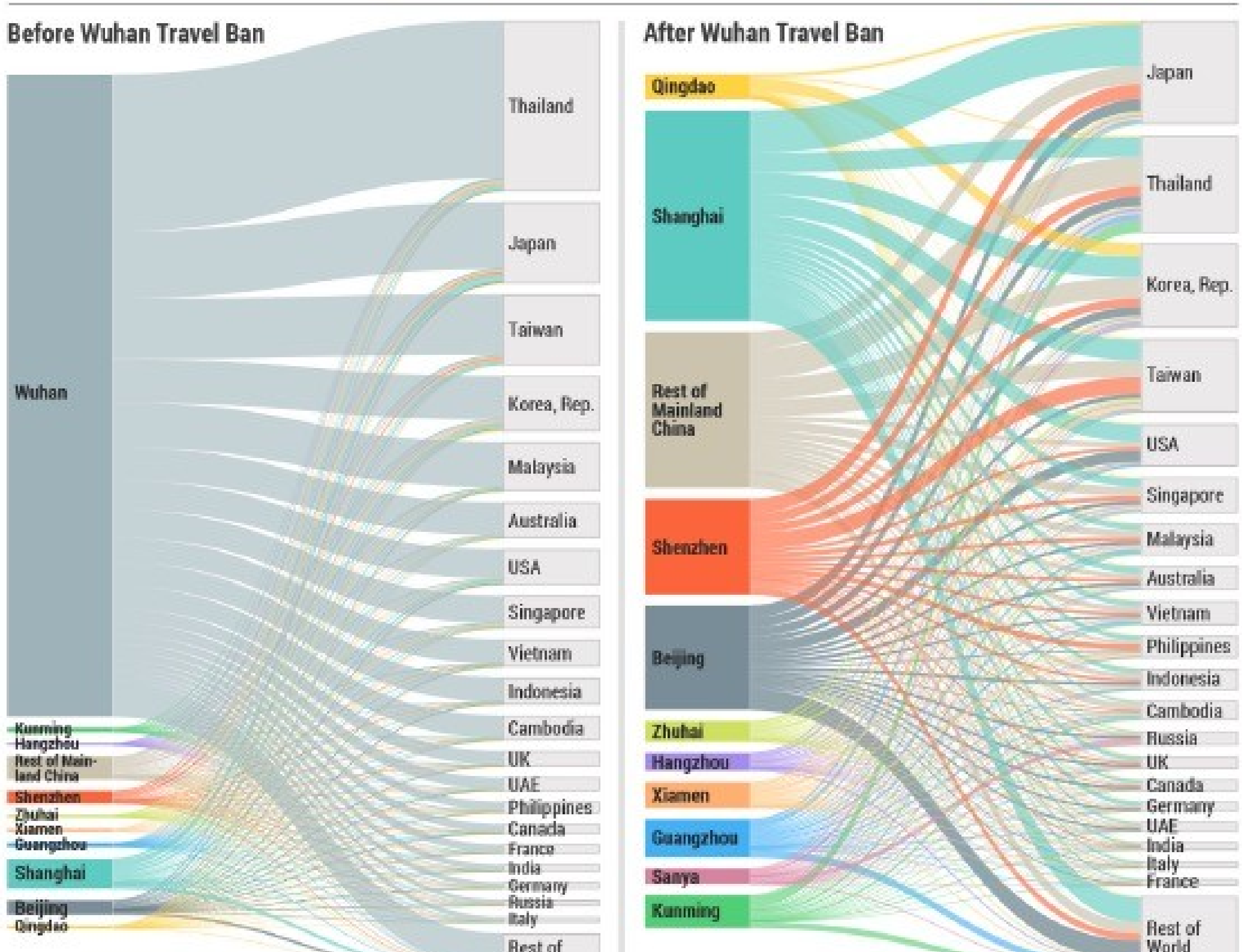
Power grid stability

2006, European power grid, cascading

Complex networks– Functionality 3

COVID-19 epidemic spreading

Will return after few slides



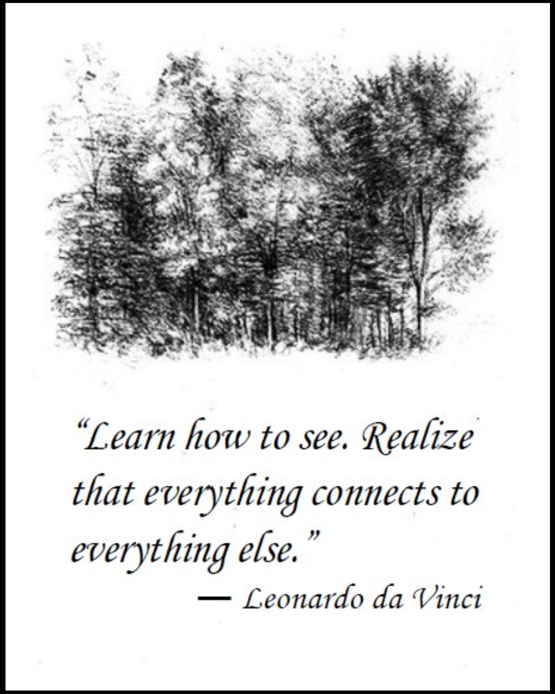
Network Science: Interdisciplinary

Complex Networks is a blend of several disciplines

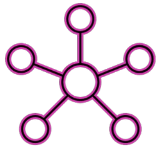
- Mathematics (Random graph theory)
- Statistics (Statistical inference)
- Computer Science (Graph algorithms)
- Physics (Statistical physics, Percolation theory)
- Sociology (Data collection and analysis)

Network Science and Complex Systems

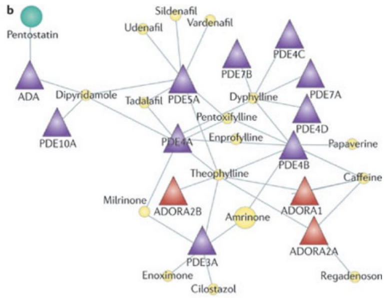
Connections

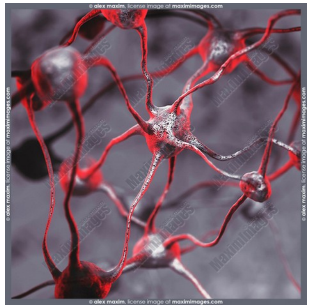


Network Science and Complex Systems



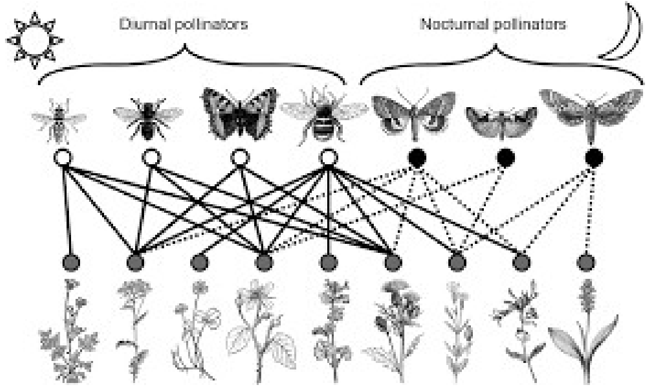
Connections

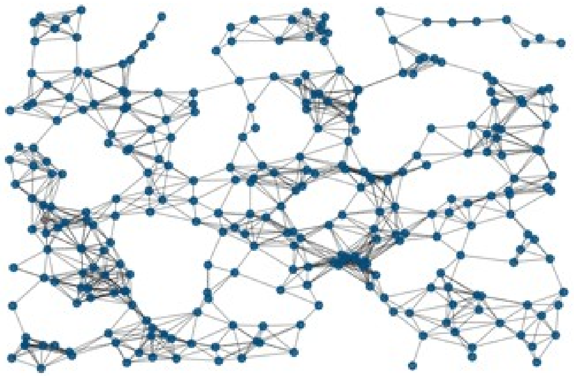




"Learn how to see. Realize that everything connects to everything else."
— *Leonardo da Vinci*

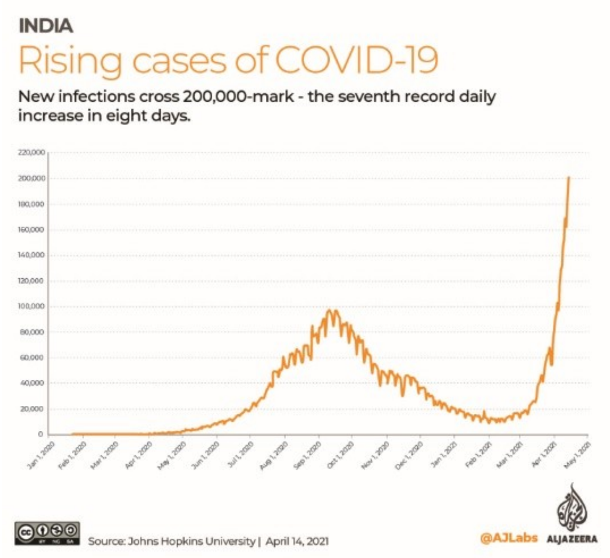
Interactions

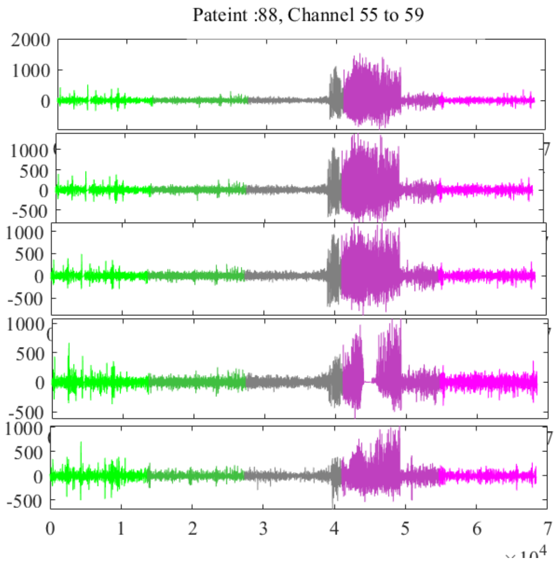


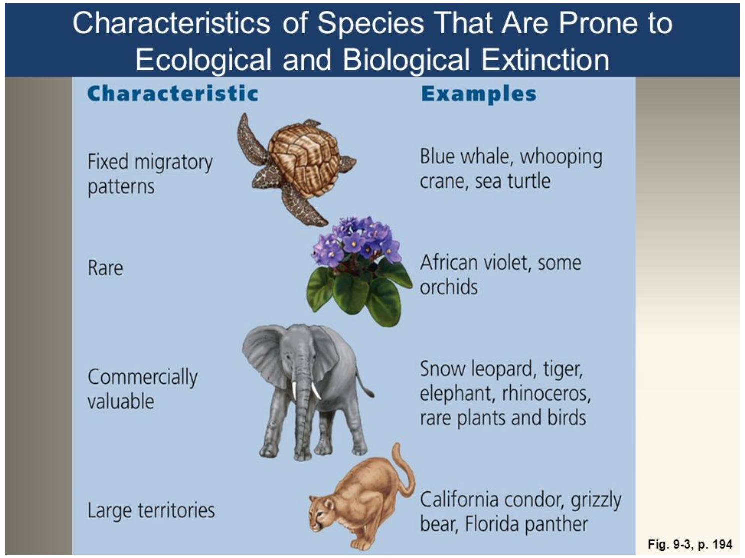


Why Network Science?

Emergence





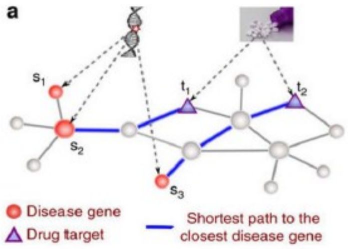


Why Network Science?

Applied sciences

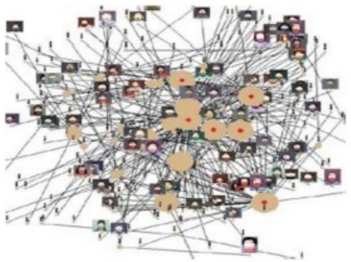
Interconnected systems exist in many applied sciences and other fields. There are numerous studies which show looking at these compex system, as a whole, gives us non trivial insights and is necessary to understand these systems.

Medicine



Disease Gene Network
Credit: Guney et al. (2016)
"the emergence of most diseases cannot be explained by single-gene defects, but involve the breakdown of the coordinated function of distinct gene groups"

Law

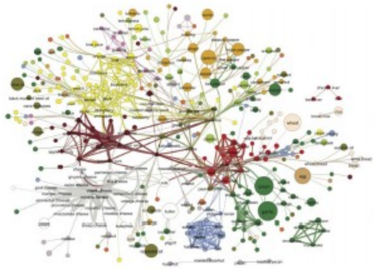


Criminal Network
Credit: Xu et al. (2005)

Economics

Trading Network
Credit: Adamic et al. (2017)
"strong feedback between the trading behaviour in buyers and sellers networks and the market conditions"

Culinary



Flavor Network
Credit: Ahn et al. (2011)

Read on food pairing theories and check out the interactive demo: https://foodgalaxy.jp/

Why Network Science?

Dissecting complex systems/connectivity using network

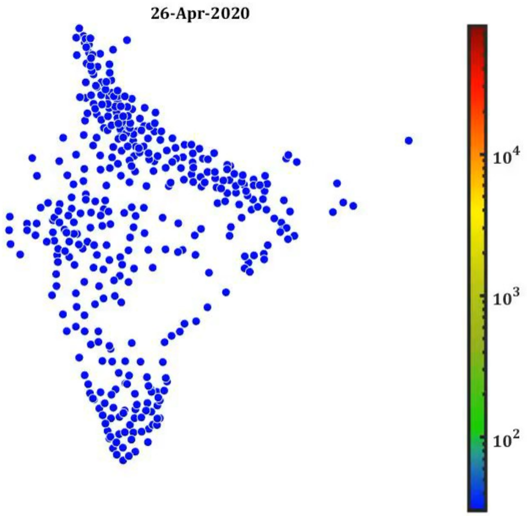
H1N1 pandemic

http://networksciencebook.com/chapter/1#societal-impact

Video courtesy of Alessandro Vespignani.

Epidemic Spreading

Epidemic Spreading



Epidemic Spreading and beyond

- Today epidemic prediction is one of the most active applications of network science, being used to foresee the spread of influenza or to contain Ebola or to understand COVID.

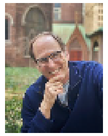
It can be used to model and predict the spread of biological, digital and social viruses (memes).
The impact of these advances are felt beyond epidemiology!!!

Indeed, in January 2010 network science tools have predicted the conditions necessary for the emergence of viruses spreading through mobile phones.

D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. BMC Medicine, 7: 45, 2009.

L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. PNAS, 101: 15124, 2004.

P. Wang, M. Gonzalez, C. A. Hidalgo, and A.-L. Barabási. Understanding the spreading patterns of mobile phone viruses. Science, 324: 1071, 2009.



Steven H. Strogatz

Schurman Professor of Applied Mathematics, Cornell University

synchronization
complex systems
networks
applied mathematics
nonlinear dynamics

	All	Since 2018
Citations	123206	42379
h-index	85	56
i10-index	169	125

0 articles — 33 articles

not available — available

Based on funding mandates

TITLE	CITED BY	YEAR
Collective dynamics of 'small-world' networks DJ Watts, SH Strogatz Nature 393, 440-442	50657	1998
Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering SH Strogatz Westview Press	16434 *	2001
Exploring complex networks SH Strogatz nature 410 (6825), 268-276	9824	2001
Random graphs with arbitrary degree distributions and their applications MEJ Newman, SH Strogatz, DJ Watts Physical review E 64 (2), 026118	4605	2001
Sync: The emerging science of spontaneous order S Strogatz Penguin UK	3655 *	2004
From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators SH Strogatz Physica D: Nonlinear Phenomena 143 (1-4), 1-20	3209	2000
Network robustness and fragility: Percolation on random graphs DS Callaway, MEJ Newman, SH Strogatz, DJ Watts Physical review letters 85 (25), 5468	2878	2000
Synchronization of pulse-coupled biological oscillators RE Mirollo, SH Strogatz SIAM Journal on Applied Mathematics 50 (6), 1645-1662	2618	1990
Random graph models of social networks MEJ Newman, DJ Watts, SH Strogatz Proceedings of the national academy of sciences 99 (suppl_1), 2566-2572	1763	2002
Synchronization of Lorenz-based chaotic circuits with applications to communications KM Cuomo, AV Oppenheim, SH Strogatz IEEE Transactions on circuits and systems II: Analog and digital signal ...	1625	1993
Chimera states for coupled oscillators DM Abrams, SH Strogatz	1379	2004

Paul Erdős

Mathematics

number theory
combinatorics
probability
set theory
mathematical analysis

Citatior
h-index
i10-ind

TITLE

- On the evolution of random graphs
  P Erdos, A Rényi
  Bull. Inst. Internat. Statist 38 (4), 343-347
- On random graphs I.
  P ERDdS, A R&WI
  Publ. Math. Debrecen 6, 290-297
- Publ. of the Math
  P Erdos, A Rényi
  Inst. of the Hung. Acad. of Sci 5, 17-61
- On the existence of a factor of degree one of a connected random gr
  P Erdős, A Rényi
  Acta Mathematica Hungarica 17 (3), 359-368
- Intersection theorems for systems of sets
  P Erdős, R Rado
  Journal of the London Mathematical Society 1 (1), 85-90
- A combinatorial problem in geometry
  P Erdős, G Szekeres
  Compositio Mathematica 2, 463-470
- A combinatorial problem
  NG de Bruijn, P Erdos
  Koninklijke Nederlandse Akademie v. Wetenschappen 49 (49), 758-764
- Intersection theorems for systems of finite sets
  P Erdos, C Ko, R Rado
  The Quarterly Journal of Mathematics 12 (1), 313-320
- Graph theory and probability
  P Erdős
  Classic Papers in Combinatorics, 276-280
- Graph theory and probability, II
  P Erdős
  Canad. J. Math 13, 346-352
- Problems and results on 3-chromatic hypergraphs and some related questions
  P Erdos, L Lovász
  Infinite and finite sets 10, 609-627
- On the strength of connectedness of a random graph
  P Erdős, A Rényi
  Acta Mathematica Hungarica 12 (1), 261-267

Albert-László Barabási

Northeastern University
, Harvard Medical School

network science
statistical physics
biological physics
physics
medicine

	All	Since 2020
Citations	312614	89106
h-index	171	109
i10-index	411	312

5 articles — 134 articles

not available — available

Based on funding mandates

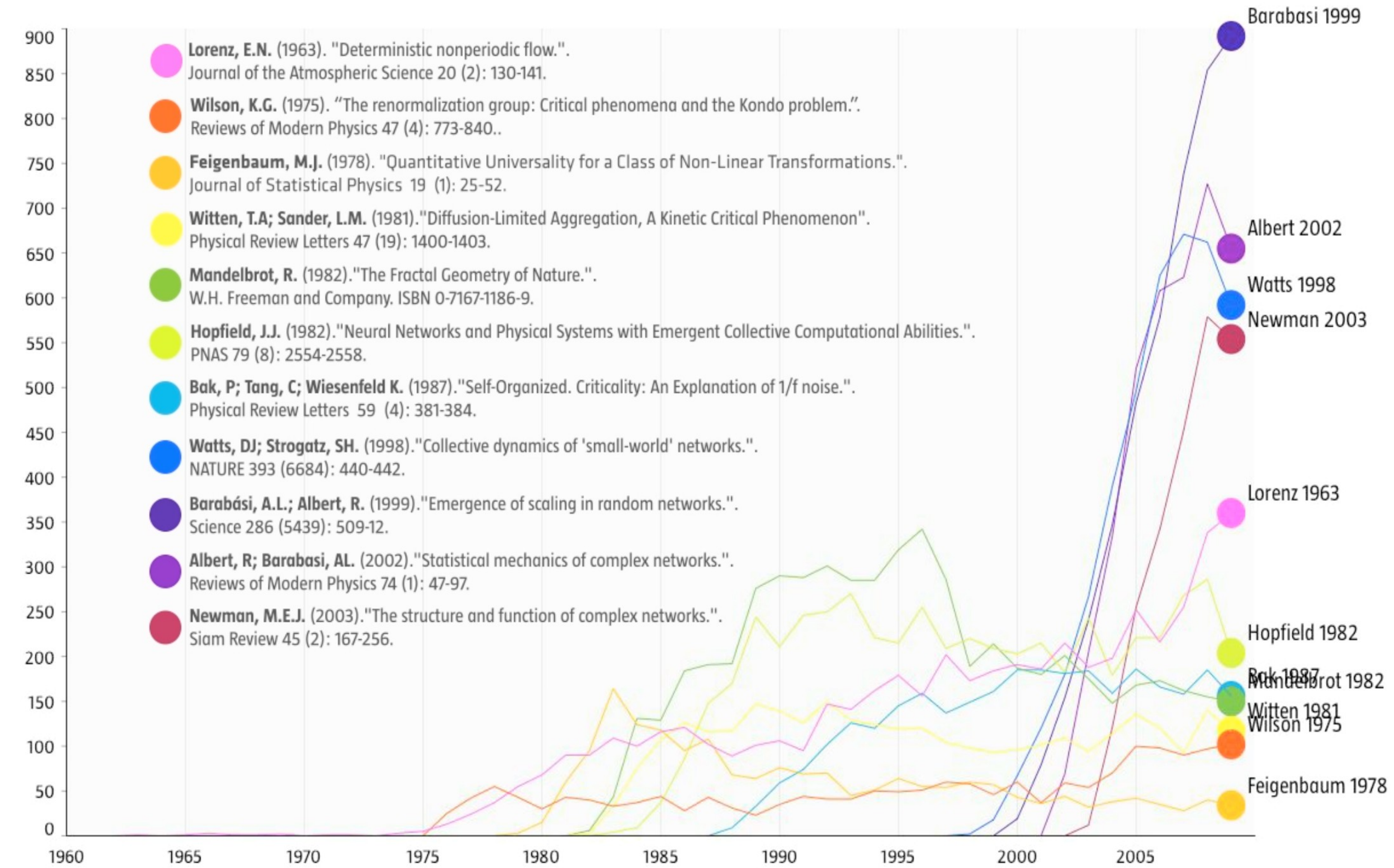
Journals' Rankings BETA — Report a Bug · Options · Vote for Next Features

Q1 94% 17 — NA 6% 1

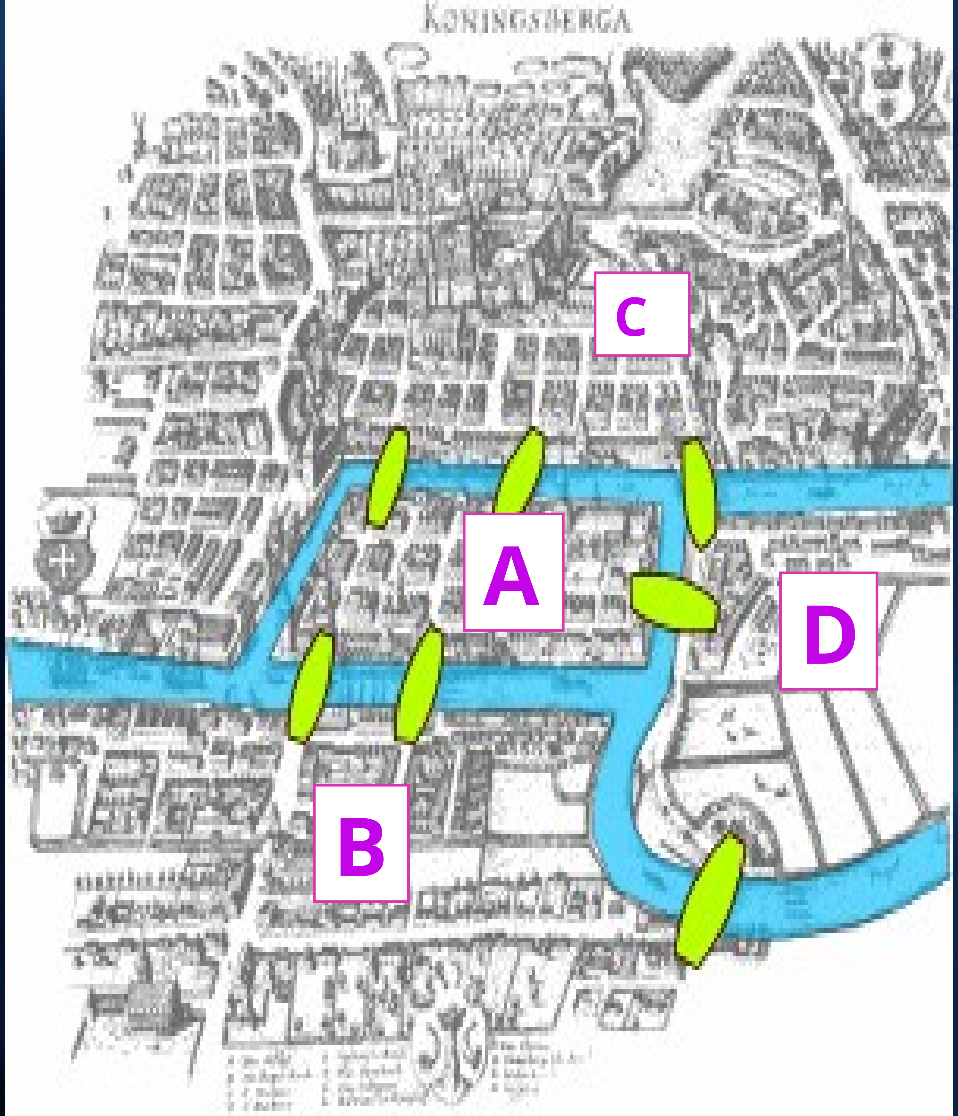
— based on 18 most cited papers in the last 10 years — ExCITATION

TITLE	CITED BY	YEAR
Emergence of scaling in random networks AL Barabási, R Albert Q1 NAA Science 286 (5439), 509-512	48708	1999
Statistical mechanics of complex networks R Albert, AL Barabasi Q1 NAA Reviews of Modern Physics 74, 47-97	29061	2002
Linked: The New Science Of Networks AL Barabási NA Basic Books	11723 *	2002
Error and attack tolerance of complex networks R Albert, H Jeong, AL Barabási Q1 NAA Nature 406, 378–482	11199	2000
Network biology: understanding the cell's functional organization AL Barabasi, ZN Oltvai Q1 NAA Nature reviews genetics 5 (2), 101-113	10132	2004
Understanding individual human mobility patterns MC Gonzalez, CA Hidalgo, AL Barabasi Q1 NAA nature 453 (7196), 779-782	7846	2008
Network science AL Barabási NA Cambridge University Press	7311	2016

Network in research domain - Acceptance

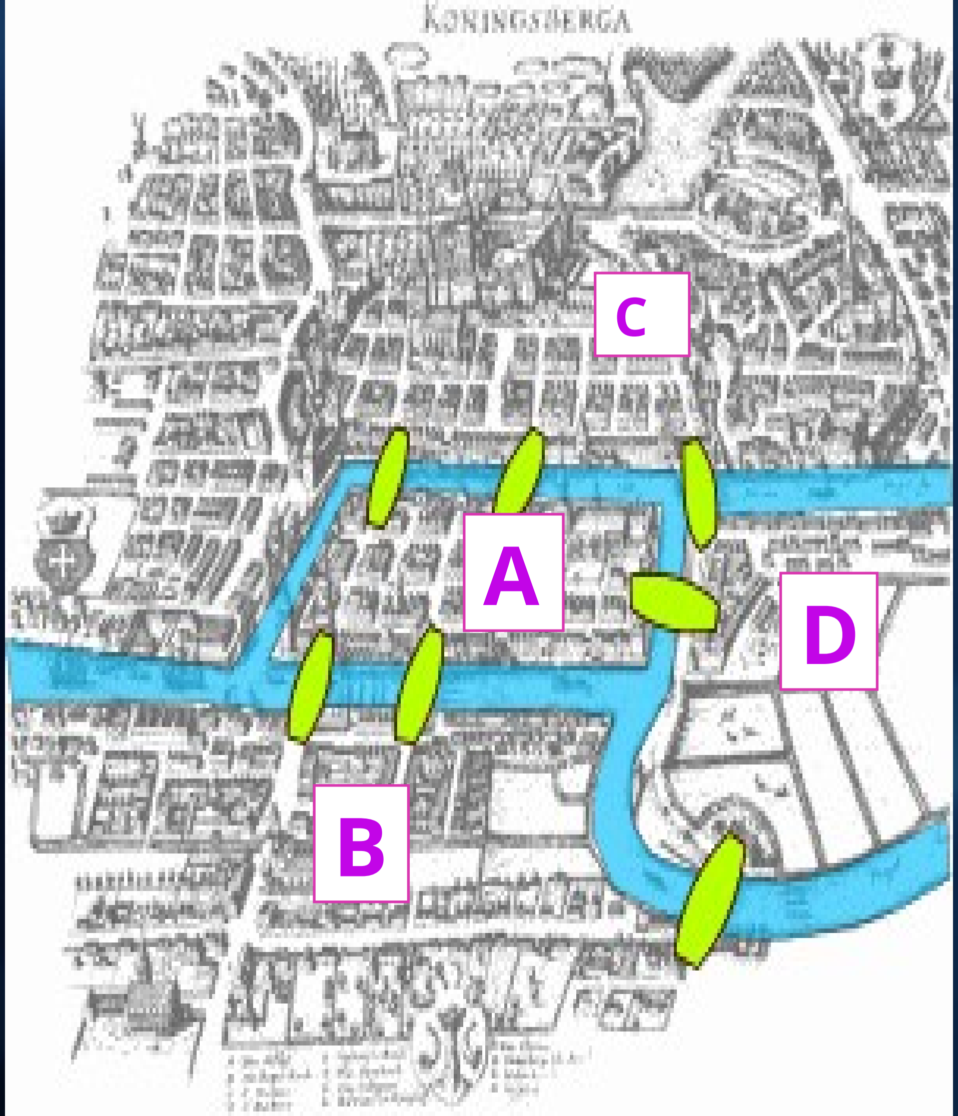


The 7 bridges of Königsberg



- Stories say that the people of Königsberg used to walk around the city on Sundays. While walking, they decided to create a game to play:
  - The initial puzzle requested a person to walk a path which crossed all bridges, crossing each bridge once and only once, and to finish at the starting point

The 7 bridges of Königsberg



- Stories say that the people of Königsberg used to walk around the city on Sundays. While walking, they decided to create a game to play:
  - The initial puzzle requested a person to walk a path which crossed all bridges, crossing each bridge once and only once, and to finish at the starting point
  - After many days and many trials, the puzzle was adjusted so that the finish point could be on any landmass, with the condition of crossing every bridge once and only once still holding

Dynamical Processes in Complex Networks(SC1.440)

- *Boccaletti, Stefano, et al. "Complex networks: Structure and dynamics."Phyics reports, 424.4-5 (2006): 175-308.*
- *Arenas, Alex, et al. "Synchronization in complex networks." Physics reports 469.3 (2008): 93-153.*
- *Pastor-Satorras, Romualdo, et al. "Epidemic processes in complex networks." Reviews of modern physics 87.3 (2015): 925.*
- *Perra, Nicola. "Non-pharmaceutical interventions during the COVID-19 pandemic: A review." Physics Reports 913 (2021): 1-52.*

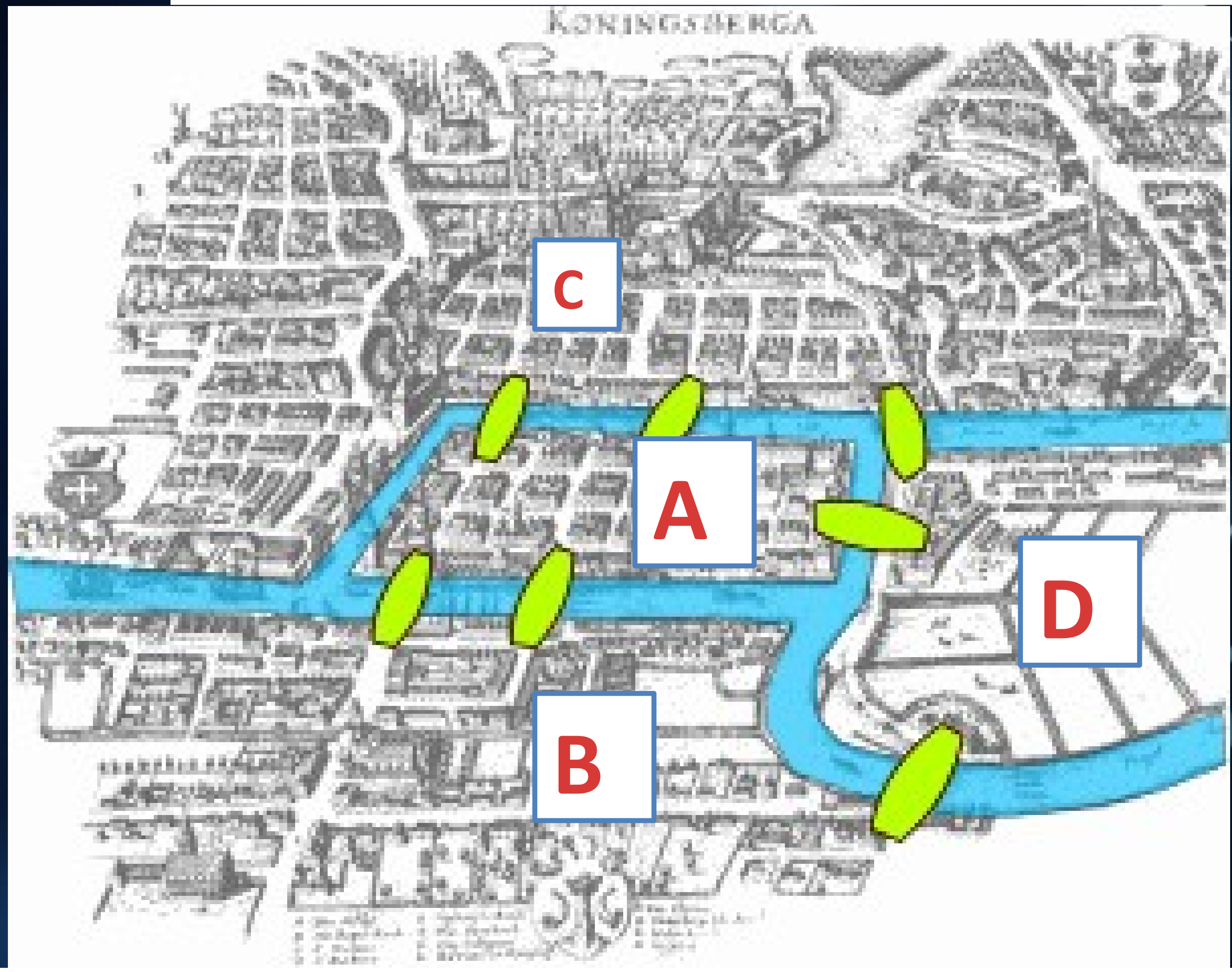
- *http://www.network-science.org/*
- *https://www.barabasilab.com/science*
- *https://www.cpt.univ-mrs.fr/~barrat/english.html*
- *https://www.mobs-lab.org/alessandro-vespignani.html*
- *https://www.stevenstrogatz.com/*
- *http://www.gitta.info/Accessibiliti/en/html/StructPropNetw_learningObject3.html*
- *https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/types-of-biological-networks/*

- *https://twitter.com/barabasi*

Dynamical Processes in Complex Networks(SC1.440): Evaluations

Type of Evaluation	Weightage (in %)
Quiz (August end) +(Surprise quiz+Assignment)	**17 +8**
Mid Sem Exam (September end)	**30**
Quiz (October end) + (Surprise quiz) +Assignment	**17 +8**
Project	**20**

The 7 bridges of Königsberg



The city of Königsberg, Germany, is divided by the Pregal River. The river forks and rejoins itself before forking off again.

4 land masses and 7 bridges

Complex networks: Outline

- *Mathematical representation of Networks: Adjacency matrix.*
- *Types of Networks: Directed/Undirected/Weighted/Regular*

❑ *Structural Characterization: Degree (in, out)*

❑ *In-degree, Out-degree, application*

❑ *Handshaking Lemma*

❑ *Average Degree, Link Density*

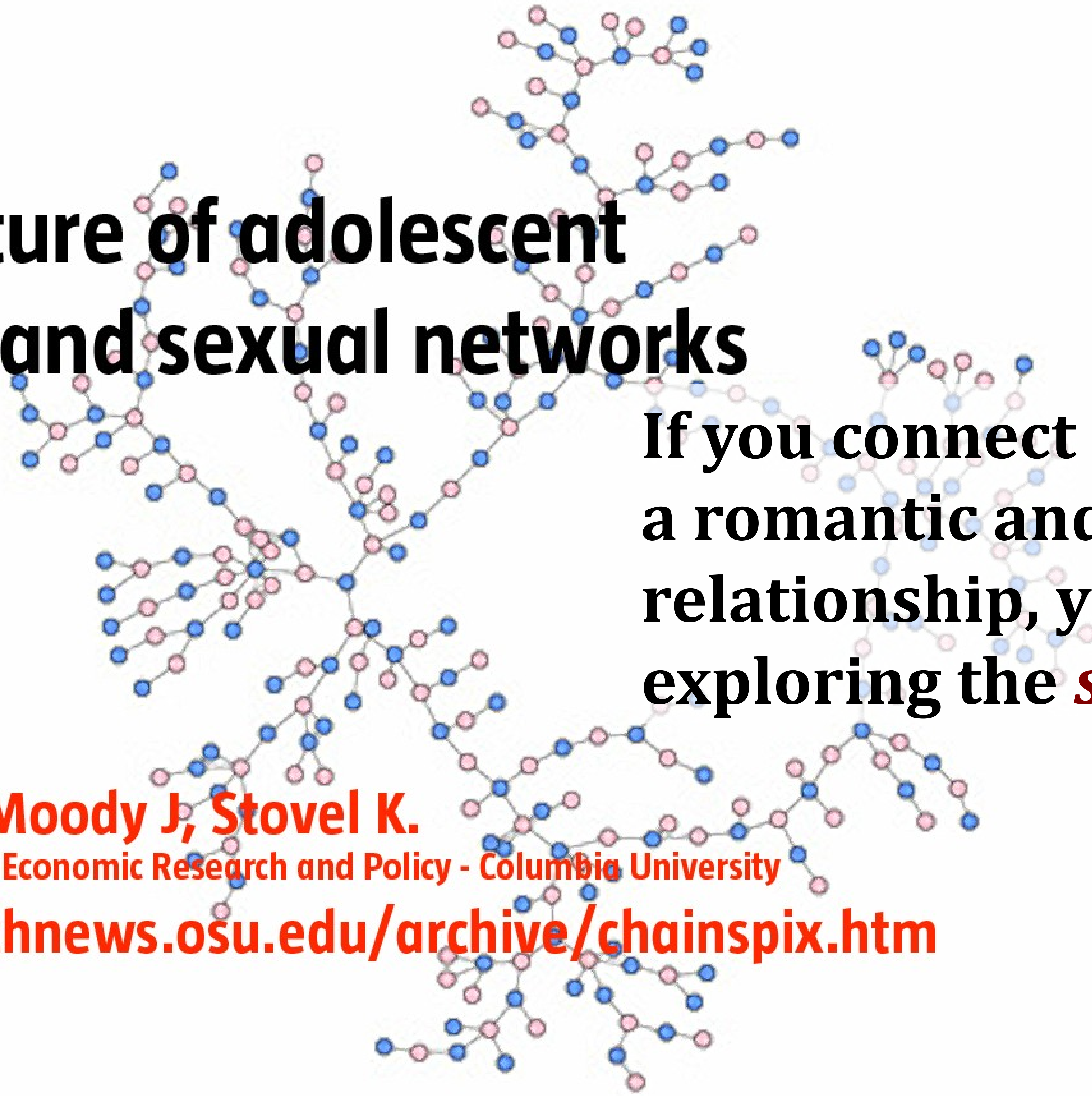
- *Walk, Trail, Path, Cycle, Shortest Path*

❑ *Eulerian and Hamiltonian Graphs*

❑ *Shortest Path, Diameter*

❑ *BFS search*

CHOOSING A PROPER REPRESENTATION



The structure of adolescent romantic and sexual networks

If you connect those that have a romantic and sexual relationship, you will be exploring the *sexual networks.*

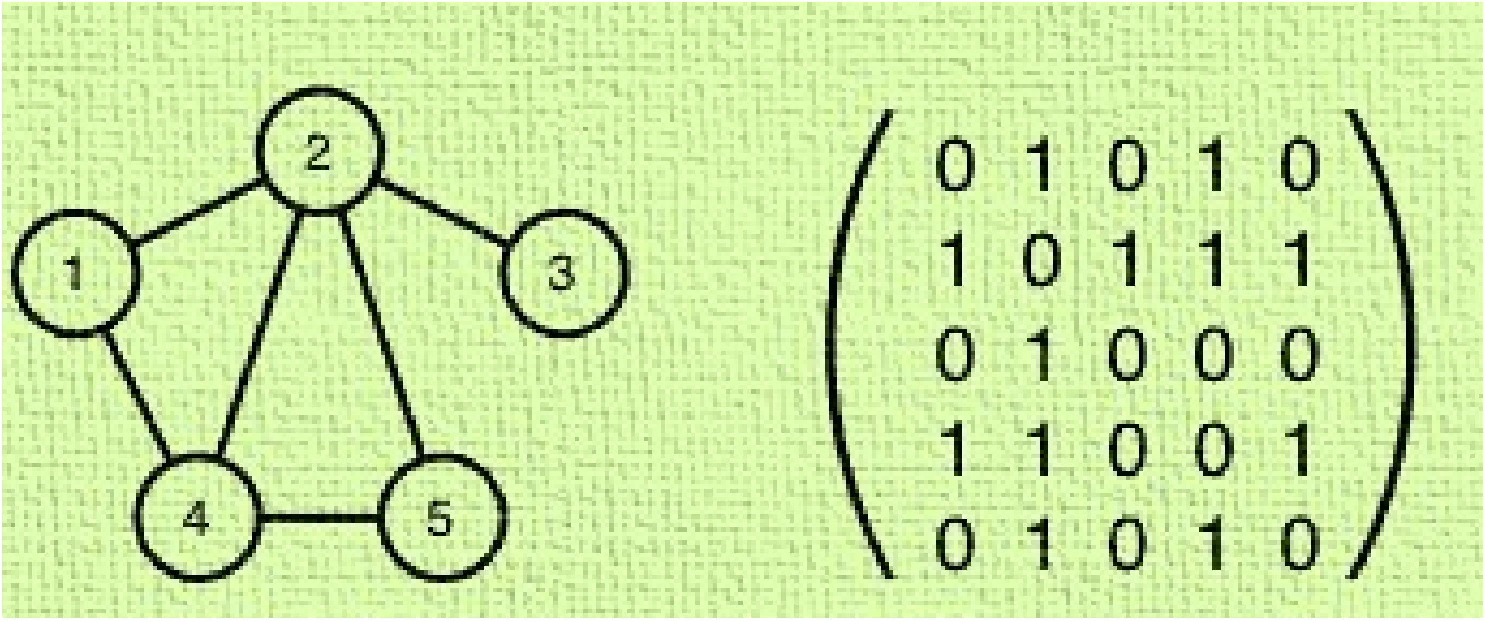
Bearman PS, Moody J, Stovel K.
Institute for Social and Economic Research and Policy - Columbia University

http://researchnews.osu.edu/archive/chainspix.htm

Degree (Structural Chracterization)

In an undirected network the degree k_i of a node i is the number of nodes connected to :

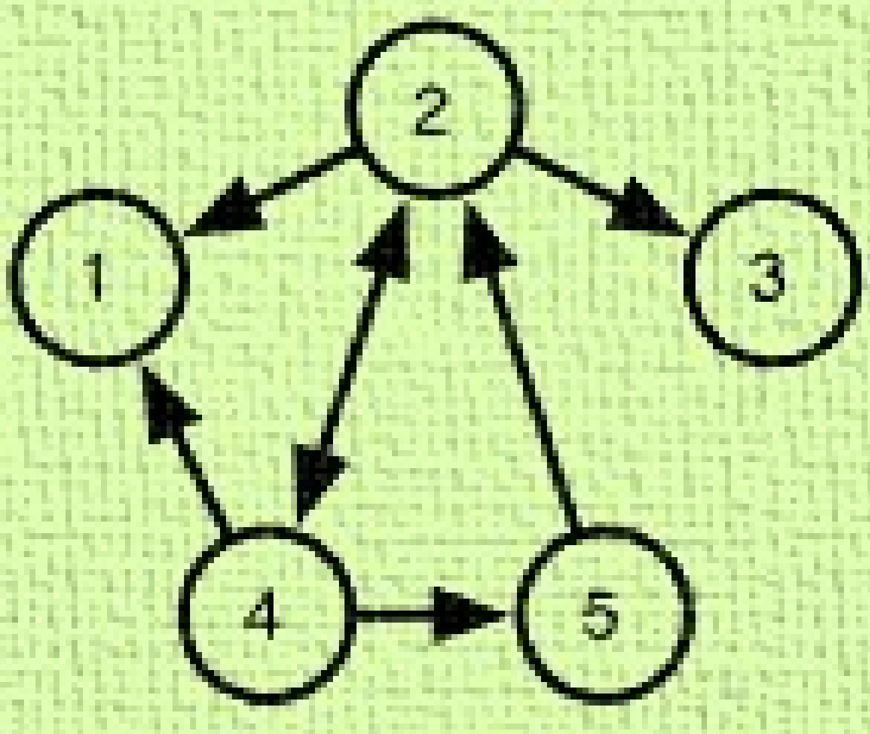
$$k_i = \Sigma_j a_{ij} = \Sigma_j a_{ji}$$



Here $k_1 = 2$, $k_2 = 4$, $k_3 = 1$, $k_4 = 3$ and $k_5 = 2$.

In-Degree and Out-degree

Thus, in a directed network, nodes can be highly connected, yet also isolated (e.g. in terms of sending or receiving information.)



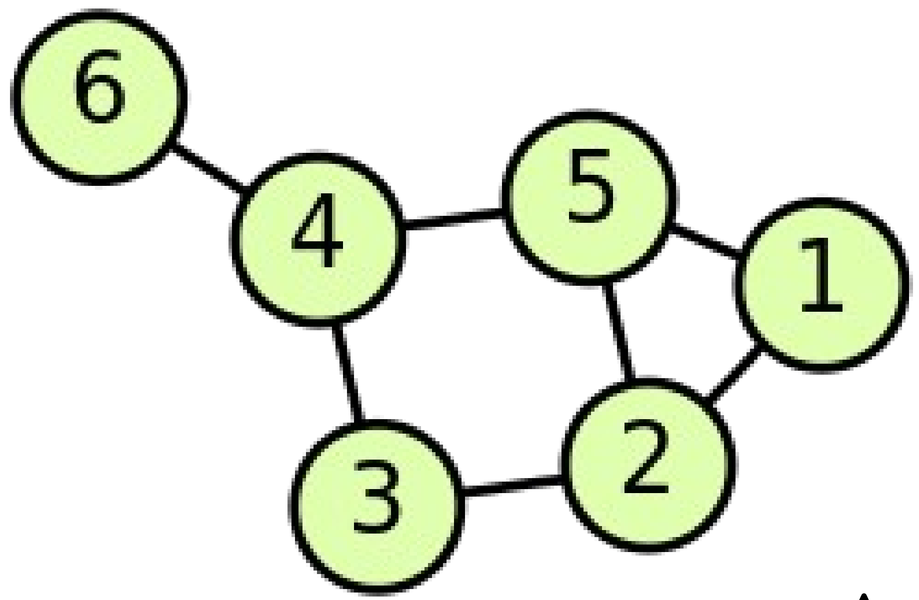
$$
\begin{array}{c c}
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0
\end{pmatrix}
&
\begin{array}{c}
\text{out-degree:} \\ \mathbf{0} \\ \mathbf{3} \\ \mathbf{0} \\ \mathbf{3} \\ \mathbf{1}
\end{array}
\\
\text{in-degree: } \mathbf{2}\ \ \mathbf{2}\ \ \mathbf{1}\ \ \mathbf{1}\ \ \mathbf{1} &
\end{array}
$$

Handshaking lemma

- *If we add up the degrees of all vertices of a graph the result is an even number -> twice the number of edges.*

$$\sum_{v \in V} \deg(v) = 2|E|$$

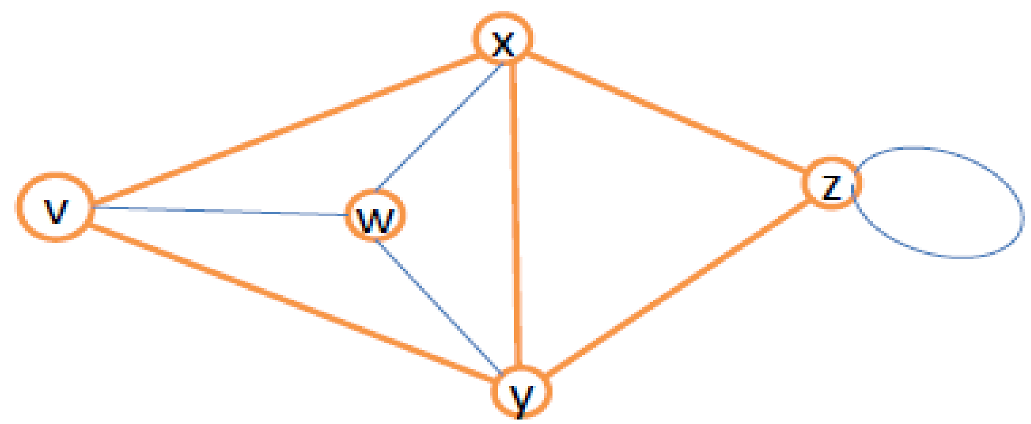
for a graph with vertex set V and edge set E



Walk, Trail, Path, Cycle

Walk: *any route through a graph from vertex to vertex along edges*

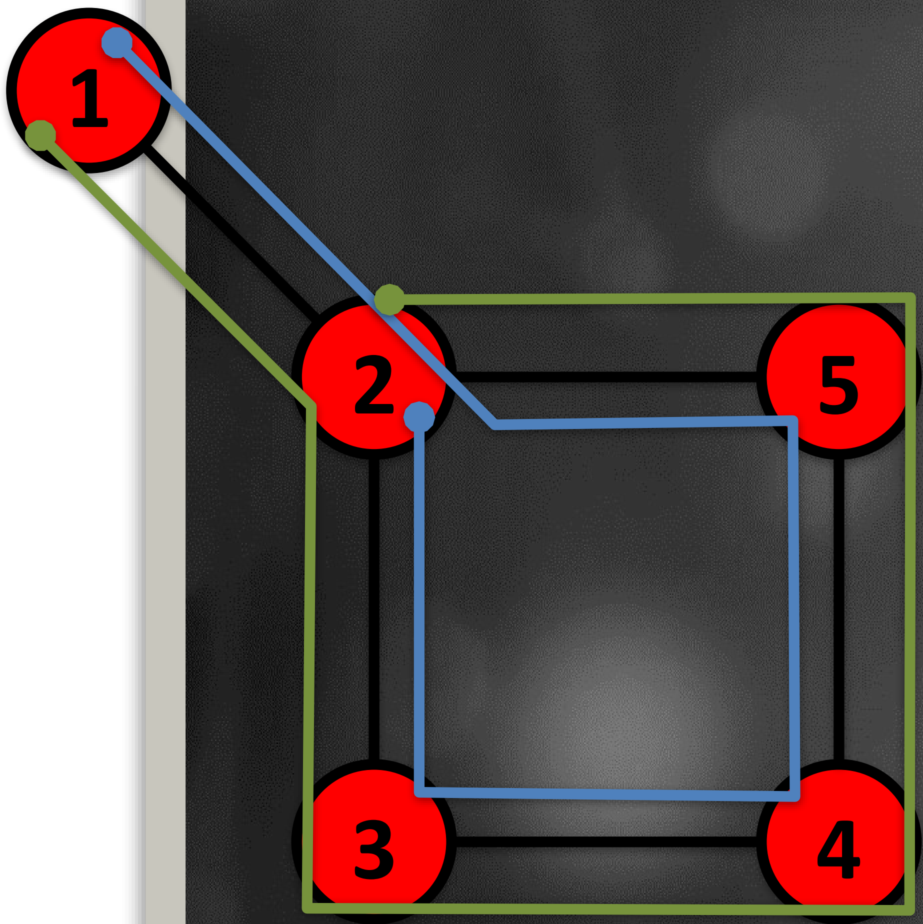
Trail: *A trail is a walk that does not pass over the same edge twice. A trail might visit the same vertex twice, but only if it comes and goes from a different edge each time.*



Trail: v->w->x->y->z->z->x

Eulerian and Hamiltonian Graph

Eulerian Trail



A path that traverses each link exactly once.

The criterion for Eulerian trail

Suppose that a graph has an Euler *trail* P.

For every vertex v other than the starting and ending vertices, the path P enters v the same number of times that it leaves v (say $\boxed{s}$ times).

Therefore, there are $\boxed{2s}$ edges having v as an endpoint.

The criterion for Eulerian trail

Suppose that a graph has an Euler *trail* P.

For every vertex v other than the starting and ending vertices, the path P enters v the same number of times that it leaves v (say $\boxed{s}$ times).

Therefore, there are $\boxed{2s}$ edges having v as an endpoint.

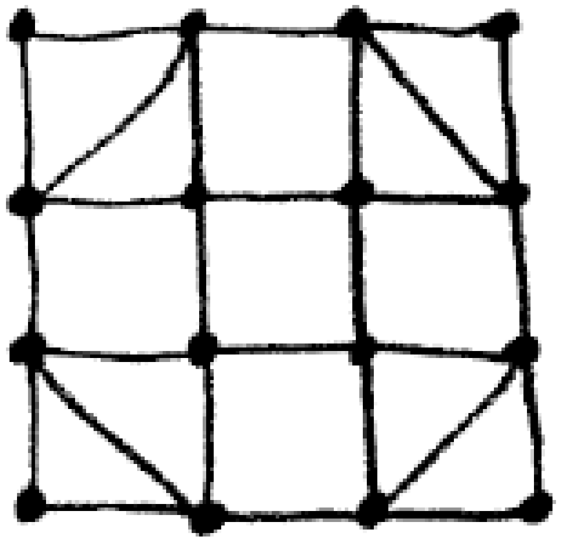
Therefore, all vertices other than the two endpoints of P must be even vertices.

The criterion for Eulerian trails

Suppose the Euler path P starts at vertex x and ends at y.

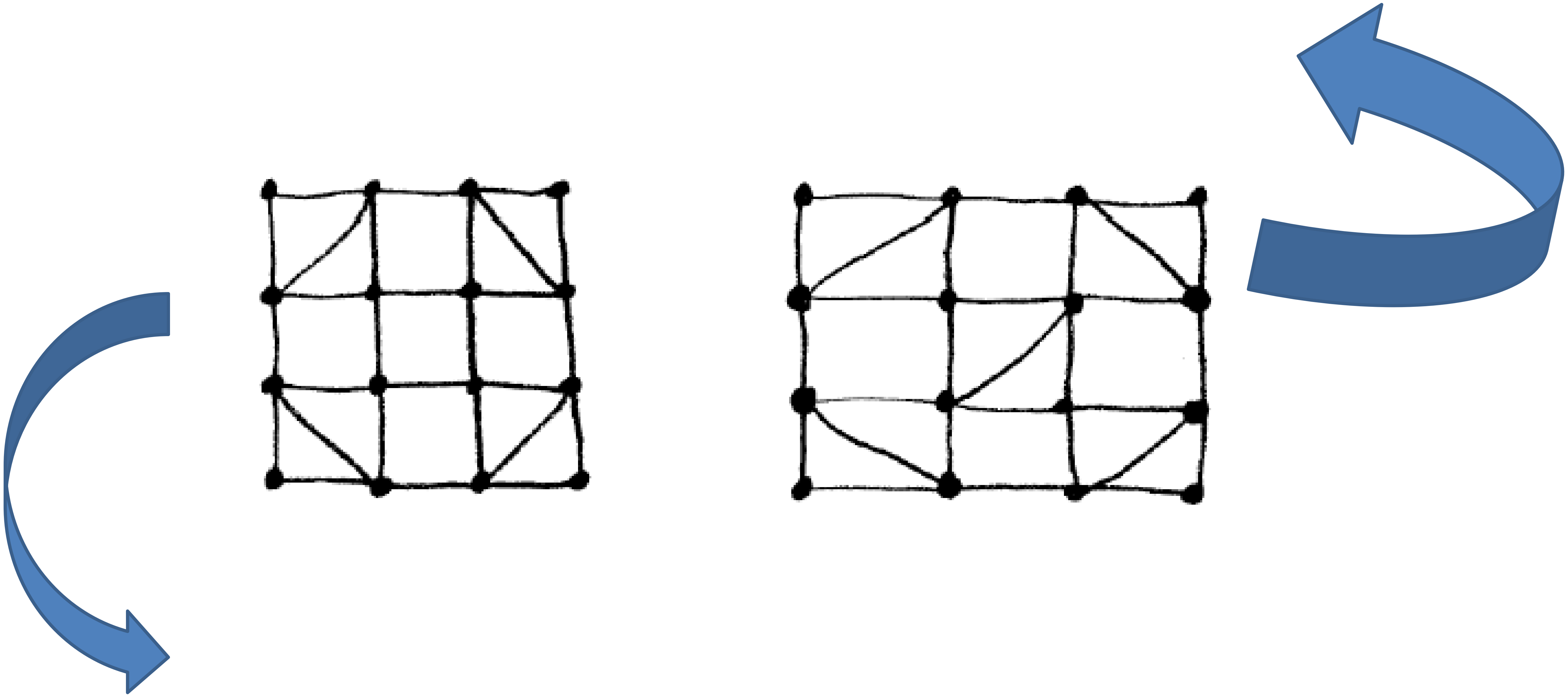
Then P leaves x one more time than it enters, and leaves y one fewer time than it enters.

Therefore, **the two endpoints of P must be odd vertices.**



- *A connected graph G is Eulerian (circuit) if and only if the degree of every vertex of G is even.*

- *A connected graph is Eulerian (trail) but not Eulerian if and only if it has exactly two vertices of odd degree.*

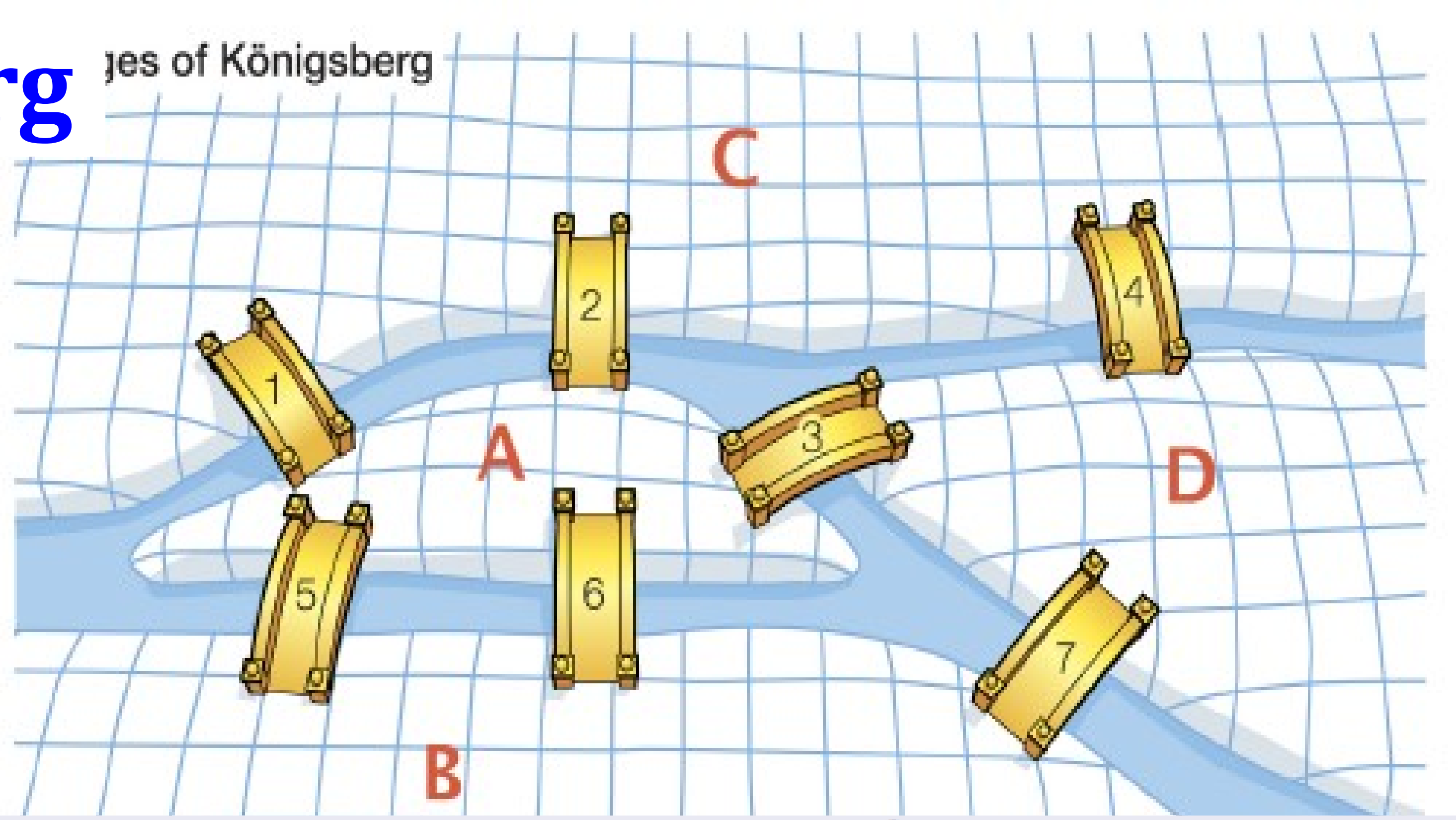


- *A connected graph G is Eulerian (circuit) if and only if the degree of every vertex of G is even.*

- Does **every** graph with **zero** odd vertices have an Euler circuit?
- Does **every** graph with **two** odd vertices have an Euler trail?
- Is it possible for a graph have just **one** odd vertex?

The 7 bridges of Königsberg

A-5 Edges; B-3 Edges
C-3 Edges; D-3 Edges



Euler gave the three guidelines that someone can use to figure out if a path exists using each bridge once and only once.

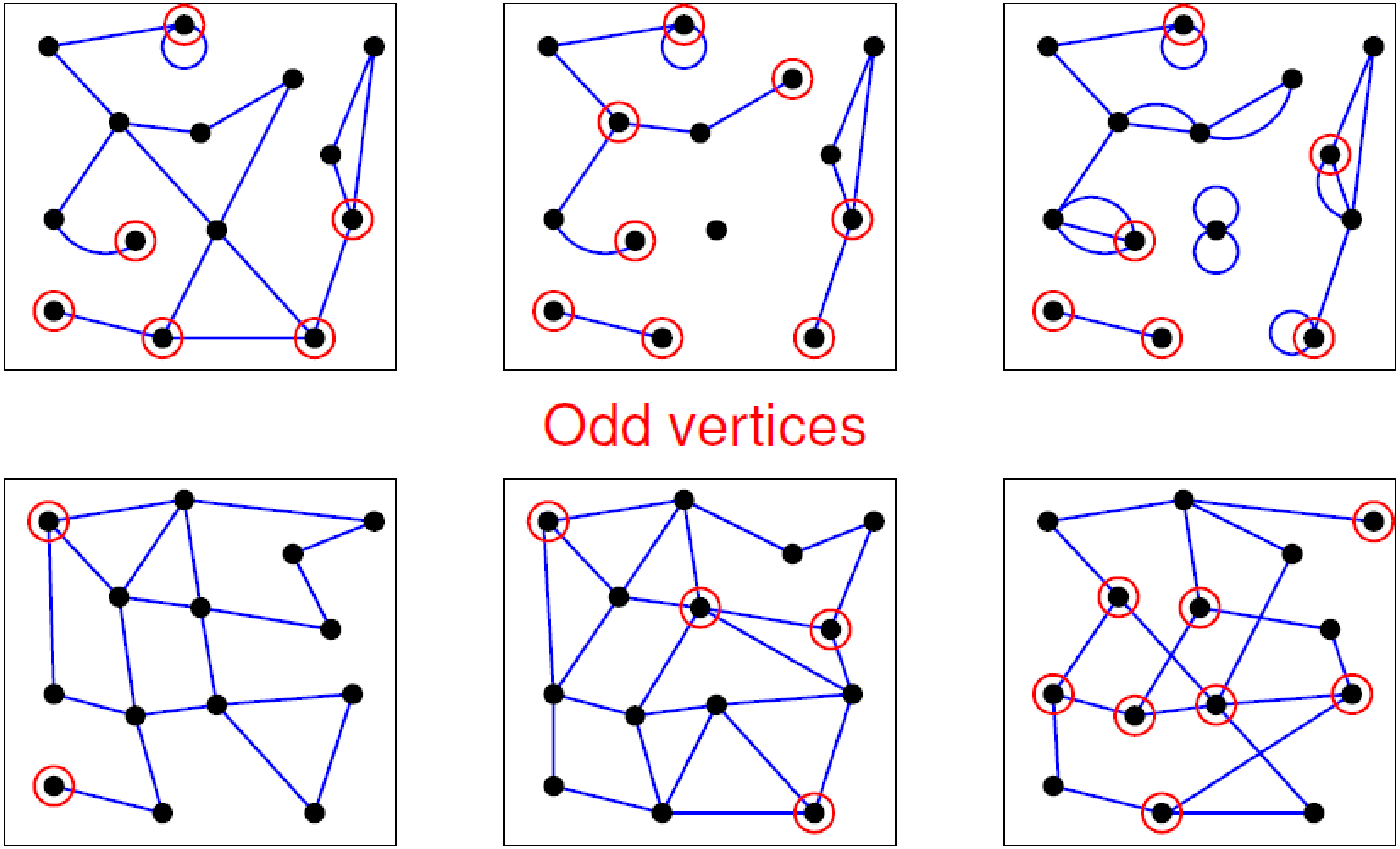
1. No journey is possible if there are **more than 2 odd landmasses**. This solves the Bridges of KönigsbergProblem since it contains 4 odd landmasses!
2. A journey is possible if there are **exactly two odd landmasses**. One can **start** in either one of the two odd landmasses and **end** on the other odd landmass.
3. A journey is possible if there are **no odd** landmasses. One starts and ends at any landmass in this case.

How to find Eulerian trail

To find an Euler trail **or an Euler circuit:**

1. Make sure the graph has either 0 or 2 odd vertices.
2. If there are 0 odd vertices, start anywhere. If there are 2 odd vertices, start at one of them.
3. Follow edges one at a time. If you have a choice between a bridge and a non-bridge, always choose the non-bridge.
4. Stop when you run out of edges.

This is called **Fleury's algorithm**, and it always works!



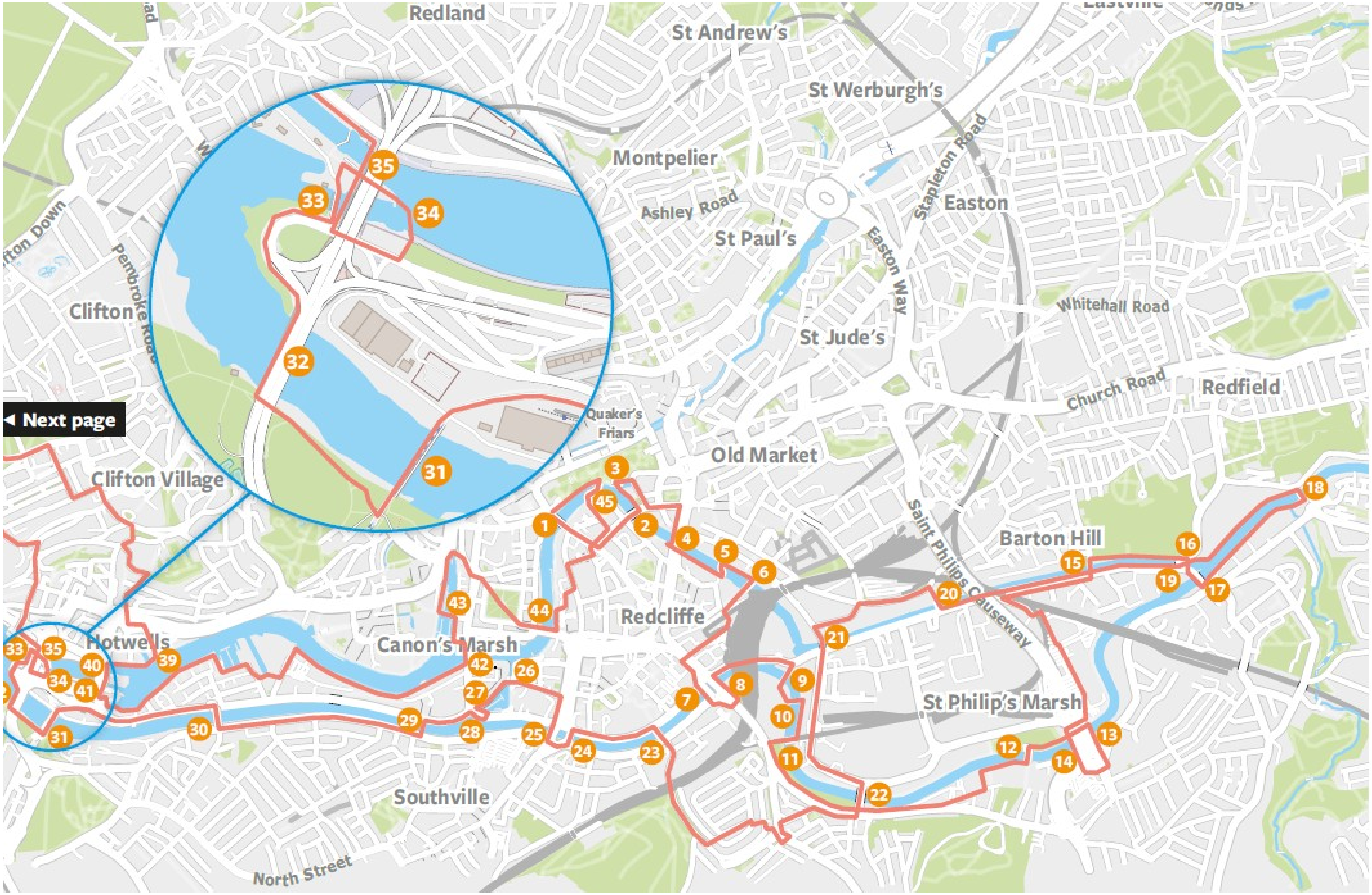
Odd vertices

THE BRISTOL BRIDGES WALK

Crossing each one only once

Query: *It doesn't matter which bridge you start at, or which direction you go in, the walk will bring you back to where you started.*

the book "From Brycgstow to Bristol in 45 Bridges" by Jeff Lucas and Thilo Gross



Redland
St Andrew's
St Werburgh's
Montpelier
Ashley Road
Stapleton Road
Easton
St Paul's
Easton Way
Whitehall Road
Clifton
Pembroke Road
St Jude's
Church Road
Redfield
Next page
Quaker's Friars
Old Market
Clifton Village
Saint Philips Causeway
Barton Hill
Redcliffe
Hotwells
Canon's Marsh
St Philip's Marsh
Southville
North Street

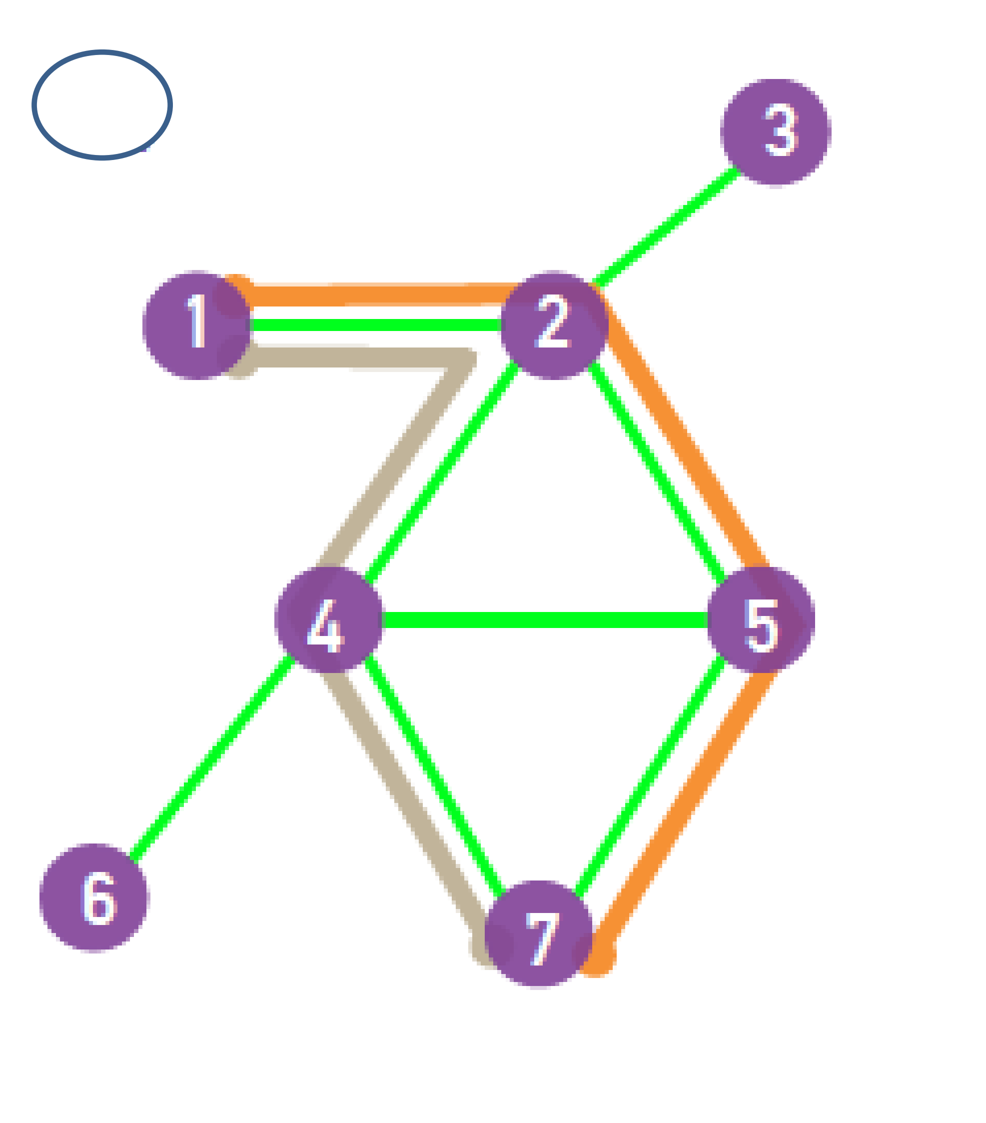
Path Length (Structural Characterization)

Physical distance vs Network Distance

Physical distance plays a key role in determining the interactions between the components of physical systems.
For example, the distance between two atoms in a crystal or between two galaxies in the universe determines the forces that act between them.

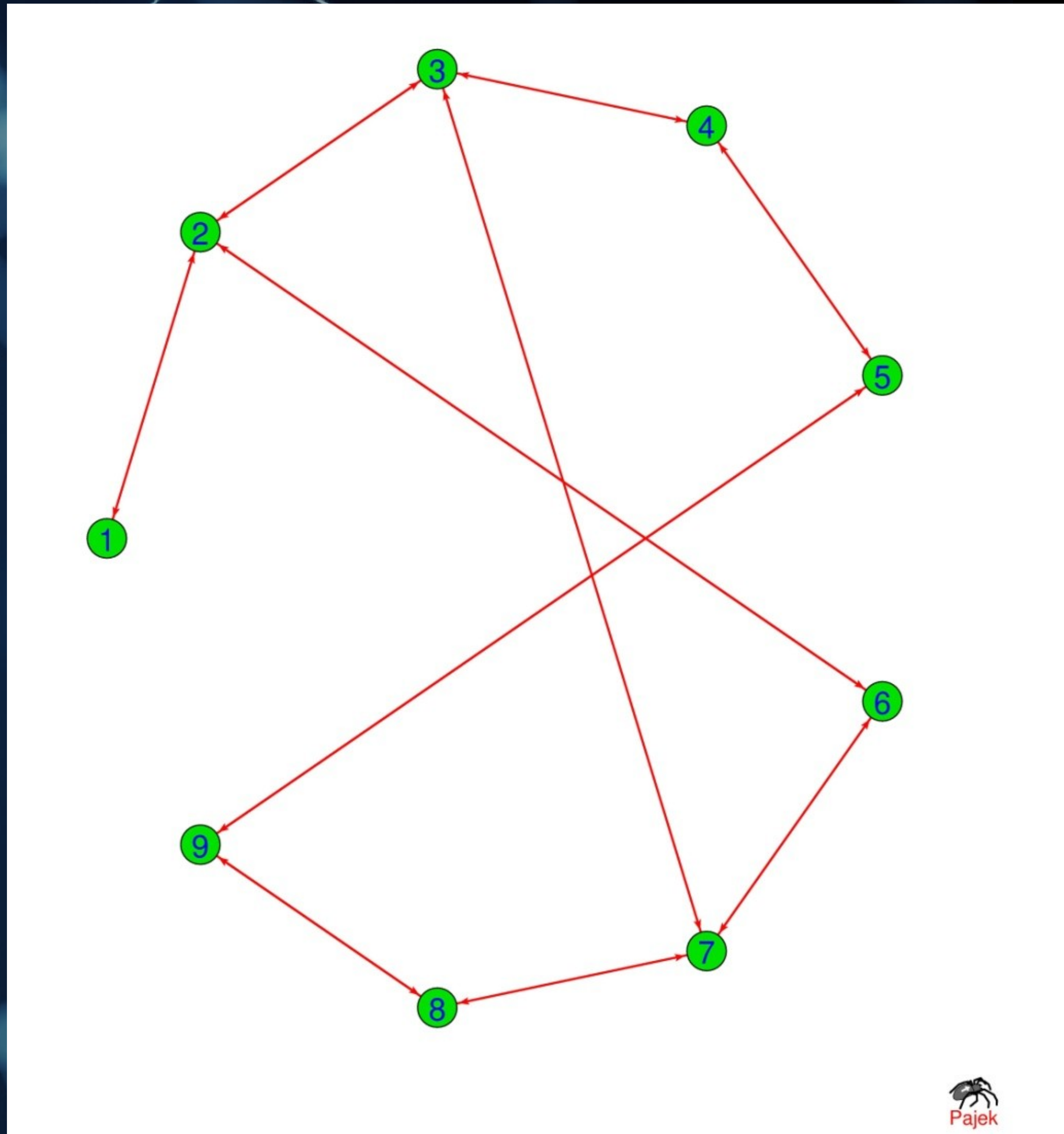
Shortest Paths

The shortest path between nodes i and j is the path with the fewest number of links The shortest path is often called the distance between nodes i and j, and is denoted by , or simply d.



Adjacency matrix

	1	2	3	4	5	6	7	8	9
1	0	1	0	0	0	0	0	0	0
2	1	0	1	0	0	1	0	0	0
3	0	1	0	1	0	0	1	0	0
4	0	0	1	0	1	0	0	0	0
5	0	0	0	1	0	0	0	0	1
6	0	1	0	0	0	0	1	0	0
7	0	0	1	0	0	1	0	1	0
8	0	0	0	0	0	0	1	0	1
9	0	0	0	0	1	0	0	1	0



MatLab script

Shortest Path : Additional notes

$d_{ij} = 2$: If there is a path of length two between i and j, then $A_{ik} A_{kj} = 1$ ($A_{ik} A_{kj} = 0$ otherwise). The number of $d_{ij} = 2$ paths between i and j is

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik} A_{kj} = \left[A^2\right]_{ij}$$

where $[\ldots]_{ij}$ denotes the $(ij)^{th}$ element of a matrix.

$d_{ij} = d$: If there is a path of length d between i and j, then $A_{ik} \ldots A_{lj} = 1$ ($A_{ik} \ldots A_{lj} = 0$ otherwise). The number of paths of length d between i and j is

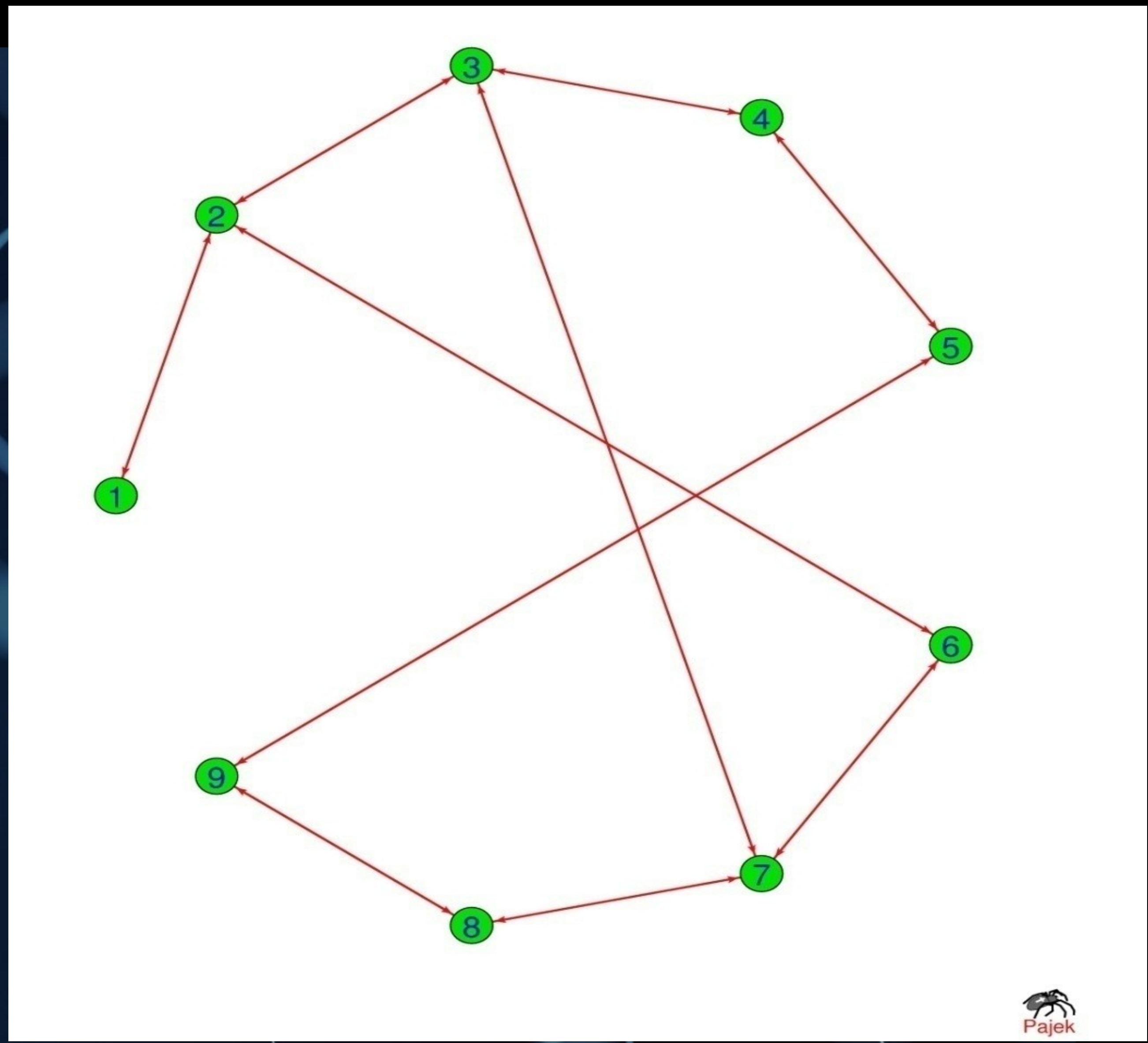
$$N_{ij}^{(d)} = \left[A^d\right]_{ij}.$$

These equations hold for directed and undirected networks. The *distance* between nodes i and j is the path with the smallest d for which $N_{ij}^{(d)} > 0$.

Shortest Path

$$\begin{array}{c|ccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
3 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
4 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
5 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
6 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
7 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
9 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{array}$$

$$A^2 = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 3 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\
1 & 0 & 3 & 0 & 1 & 2 & 0 & 1 & 0 \\
0 & 1 & 0 & 2 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 2 & 0 & 0 & 1 & 0 \\
1 & 0 & 2 & 0 & 0 & 2 & 0 & 1 & 0 \\
0 & 2 & 0 & 1 & 0 & 0 & 3 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 2 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2
\end{bmatrix}$$



Edge list and node list

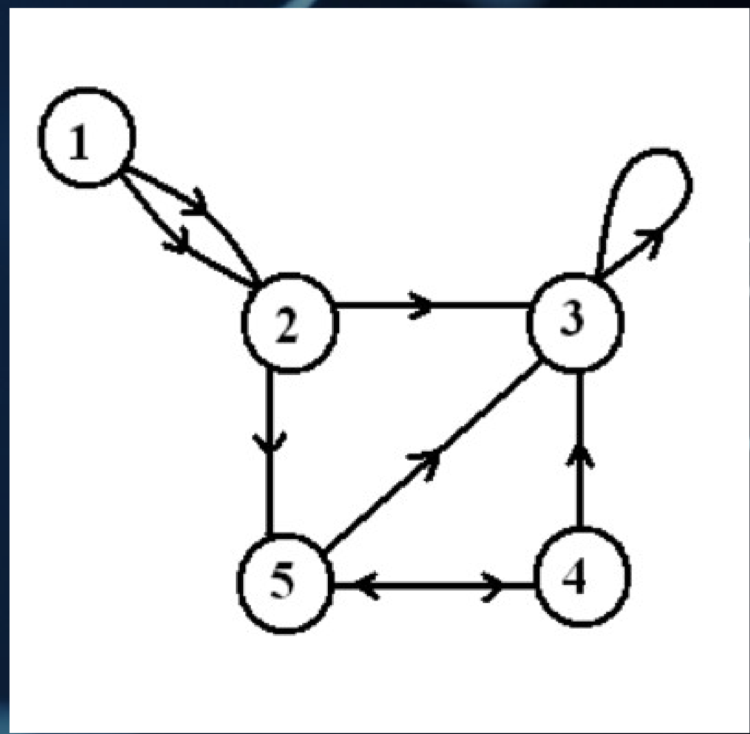
Representation (List)

- Edge List
  - pairs (ordered if directed) of vertices
  - Optionally weight and other data
- Adjacency List (node list)

Edge list and node list

Representation (List)

- Edge List
  - pairs (ordered if directed) of vertices
  - Optionally weight and other data
- Adjacency List (node list)



Edge List

1	2
1	2
2	3
2	5
3	3
4	3
4	5
5	3
5	4

Network :structural chracteristics

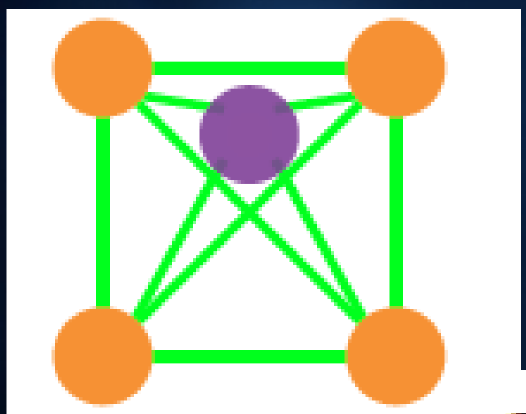
- Clustering Coefficients
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Centrality
- Gephi

Clustering Coefficients

The clustering coefficient captures the degree to which the neighbors of a given node link to each other. For a node with degree the local clustering coefficient is defined as

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

CC of the center node (purple)

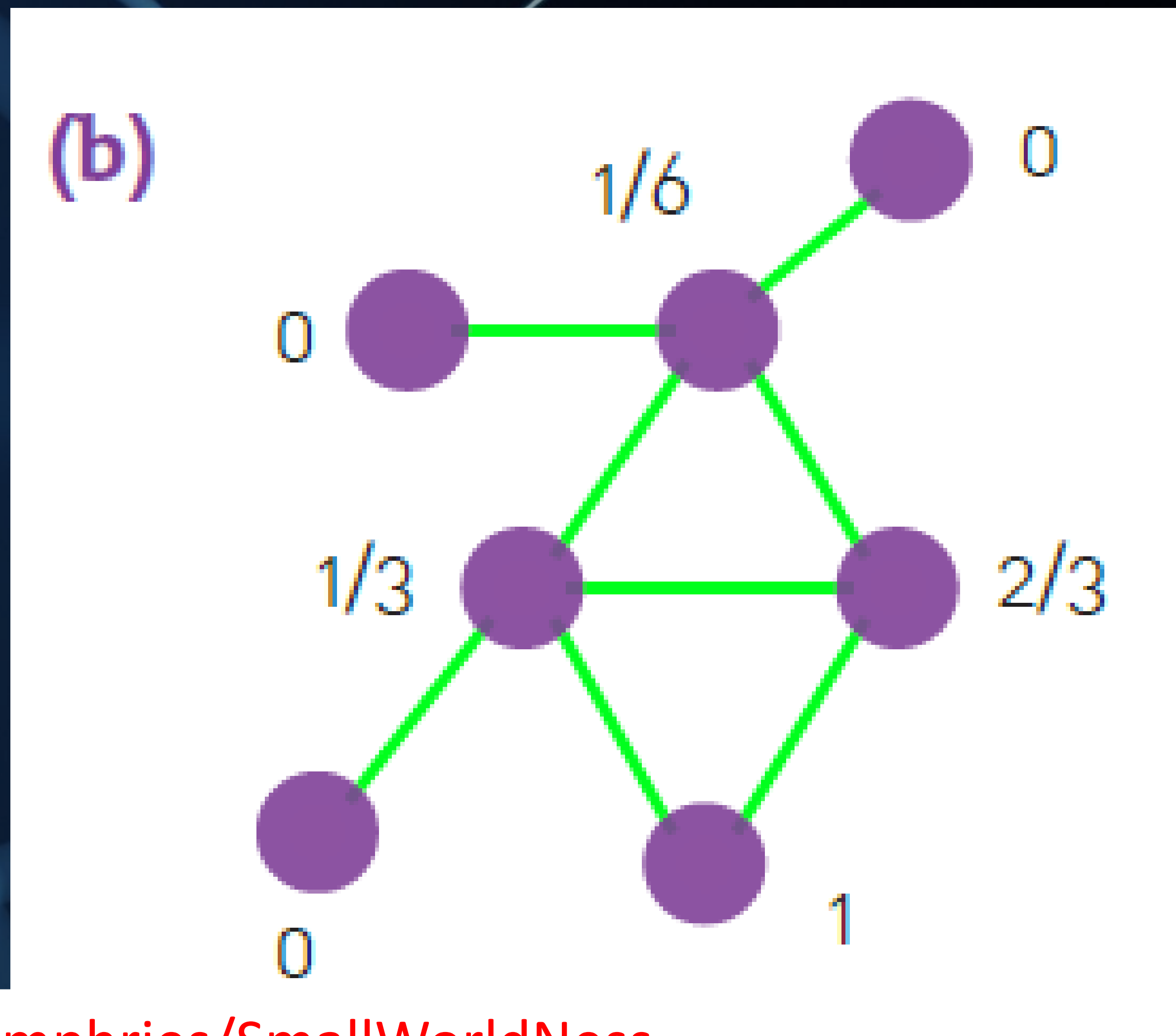


C_i=1

Clustering Coefficients

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

```
n=length(G);
C=zeros(n,1);
for u=1:n
V=find(G(u,:));
k=length(V);
if k>=2 %degree must be at least 2
S=G(V,V);
C(u)=sum(S(:))/(k^2-k);
end
end
```



https://github.com/mdhumphries/SmallWorldNess

Clustering Coefficients

The degree of clustering of a whole network is captured by the *average clustering coefficient*, $\langle C \rangle$, representing the average of C_i over all nodes i = 1, ..., N [12],

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i .$$

Network Analysis: Centrality measures

Centrality measures indicate the most important vertices within a graph:

- The most influential person in a social network.
- The most critical nodes in a infrastructure (such as Internet).
- The highest spreaders of disease.

- Degree centrality.
- Closeness centrality.
- Betweenness centrality.
- Eigenvector centrality.
- ...

Help Barabasi

Different types of centralities

What makes a vertex central in a network? (one or more ideas)	*How do you describe it mathematically?*	*When is it appropriate to use it?*	*How can we capture it?*

Different types of centralities

What makes a vertex central in a network? (one or more ideas)	*How do you describe it mathematically?*	*When is it appropriate to use it?*	*How can we capture it?*
Lots of one-hop connections from	*The number of vertices that influences directly*	*Local influence matters Small diameter*	**Degree centrality** *(or simply the)*
Lots of one-hop connections from relative to the size of the graph	*The proportion of the vertices that influences directly*	*Local influence matters Small diameter*	**Normalized degree centrality**
In the "middle" of the graph - **closeness centrality**	*Close to everyone at the same time*	*The efficiency of a vertex of reaching everyone quickly (spreading news or a virus for example)*	$C_u = 1 / \sum_{v=1}^{n} d(u,$

npj | climate and atmospheric science

Article

Published in partnership with CECCR at King Abdulaziz University

https://doi.org/10.1038/s41612-025-01146-1

Recent changes in spatiotemporal patterns of heat extremes in South Asia

Check for updates

Abhirup Banerjee[1,11], Shraddha Gupta[2,3,11], Pranava Priyanshu[4], Ankan Kar[5], Ruby Saha[6], Tanujit Chakraborty[7,8], Dibakar Ghosh[4], Jürgen Kurths[3,9] & Chittaranjan Hens[10]

The likelihood of intense heatwaves in South Asia is increasing due to climate change, highlighting the need to understand their evolving spatiotemporal patterns. Using a complex network-based approach, we analyze synchronous extreme heat events across South and West Asia over three 30-year periods: two historical phases (1960–1989, 1990–2019) and a near-future projection (2020–2049) under the SSP2-4.5 scenario. Our findings reveal a shift in heatwave synchronization from western and central Asia before 1990 towards Pakistan, northwest India, and the southwestern Tibetan Plateau by the mid-21st century. This shift is primarily driven by increased surface sensible heat flux, which enhances atmospheric diabatic heating and strengthens the early-summer circumglobal teleconnection. Additionally, atmospheric conditions over the North Atlantic-Greenland sector modulate South Asian heatwave synchronization. Our study provides novel insights into the evolving land-atmosphere interactions driving extreme heat events, with implications for heatwave predictability and risk assessment in a warming world.

Introduction

- Climate change is increasing the frequency of extreme heat events, especially in vulnerable regions like South Asia.
- Understanding the interconnectedness of heatwaves is critical for predicting and preparing for their impacts.
- This study explores how heatwaves across South and West Asia are linked and how these connections evolve.

Event synchronization

Methodology

- Analyzed heatwave patterns using climate networks, connecting locations with synchronized heatwaves.
- Quantified heatwave similarity using event synchronization based on temperature data from 1960-2019.
- Projected heatwave patterns from 2020-2049 using climate models to assess near-future changes.
- Defined heatwaves as periods with temperatures exceeding one standard deviation above average for at least five days.

Fig. 1 | Number of heatwaves during over South Asia during the pre-monsoon season across historical periods. Heatwaves during March-April-May (see definition in Heatwave identification in "Methods") were identified using ERA5 daily 2m-temperature for the periods: **a** 1960–1989 and **b** 1990–2019. Each grid point shows the total number of heatwave events over the respective 30-year span. Warmer colors indicate higher frequency.

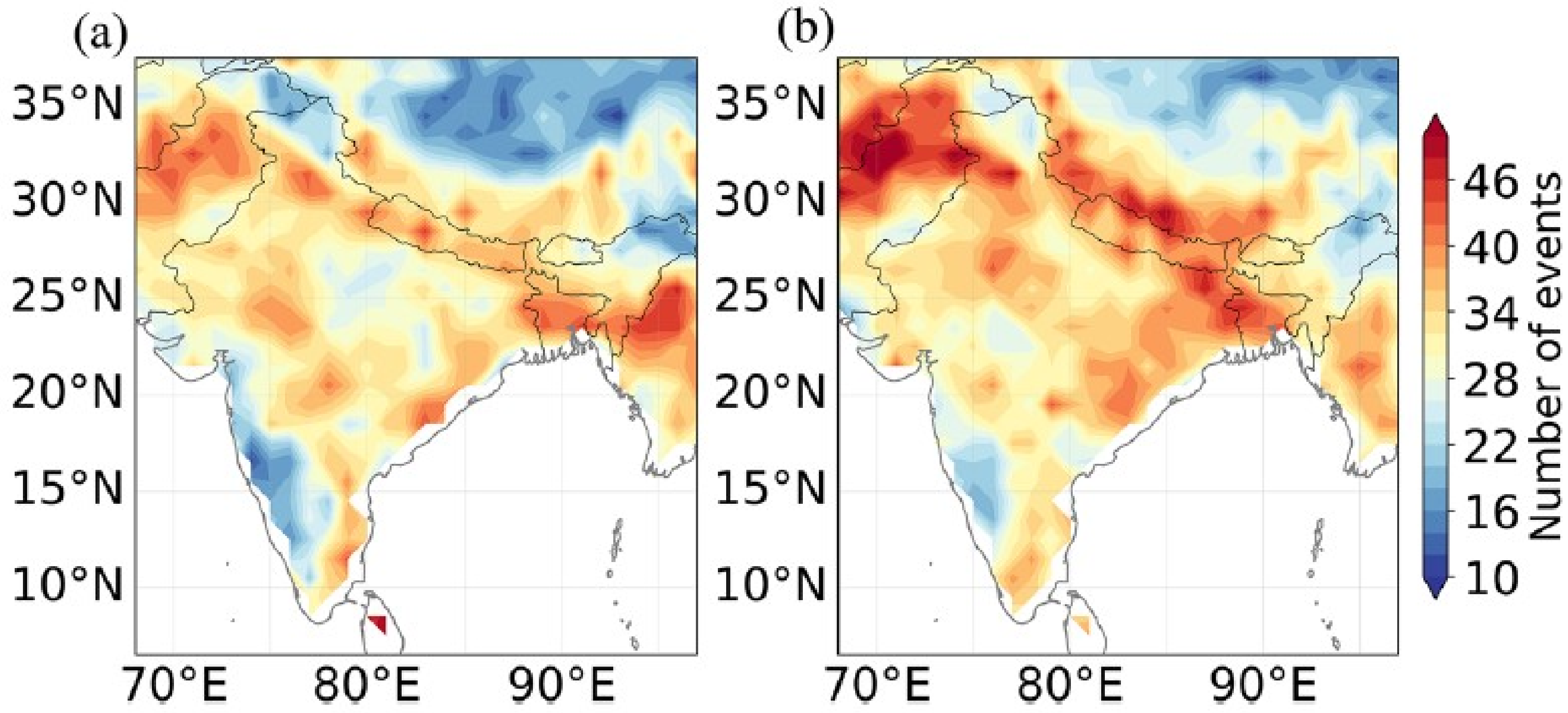
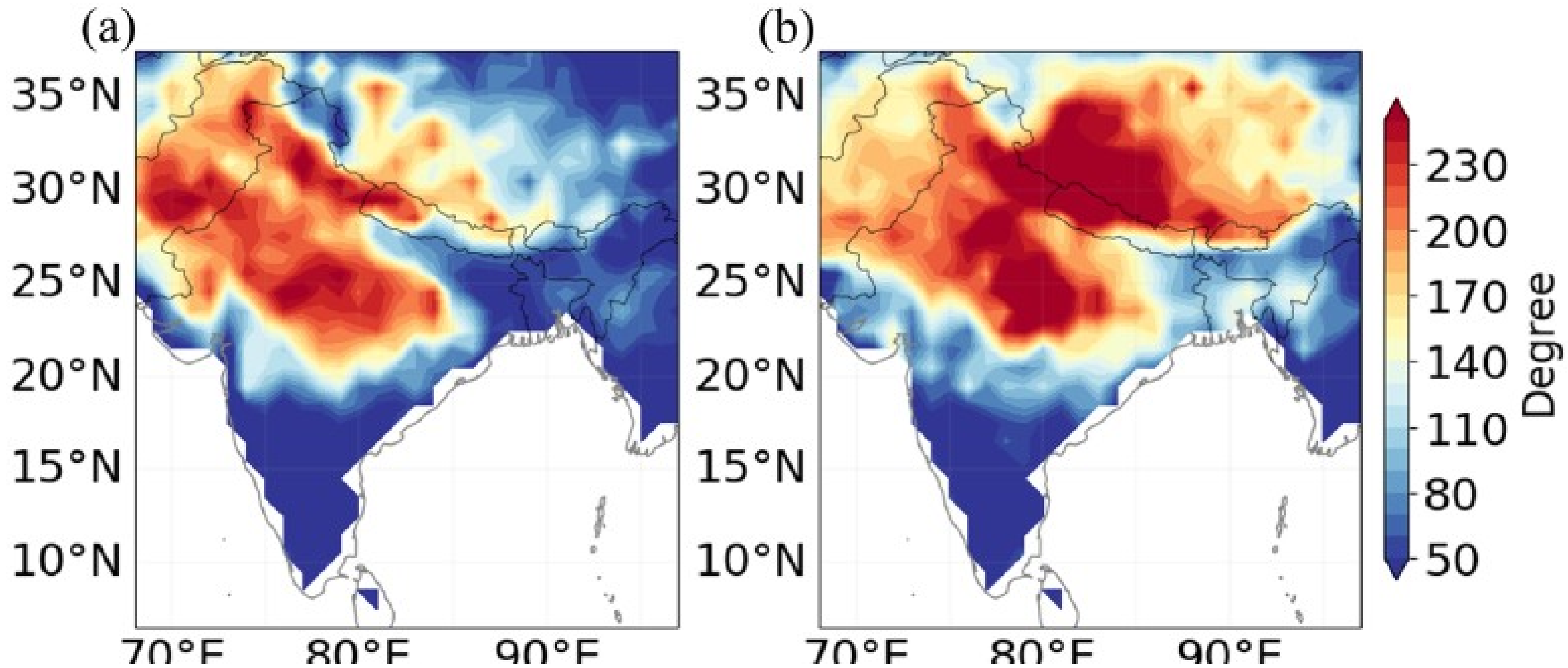


Fig. 2 | Network degree spatial patterns of synchronous heatwaves over South Asia during the pre-monsoon season. Network degree maps derived from ERA5 daily 2m-temperature during March-April-May, capturing synchronous heatwave occurrences with maximum allowed temporal delay of $\tau_{max} = 5$ days for the span: **a** 1960–1989, and **b** 1990–2019. Higher degree values indicate regions with more frequent synchronous heatwave connections to other grid points in the network.



- Heatwave synchronization shifted from western/central Asia towards Pakistan, northwest India, and the southwestern Tibetan Plateau.
- South Asia exhibits increased node centrality due to enhanced surface sensible heat flux (SHF) post-1990.
- SHF became the dominant driver of heat exchange between land and atmosphere over South Asia after 1990.

Closeness Centrality

Reciprocal *of the total distance from a node v to all the other nodes in a network*

$$C_u = 1 / \sum_{v=1}^{n} d(u,v)$$

d(u,v) is the distance between node u and v.

Closeness centrality can be viewed as the efficiency of a vertex in spreading information to all other vertices.

Eigenvector Centrality

- *A natural extension of degree centrality is* **eigenvector centrality**.
- *In-degree centrality awards one centrality point for every link a node receives. But not all vertices are equivalent: some are more relevant than others, and, reasonably, endorsements from important nodes count more.*

Eigenvector Centrality

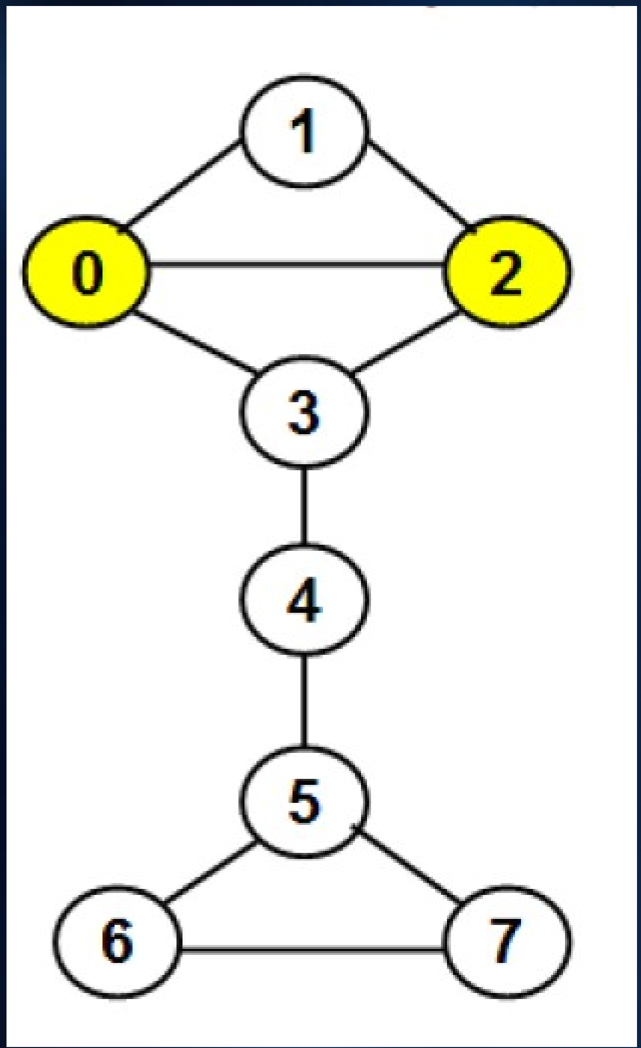
Definition of Dominant Eigenvalue and Dominant Eigenvector

Let $\lambda_1, \lambda_2, \ldots,$ and λ_n be the eigenvalues of an $n \times n$ matrix A. λ_1 is called the **dominant eigenvalue** of A if

$$|\lambda_1| > |\lambda_i|, \quad i = 2, \ldots, n.$$

The eigenvectors corresponding to λ_1 are called **dominant eigenvectors** of A.

Eigenvector Centrality



Iteration 3

0 1 1 1 0 0 0 0
1 0 1 0 0 0 0 0
1 1 0 1 0 0 0 0
1 0 1 0 1 0 0 0
0 0 0 1 0 1 0 0
0 0 0 0 1 0 1 1

x

0.428
0.321
0.428
0.428
0.321
0.321

=

1.177
0.856
1.284
1.177
0.749
0.857
0.589
0.589

≡

0.441
0.321
0.481
0.441
0.281
0.321
0.221
0.221

= 2.67

Note that we typically stop when the EigenVector values converge.
For exam purposes, we will Stop when the Normalized value converges.

0 0 0 0 0 1 0 1
0 0 0 0 0 1 1 0

0.221
0.221

1.243
0.922
1.203
1.203
0.762
0.723
0.542
0.542

≡

0.471
0.349
0.456
0.456
0.289
0.274
0.205
0.205

Normalized Value = 2.64

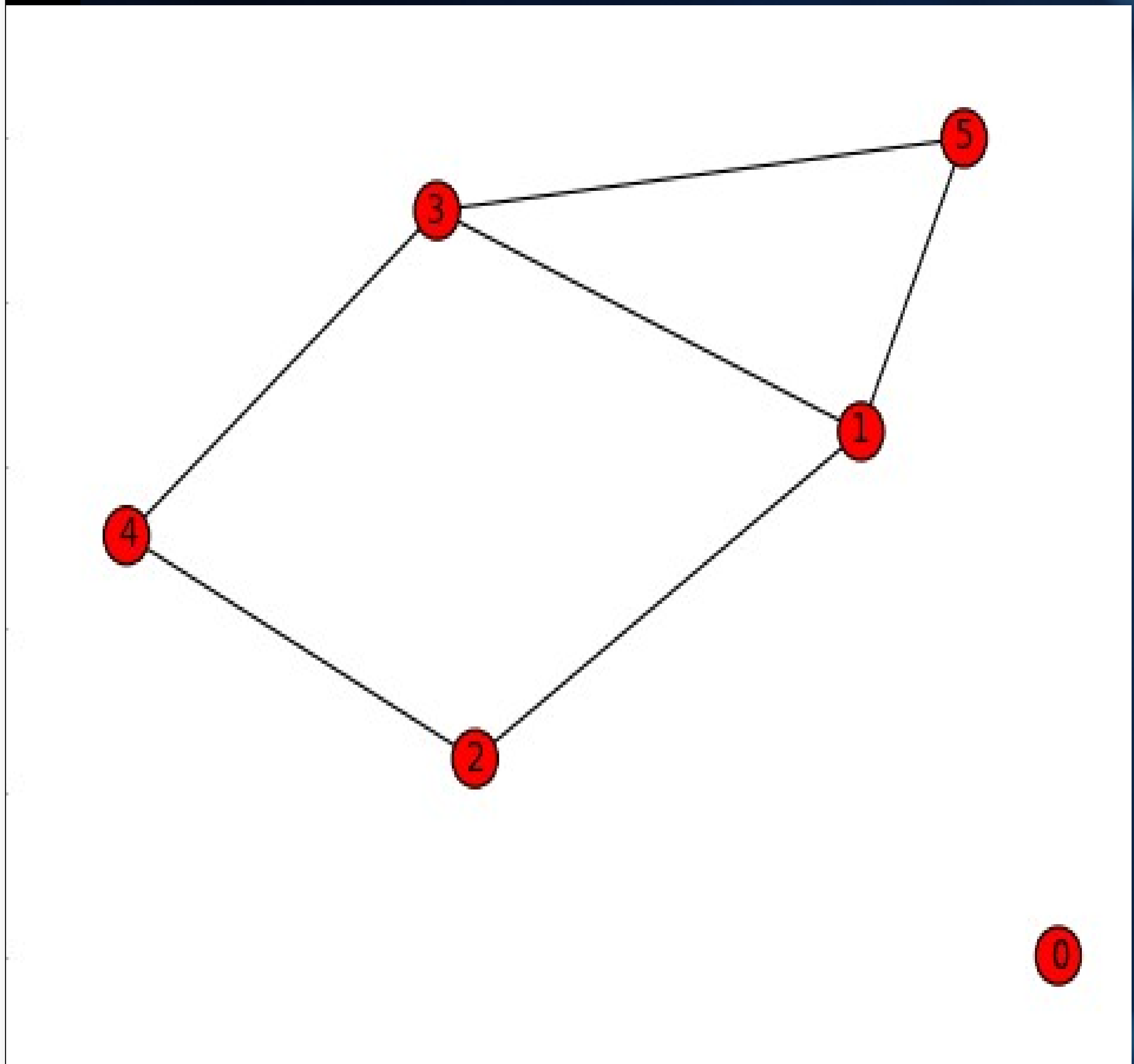
Normalized Value = 2.64

After 7 iterations

Vertex ID	0	1	2	3	4	5	6	7
Principal Eigenvector	0.489	0.364	0.489	0.467	0.264	0.232	0.155	0.155

Raluca Gera, NPS

Eigenvector Centrality




```
clear all
A=zeros(6,6);
A(5,[1 3])=1;
A(2,[1 4])=1;
A([1 4],3)=1;


A=A+A';
[u lam]=eig(A);


evc_large=u(:,6);
```

Node Betweenness

Betweenness centrality counts the number of times a node is part of the shortest path between each other pair of nodes in the graph.

In other words, the betweenness centrality of a node v is the ratio of all shortests paths from one node to any other node in the graph that pass through v.

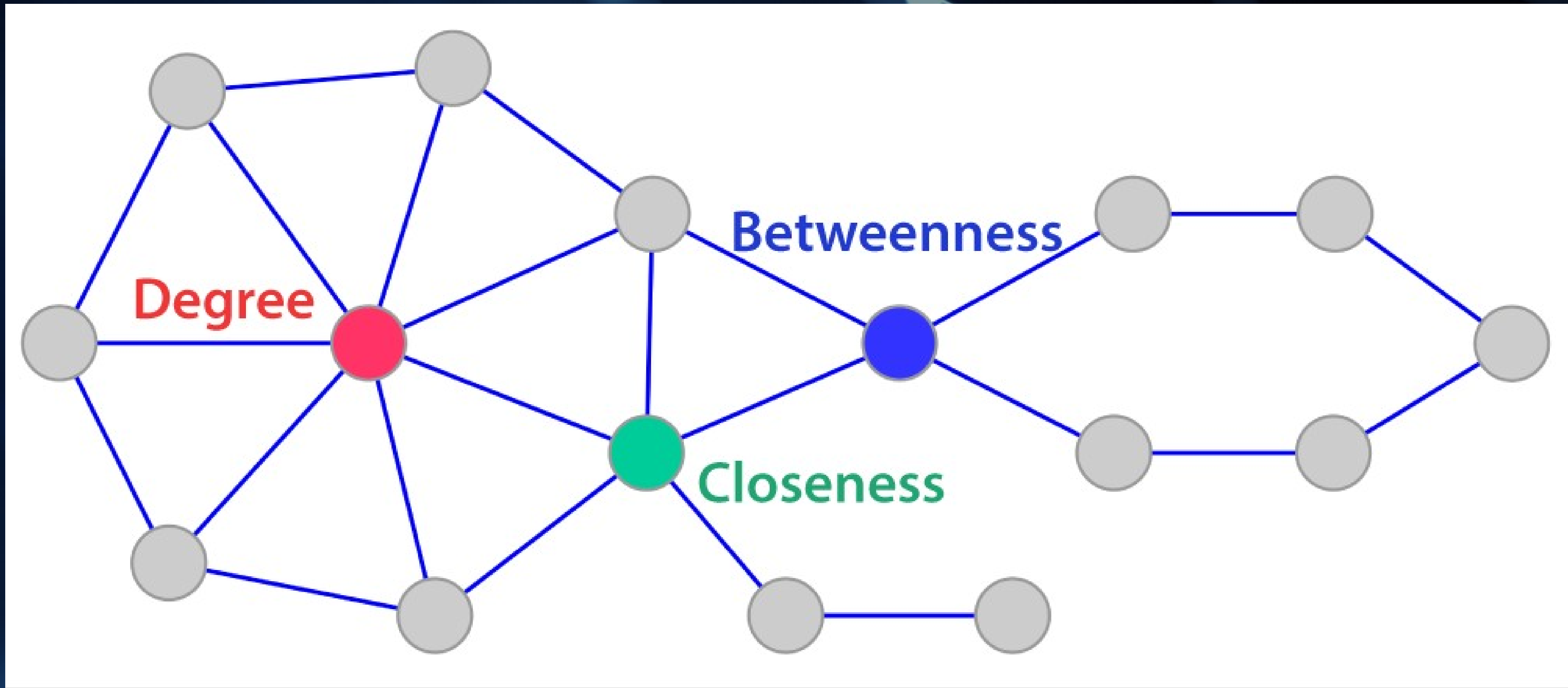
Mathematically:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

The sum of all the shortest paths from s to t that pass through v divided by all the shortest paths from s to t.

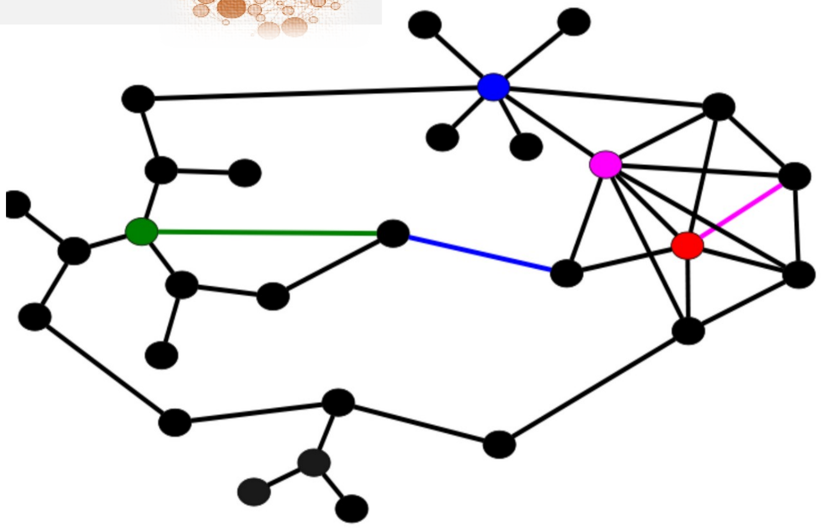
Shortest path based betweenness centrality

Centralities



Degree	ClosenessCentrality	BetweennessCentrality
7	0.45454545	0.29047619
5	0.51724138	0.42380952
4	0.48387097	0.4952381

Node-based and edge-based centralites



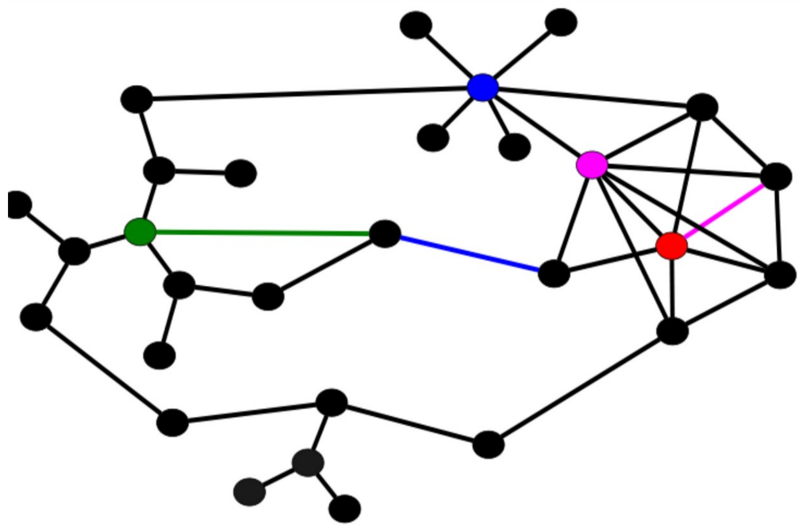
; **Strength centrality**

; **Nearest-neighbor edge centrality**

where is an edge and nodes that connect the edge and Is the strength of that edge.

- *Centrality-based identification of important edges in complex networks T Bröhl, K Lehnertz, Chaos: An Interdisciplinary Journal of Nonlinear Science 29 (3)*
- *A straightforward edge centrality concept derived from generalizing degree and strength, T Bröhl, K Lehnertz, Scientific Reports 12 (1), 4407*

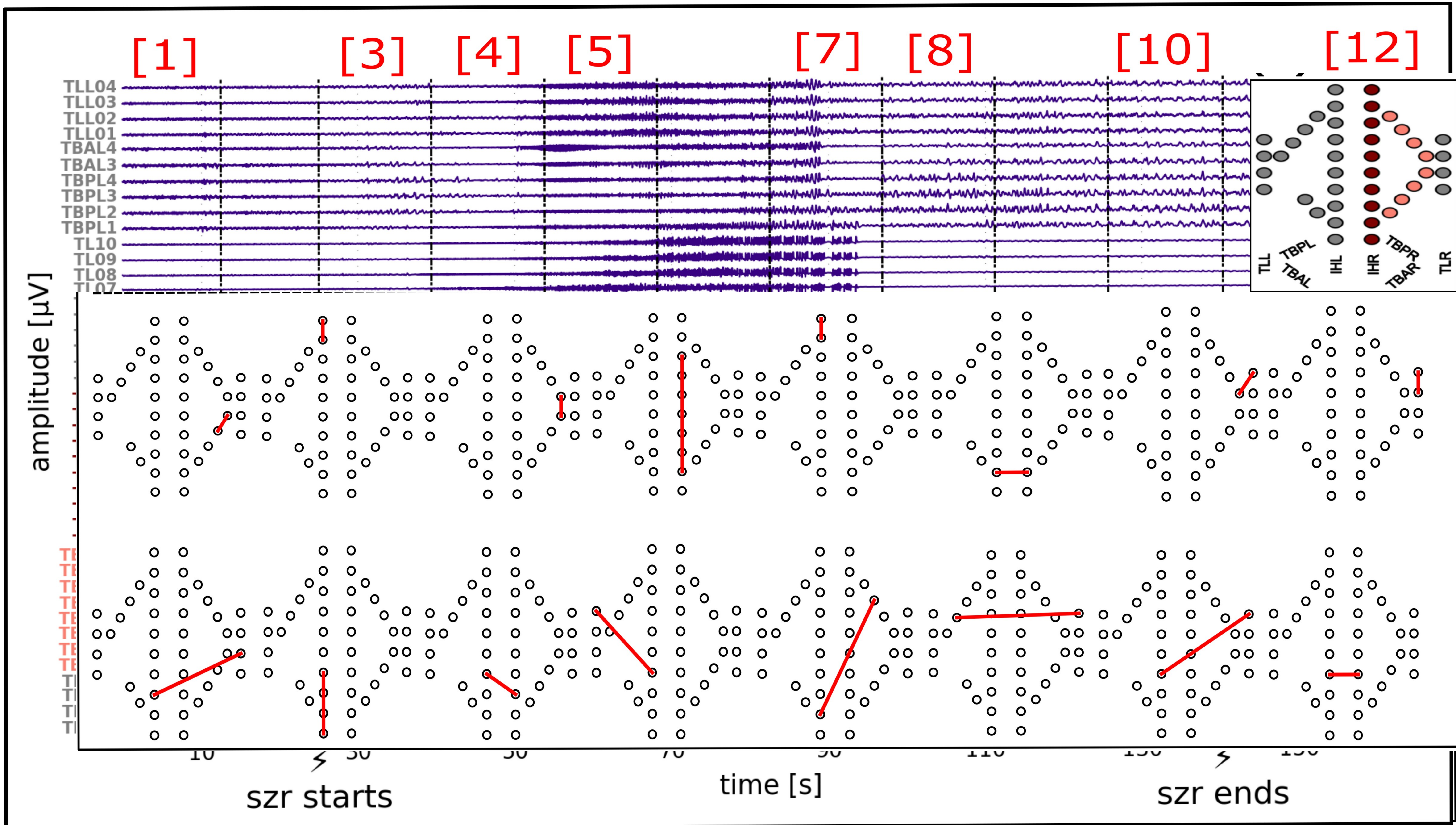
Node-based and edge-based centralites



- *(for vertices), resp.* $q \in \{1, \ldots, E\}$ *(for edges), with*
- *denotes the number of shortest paths between vertices* x and y *that pass through the vertex/edge q,*
- M_{xy} *represents the total number of shortest paths between* x *and* y.

- *The normalization factor is given by in case of vertices and* $A = V(V - 1)$ *for edge betweenness.*

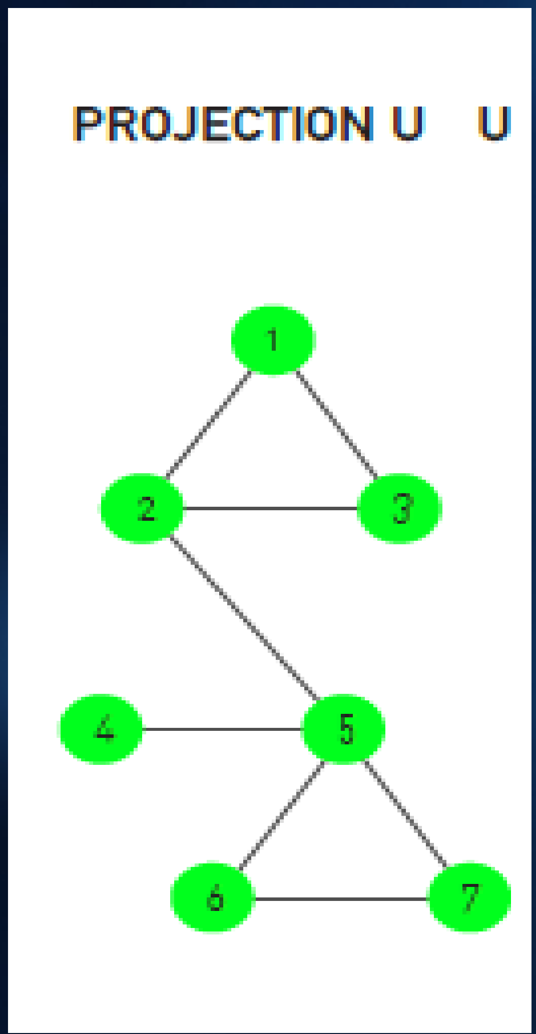
- *In general, a high (low) centrality value indicates a high (low) importance of a constituent within a network.*

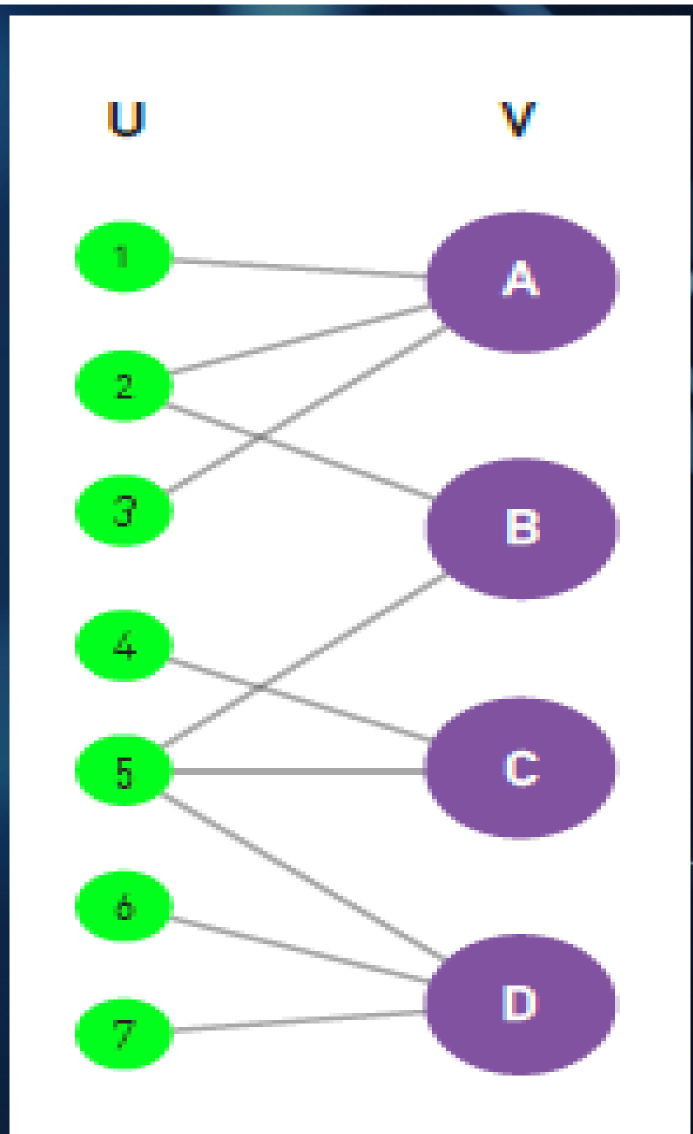


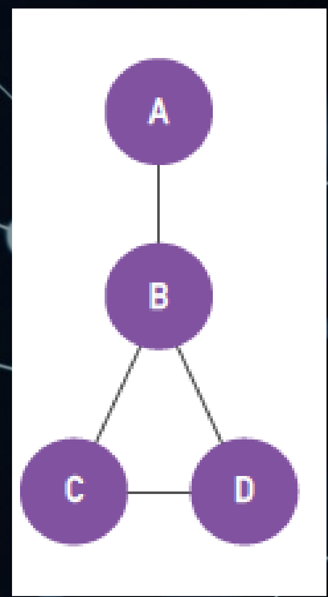
[1]
[3]
[4]
[5]
[7]
[8]
[10]
[12]
TLL04
TLL03
TLL02
TLL01
TBAL4
TBAL3
TBPL4
TBPL3
TBPL2
TBPL1
TL10
TL09
TL08
TLL
TBPL
TBAL
IHL
IHR
TBPR
TBAR
TLR
amplitude [μV]
10
30
50
70
90
110
150
time [s]
szr starts
szr ends

Bipartite Graphs: Subprojections

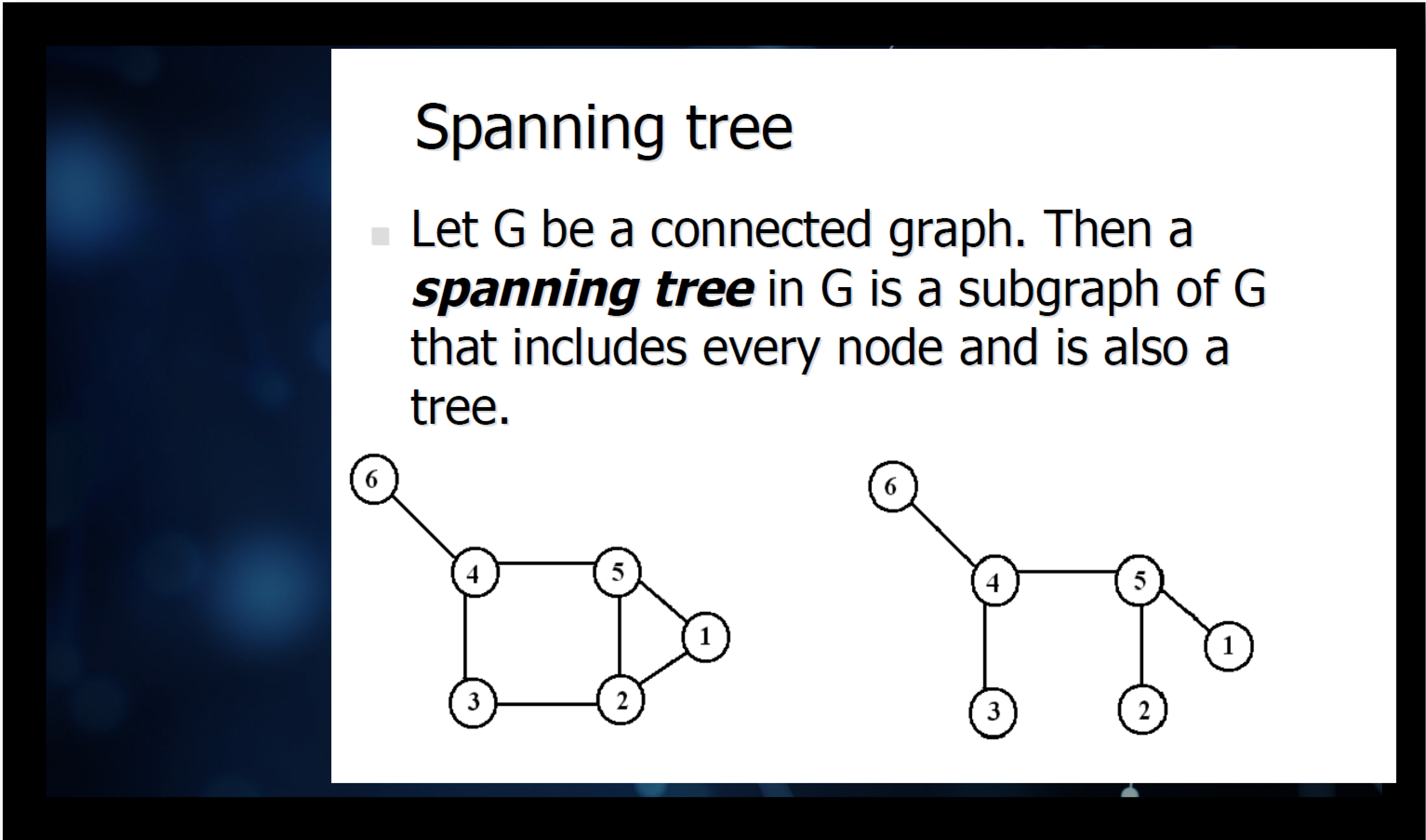
A bipartite graph (or bigraph) is a network whose nodes can be divided into two disjoint sets U and V such that each link connects a U-node to a V-node. In other words, if we color the U-nodes green and the V-nodes purple, then each link must connect nodes of different colors.







Spanning Tree



Help Barabasi

Numerical Solution with Euler Method

Basics: Numerical techniques (Derivative of a function)

$$f(x_0+h)=f(x_0)+h\,f'(x_0)+\frac{h^2}{2!}f''(x_0)+\ldots+\frac{h^n}{n!}f^n(x_0)+R_n(x)$$

Denotes the difference between Talylor Polynomial of degree and the original function

For the first derivative

$$f(x_0+h)=f(x_0)+h\,f'(x_0)+R_1(x)$$

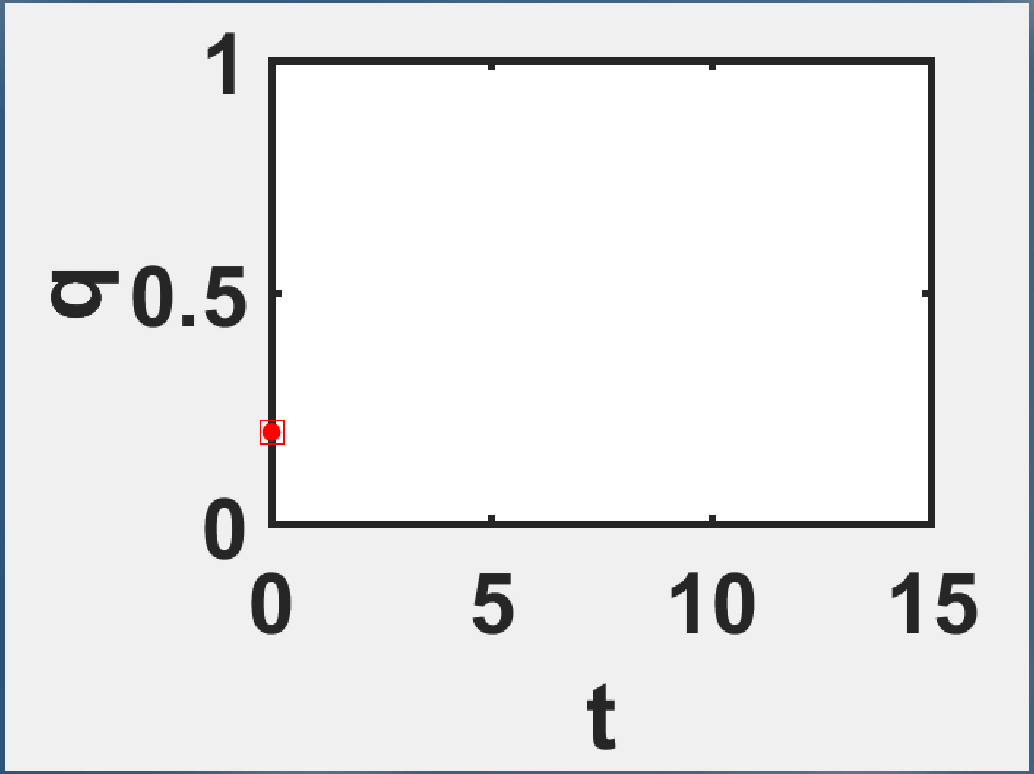
Samll

$$Lt_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h}=f'(x_0)$$

Numerical errors

- *Discretization error:*
  - *Error resulting from the computation of quantities that have been calculated using an approximation to the true solution*
    - *We may neglect higher order terms for example*
    - *This error is sometimes called truncation error*
  - *Unavoidable problem in numerical integration work*
    - *Would occur even in the presence of infinite precision*
- *Round-off error*
  - *Result of finite precision arithmetic*

Euler’s Method: Model 1

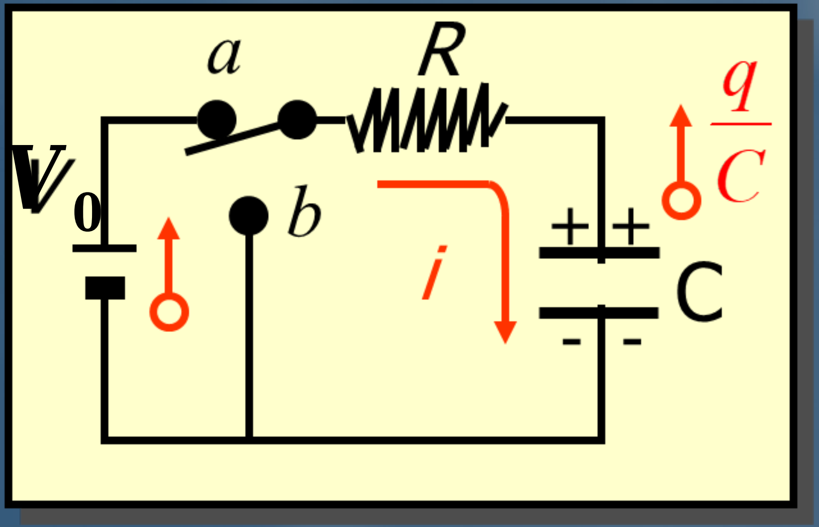



```
x0=0.2;
y0=[x0];
%[t,y]=ode45(@(t,y) stability_ex1(t,y,a,b),
[0,tspan], y0);%[ 2.5 0 1 0]);
init = [x0 y0]; % y0 = [v0, w0]
t = linspace(0,10,500)';
% Runge-Kutta 4th-Order Algorithm
y_euler = zeros(length(t), 2);
y_euler(1, :) = init;
h = t(2)-t(1);
for i = 2:length(t)
k1 =stability_ex1(t(i-1), y_euler(i-1, :),a,b);
y_euler(i, :) = y_euler(i-1, :)+h*k1;
End

function [yprime] = stability_ex1(t,y,a,b)
y_prime=zeros(1,1);
yprime(1) = a-b*y(1);
yprime=yprime';
end
```

Interpreting a differential equation as a vector field Flows on the Line

RC Circuit: Charging Capacitor (Model 2)



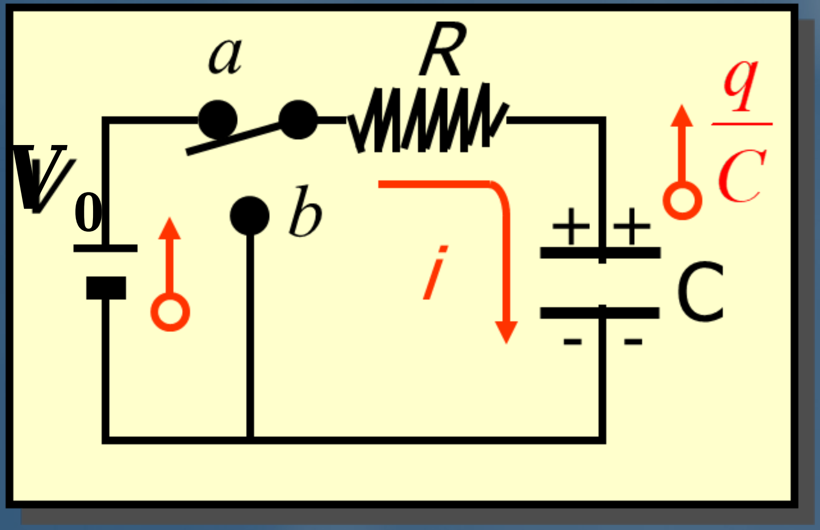
$$V_0 - \frac{q}{C} = iR$$

$$R\frac{dq}{dt} = V_0 - \frac{q}{C}$$

Q: What will be the asymptotic value of ?

Interpreting a differential equation as a vector field Flows on the Line

RC Circuit: Charging Capacitor



$$V_0 - \frac{q}{C} = iR$$

$$R\frac{dq}{dt} = V_0 - \frac{q}{C}$$

$$\int_0^q \frac{dq}{(CV_0 - q)} = \int_o^t \frac{dt}{RC}$$

$$\ln(\text{¿}CV_0 - q) - \ln(\text{¿}CV_0) = \frac{-t}{RC}\text{¿¿}$$

$$\ln\frac{(CV_0 - q)}{CV_0} = \frac{-t}{RC}$$

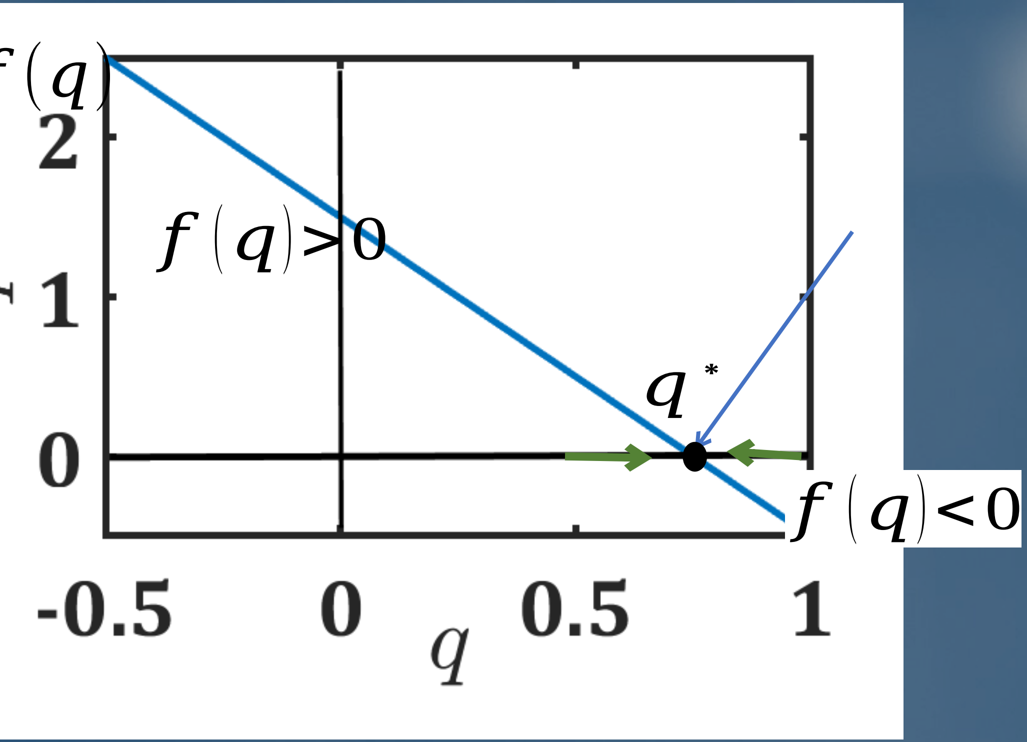
At time

At time

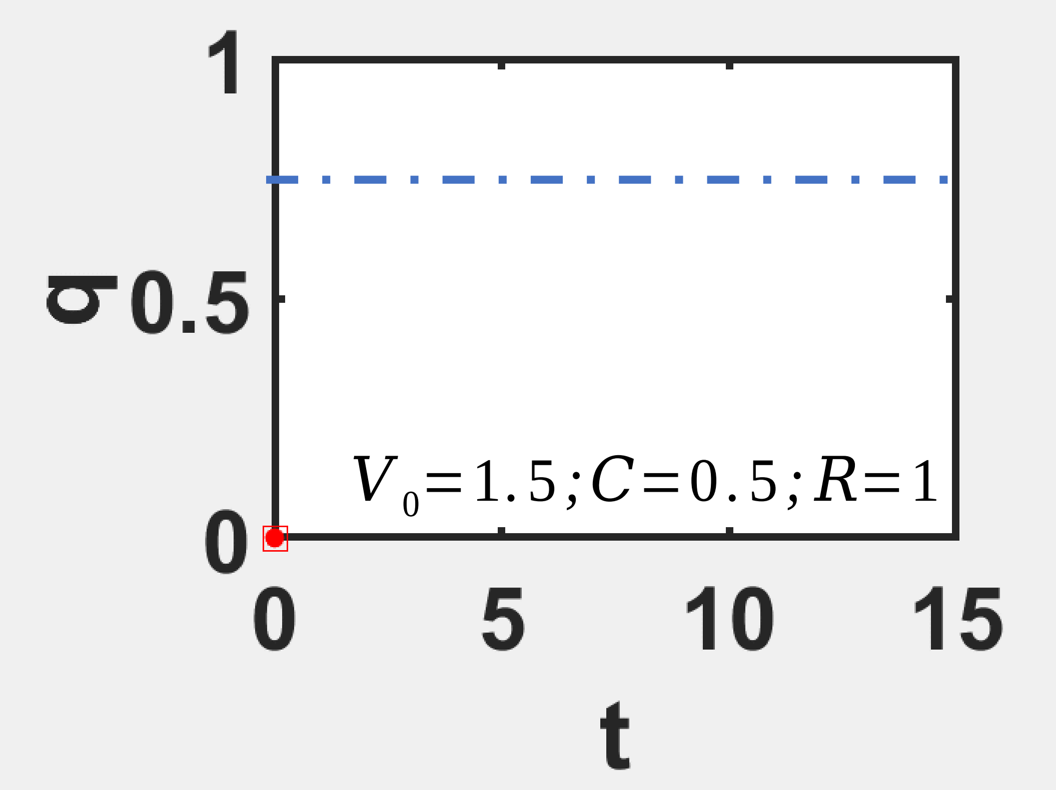
$$q = CV_0\left(1 - e^{-t/RC}\right)$$

The charge q rises from zero initially to its maximum value

Interpreting a differential equation as a vector field Flows on the Line



Solution



Interpreting a differential equation as a vector field Flows on the Line

Logistic equation (Model 3)

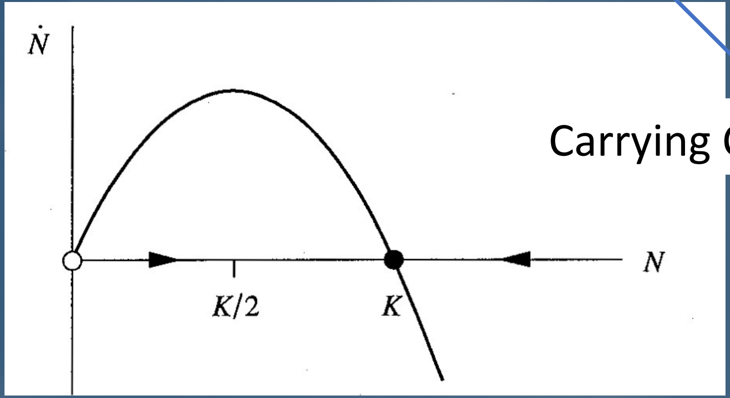
Analysis of Fixed points

$$\dot{N} = rN\left(1 - \frac{N}{K}\right)$$

Carrying Capacity

Per Capita Growth Rate

$\dot{N}/N$



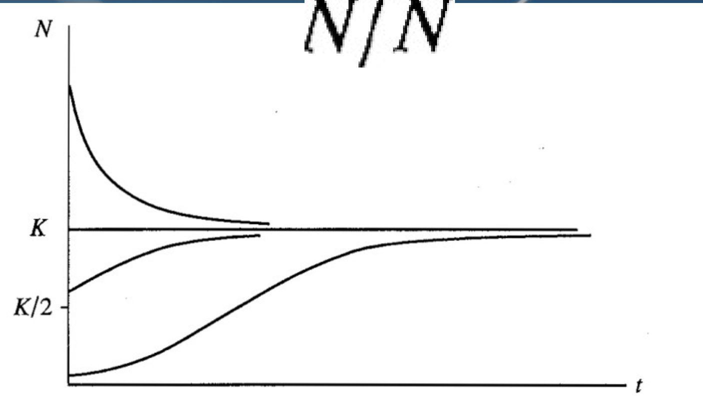
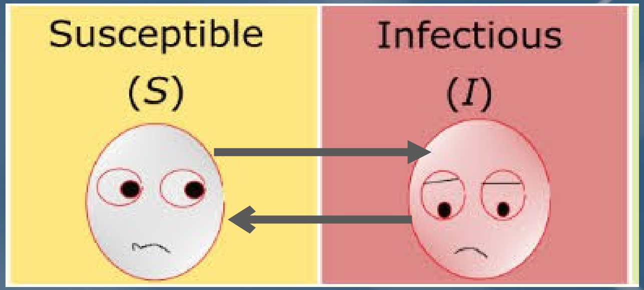


Figure 2.3.3 also allows us to deduce the qualitative shape of the solutions. For example, if $N_0 < K/2$, the phase point moves faster and faster until it crosses $N = K/2$, where the parabola in Figure 2.3.3 reaches its maximum. Then the phase point slows down and eventually creeps toward $N = K$. In biological terms, this means that the population initially grows in an accelerating fashion, and the graph of $N(t)$ is concave up. But after $N = K/2$, the derivative $\dot{N}$ begins to decrease, and so $N(t)$ is concave down as it asymptotes to the horizontal line $N = K$ (Figure 2.3.4). Thus the graph of $N(t)$ is S-shaped or *sigmoid* for $N_0 < K/2$.

$N(t)$ is S-shaped or *sigmoid* for $N_0 < K/2$.

Epidemic Spreading : Modelling

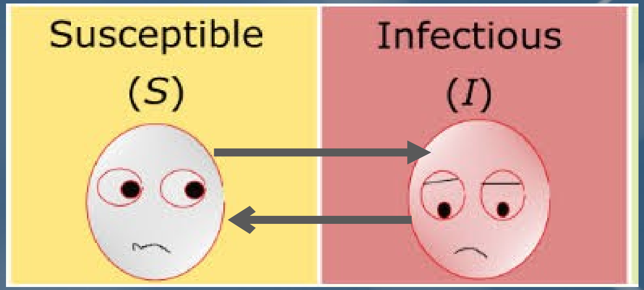
SIS Dynamics



- *Large population*
- *Single initial case*
- *Rest of the population is fully susceptible*

Epidemic Spreading : Modelling

SIS Dynamics



$s \rightarrow i$ β *(infection rate)*
$i \rightarrow r$ *(recovery rate)*

- *Large population*
- *Single initial case*
- *Rest of the population is fully susceptible*

Examples:

- *Chlamydia*
- *Human papilloma virus*
- *(Respiratory syncytial virus / influenza)*

$$\frac{di}{dt} = \beta s i - \gamma i$$

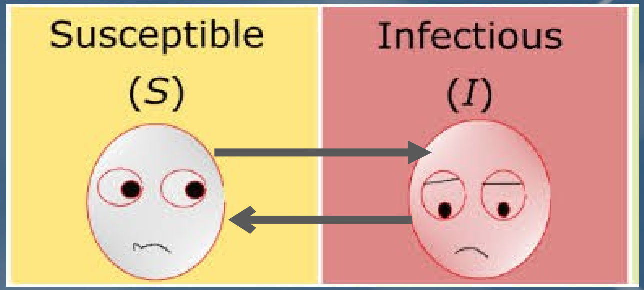
(Constant)

Assumptions:

Homogeneous mixing
Constant recovering rate

Epidemic Spreading : Modelling

SIS Dynamics



$$\frac{di}{dt} = \beta s i - \gamma i$$

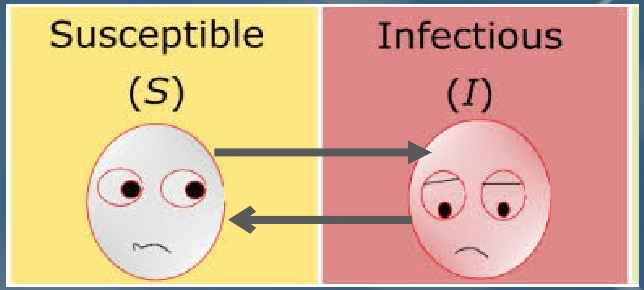
(Constant)

$$\frac{di}{dt} = \beta(1-i)i - \gamma i = (\beta - \gamma)i - \beta i^2$$

$$\int_{x_0}^{x} \frac{dx}{px - qx^2} = \int_0^t dt$$

$$p = (\beta - \gamma);\ q = \beta$$

Epidemic Spreading : Modelling



SIS Dynamics

$$\frac{di}{dt} = \beta s i - \gamma i$$

(Constant)

$$\frac{di}{dt} = \beta(1-i)i - \gamma i = (\beta - \gamma) i - \beta i^2$$

$$\int_{x_0}^{x} \frac{dx}{px - qx^2} = \int_0^t dt$$

$$p = (\beta - \gamma)\,;\; q = \beta$$

$$\frac{1}{p}\left(\log\left(\frac{p - q x_{\square}}{x_{\square}}\right) - \log\left(\frac{p - q x_0}{x_0}\right)\right) = -t$$

SIR Modelling

SIR Dynamics

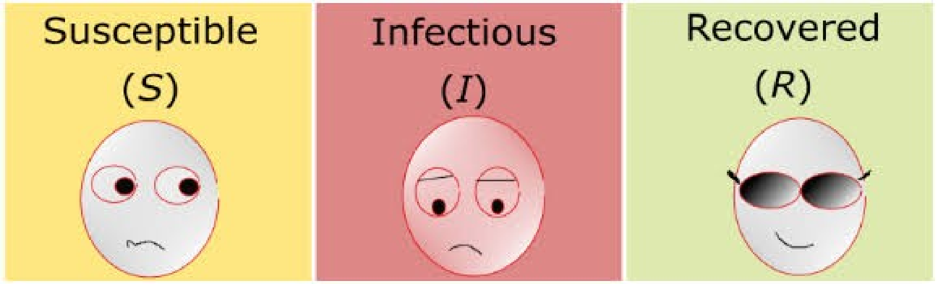
$s=S/N$: *density* of Susceptible individuals
$i=I/N$: *density* of Infective individuals
$r=R/N$: *density* of Removed (recovered/dead) individuals

$$s \to i \quad \beta \ \text{(infection rate)}$$
$$i \to r \quad \ \ \text{(recovery rate)}$$

$$\left(\frac{ds}{dt}+\frac{di}{dt}+\frac{dr}{dt}=0\right)$$
$$s+i+r=1$$

Assumptions:

- *Homogeneous mixing*
- *Constant recovering rate*



SIR Modelling

SIR Dynamics

$s=S/N$: *density* of Susceptible individuals
$i=I/N$: *density* of Infective individuals
$r=R/N$: *density* of Removed (recovered/dead) individuals

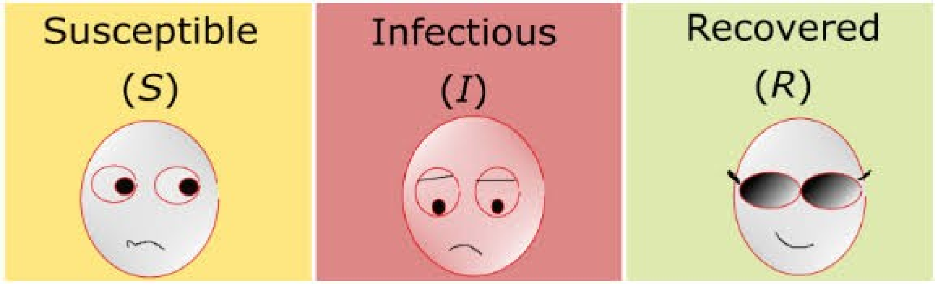
$$\left(\frac{ds}{dt}+\frac{di}{dt}+\frac{dr}{dt}=0\right)$$

$$s+i+r=1$$

$s \rightarrow i \quad \beta$ (infection rate)
$i \rightarrow r \quad$ (recovery rate)

$$\frac{di}{dt}=\beta s i-\gamma i$$

$$\frac{dr}{dt}=\gamma i$$



SIR Modelling

SIR Dynamics

$s = S/N$: *density* of Susceptible individuals
$i = I/N$: *density* of Infective individuals
$r = R/N$: *density* of Removed (recovered/dead) individuals

$$\left(\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0\right)$$

$$s + i + r = 1$$

$s \to i \quad \beta$ (infection rate)
$i \to r \quad$ (recovery rate)

$$\frac{di}{dt} = \beta s i - \gamma i$$

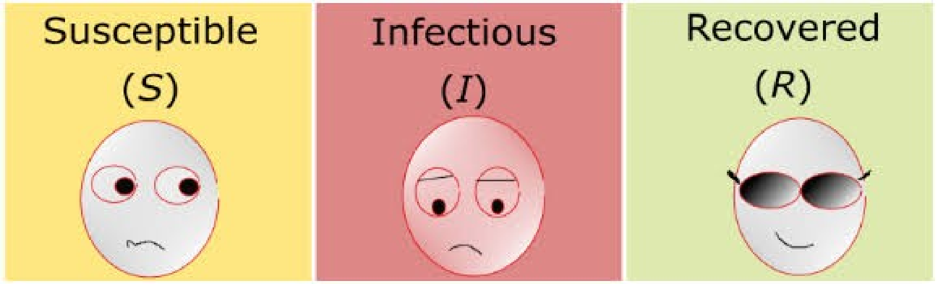
$$\frac{dr}{dt} = \gamma i$$

$$\boxed{\frac{di(t)}{dt} = (\beta s - \gamma) i > 0}$$

$$(s(0) \approx 1)$$

$$\boxed{\beta s(0) - \gamma > 0}$$

$$\boxed{R_o \equiv \frac{\beta}{\gamma} > 1}$$

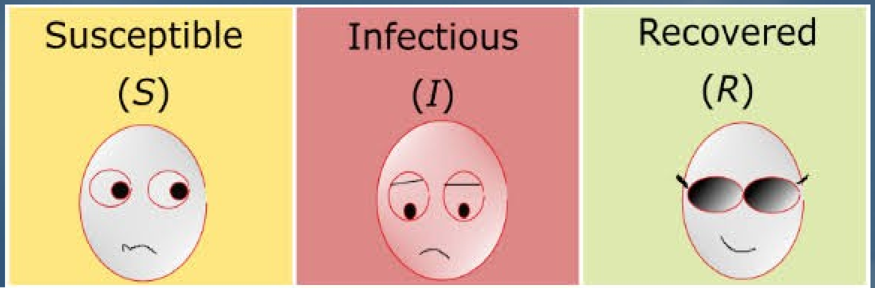


SIR Modelling

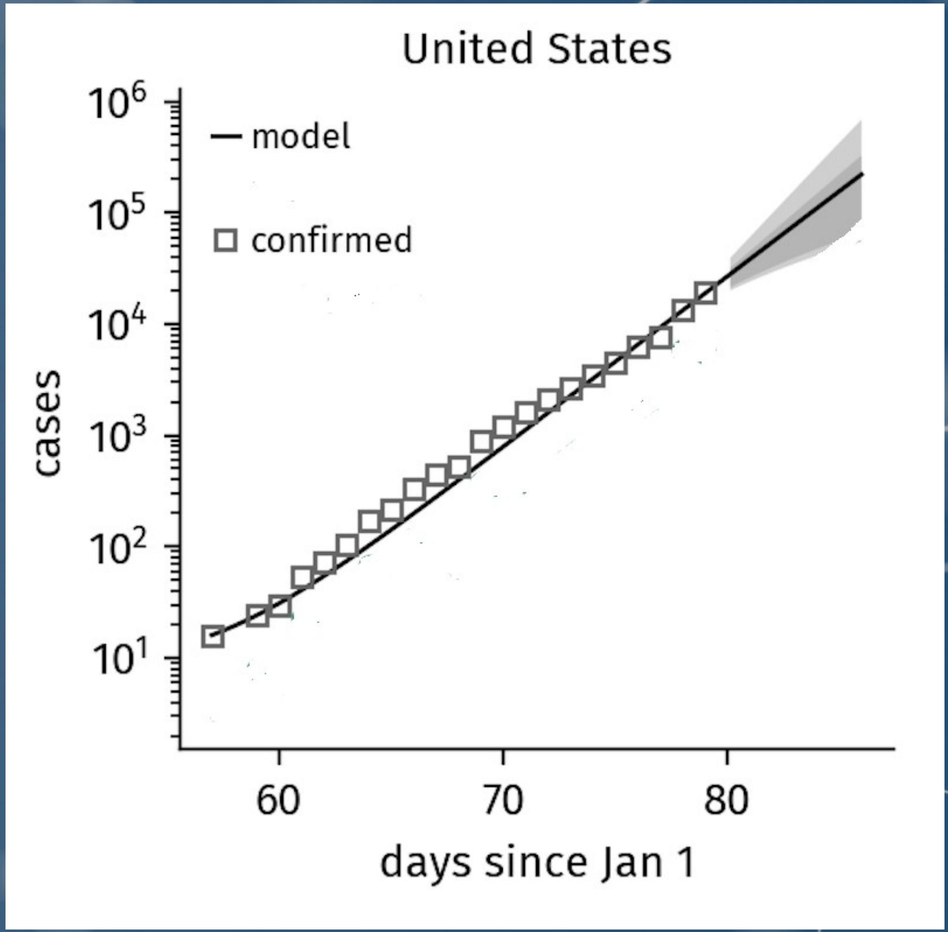
SIR Dynamics

$$\frac{d\,i}{dt} = \beta\, s\, i - \gamma\, i$$

$$\frac{d\,r}{dt} = \gamma\, i$$



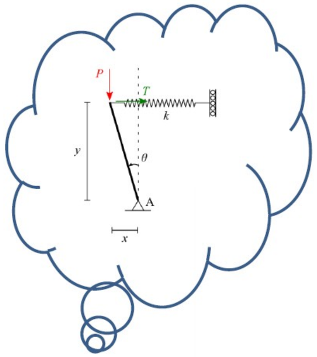
http://rocs.hu-berlin.de/corona/docs/forecast/model/

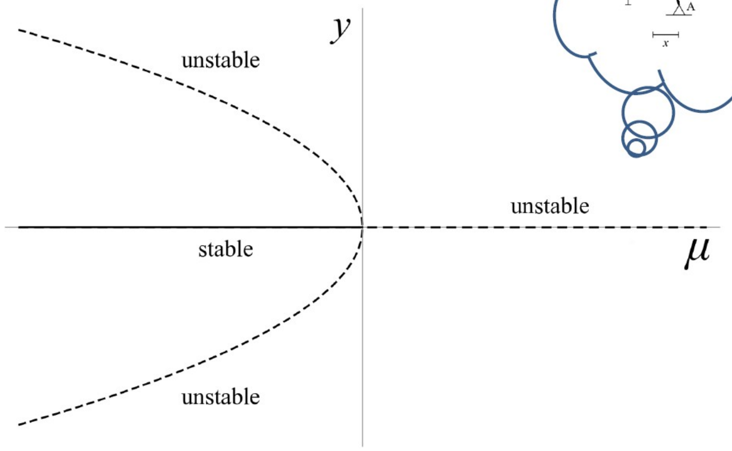


Science 368, 742–746 (2020), Maier et al

Subcritical pitchfork bifurcation

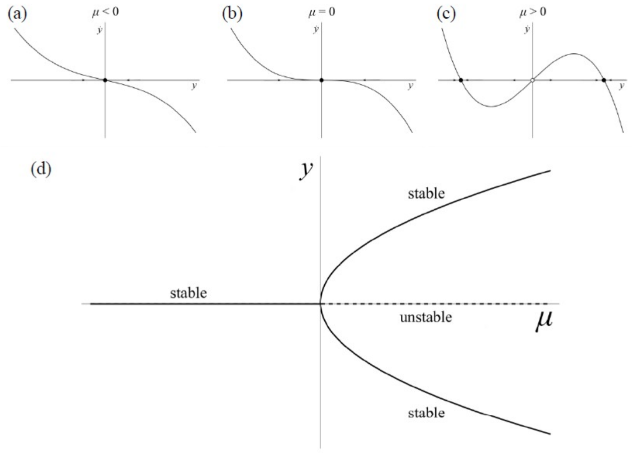
Normal form: $\dot{y} = \mu y + y^3$





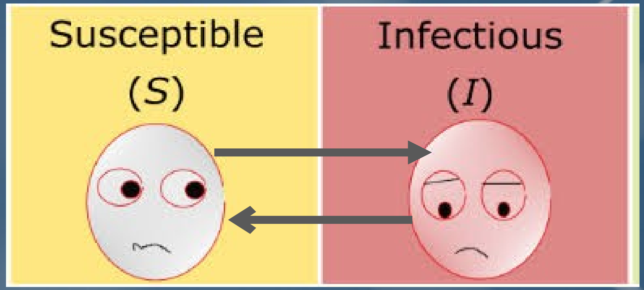
Supercritical pitchfork bifurcation

Normal form: $\dot{y} = \mu y - y^3$



Epidemic Spreading : Modelling

SIS Dynamics



$s \rightarrow i \quad \beta$ *(infection rate)*

$i \rightarrow r \quad$ *(recovery rate)*

$$\frac{di}{dt} = \beta s i - \gamma i$$

(Constant)

- *Large population*
- *Single initial case*
- *Rest of the population is fully susceptible*

Examples:

- *Chlamydia*
- *Human papilloma virus*
- *(Respiratory syncytial virus / influenza)*

Assumptions:

Homogeneous mixing

Constant recovering rate

SIR Modelling

SIR Dynamics

$s=S/N$: *density* of Susceptible individuals
$i=I/N$: *density* of Infective individuals
$r=R/N$: *density* of Removed (recovered/dead) individuals

$$\left(\frac{ds}{dt}+\frac{di}{dt}+\frac{dr}{dt}=0\right)$$

$$s+i+r=1$$

$$s \rightarrow i \quad \beta \ \text{(infection rate)}$$
$$i \rightarrow r \quad \text{(recovery rate)}$$

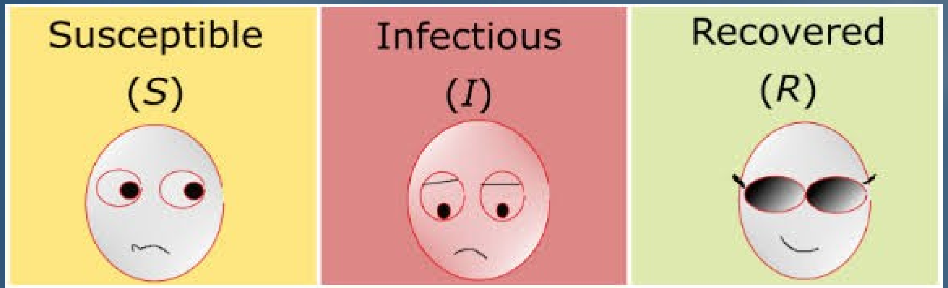
$$\frac{di}{dt}=\beta s i-\gamma i$$
$$\frac{dr}{dt}=\gamma i$$

$$\boxed{\frac{di(t)}{dt}=(\beta s-\gamma)i>0}$$

$$(s(0)\approx 1)$$

$$\boxed{\beta s(0)-\gamma>0}$$

$$\boxed{R_o\equiv\frac{\beta}{\gamma}>1}$$



Graph Laplacian

- Alternatively, we can write
- δ_{ij} is the Kronecker delta, which is 1 for $i = j$ and 0 otherwise

$$L_{ij} = \delta_{ij}k_i - A_{ij} \qquad (51)$$

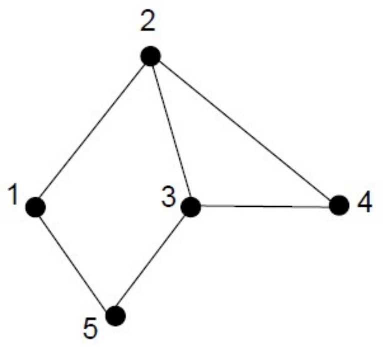


Figure: $L = D - A$

Graph Laplacian

- Alternatively, we can write
- δ_{ij} is the Kronecker delta, which is 1 for $i = j$ and 0 otherwise

$$L_{ij} = \delta_{ij} k_i - A_{ij} \tag{51}$$

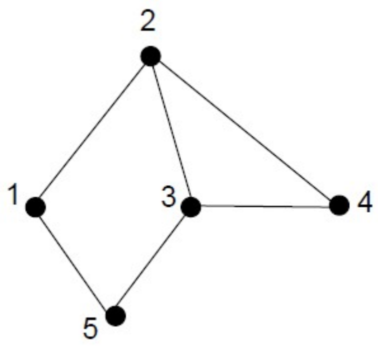


Figure: $L = D - A$

$$\mathbf{L} = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & -1 & -1 & 2 & 0 \\ -1 & 0 & -1 & 0 & 2 \end{pmatrix}$$

Graph Laplacian

- Alternatively, we can write
- δ_{ij} is the Kronecker delta, which is 1 for $i = j$ and 0 otherwise

$$L_{ij} = \delta_{ij} k_i - A_{ij} \tag{51}$$

The eigenvalues of the graph Laplacian are its most interesting property

The Laplacian is a symmetric matrix → it has real eigenvalues

We can even show that all of its eigenvalues are non-negative

Also, we can show that its smallest eigenvalue $\lambda_1 = 0$

Graph Laplacian

- The vector $\mathbf{1}$ is always an eigenvector of $\mathbf{L}$ with eigenvalue 0
- There are no negative eigenvalues, thus this is the lowest eigenvalue
- Convention: $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$
- We always have $\lambda_1 = 0$

Connected Components and the algebraic connectivity

Suppose we have a network with different components

The components have sizes

$$\mathbf{L} = \begin{pmatrix} \boxed{} & 0 & \cdots \\ 0 & \boxed{} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Connected Components and the algebraic connectivity

Suppose we have a network with different components

The components have sizes

$$\mathbf{L} = \begin{pmatrix} \boxed{} & 0 & \cdots \\ 0 & \boxed{} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

- We have n_1 ones and this is an eigenvector with eigenvalue 0
- We have c such eigenvectors

Connected Components and the algebraic connectivity

- In a network with c components c eigenvalues are equal to 0
- The second eigenvalue λ_2 of the graph Laplacian is non-zero iff the network is connected
- The second eigenvalue of the Laplacian is called *algebraic connectivity*
- It is a measure of how connected is a network, i.e. how difficult is to divide that network

Dynamics on Networks: General Model

$$\dot{x}_i = f_i(x_i) + \sum_j A_{ij} g_{ij}(x_i, x_j)$$

- f_i specifies the intrinsic dynamics of a node - it specifies how x_i would evolve without any connections
- g_{ij} specifies the contribution from the neighbors

Denis Helic

Dynamics on Networks: Diffusion Dynamics

Suppose that this commodity moves from one node to another along the links at a constant rate C

- E.g. if i and j are adjacent to each other then the amount flowing from j to i in some small time interval dt:

$$C(x_j - x_i)$$

- Similarly, the amount flowing from i to j is given by (which is the negative amount from above):

$$C(x_i - x_j)$$

Dynamics on Networks: Diffusion Dynamics

Suppose that this commodity moves from one node to another along the links at a constant rate C

- E.g. if i and j are adjacent to each other then the amount flowing from j to i in some small time interval dt:

$$C(x_j - x_i)$$

- Similarly, the amount flowing from i to j is given by (which is the negative amount from above):

$$C(x_i - x_j)$$

- Then the total amount of commodity flowing to i is given by the same of flows over all of i's neighbors:

$$\dot{x}_i = C \sum_j A_{ij}(x_j - x_i)$$

Denis Helic

Dynamics on Networks: Diffusion Dynamics

- Then the total amount of commodity flowing to i is given by the same of flows over all of i's neighbors:

$$\dot{x}_i = C \sum_j A_{ij}(x_j - x_i)$$

- By splitting the terms on the right side we obtain:

$$\begin{aligned} \dot{x}_i &= C \sum_j A_{ij} x_j - C x_i \sum_j A_{ij} = C \sum_j A_{ij} x_j - C x_i k_i \\ &= C \sum_j (A_{ij} - \delta_{ij} k_i) x_j \end{aligned}$$

- In matrix form we obtain a so-called diffusion equation:

$$\dot{\mathbf{x}} = C(\mathbf{A} - \mathbf{D})\mathbf{x} = -C\mathbf{L}\mathbf{x}$$

Denis Helic

Dynamics on Networks: Diffusion Dynamics

- In matrix form we obtain a so-called diffusion equation:

$$\dot{\mathbf{x}} = C(\mathbf{A} - \mathbf{D})\mathbf{x} = -C\mathbf{L}\mathbf{x}$$

- Linear system, which we know how to solve!
- Let us write $\mathbf{x}$ as the linear combination of the eigenvectors of the Graph Laplacian
- $\mathbf{v}_r$ is the eigenvector with eigenvalue λ_r

$$\mathbf{x}(t) = \sum_{r=1}^{n} a_r(t)\mathbf{v}_r$$

Denis Helic

Dynamics on Networks: Diffusion Dynamics

- Linear system, which we know how to solve!
- Let us write $\mathbf{x}$ as the linear combination of the eigenvectors of the Graph Laplacian
- $\mathbf{v}_r$ is the eigenvector with eigenvalue λ_r

$$\mathbf{x}(t) = \sum_{r=1}^{n} a_r(t)\mathbf{v}_r$$

- Now coefficients $a_r(t)$ vary over time
- Substituting this form in the diffusion equation:

$$\sum_{r=1}^{n} \frac{da_r(t)}{dt}\mathbf{v}_r = -C\mathbf{L}\sum_{r=1}^{n} a_r(t)\mathbf{v}_r$$

Denis Helic

Dynamics on Networks: Diffusion Dynamics

- Now coefficients $a_r(t)$ vary over time
- Substituting this form in the diffusion equation:

$$\sum_{r=1}^{n} \frac{da_r(t)}{dt}\mathbf{v}_r = -\mathrm{C}\mathbf{L}\sum_{r=1}^{n} a_r(t)\mathbf{v}_r$$

- We have $\mathbf{L}\mathbf{v}_r = \lambda_r \mathbf{v}_r$:

$$\sum_{r=1}^{n} \frac{da_r(t)}{dt}\mathbf{v}_r = -\mathrm{C}\sum_{r=1}^{n} a_r(t)\lambda_r\mathbf{v}_r$$

- Now we multiply both sides with any eigenvector v_s of the Laplacian
- Recollect that since $\mathbf{L}$ is symmetric its eigenvectors are orthogonal to each other

Denis Helic

Dynamics on Networks: Diffusion Dynamics

$$\sum_{r=1}^{n} \frac{da_r(t)}{dt}\mathbf{v}_r\mathbf{v}_s = -C\sum_{r=1}^{n} a_r(t)\lambda_r\mathbf{v}_r\mathbf{v}_s$$

- $\mathbf{v}_r\mathbf{v}_s = 0$ if $s \neq r$
- $\mathbf{v}_r\mathbf{v}_s = 1$ if $s = r$

$$\frac{da_r(t)}{dt} = -Ca_r(t)\lambda_r$$

Dynamics on Networks: Diffusion Dynamics

- The last equation has the solution (by separating variables and integrating):

$$a_r(t) = a_r(0)e^{-C\lambda_r t}$$

- The vector $\mathbf{1}$ is always an eigenvector of $\mathbf{L}$ with eigenvalue 0
- There are no negative eigenvalues, thus this is the lowest eigenva
- Convention: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$
- We always have $\lambda_1 = 0$

Dynamics on Networks: SI model

Dynamics on Networks: SI model

- For example, the network version of the SI model:

$$\frac{dx_i}{dt} = \dot{x}_i = \beta(1 - x_i) \sum_j A_{ij} x_j$$

Infected dynamics

- This equation has only terms involving pairs of variables connected by links

$$\dot{s}_i = -\beta s_i \sum_j A_{ij}(1 - s_j)$$

$s_i + x_i = 1$

Denis Helic

Dynamics on Networks: SI model

- For example, the network version of the SI model:

$$\frac{dx_i}{dt} = \dot{x}_i = \beta(1 - x_i) \sum_j A_{ij} x_j$$

Infected dynamics

Denis Helic

Dynamics on Networks: SI model

- *We start with a single infected node chosen uniformly at random, or a small fraction of nodes*
- *In the limit of large network size the initial conditions are:*
- *The equations are coupled set of* n *non-linear differential equations.*
- *Can not be solved in closed form for general* A_{ij}.
- *Let us therefore consider suitable limits.*

- *For large and assuming initial conditions as above will be small.*
- *We can ignore terms of quadratic order in small quantities.*

$$\frac{dx_i}{dt} = \dot{x}_i = \beta(1 - x_i)\sum_j A_{ij}x_j$$

$$\dot{x}_i = \beta\sum_j A_{ij}x_j$$

Denis Helic

Dynamics on Networks: SI model

$$\dot{x}_i = \beta \sum_j A_{ij} x_j$$

$$\dot{\mathbf{x}} = \beta \mathbf{A} \mathbf{x}$$

$$\mathbf{x}(t) = \sum_{r=1}^{n} a_r(t) \mathbf{v}_r$$

$$\dot{\mathbf{x}} = \sum_{r=1}^{n} \frac{da_r}{dt} \mathbf{v}_r = \beta \mathbf{A} \sum_{r=1}^{n} a_r(t) \mathbf{v}_r = \beta \sum_{r=1}^{n} \kappa_r a_r(t) \mathbf{v}_r$$

- Then comparing terms in $\mathbf{v}_r$:

$$\dot{a}_r = \beta \kappa_r a_r$$

Denis Helic

Dynamics on Networks: SI model

$$\mathbf{x}(t) = \sum_{r=1}^{n} a_r(t)\mathbf{v}_r$$

$$\dot{\mathbf{x}} = \sum_{r=1}^{n} \frac{da_r}{dt}\mathbf{v}_r = \beta\mathbf{A}\sum_{r=1}^{n} a_r(t)\mathbf{v}_r = \beta\sum_{r=1}^{n} \kappa_r a_r(t)\mathbf{v}_r$$

- Then comparing terms in $\mathbf{v}_r$:

$$\dot{a}_r = \beta\kappa_r a_r$$

Substituting in the previous equation:

$$\mathbf{x}(t) = \sum_{i=1}^{n} a_r(0)e^{\beta\kappa_r t}\mathbf{v}_r$$

The fastest growing term corresponds to κ_1 and assuming that it dominates over the others:

$$\mathbf{x}(t) \sim e^{\beta\kappa_1 t}\mathbf{v}_1$$

Denis Helic

Dynamics on Networks: SI model

- We expect the number of I individuals to grow exponentially
- Similarly to the fully mixed model
- Now, the exponential constant depends not just on β
- It also depends on the leading eigenvalue of the adjacency matrix (structure of the network)
- Moreover, the probability infection in this early period varies from node to node roughly as the corresponding element of the leading eigenvector

Ring : $k=4$ neighbours.

$$P(k) = \delta(k-4)$$

Now, choose a random node, i.
In the first hop ($l=1$) it has 4 nodes
In the second hop ($l=2$) it has connect to next 4 nodes. Let's assume there is maxhop: l_{max} hop (highest distance)

So, $$1 + \sum_{l=1}^{l_{max}} 4 \approx N \quad \text{(why approx.)} \longrightarrow ①$$

$$\Rightarrow \quad 1 + 4 l_{max} = N$$
$$l_{max} \approx \frac{N}{4} \longrightarrow ②$$

For one particular node (Sum of all distances for $N \to \infty$)

$$4\left(1+2+\cdots+\frac{N}{4}\right) = 4 \times \frac{\frac{N}{4}\left(\frac{N}{4}+1\right)}{2}$$

$$\therefore \langle l \rangle = \frac{N}{2} \; \frac{4 \times \frac{N}{4}\left(\frac{N}{4}+1\right)}{2 \times \frac{N(N-1)}{2}} = \frac{1}{4} \; \frac{N(N+4)}{N(N-1)} \times \frac{N}{2}$$

$$\approx \frac{N}{2 \times 4}$$

For N nodes

For double counts

(See pages 2 to 4 for ring)

if N is even

For 4th move (it will not be multiplied by 2)

$$2\times\left(1+2+3+\dots+\frac{N}{2}-1\right)+\frac{N}{2}$$

1-2 or 1-6   1-3 or 1-5   1 to 4

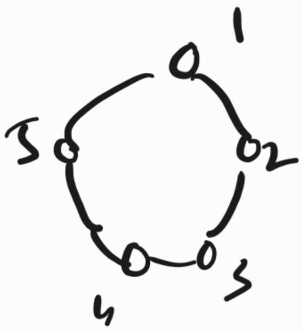
(F, S 1)

$$=2\times\frac{\left(\frac{N}{2}-1\right)\left(\frac{N}{2}-1+1\right)}{2}+\frac{N}{2}$$

$$=\frac{N(N-2)}{\cancel{2}\,4}+\frac{N}{2}$$

$$=\frac{N^2-2N+2N}{4}=\frac{N^2}{4}$$

If N is odd



$$2\times\left(1+2+3+\dots+\frac{N-1}{2}\right)$$

$$=\frac{2\times\left(\frac{N-1}{2}\right)\left(\frac{N-1}{2}+1\right)}{2}=\frac{(N-1)(N+1)}{4}=\frac{N-1}{4}$$

For odd $\quad \dfrac{N(N^2-1) \times 2}{8 \times N(N-1)} = \dfrac{N+1}{4}$

$$\sim \frac{N}{4}$$

What about degree : k

$$\boxed{\langle l \rangle \approx \frac{N}{2K}}$$

Avg. pathlength of chain

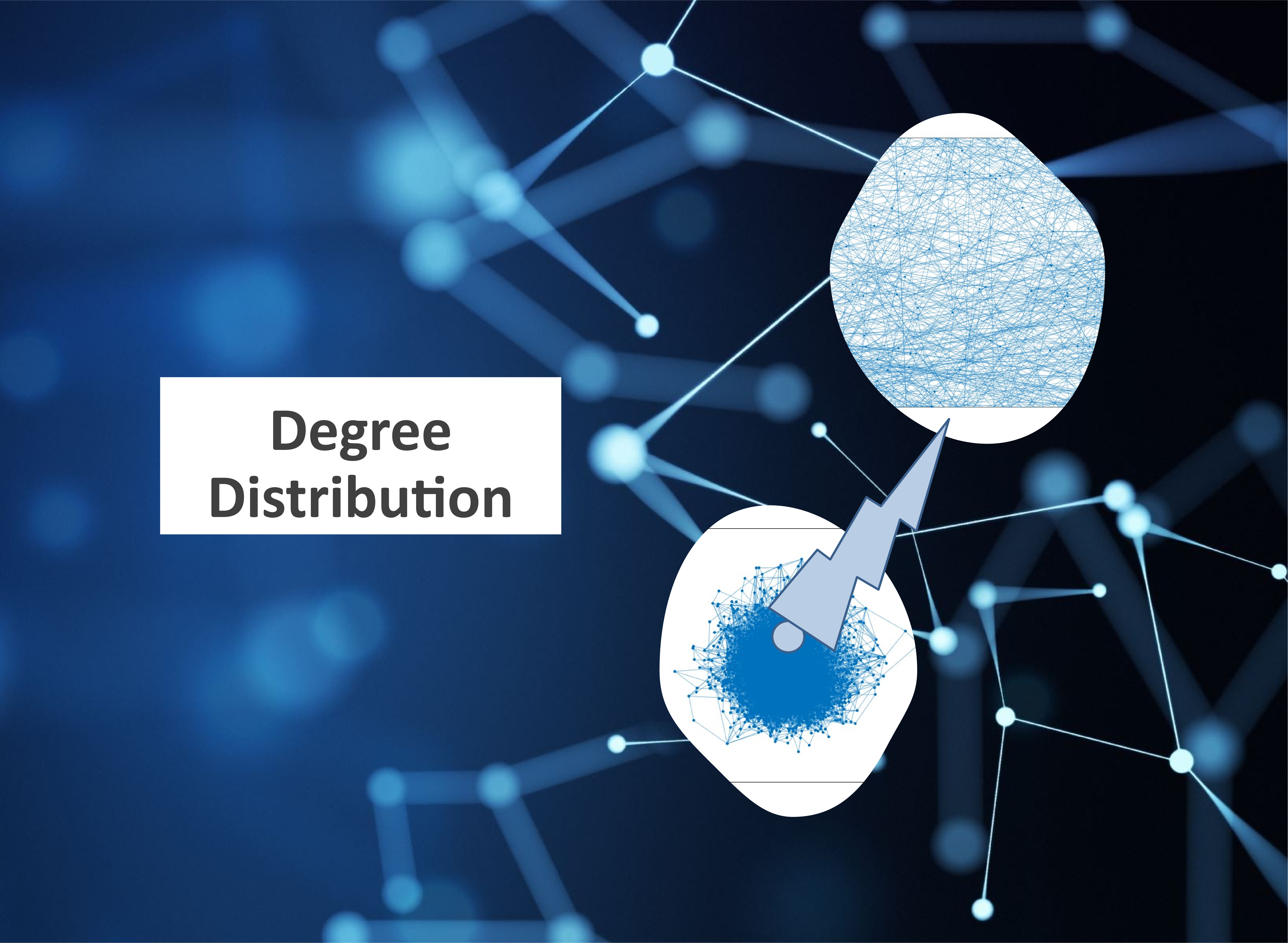
$A - B - C - D - E \qquad N=5$

pathlength of 1 : $\quad AB \,/\, BC \,/\, CD \,/\, D\text{-}E$

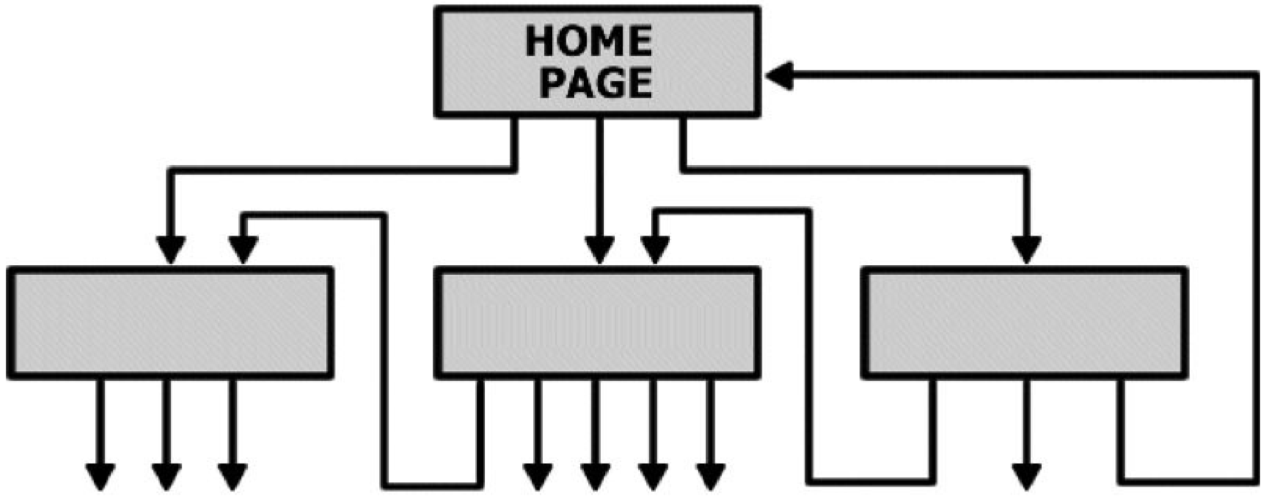
$L_1 : \quad N-1$

Pathlength of 2 : $\quad \overline{AC} \,/\, \overline{BD} \,/\, \overline{CD}$

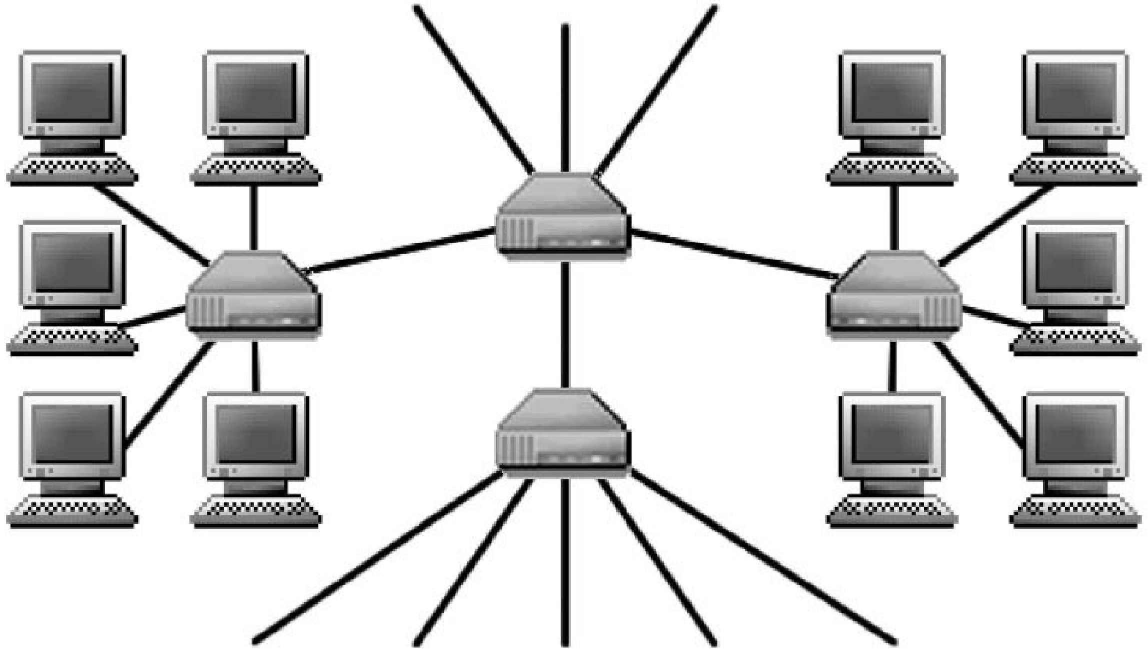
$L_2 : \quad N-2$


Degree
Distribution

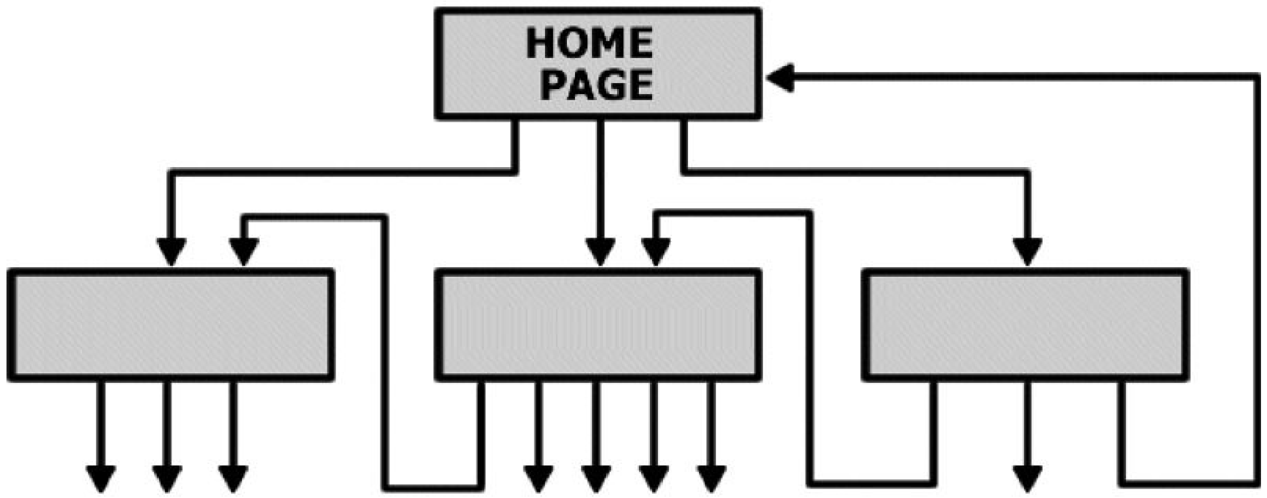
WORLD-WIDE WEB



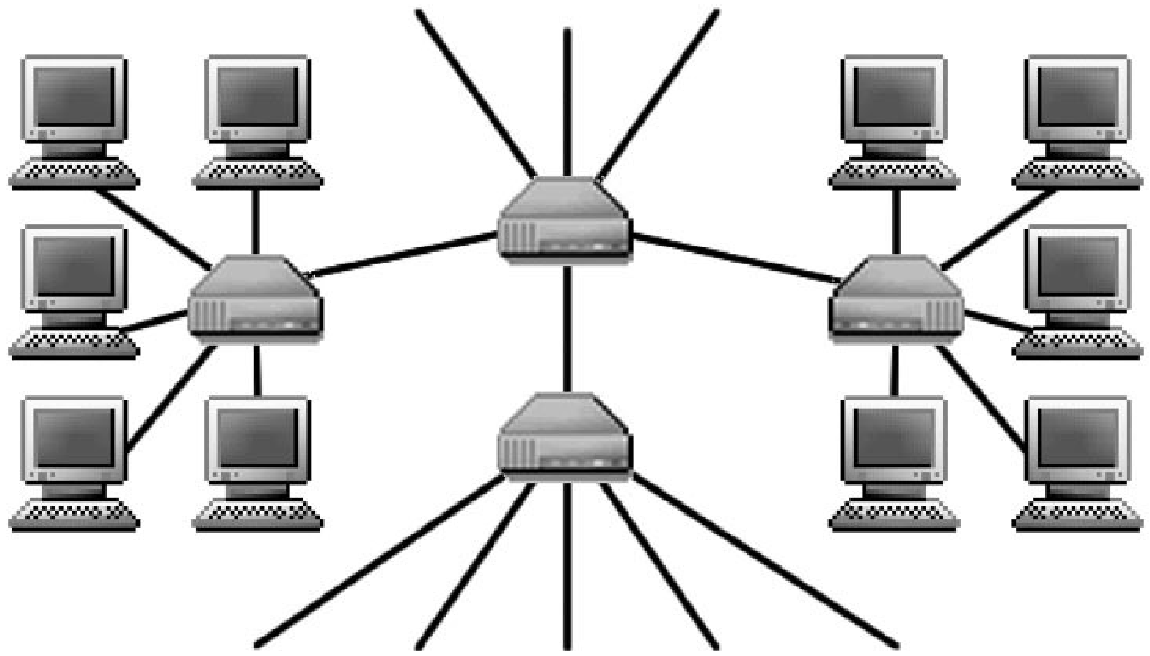
INTERNET



WORLD-WIDE WEB



INTERNET



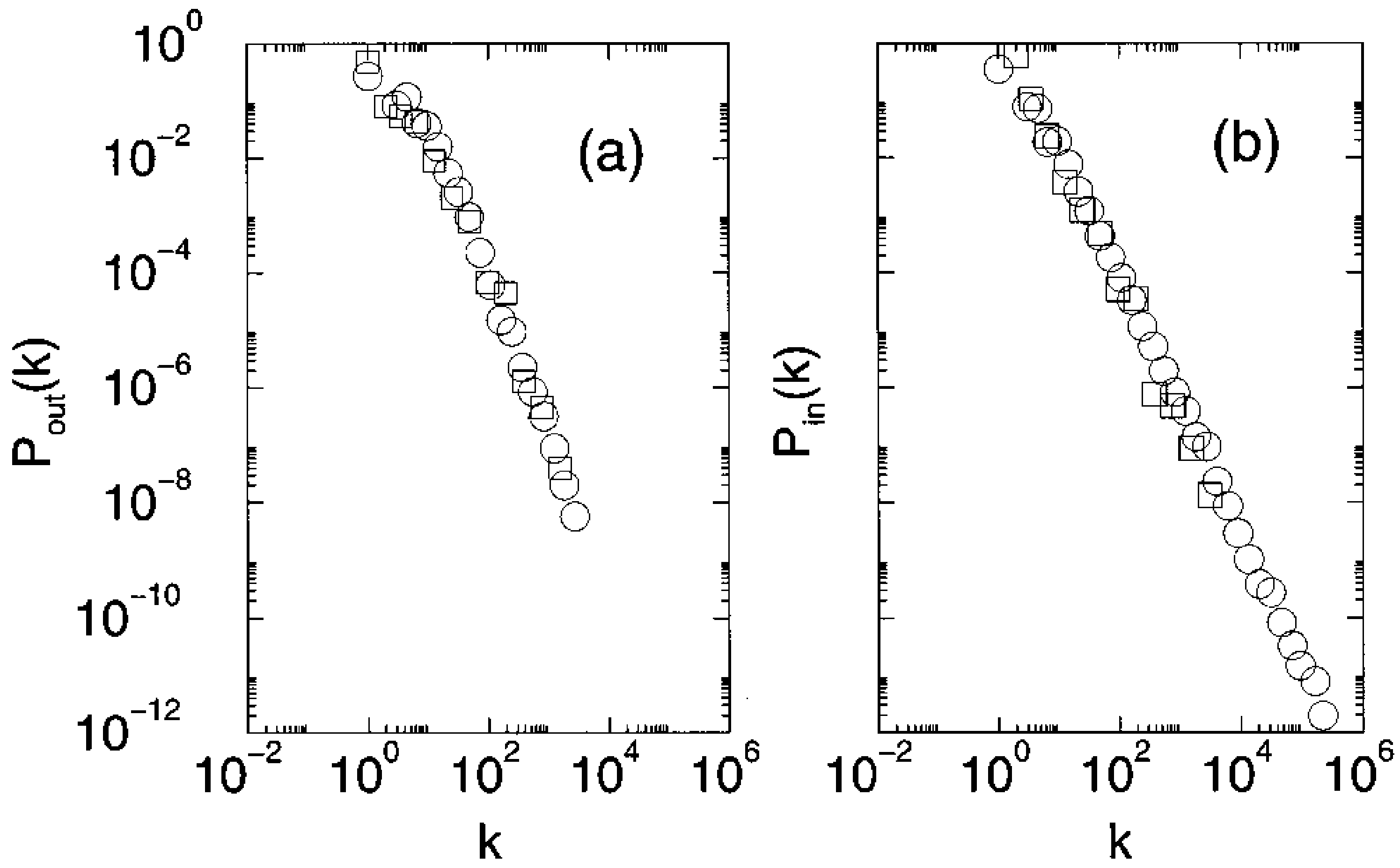
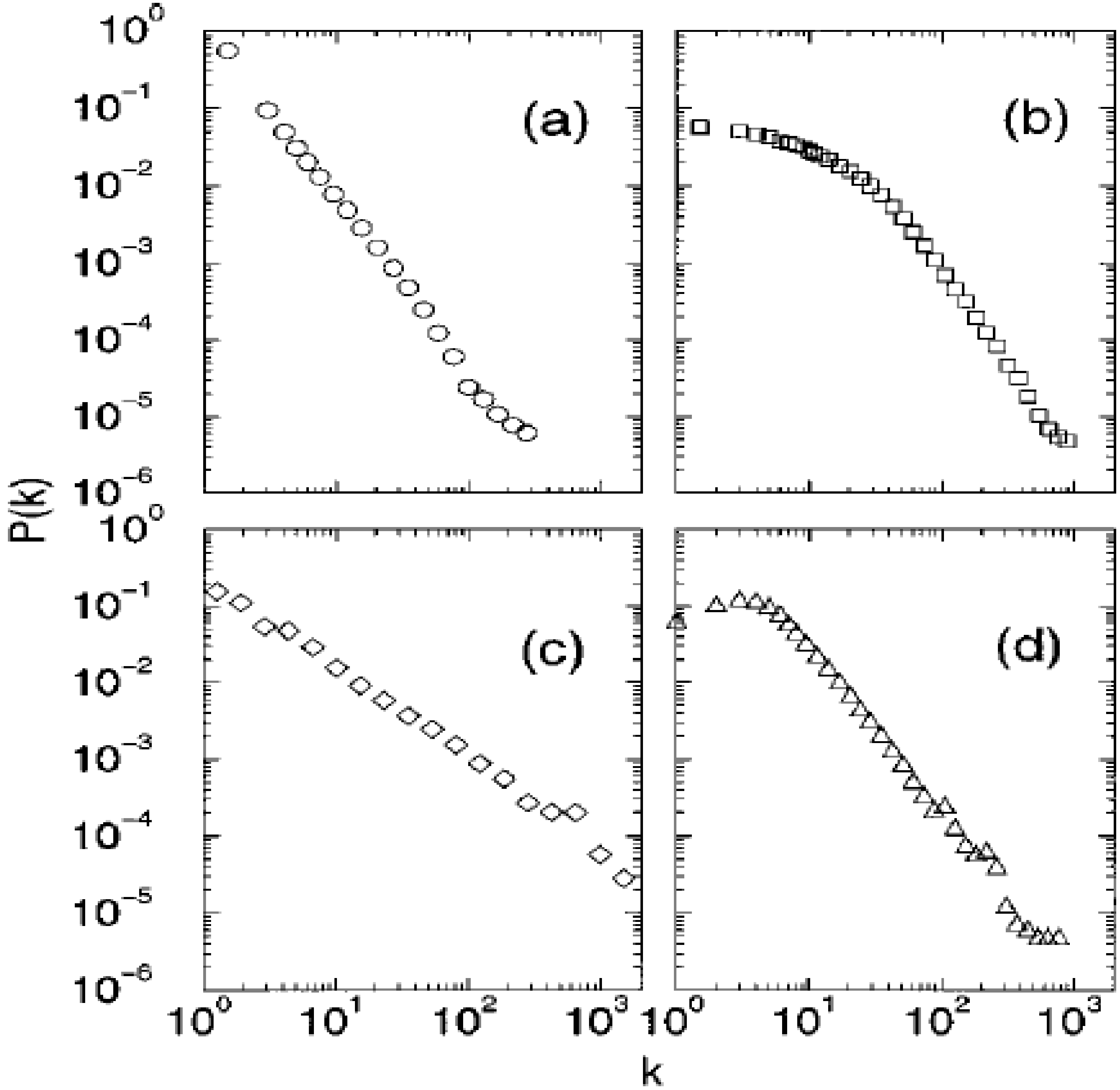


FIG. 2. Degree distribution of the World Wide Web from two different measurements: □, the 325 729-node sample of Albert *et al.* (1999); ○, the measurements of over 200 million pages by Broder *et al.* (2000); (a) degree distribution of the outgoing edges; (b) degree distribution of the incoming edges. The data have been binned logarithmically to reduce noise. Courtesy of Altavista and Andrew Tomkins. The authors wish to thank Luis Amaral for correcting a mistake in a previous version of this figure (see Mossa *et al.*, 2001).



P(k)
k
(a)
(b)
(c)
(d)

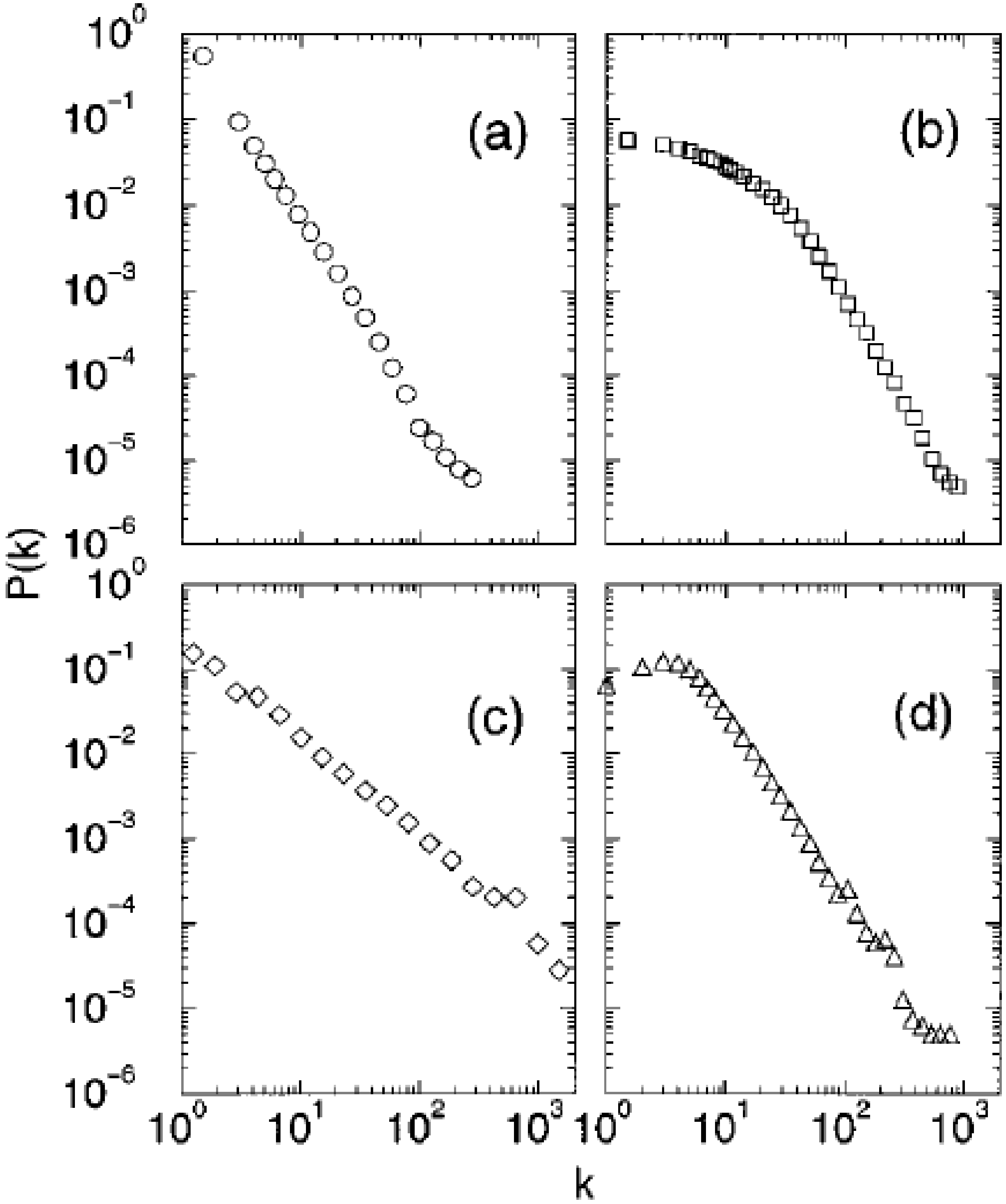


FIG. 3. The degree distribution of several real networks: (a) Internet at the router level. Data courtesy of Ramesh Govindan; (b) movie actor collaboration network. After Barabási and Albert 1999. Note that if TV series are included as well, which aggregate a large number of actors, an exponential cutoff emerges for large *k* (Amaral *et al.*, 2000); (c) co-authorship network of high-energy physicists. After Newman (2001a, 2001b); (d) co-authorship network of neuroscientists. After Barabási *et al.* (2001).

RMP-Barabasi

Power laws, Pareto distributions and Zipf's law

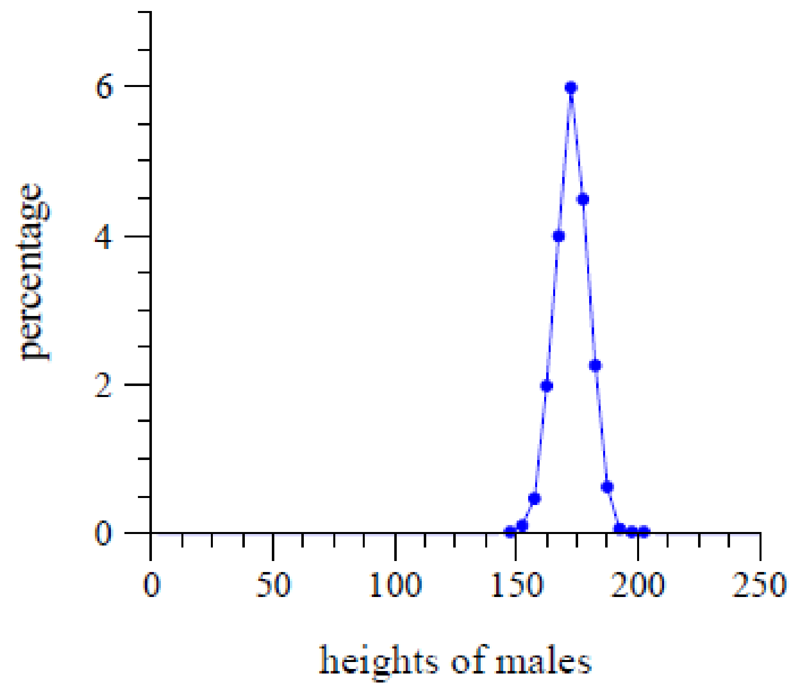
M. E. J. Newman

Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109. U.S.A.

- *Many of the things that scientists measure have a typical size or "scale"—a typical value around which individual measurements are centred.*

- *Many of the things that scientists measure have a typical size or "scale"—a typical value around which individual measurements are centred.*

- **Histogram** *of people's heights*



Heights in centimetres of American males. Data from the National Health Examination Survey, 1959–1962 (US Department of Health and Human Services).

https://aaronclauset.github.io/powerlaws/data.htm

Histogram : Linear binning

Key Steps in Linear Binning

1. Define the Range*: Determine the minimum and maximum values of the dataset.*

2. Divide into Bins*: Split the range into equal-sized intervals or bins.*

3. Assign Data Points*: Place each data point into the corresponding bin based on its value.*

4. Aggregate*: Optionally, compute aggregate statistics (e.g., count, sum, or average) for each bin*

Histogram : Linear binning

Key Steps in Linear Binning

1. **Define the Range***: Determine the minimum and maximum values of the dataset.*
2. **Divide into Bins***: Split the range into equal-sized intervals or bins.*
3. **Assign Data Points***: Place each data point into the corresponding bin based on its value.*
4. **Aggregate***: Optionally, compute aggregate statistics (e.g., count, sum, or average) for each bin*

Suppose you have a dataset of values:

- **Range***: Minimum = 2, Maximum = 15*
- **Number of Bins***: 3*
- **Bin Width***: (15−2)/3=4.33 (round to 5 for simplicity)*

Histogram : Linear binning

Key Steps in Linear Binning

1. Define the Range*: Determine the minimum and maximum values of the dataset.*

2. Divide into Bins*: Split the range into equal-sized intervals or bins.*

3. Assign Data Points*: Place each data point into the corresponding bin based on its value.*

4. Aggregate*: Optionally, compute aggregate statistics (e.g., count, sum, or average) for each bin*

Suppose you have a dataset of values:

- **Range***: Minimum = 2, Maximum = 15*
- **Number of Bins***: 3*
- **Bin Width***: (15−2)/3=4.33 (round to 5 for simplicity)*

Data points in bins:

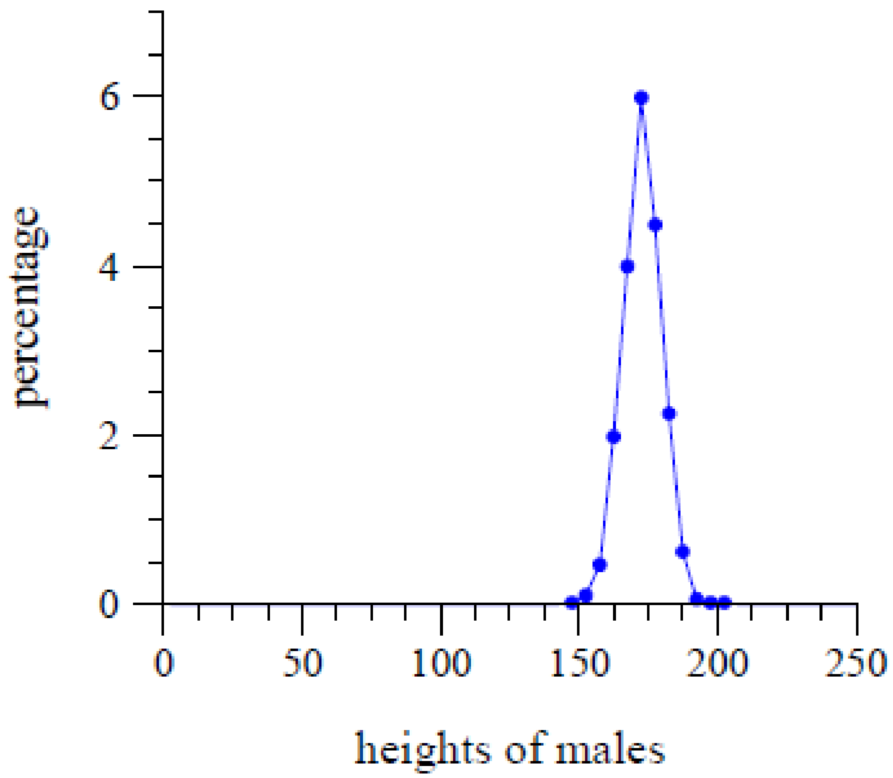
- *Bin 1: [2, 5]*
- *Bin 2: [7.3]*
- *Bin 3: [13, 15]*

Histogram : Linear binning

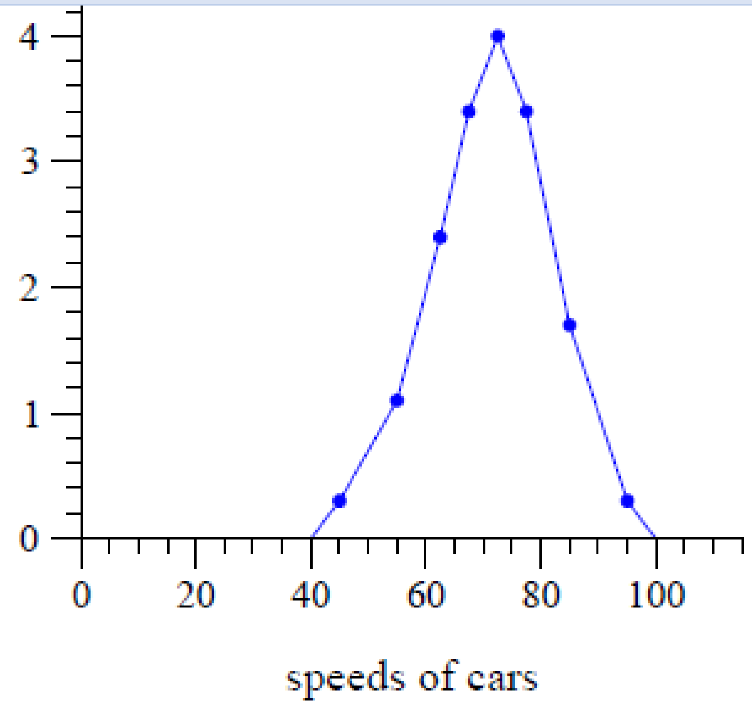
Applications

- **Histograms***: To visualize the distribution of data.*
- **Data Compression***: Reducing the number of data points by grouping them into bins.*
- **Signal Processing***: Discretizing continuous signals.*
- **Numerical Methods***: Approximating functions or integrals.*

- *Many of the things that scientists measure have a typical size or "scale"—a typical value around which individual measurements are centred.*



Again the histogram of speeds is strongly peaked, in this case around 75mph.



Right: histogram of speeds in miles per hour of cars on UK motorways. Data from Transport Statistics 2003 (UK Department for Transport).

Power laws, Pareto distributions and Zipf's law

M. E. J. Newman

Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109. U.S.A.

But not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range, sometimes many orders of magnitude!!!!!

- **frequency table** *of words in a text.*

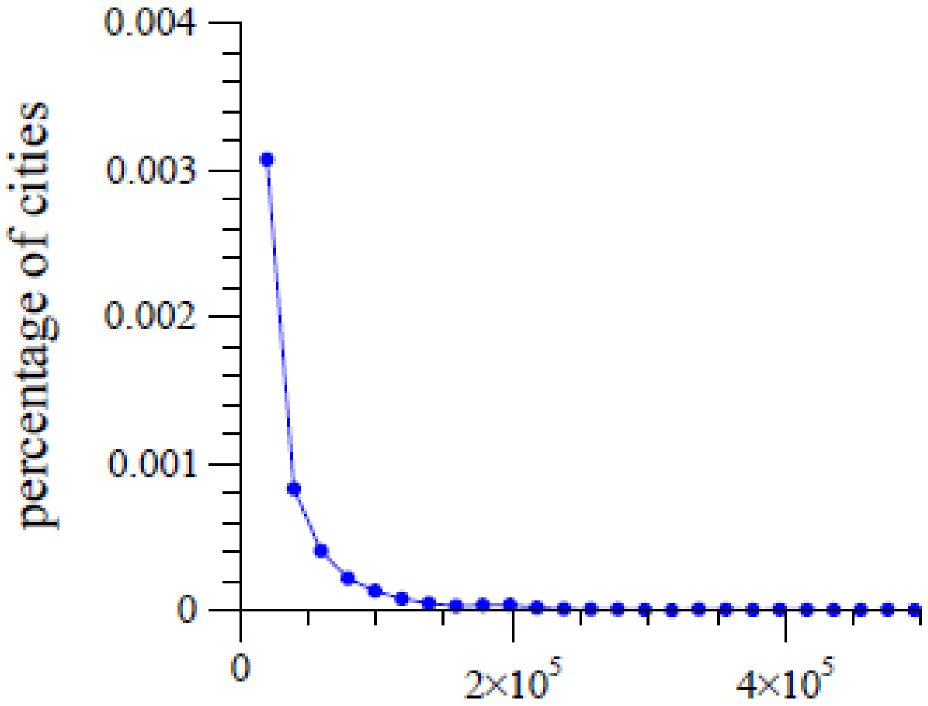
$$\text{word frequency} \propto \frac{1}{\text{word rank}}.$$

Zipf's law *(/zɪf/; German pronunciation: [tsɪpf]) is an empirical law stating that when a list of measured values is sorted in decreasing order, the value of the n-th entry is often approximately inversely proportional to n.*

George Kingsley Zipf (Statistics, linguistics)

But not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range, sometimes many orders of magnitude!!!!!

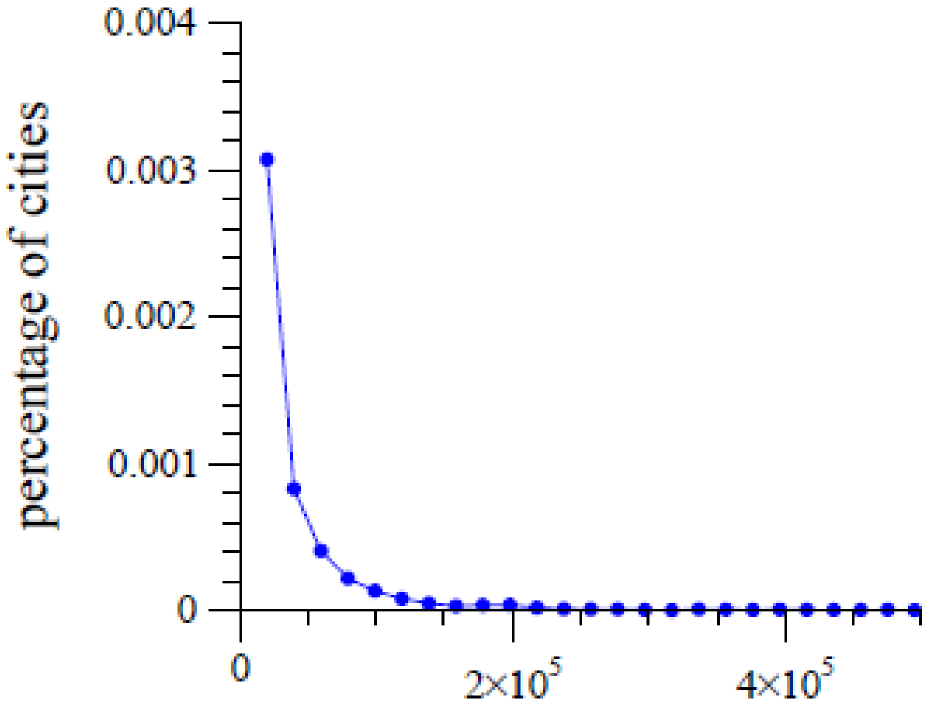
But not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range, sometimes many orders of magnitude!!!!!



Histogram of the populations of all US cities with population of 10000 or more.

The largest city in the US is New York City, with over 8.5 million residents. Los Angeles and Chicago follow, each with more than 2.5 million residents, and southern US cities Houston and Phoenix round out the top five with populations of almost 2.3 million and 1.6 million, respectively.

But not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range, sometimes many orders of magnitude!!!!!

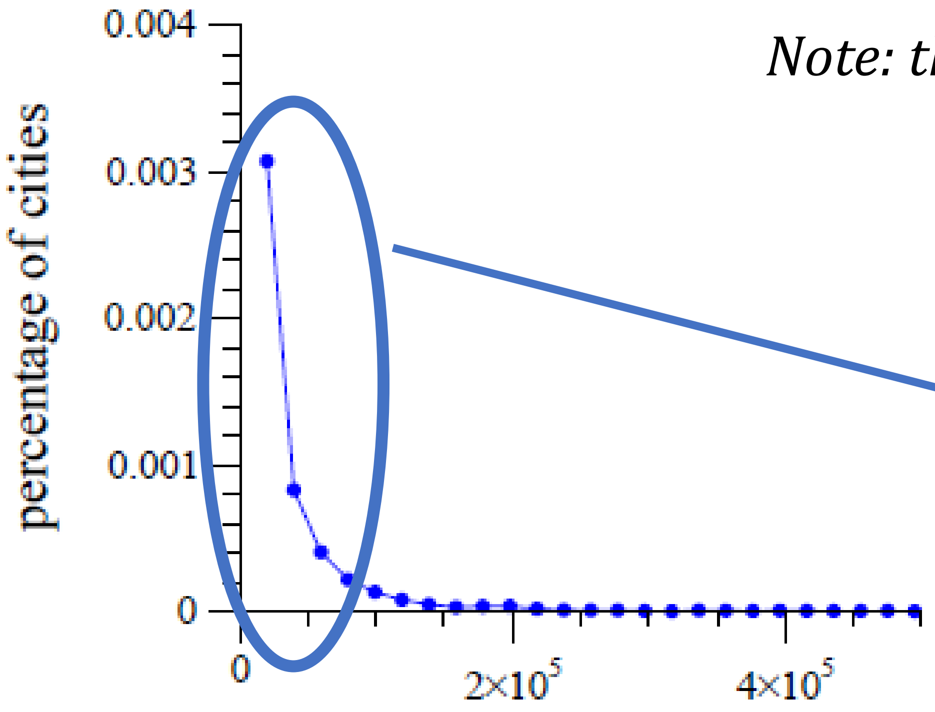


Histogram of the populations of all US cities with population of 10000 or more.

Monowi, Nebraska

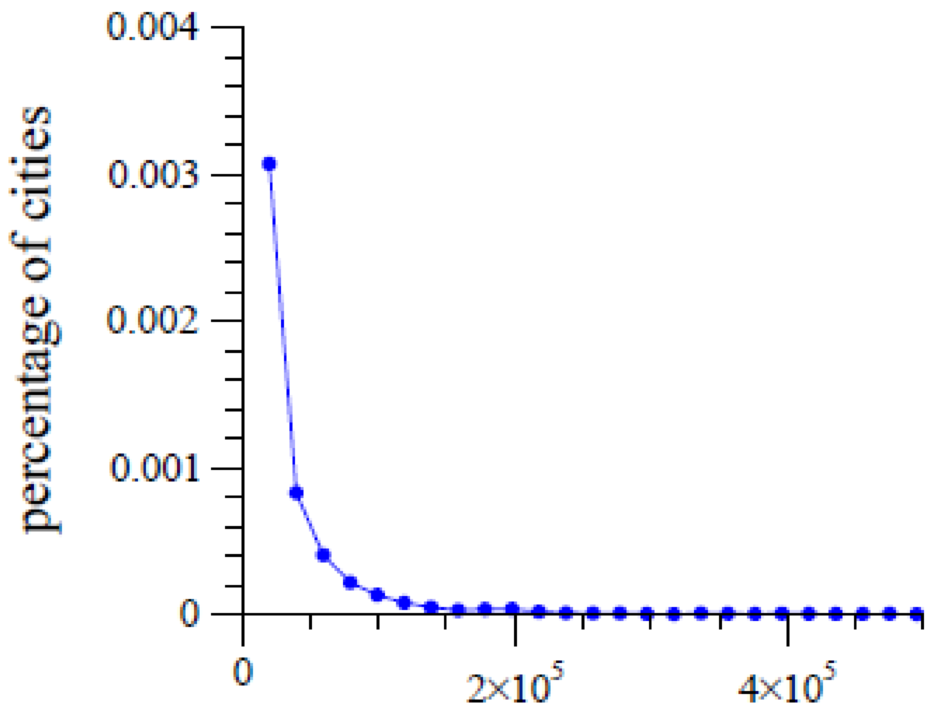
During the 2000 census, the village had a total population of two; only one married couple, Rudy and Elsie Eiler, lived there. Rudy died in 2004, leaving his wife as the only resident in the village. In this capacity, she acts as mayor, and granted herself a liquor license. She is required to produce a municipal road plan every year in order to secure state funding for the village's four street lights.

But not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range, sometimes many orders of magnitude!!!!!

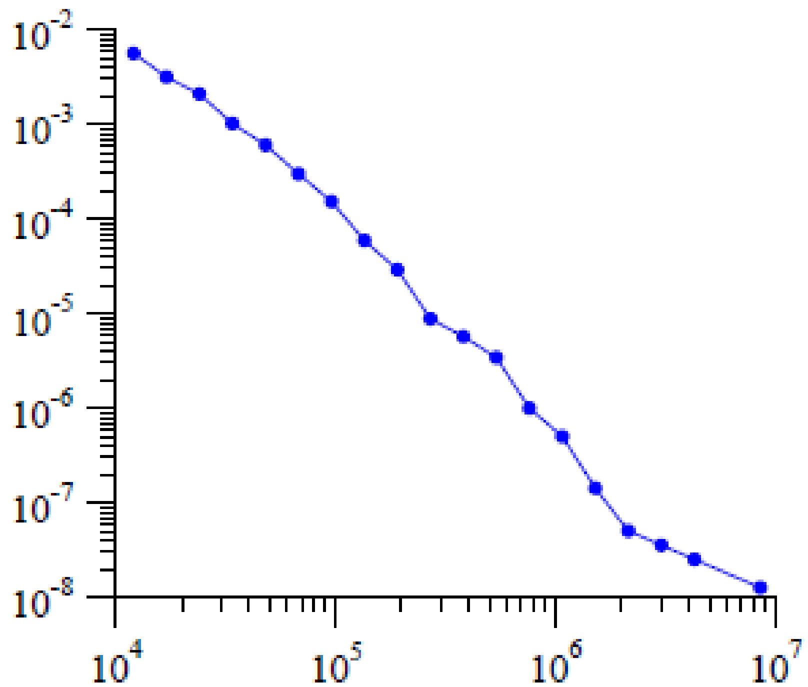


Note: the ratio of largest to smallest population is at least .

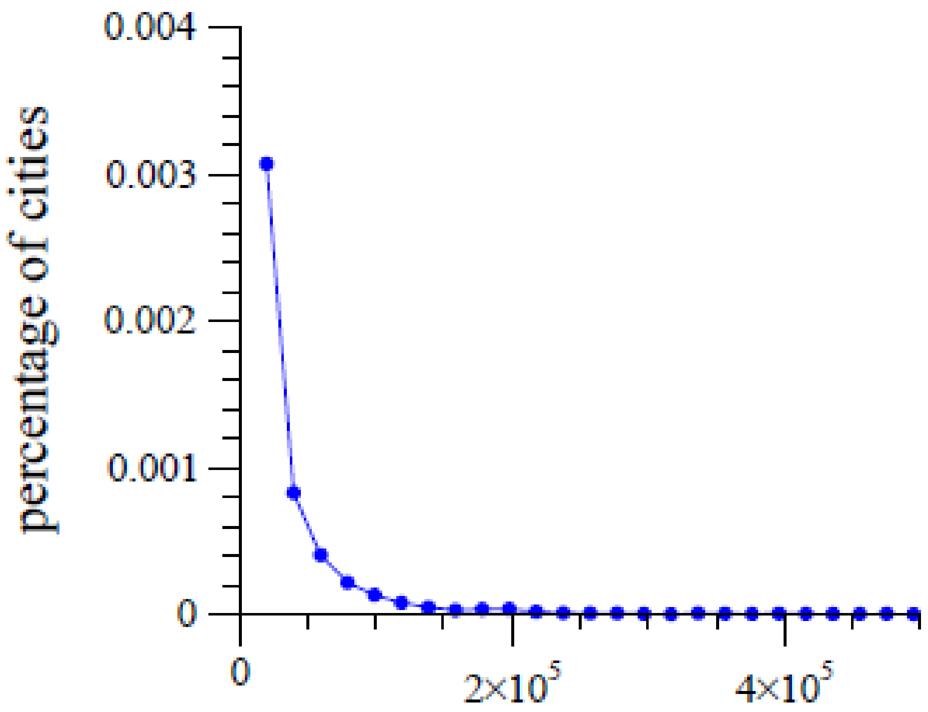
The histogram is highly right-skewed, meaning that while the bulk of the distribution occurs for fairly small sizes most US cities have small populations—there is a small number of cities with population much higher than the typical value, producing the long tail to the right of the histogram.

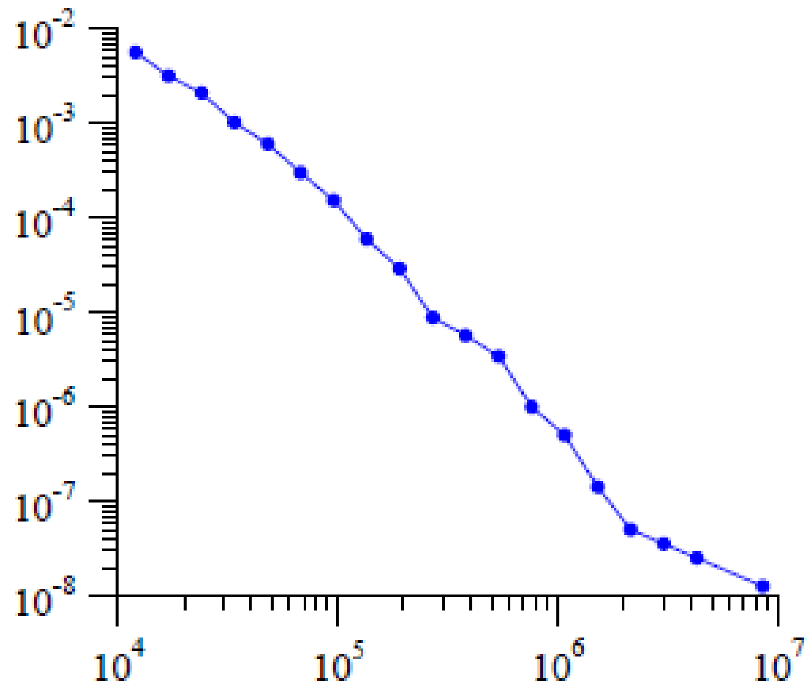


This right-skewed form is qualitatively quite different from the histograms of people's heights, but is not itself very surprising. Given that we know there is a large dynamic range from the smallest to the largest city sizes, we can immediately deduce that there can only be a small number of very large cities



The histogram of city sizes again, but this time replotted with logarithmic horizontal and vertical axes. Now a remarkable pattern emerges: the histogram, when plotted in this fashion, follows quite closely a straight line.

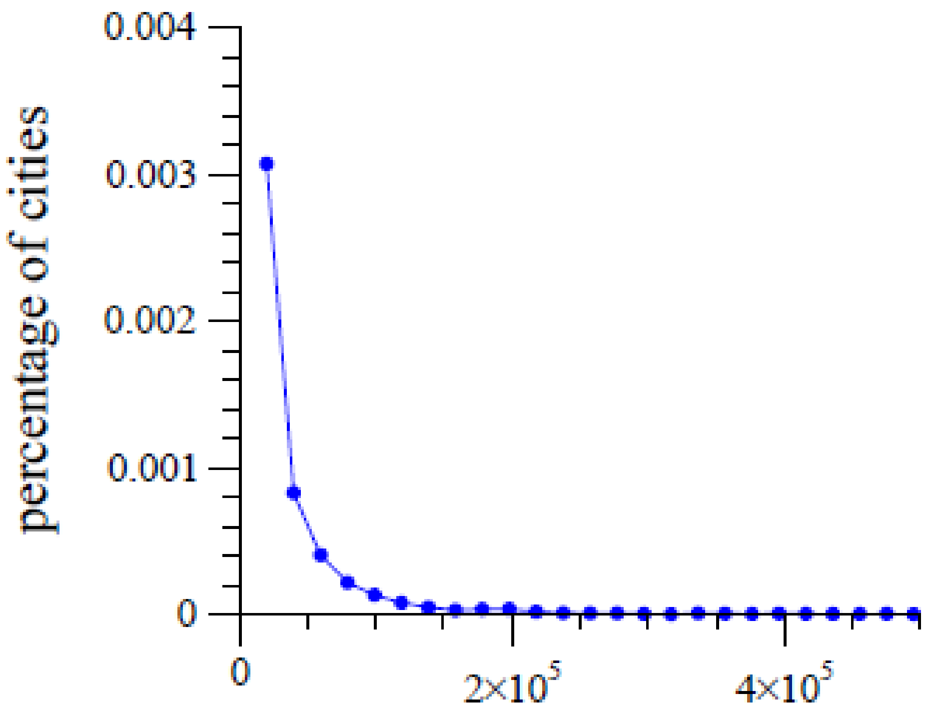


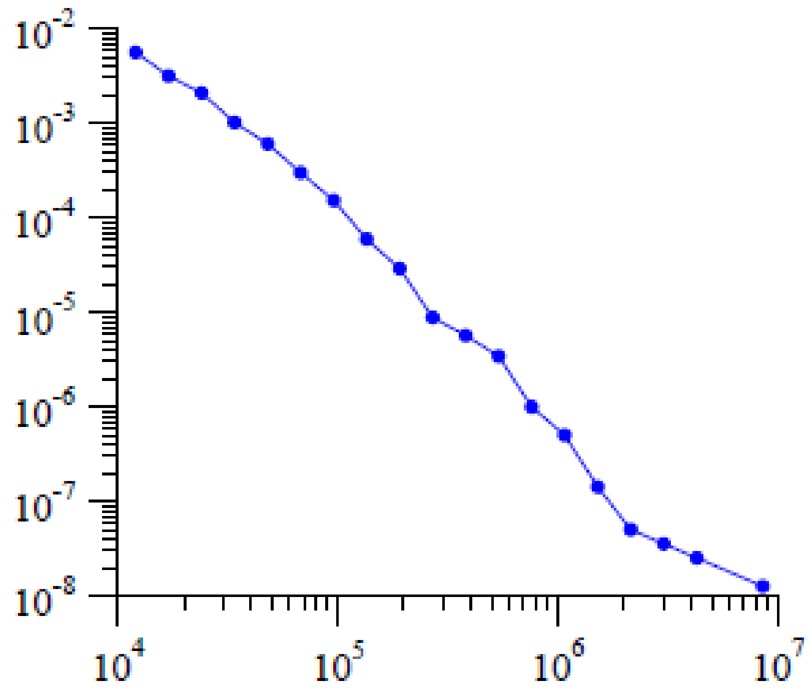


This observation seems first to have been made by Auerbach, although it is often attributed to *Zipf.*
What does it mean?

Let be the fraction of cities with population between and
If the histogram is a straight line on log-log scale, then where and are constants.

- *The slope is negative here.*
- *This is actually , where ,*
- *This follows a power law.*
- *(sometimes called as scale-free). Why?*

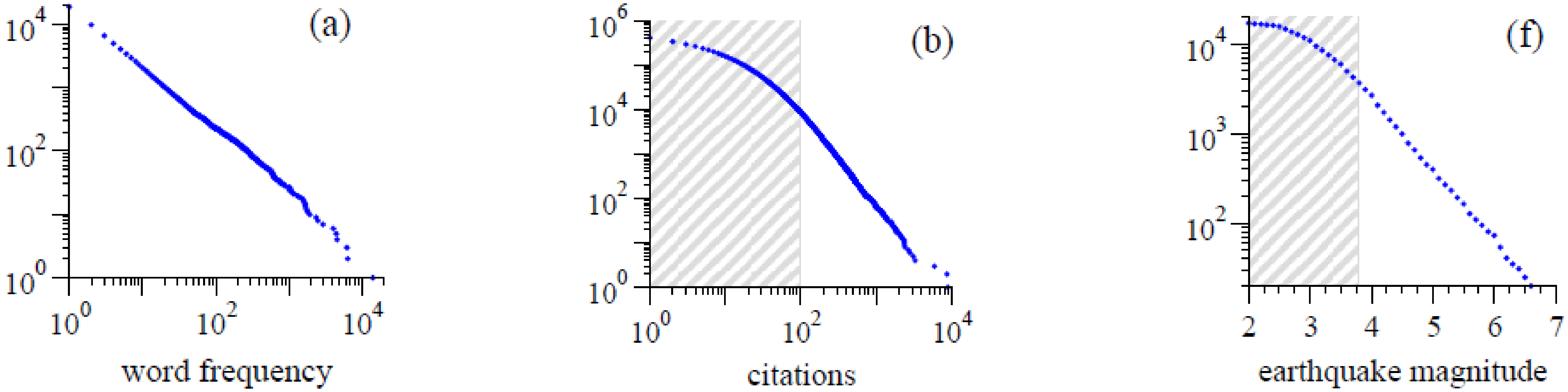




This observation seems first to have been made by Auerbach, although it is often attributed to Zipf. What does it mean?

Let be the fraction of cities with population between and
If the histogram is a straight line on log-log scale, then where and are constants.

- *The slope is negative here.*
- *This is actually , where ,*
- *This follows a power law.*
- *(sometimes called as scale-free). Why?*



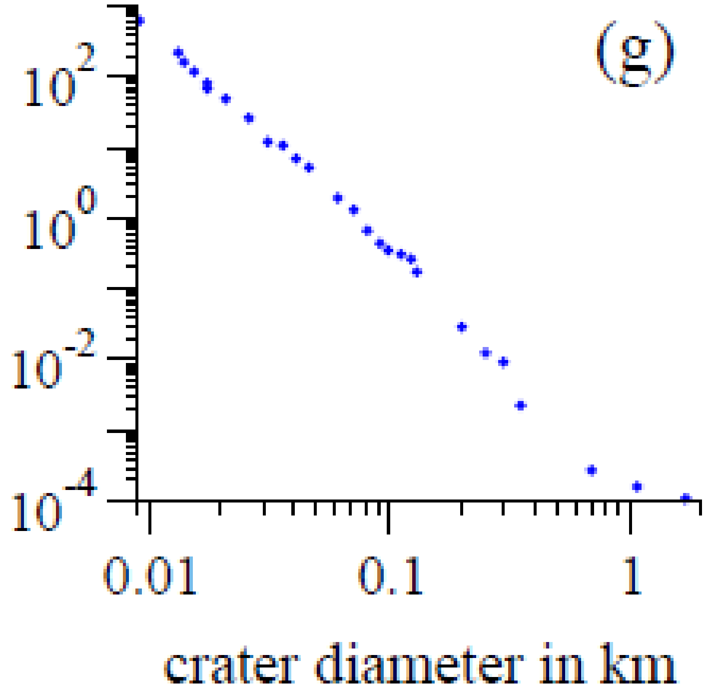
Cumulative distributions or "rank/frequency plots"

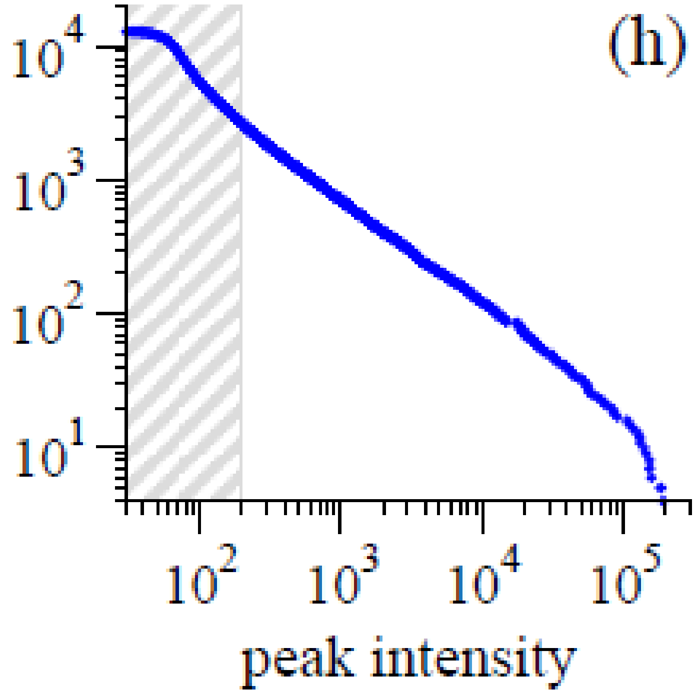
(a) Numbers of occurrences of unique words in the novel Moby Dick by Hermann Melville.

(b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997.

(f) Magnitude of earthquakes in California between January 1910 and May 1992. The Richter magnitude is defined as the logarithm, base 10, of the maximum amplitude of motion detected in the earthquake, and hence the horizontal scale in the plot, which is drawn as linear, is in effect a logarithmic scale of amplitude. The power law relationship in the earthquake distribution is thus a relationship between amplitude and frequency of occurrence. The data are from the National Geophysical Data Center, www.ngdc.noaa.gov.

https://aaronclauset.github.io/powerlaws/data.html
https://physics.bu.edu/~redner/projects/citation/isi.html





	quantity	minimum $x_{\min}$	exponent α
(a)	frequency of use of words	1	2.20(1)
(b)	number of citations to papers	100	3.04(2)
(c)	number of hits on web sites	1	2.40(1)
(d)	copies of books sold in the US	2 000 000	3.51(16)
(e)	telephone calls received	10	2.22(1)
(f)	magnitude of earthquakes	3.8	3.04(4)
(g)	diameter of moon craters	0.01	3.14(5)
(h)	intensity of solar flares	200	1.83(2)
(i)	intensity of wars	3	1.80(9)
(j)	net worth of Americans	\$600m	2.09(4)
(k)	frequency of family names	10 000	1.94(1)
(l)	population of US cities	40 000	2.30(5)

(g) Diameter of craters on the moon. Vertical axis is measured per square kilometre.

(h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989.

The observations were made between 1980 and 1989 by the instrument known as the Hard X-Ray Burst Spectrometer aboard the Solar Maximum Mission satellite launched in 1980. The spectrometer used a CsI scintillation detector to measure gamma-rays from solar flares and the horizontal axis in the figure is calibrated in terms of scintillation counts per second from this detector. The data are from the NASA Goddard Space Flight Center, umbra.nascom.nasa.gov/smm/hxrbs.html.

But these plots are cumulative!!

Power-law distributions occur in an extraordinarily diverse range of phenomena. In addition to city populations,

- *The sizes of earthquakes [3],*
- *Moon craters [4],*
- *Solar flares [5],*
- *Computer files [6] and wars [7],*
- *The frequency of use of words in any human language [2, 8],*
- *The frequency of occurrence of personal names in most cultures [9], the numbers of papers scientists write [10], the number of citations received by papers [11],*
- *The number of hits on web pages [12],*
- *The sales of books, music recordings and almost every other branded commodity [13, 14],*
- *The numbers of species in biological taxa [15],*
- *People's annual incomes [16] follows power law distributions.*

[3] B. Gutenberg and R. F. Richter, Frequency of earthquakes in california. Bulletin of the Seismological Society of America 34, 185–188 (1944).
[4] G. Neukum and B. A. Ivanov, Crater size distributions and impact probabilities on Earth from lunar, terrestrial planet, and asteroid cratering data. In T. Gehrels (ed.), Hazards Due to Comets and Asteroids, pp. 359–416, University of Arizona Press, Tucson, AZ (1994).
[5] E. T. Lu and R. J. Hamilton, Avalanches of the distribution of solar flares. Astrophysical Journal 380, 89–92 (1991).
[6] M. E. Crovella and A. Bestavros, Self-similarity in World Wide Web traffic: Evidence and possible causes. In B. E. Gaither and D. A. Reed (eds.), Proceedings of the 1996 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems, pp. 148–159, Association of Computing Machinery, New York (1996).
[7] D. C. Roberts and D. L. Turcotte, Fractality and sel forganized criticality of wars. Fractals 6, 351–357 (1998).
[8] J. B. Estoup, Gammes Stenographiques. Institut Stenographique de France, Paris (1916).
[9] D. H. Zanette and S. C. Manrubia, Vertical transmission of culture and the distribution of family names. Physica A 295, 1–8 (2001).
[10] A. J. Lotka, The frequency distribution of scientific production. J. Wash. Acad. Sci. 16, 317–323 (1926).
[11] D. J. de S. Price, Networks of scientific papers. Science 149, 510–515 (1965).
[12] L. A. Adamic and B. A. Huberman, The nature of markets in the World Wide Web. Quarterly Journal of Electronic Commerce 1, 512 (2000).
[13] R. A. K. Cox, J. M. Felton, and K. C. Chung, The concentration of commercial success in popular music: an analysis of the distribution of gold records. Journal of Cultural Economics 19, 333–340 (1995).
[14] R. Kohli and R. Sah, Market shares: Some power law results and observations. Working paper 04.01, Harris School of Public Policy, University of Chicago (2003).
[15] J. C. Willis and G. U. Yule, Some statistics of evolution and geographical distribution in plants and animals, and their significance. Nature 109, 177–179 (1922).
[16] V. Pareto, Cours d'Economie Politique. Droz, Geneva (1896).

Histogram : Linear binning

Key Steps in Linear Binning

1. Define the Range*: Determine the minimum and maximum values of the dataset.*

2. Divide into Bins*: Split the range into equal-sized intervals or bins.*

3. Assign Data Points*: Place each data point into the corresponding bin based on its value.*

4. Aggregate*: Optionally, compute aggregate statistics (e.g., count, sum, or average) for each bin*

Histogram : Linear binning

Key Steps in Linear Binning

1. **Define the Range**: *Determine the minimum and maximum values of the dataset.*
2. **Divide into Bins**: *Split the range into equal-sized intervals or bins.*
3. **Assign Data Points**: *Place each data point into the corresponding bin based on its value.*
4. **Aggregate**: *Optionally, compute aggregate statistics (e.g., count, sum, or average) for each bin*

Suppose you have a dataset of values:

- **Range**: *Minimum = 2, Maximum = 16*
- **Number of Bins**: *3*
- **Bin Width**: *(15−2)/3=4.6 (round to 5 for simplicity)*

Histogram : Linear binning

Key Steps in Linear Binning

1. Define the Range: *Determine the minimum and maximum values of the dataset.*
2. Divide into Bins: *Split the range into equal-sized intervals or bins.*
3. Assign Data Points: *Place each data point into the corresponding bin based on its value.*
4. Aggregate: *Optionally, compute aggregate statistics (e.g., count, sum, or average) for each bin*

Suppose you have a dataset of values:

- **Range**: *Minimum = 2, Maximum = 16*
- **Number of Bins**: *3*
- **Bin Width**: *(15−2)/3=4.6 (round to 5 for simplicity)*

Data points in bins:

- *Bin 1: [2, 5]*
- *Bin 2: [7.3]*
- *Bin 3: [13, 15, 16]*

Histogram : Linear binning

Applications

- **Histograms**: *To visualize the distribution of data.*
- **Data Compression**: *Reducing the number of data points by grouping them into bins.*
- **Signal Processing**: *Discretizing continuous signals.*
- **Numerical Methods**: *Approximating functions or integrals.*

Histogram : Log binning

- *Log binning is a technique used in data analysis, particularly for dealing with data that spans several orders of magnitude, such as distributions that follow* **power laws**.
- *Instead of using linear bin widths, the* **bins are spaced logarithmically***, allowing better representation of data in cases where the frequency of occurrence decreases exponentially with increasing values.*

Histogram : Log binning

Key Features of Log Binning:

1.Logarithmic Bin Widths*:*

1. *The width of the bins increases logarithmically, meaning the difference between bin boundaries grows as the values increase.*
2. *For example, bin edges might be spaced like 1,10, 100, 1000 rather than 1,2,3,4.*

2.Preserves Detail Across Scales*:*

1. *Small bins are used for low-value ranges, capturing fine details.*
2. *Larger bins are used for high-value ranges, ensuring sparsely populated regions are still well-represented.*

Histogram : Log binning

Key Features of Log Binning:

1.Logarithmic Bin Widths*:*

1. *The width of the bins increases logarithmically, meaning the difference between bin boundaries grows as the values increase.*
2. *For example, bin edges might be spaced like 1,10, 100, 1000 rather than 1,2,3,4.*

2.Preserves Detail Across Scales*:*

1. *Small bins are used for low-value ranges, capturing fine details.*
2. *Larger bins are used for high-value ranges, ensuring sparsely populated regions are still well-represented.*

3.Reduces Noise*:*

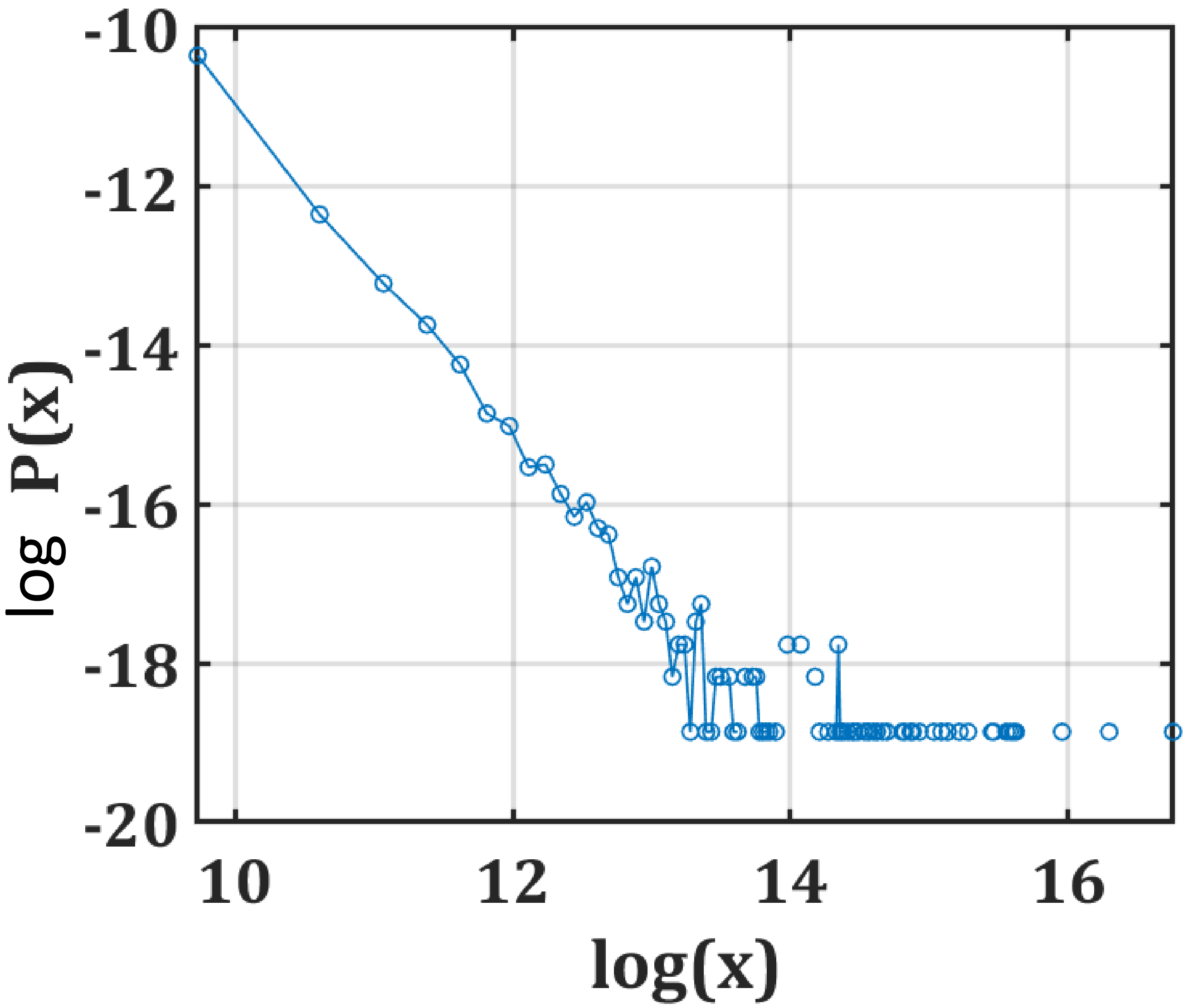
1. *In distributions with a long tail (e.g., power-law distributions), linear binning can lead to sparse or empty bins for large values. Log binning aggregates these values more effectively, reducing statistical noise.*

4.Logarithmic Representation*:*

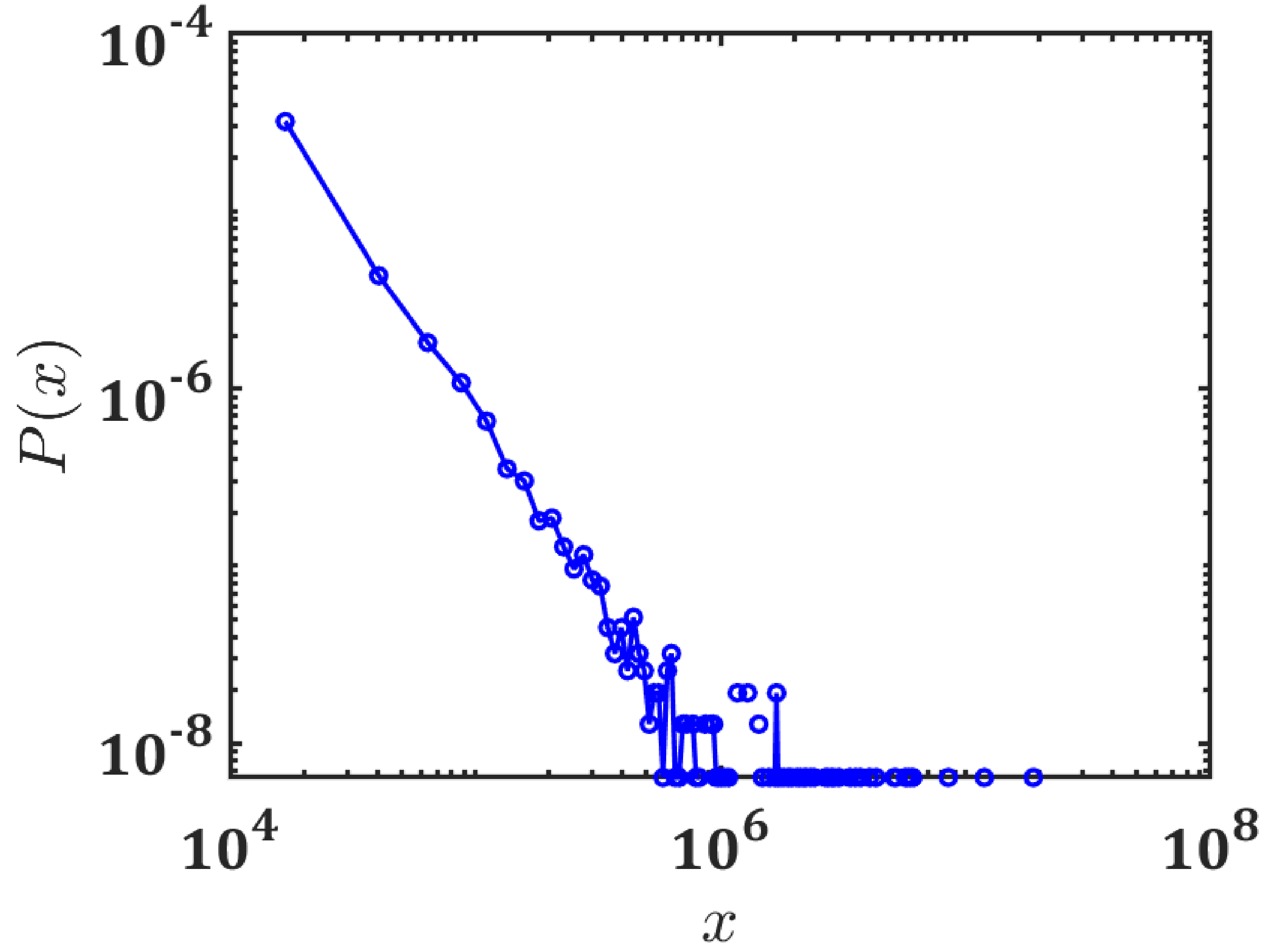
1. *Data is often plotted on a log-log scale for visualization. The x-axis represents the bin range, and the y-axis shows the frequency or probability density.*

Histogram (US population)

Linear binning and log of the binned data

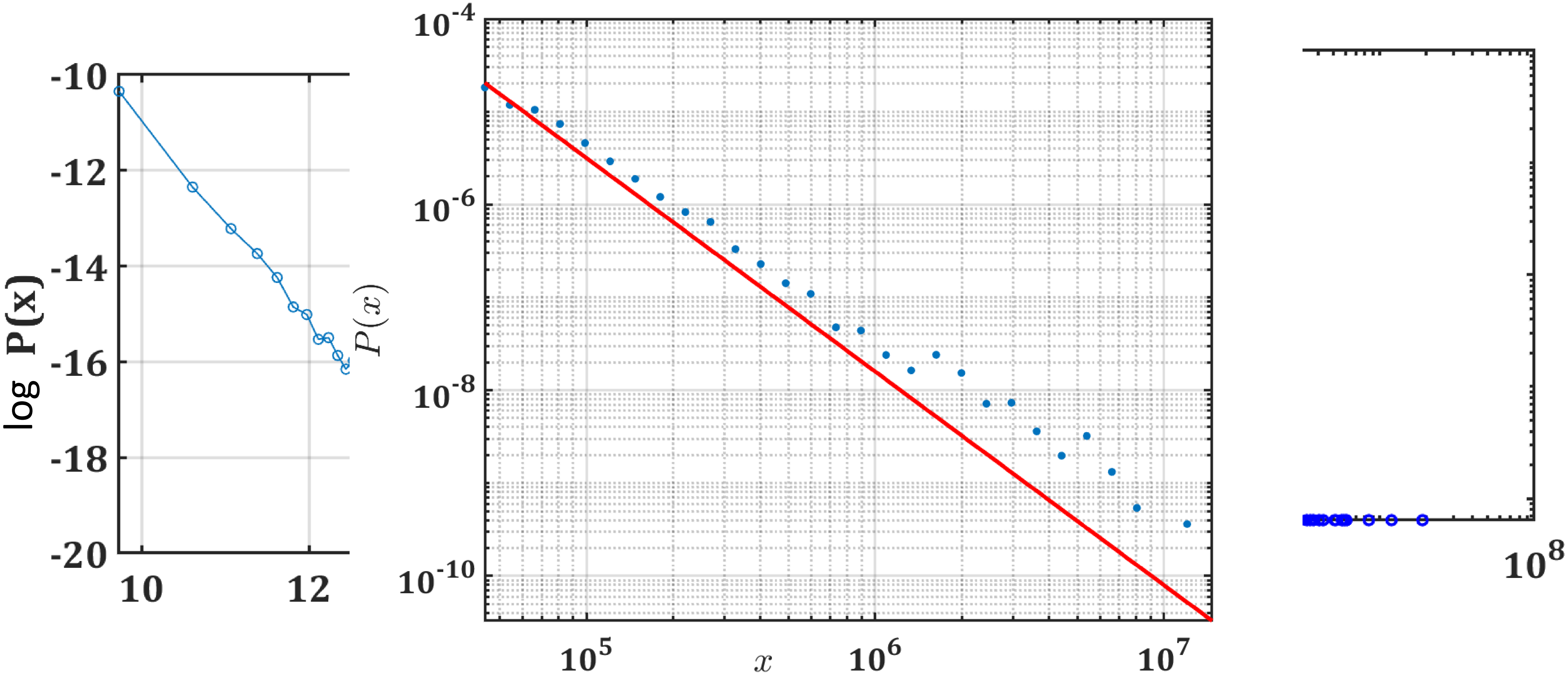


Linear binning and log-scale



Histogram (US population)

Log binning, and log scale



Histogram : Log binning

1.Choose the Base of the Logarithm*: Decide the base for your logarithmic binning (e.g., base-2, base-10, or natural log). This determines the width of the bins.*
2.Determine the Range of Data*: Identify the minimum and maximum values of the dataset. These will define the range for your bins.*
3.Define the Bin Edges*: Create bin edges using logarithmic intervals. For instance, if the base is 10:*
Bin edges = {
Or, for smaller steps:
Bin edges = {
4. Assign Data Points to Bins*: For each data point, determine which bin it belongs to by comparing its value to the bin edges. In practice, you can use algorithms for efficient bin allocation.*
5. Count Data Points in Each Bin*: After assigning all data points to bins, count the number of points in each bin.*
6. Normalize Bin Counts *: Normalize the bin counts by the bin width to account for the fact that the bin sizes grow logarithmically.*
7. Plot the Results*: Plot the histogram, often using a log-log scale to reveal power-law behavior or other patterns.*

Histogram : Linear binning and Log binning

Task : Histogram of the set of random numbers which will have a power-law distribution with exponent α = 2.5.

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}.$$

Use transformation method.

Generate a random real number uniformly distributed in the range ,
then is a random power-law-distributed real number
in the range with exponent .

Note that there has to be a lower limit on the range; the power-law distribution diverges as

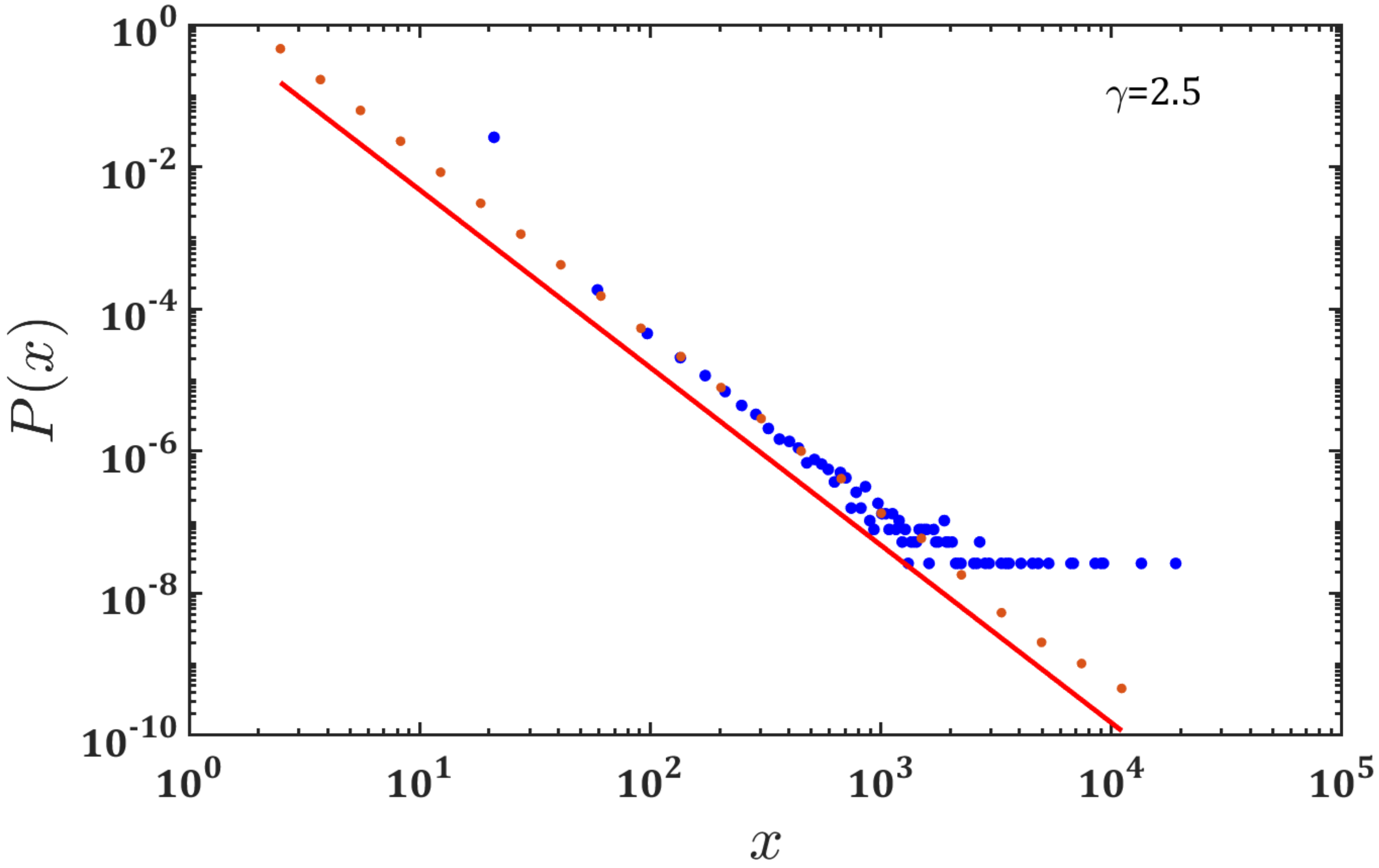
Histogram : Linear binning vs Log binning (Synthetic data)

Task : Histogram of the set of random numbers which will have a power-law distribution with exponent α = 2.5.

Use transformation method.

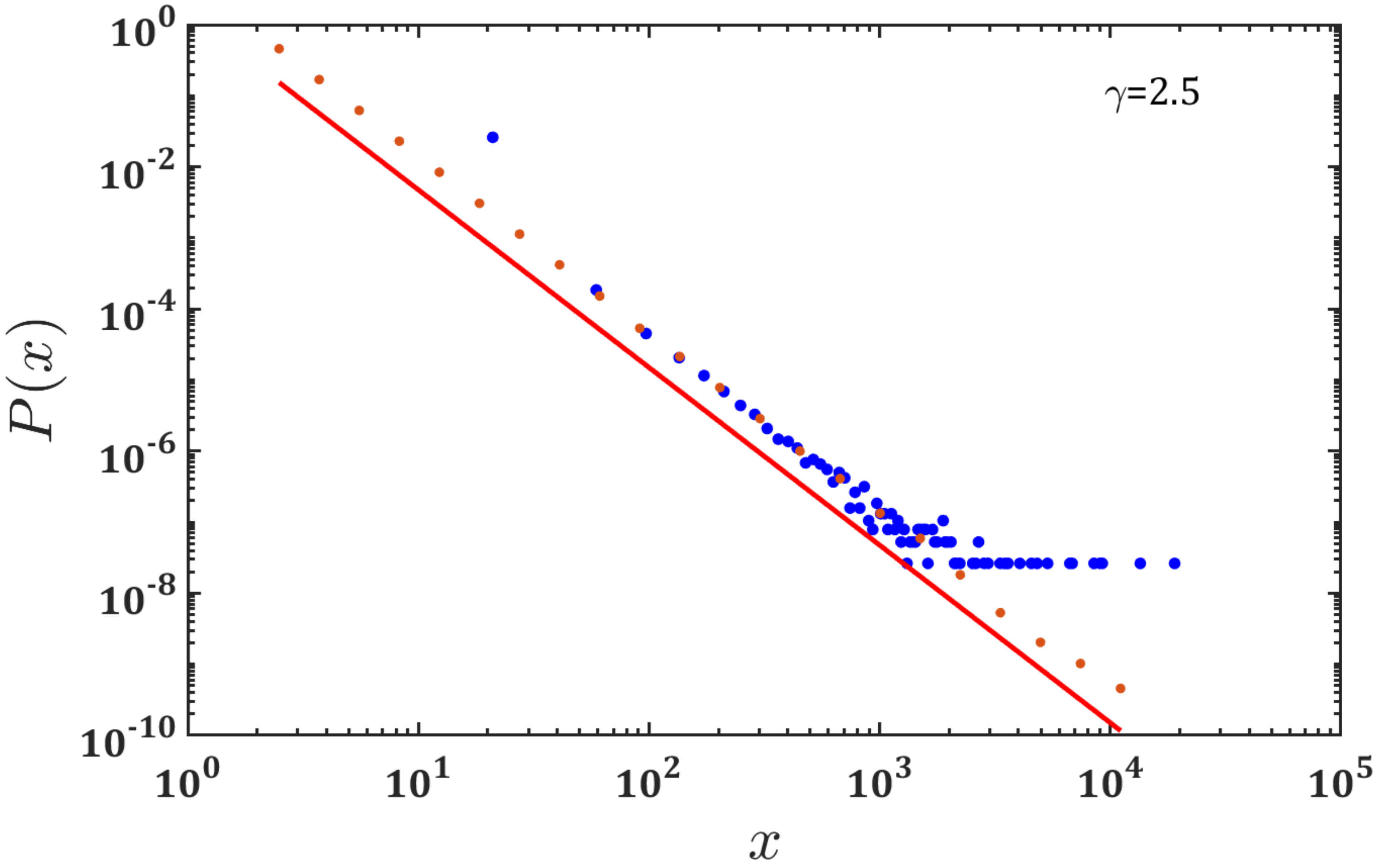
Generate a random real number uniformly distributed in the range ,
then is a random power-law-distributed real number
in the range with exponent .

Note that there has to be a lower limit on the range; the power-law distribution diverges as



Histogram : Linear binning vs Log binning (Synthetic data)

Task : *Histogram of the set of random numbers which will have a power-law distribution with exponent α = 2.5.*



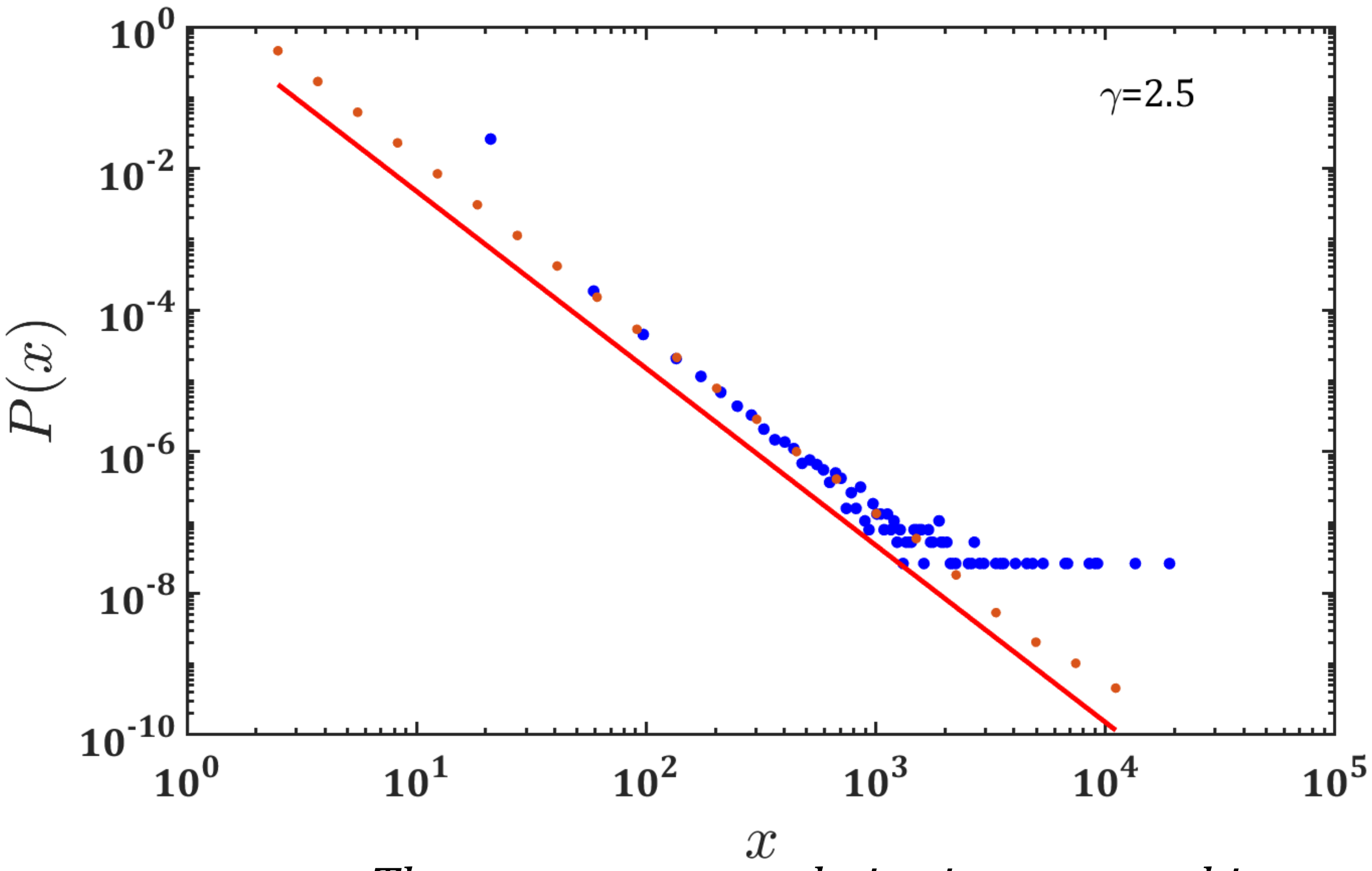
- *To reveal the power-law form of the distribution it is better, as we have seen, to plot the histogram on logarithmic scales, and when we do this for the current data we see the characteristic straight-line form of the power-law distribution (blue dots).*

- *However, the plot is in some respects not a very good one. In particular the right-hand end of the distribution is noisy because of sampling errors.*

- *The power-law distribution dwindles in this region, meaning that each bin only has a few samples in it, if any. So the fractional fluctuations in the bin counts are large and this appears as a noisy curve on the plot.*

Histogram : Linear binning vs Log binning

Task : Histogram of the set of random numbers which will have a power-law distribution with exponent α = 2.5.



- *Solution: Vary the width of the bins in the histogram. If we are going to do this, we must also normalize the sample counts by the width of bins they fall in. That is, the number of samples in a bin of width should be divided by to get a count per unit interval of .*

- *Then the normalized sample count becomes independent of bin width on average and we are free to vary the bin widths as we like.*

The most common choice is to create bins such that each is a fixed multiple wider than the one before it. This is known as logarithmic binning.

A continuous real variable with a power-law distribution has a probability $p(x)\,\mathrm{d}x$ of taking a value in the interval from x to $x+\mathrm{d}x$, where

$$p(x) = Cx^{-\alpha},$$

with $\alpha > 0$.

The constant C is given by the normalization requirement that

$$1 = \int_{x_{\min}}^{\infty} p(x)\,\mathrm{d}x = C\int_{x_{\min}}^{\infty} x^{-\alpha}\,\mathrm{d}x = \frac{C}{1-\alpha}\left[x^{-\alpha+1}\right]_{x_{\min}}^{\infty}.$$

$$C = (\alpha-1)x_{\min}^{\alpha-1},$$

$$p(x) = \frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha}.$$

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}.$$

The mean value of our power-law distributed quantity x is given by

$$\begin{aligned} \langle x \rangle &= \int_{x_{\min}}^{\infty} x p(x) \, \mathrm{d}x = C \int_{x_{\min}}^{\infty} x^{-\alpha+1} \, \mathrm{d}x \\ &= \frac{C}{2-\alpha} \left[x^{-\alpha+2} \right]_{x_{\min}}^{\infty}. \end{aligned}$$

Note that this expression becomes infinite if Power laws with such low values of have no finite mean. The distributions of sizes of solar flares and wars in are examples of such power laws.

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}.$$

For $\alpha > 2$ however, the mean is perfectly well defined, with a value given by

$$\langle x \rangle = \frac{\alpha - 1}{\alpha - 2} x_{\min}.$$

We can also calculate higher moments of the distribution $p(x)$. For instance, the second moment, the mean square, is given by

$$\langle x^2 \rangle = \frac{C}{3 - \alpha} \left[x^{-\alpha+3} \right]_{x_{\min}}^{\infty}.$$

For $\alpha > 2$ however, the mean is perfectly well defined, with a value given by Eq. (11) of

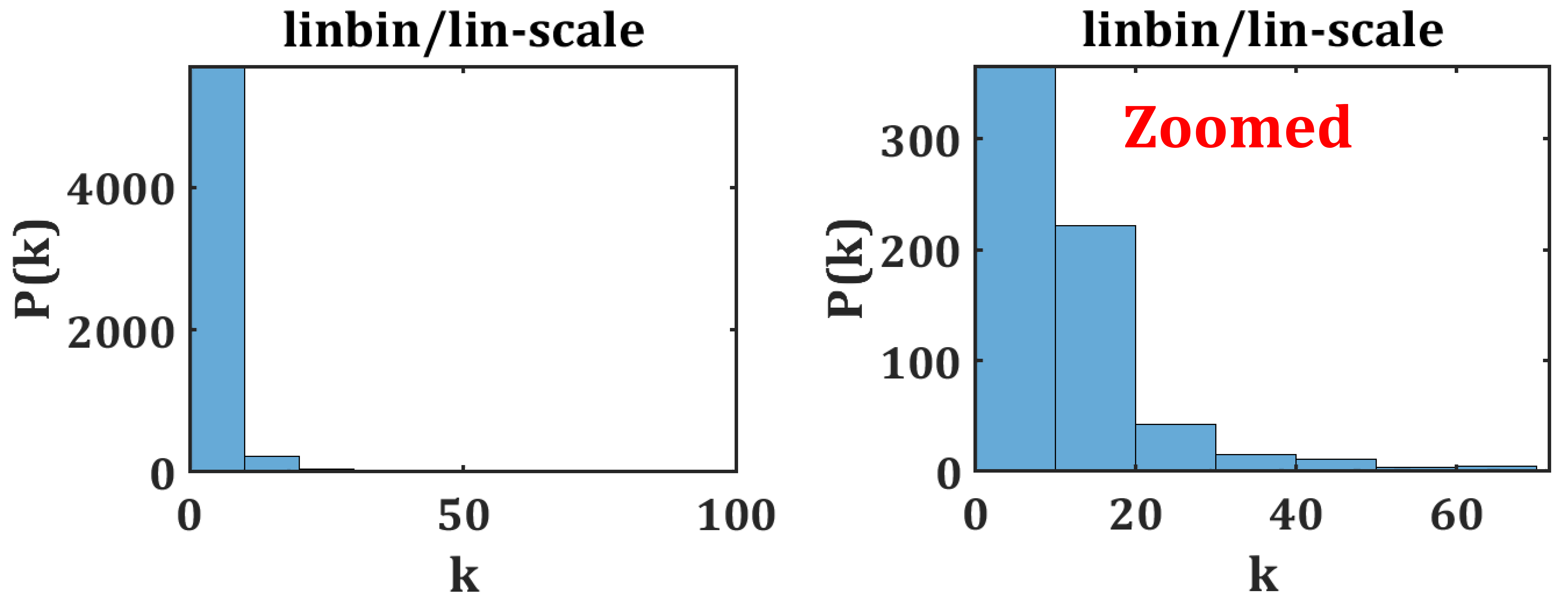
$$\langle x \rangle = \frac{\alpha - 1}{\alpha - 2} x_{\min}.$$

We can also calculate higher moments of the distribution $p(x)$. For instance, the second moment, the mean square, is given by

$$\langle x^2 \rangle = \frac{C}{3 - \alpha} \left[x^{-\alpha+3} \right]_{x_{\min}}^{\infty}.$$

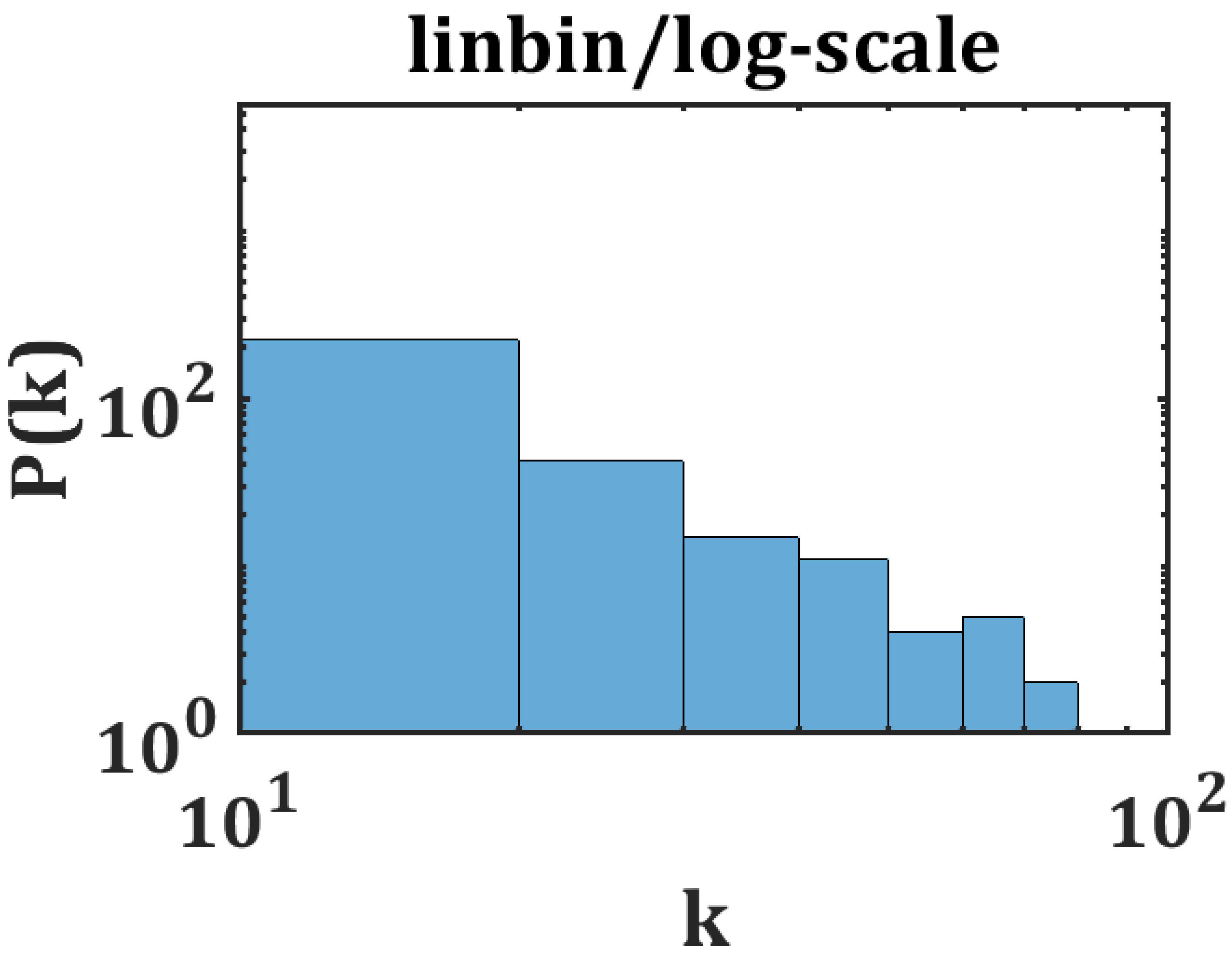
If , then the, second moment is finite and well-defined, taking the value

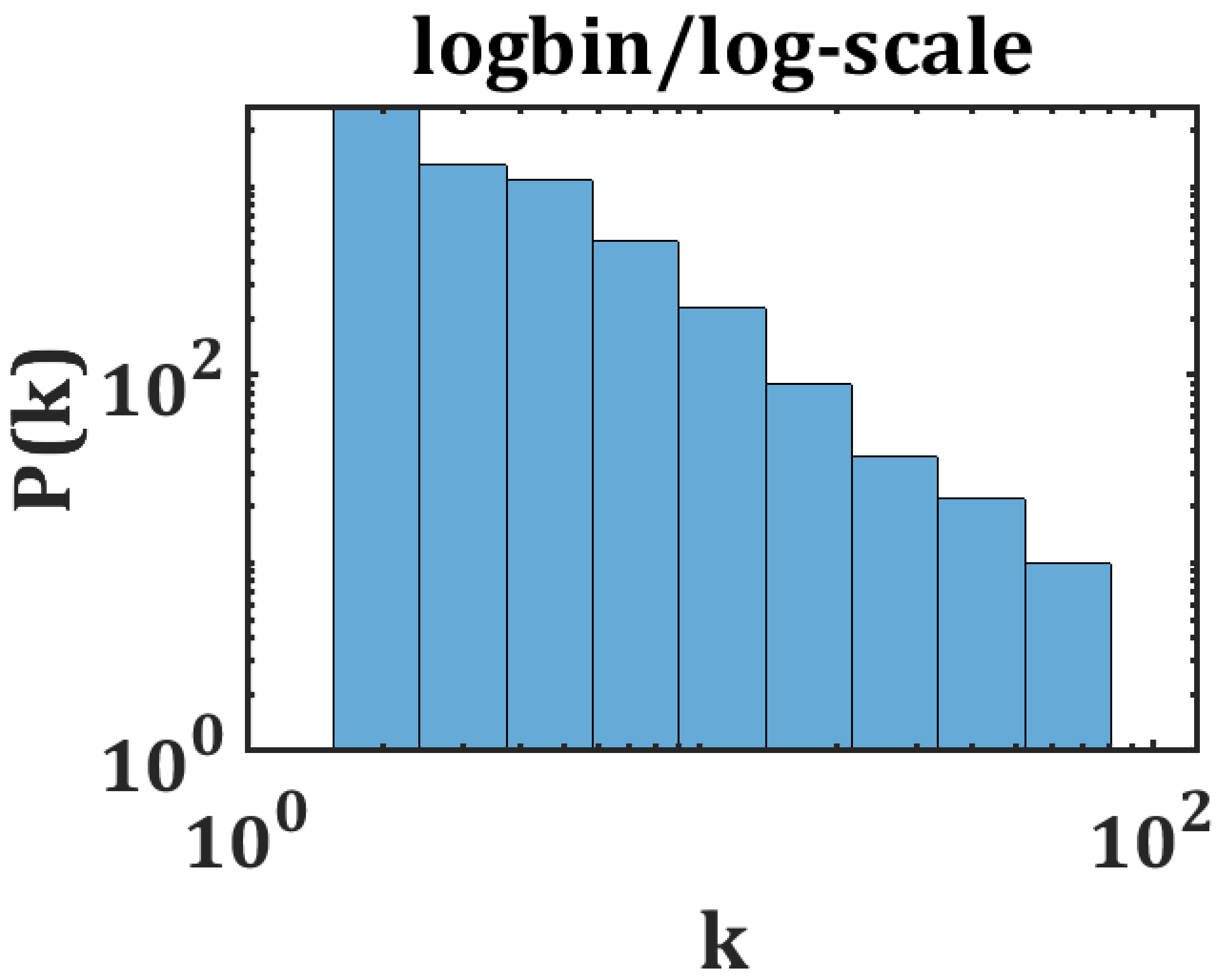
$$\langle x^2 \rangle = \frac{\alpha - 1}{\alpha - 3} x_{\min}^2.$$


linbin/lin-scale
P(k)
4000
2000
0
0
50
100
k
linbin/lin-scale
Zoomed
P(k)
300
200
100
0
0
20
40
60
k

Degree Distribution

Log/lin-binning , log scale





Degree Distribution

Matlab Script

```
load A_7k.mat;
deg=sum(A);
edges = 0.0:10:106;
edges2 = 3.^(0:0.4:4.4);
% %edges=10.^[-1:1:3];
figure(1); histogram(deg,edges);
axis tight;%%% Linear
set(findall(gcf,'-property','FontSize'),'FontName','Cambria',...
'FontSize',24,'linewidth',2.0,'fontweight','b') ;
set(gca, "Xscale", "log")
set(gca, "YScale", "log")
xlabel('k');
ylabel('P(k)');
title('linbin/log-scale')
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
figure(2); histogram(deg,edges2);
axis tight;
set(findall(gcf,'-property','FontSize'),'FontName','Cambria',...
'FontSize',24,'linewidth',2.0,'fontweight','b')
set(gca, "Xscale", "log");
set(gca, "YScale", "log");
xlabel('k');
ylabel('P(k)');
title('logbin/log-scale')
```

Degree Distribution (log/lin binning in log-log scale)

6k node

```
load('A_6k.mat');
% A=BA(9000,4);
%load A_7k.mat;
deg=sum(A);
ind=find(deg==1);
deg(ind)=2;
edges = 1.1:5:150;
edges2 = 3.^(0.01:0.5:4.7);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[Lob, Lob2] =histcounts(deg, edges2);
ds = diff(edges2);
sl = (edges2(1:end-1)+edges2(2:end))*0.5;
nsl = Lob(1:end)./(sum(Lob)*ds);
loglog(sl,nsl,'g.','MarkerSize',30); hold on;
%%%%% Linear bin %%%%%%%%%%%%%%%
[LiNb,LiNb1] = histcounts(deg,edges);
ds1 = diff(edges);
slin = (edges(1:end-1)+edges(2:end))*0.5;
loglog(slin,LiNb./(sum(LiNb)*ds1),'b.','MarkerSize',30);
```

```
hold on;
% loglog(Lob2(2:end-1),Lob2(2:end-1).^-3,'-','linewidth',3);
loglog(sl,8*sl.^-3,'-','linewidth',3);
set(gcf,'InvertHardCopy','off','Color','white');
set(gcf, 'PaperPositionMode', 'auto','position', [0, 0, 700, 700]);
set(legend,'color','none');
set(legend, 'Box', 'on');
xlabel('$k$','Interpreter','LaTeX','FontSize',30);
ylabel('$P(k)$','Interpreter','LaTeX','FontSize',30);
 xlim([0.8 200]);
  ylim([10^-6 10]);
 set(findall(gcf,'-property','FontSize'),'FontName','Cambria',...
     'FontSize',24,'linewidth',2.0,'fontweight','b')
```

Degree Distribution

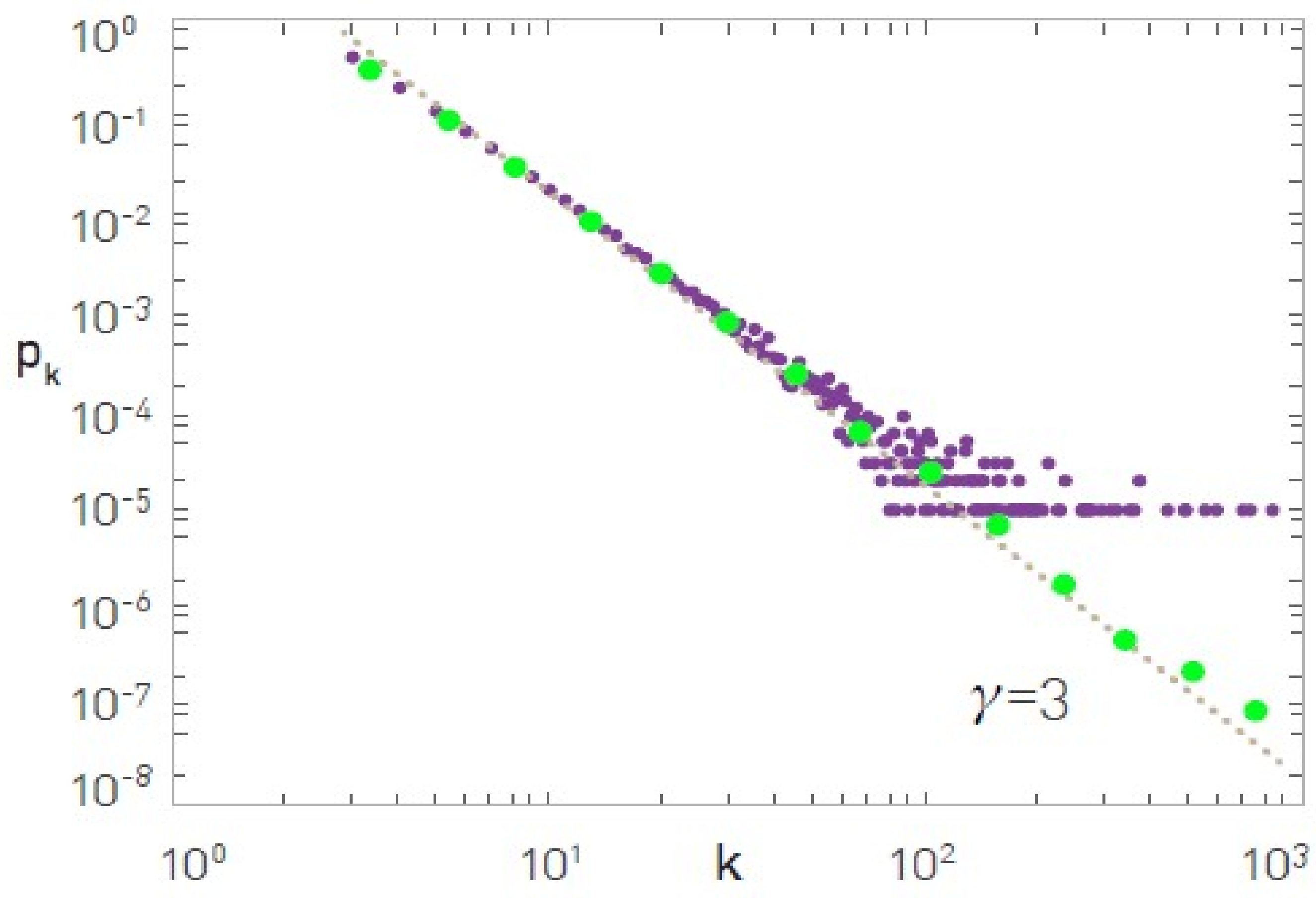


Figure 5.4

The Degree Distribution

The degree distribution of a network generated by the Barabási-Albert model. The figure shows p_k for a single network of size N=100,000 and m=3. It shows both the linearly-binned (purple) and the log-binned version (green) of p_k. The straight line is added to guide the eye and has slope γ=3, corresponding to the network's predicted degree exponent.

Regular Graph (Open Chain)

A — B

B — C

C — D

Length 1

4-1=3

A — C

B — D

Length 2

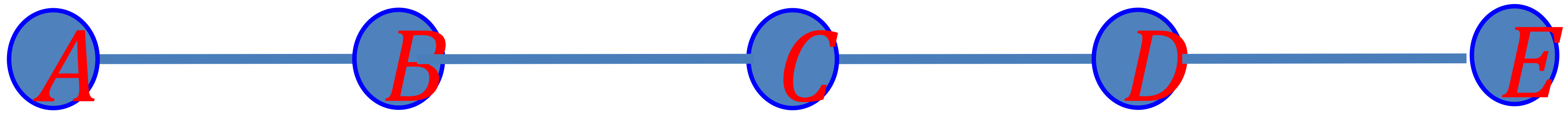
4-2=2

A — D

Length 3

4-3=1

Regular Graph (Open Chain)



A — B

B — C

C — D

D — E

Length 1

5-1=4

A — C

B — D

C — E

Length 2

5-2=3

A — D

B — E

Length 3

5-3=2

A — E

Length 4

5-4=1

$$\frac{2}{(n(n-1))}\sum_{j=1}^{n-1} j(n-j) = ¿$$

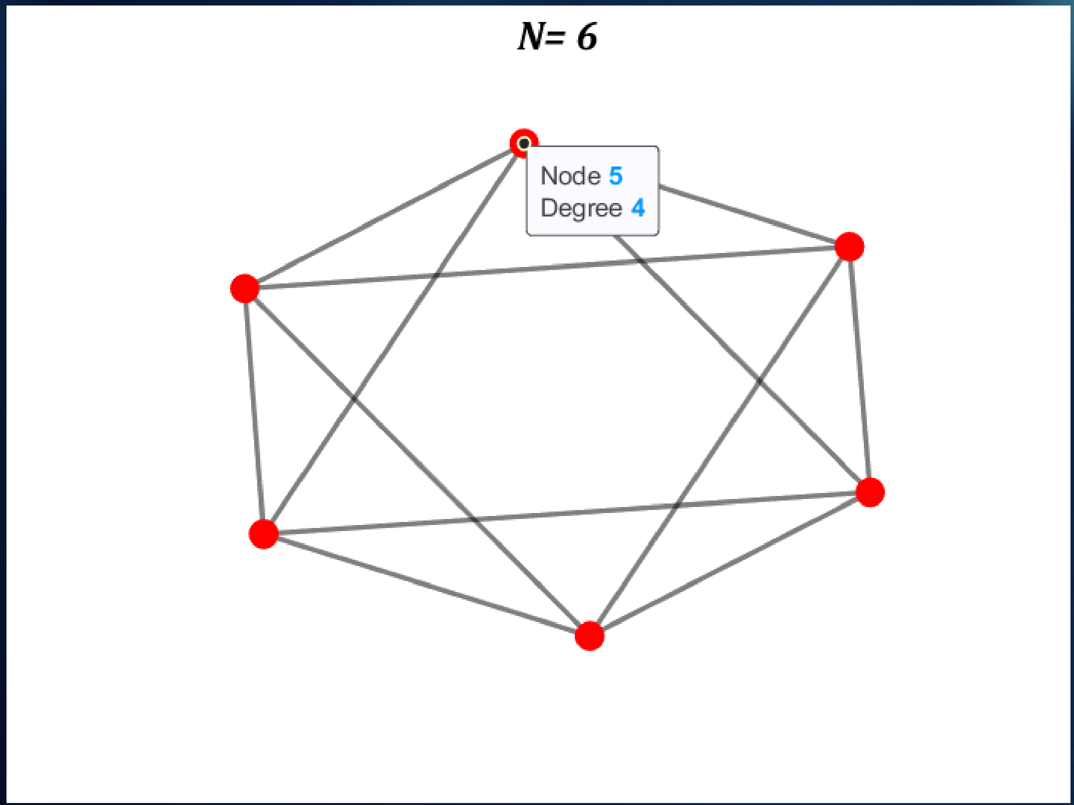
+

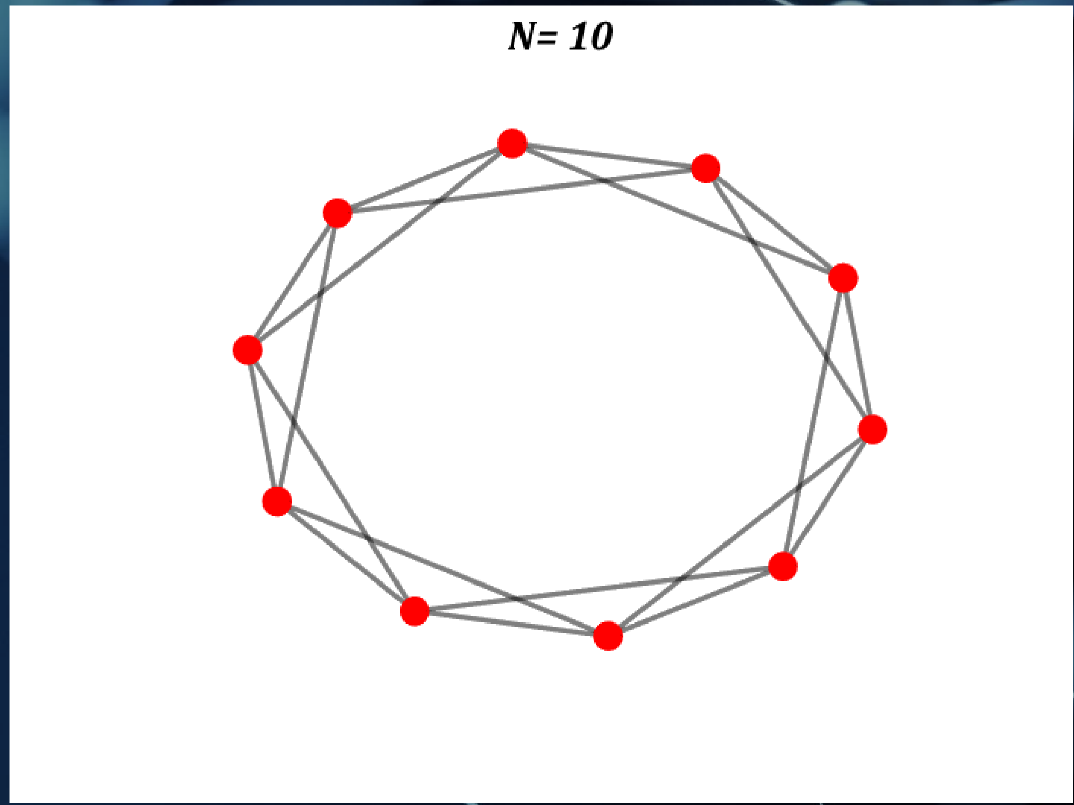
$$¿\left(n-\frac{(2n-1)}{3}\right)=\frac{\boldsymbol{n+1}}{\boldsymbol{3}}$$

Star Graph

Q: Find the average pathlength, clustering coefficients of start graph. Change network size , derive mathematically and explain.

Regular Graph (Ring)



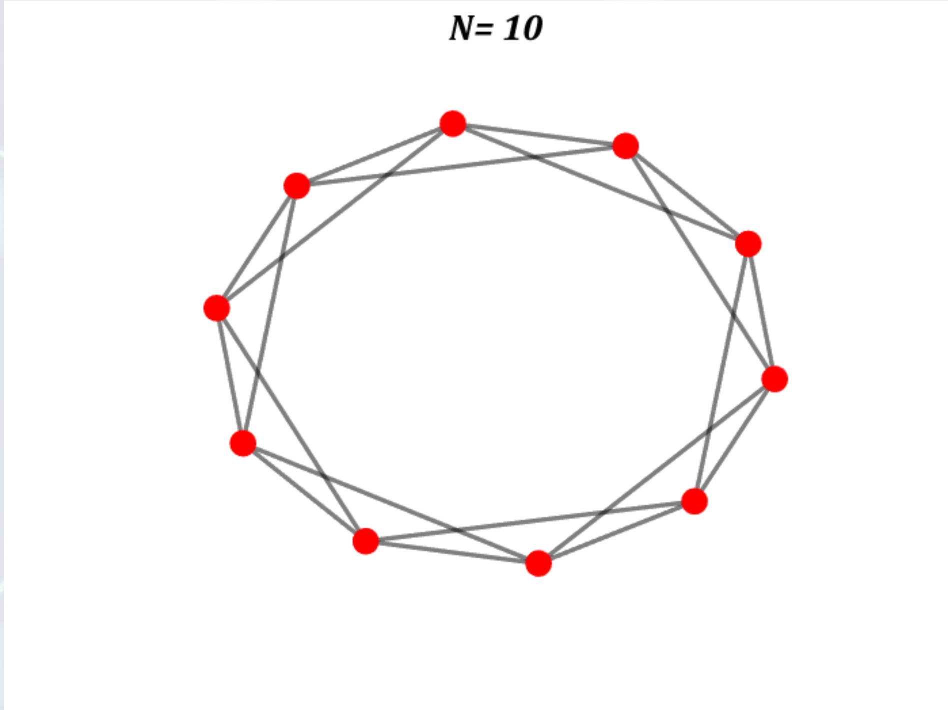


Q: What will be the average pathlength?

ONE DIMENSIONAL LATTICE: nodes on a ring

$P(k) = \delta(k-4)$ $k = 4$ for each node

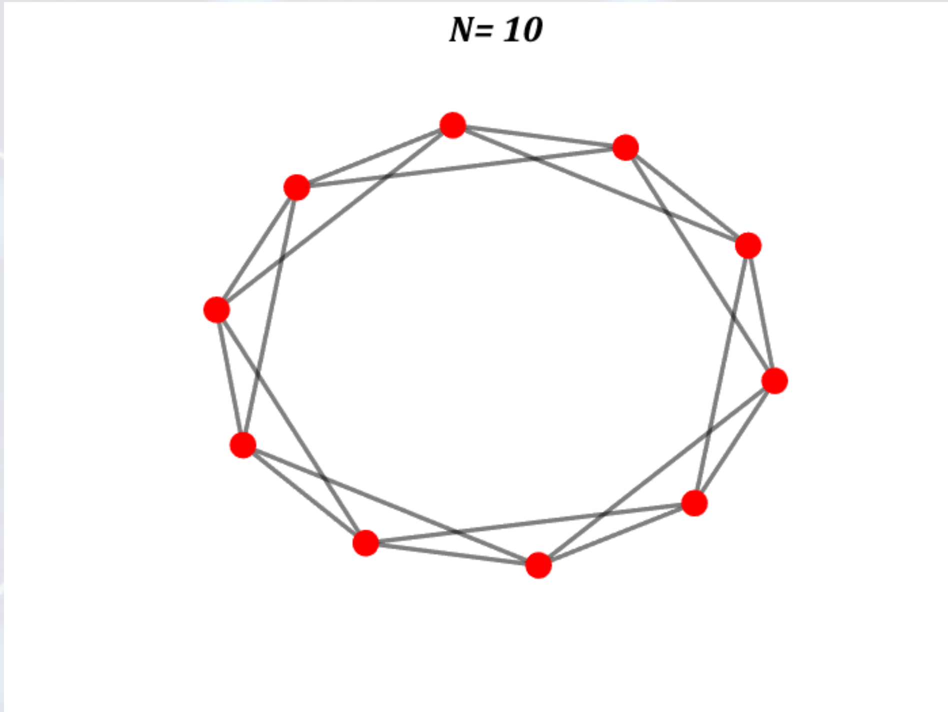
$C = \frac{1}{2}$ for each node if N > 6



ONE DIMENSIONAL LATTICE: nodes on a ring

$P(k) = \delta(k-4)$ $\quad k = 4$ for each node

$C = \frac{1}{2}$ for each node if N > 6



Thus,

$$\langle l \rangle = \frac{4\sum_{l=1}^{l_{max}} l}{N} = \frac{4\left(1+2+\ldots+\frac{N}{4}\right)}{N} \approx \frac{N}{2 * 4}$$

The average path-length varies as

$$\langle l \rangle \approx N \; ; \; N \gg 1$$

Constant degree, constant clustering coefficient.

TWO DIMENSIONAL LATTICE

Example 3 *2D square lattice graph. We can label vertices $v_{i,j}$ where i and j are integers (consider i, j are between 1 and n, so there are $p = n^2$ vertices). The distance between two vertices is $d\left(v_{i,j}, v_{k,\ell}\right) = |i-k| + |j-\ell|$. There are $\frac{n^2!}{2(n^2-2)!} = \frac{n^2(n^2-1)}{2}$ pairs of vertices. Thus the average path length is*

$$\begin{aligned}
\frac{1}{n^2\left(n^2-1\right)} \sum_{i,j,k,\ell=1}^{n} \left(|i-k| + |j-\ell|\right) &= \frac{1}{n^2\left(n^2-1\right)} \left(2n^2 \sum_{i,k=1}^{n} |i-k|\right) \\
&= \frac{2n^2}{n^2\left(n^2-1\right)} \left(\frac{1}{3}\left(n+1\right) n \left(n-1\right)\right) \\
&= \frac{1}{3} n = \frac{1}{3} p^{1/2}.
\end{aligned}$$

Other than at the edges, each vertex is adjacent to 4 vertices, none of which are adjacent to each other, so it has clustering 0.

Complete Graphs plot

Complete Graph

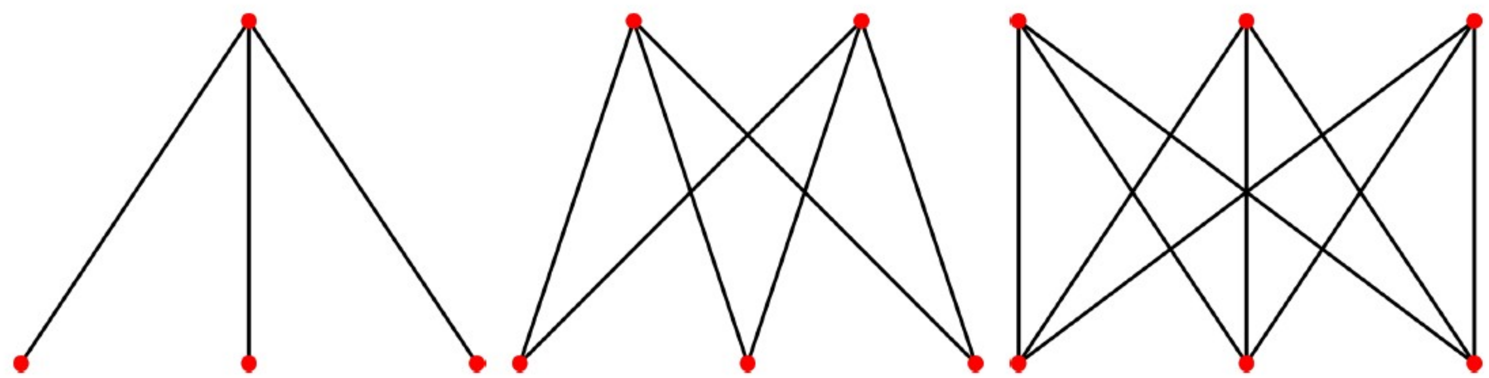
```
for N= [3 5 10 50 1000]
A=ones(N,N);
A(1:N+1:N*N)=0;
G=graph(A);
figure; plot(G);
deg = degree(G);
%deg = degree(h2);
nSizes = 8+4*sqrt(deg-min(deg)+0.5);
% NC=cell2mat(data(j,7));
nColors = rand(N,1)*0+3;%NC(:,3);% deg;
%%
%p=plot(G,'MarkerSize',nSizes,'NodeCData',...
% nColors,'EdgeAlpha',0.5);
p=plot(G,'MarkerSize',nSizes,'EdgeAlpha',0.5);
p.NodeFontSize=0.3;
p.NodeFontWeight='b';
p.EdgeColor='k';
```

Complete bipartite Graphs

Complete bipartite Graph

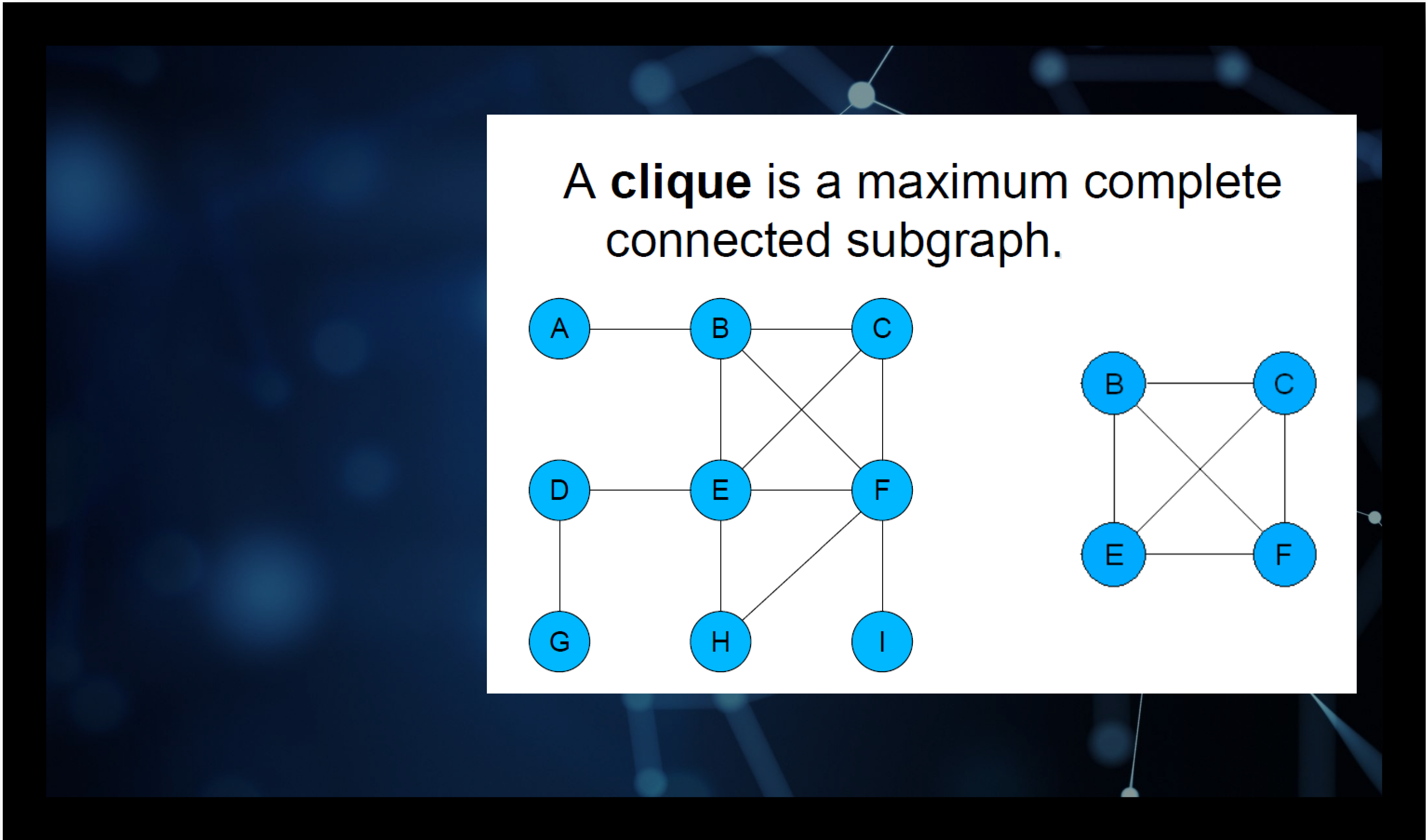
Complete Bipartite Graph

- Bipartite Variation of Complete Graph
- Every node of one set is connected to every other node on the other set



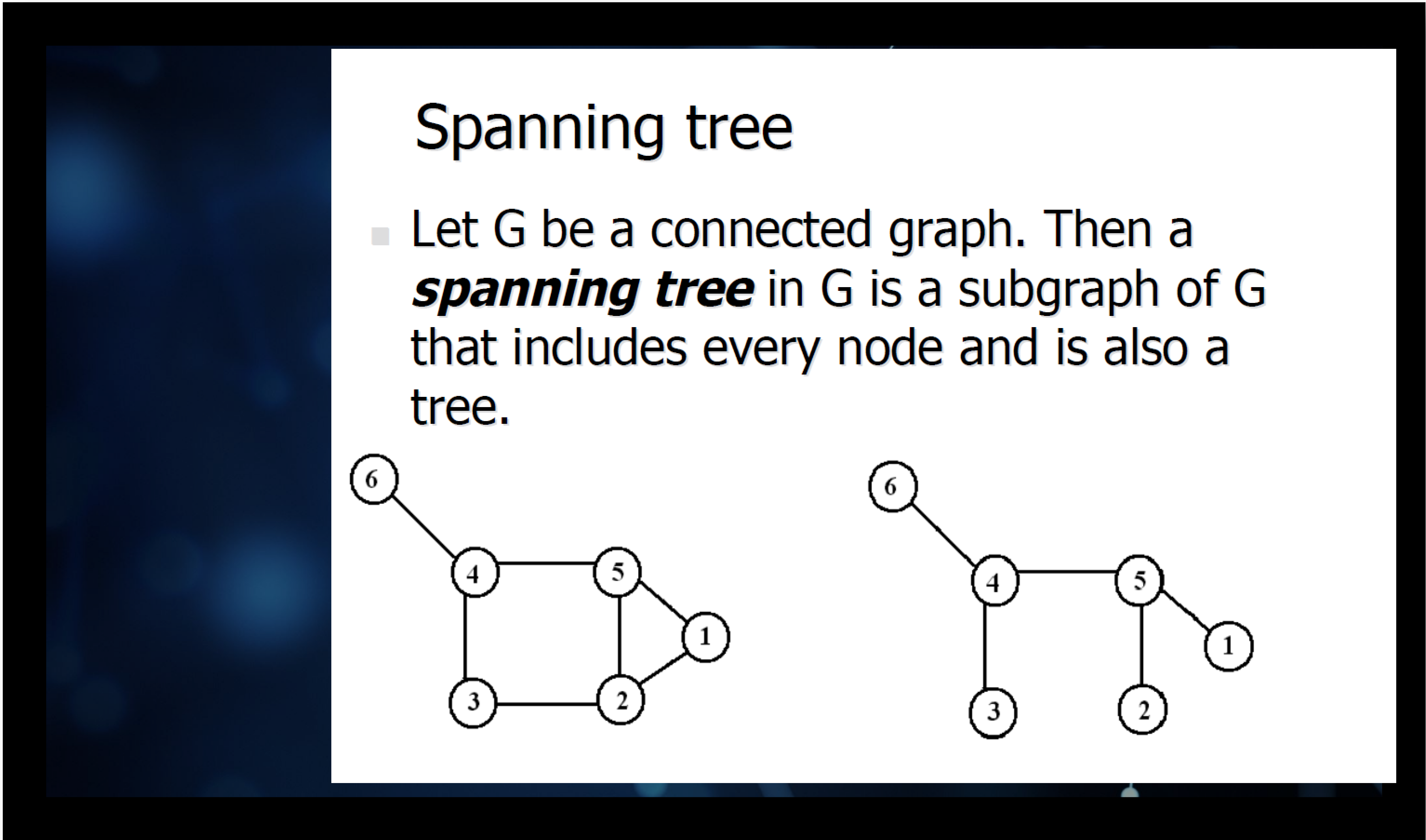
Stars

Cliques



Help Barabasi

Spanning Tree



Help Barabasi

Cycles (Calculation of total number of cycles of length)

- *Paths that start and end at are cycles in a network*
- *The number of cycles of length   is*
- *Then the total number of cycles of length   in a network*

Cycles (Calculation of total number of cycles of length)

- *Paths that start and end at are cycles in a network*
- *The number of cycles of length is*
- *Then the* *total number of cycles* *of length in a network*

$$L_r = \sum [A^r]_{ii} = Tr\, A^r$$

Tr is a trace of a matrix, the sum of elements of the main diagonal.

- *Trace can be in terms of the eigenvalues of the adjacency matrix*
- *For undirected graphs the adjacency matrix is symmetric*
- *The adjacency matrix has* n *real eigenvalues*

- *Trace can be in terms of the eigenvalues of the adjacency matrix*
- *For undirected graphs the adjacency matrix is symmetric*
- *The adjacency matrix has* n *real eigenvalues*
- *The eigenvectors have real elements*
- *can be written as*

- *Trace can be in terms of the eigenvalues of the adjacency matrix*
- *For undirected graphs the adjacency matrix is symmetric*
- *The adjacency matrix has n real eigenvalues*
- *The eigenvectors have real elements*
- *can be written as*

Q is orthogonal matrix whose columns are the real, orthonormal eigenvectors of . And is a diagonal matrix whose entries are the eigenvalues of

- *can be written as*
  Q is *orthogonal* *matrix whose columns are the real,* *orthonormal* *eigenvectors* *of . And is a diagonal matrix whose entries are the* *eigenvalues* *of*

say r=2 :

- can be written as
  Q is orthogonal matrix whose columns are the real, orthonormal eigenvectors of . And is a diagonal matrix whose entries are the eigenvalues of

For instance, r=2 :

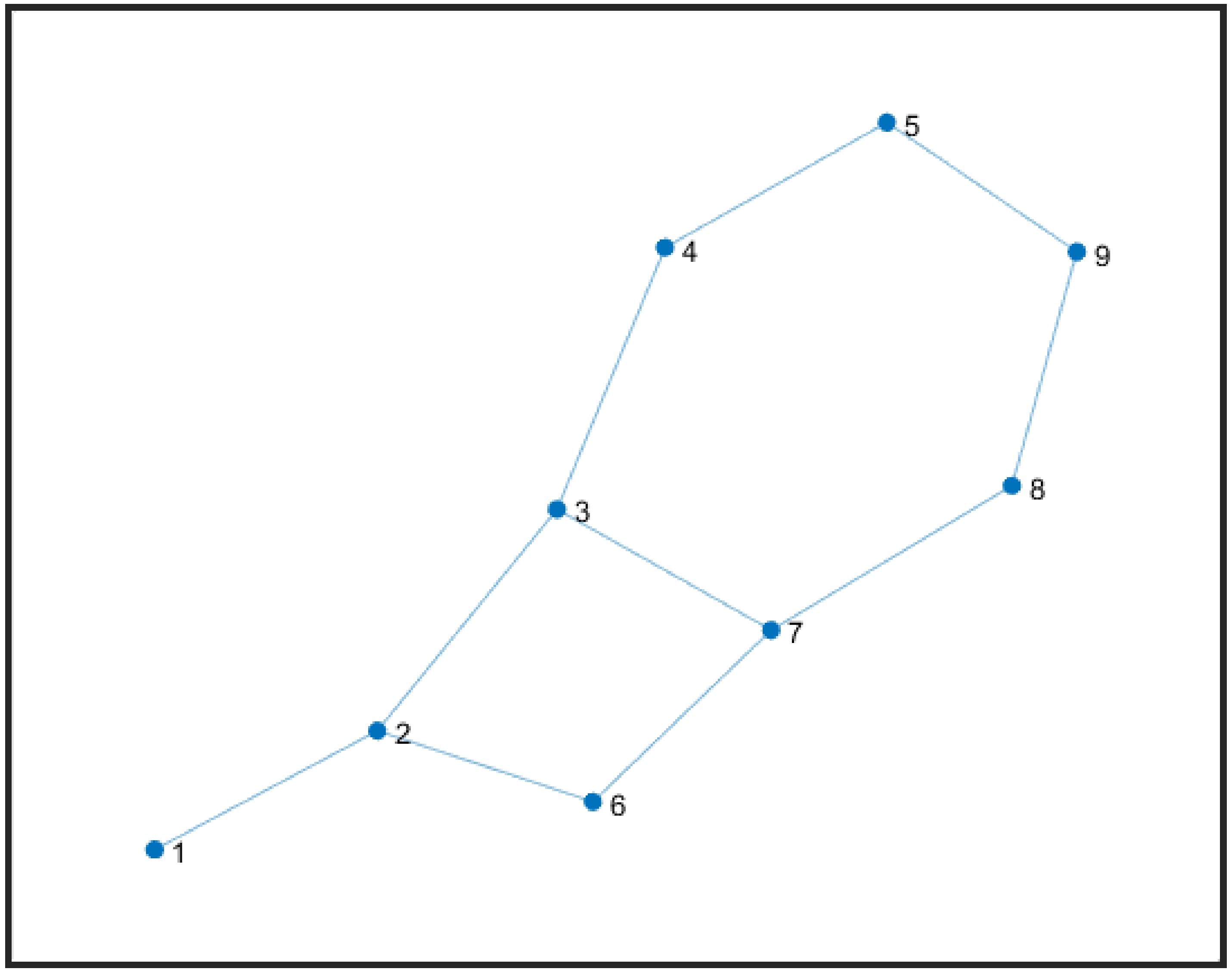
$$A^r = Q \Lambda^r Q^T$$

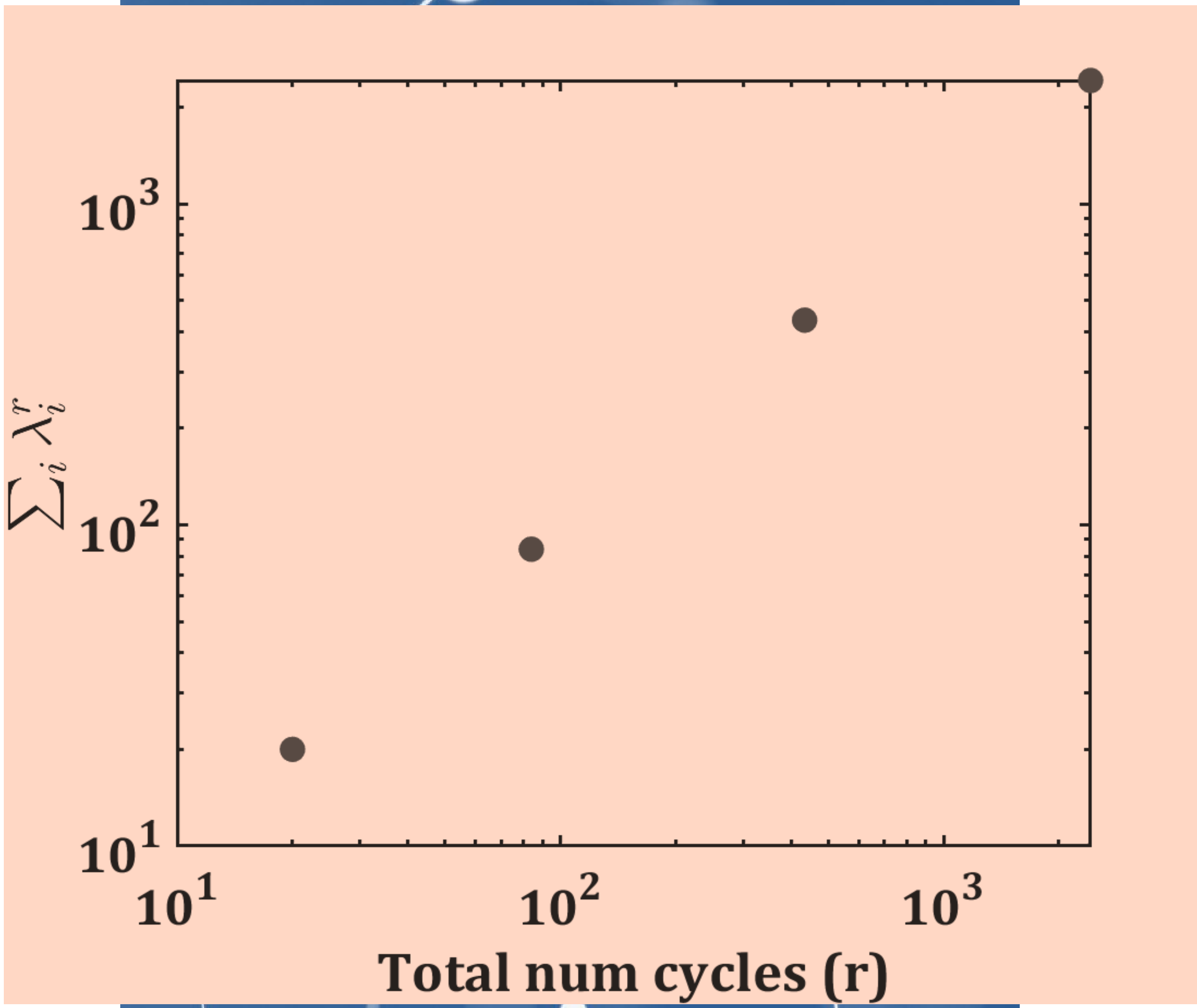
Cycles (Calculation of total number of cycles of length)

$$A^r = Q \Lambda^r Q^T$$

The total number of cycles of length in a network

Cycles (Calculation of total number of cycles of length)

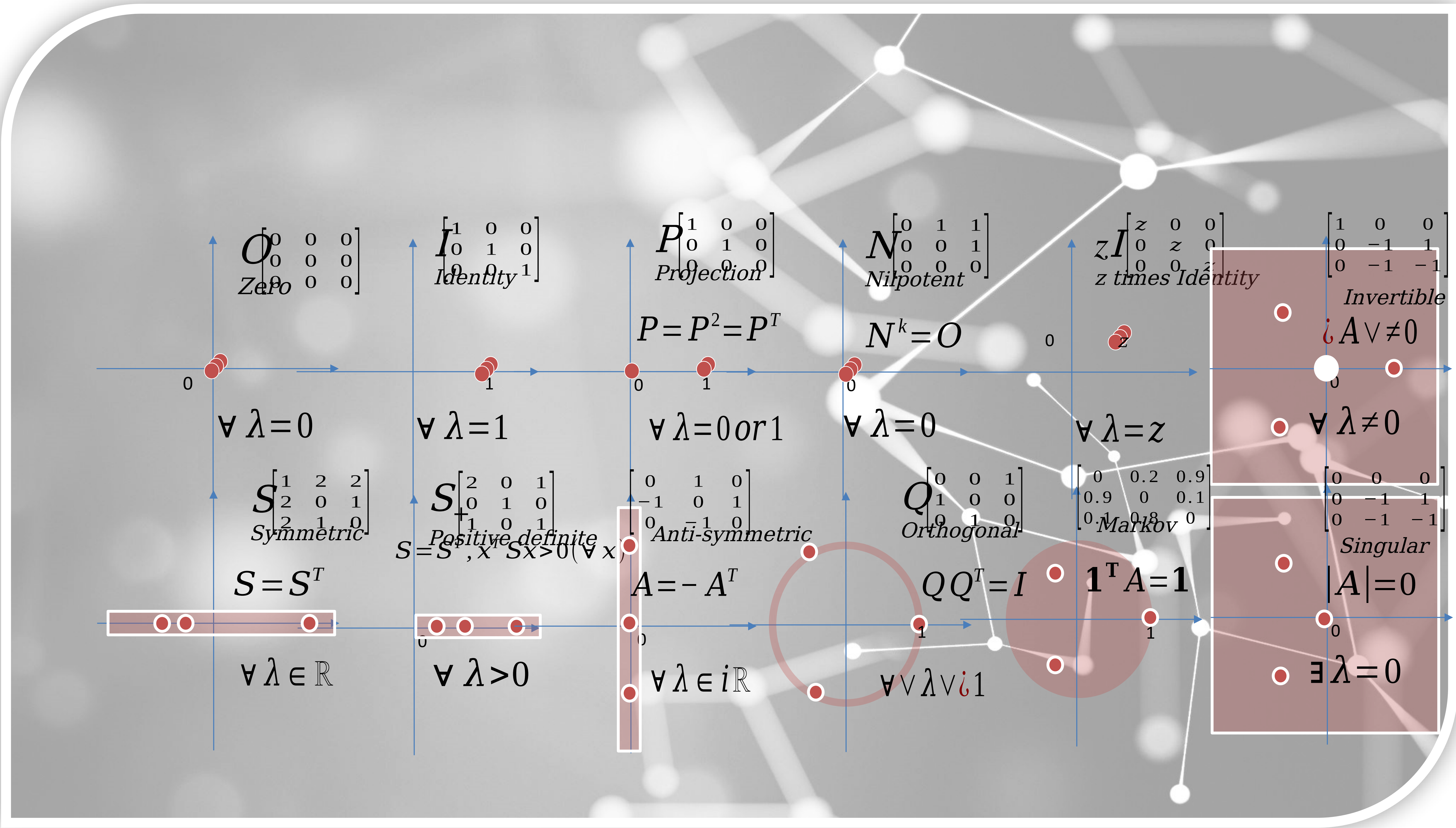




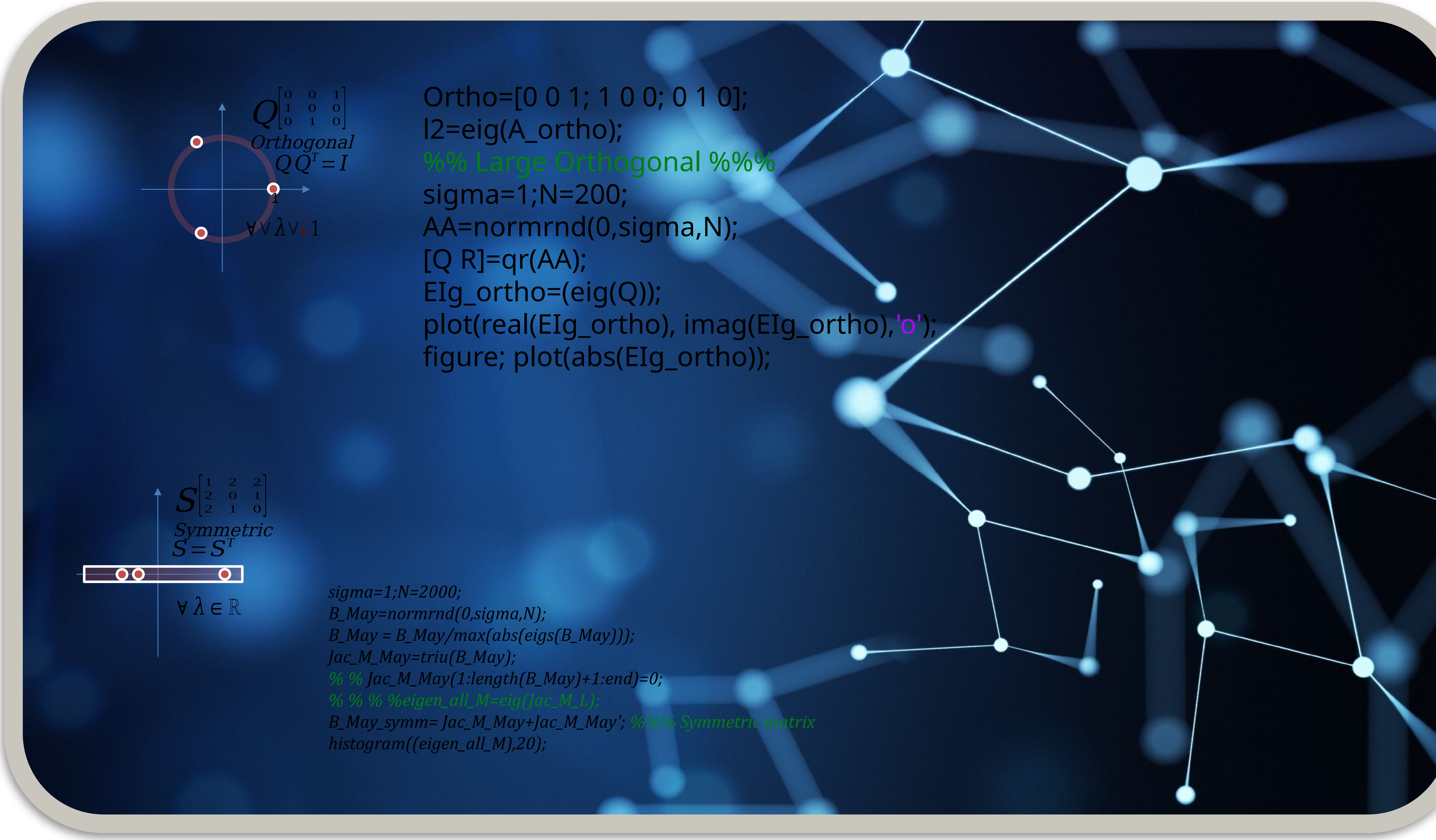
Advanced studies (First half)

- *Eigen value distribution of random graph*
- *Largest eigenvalue of adjacency matrix*

Map of Eigenvalues for real $n \times n$ square matrices

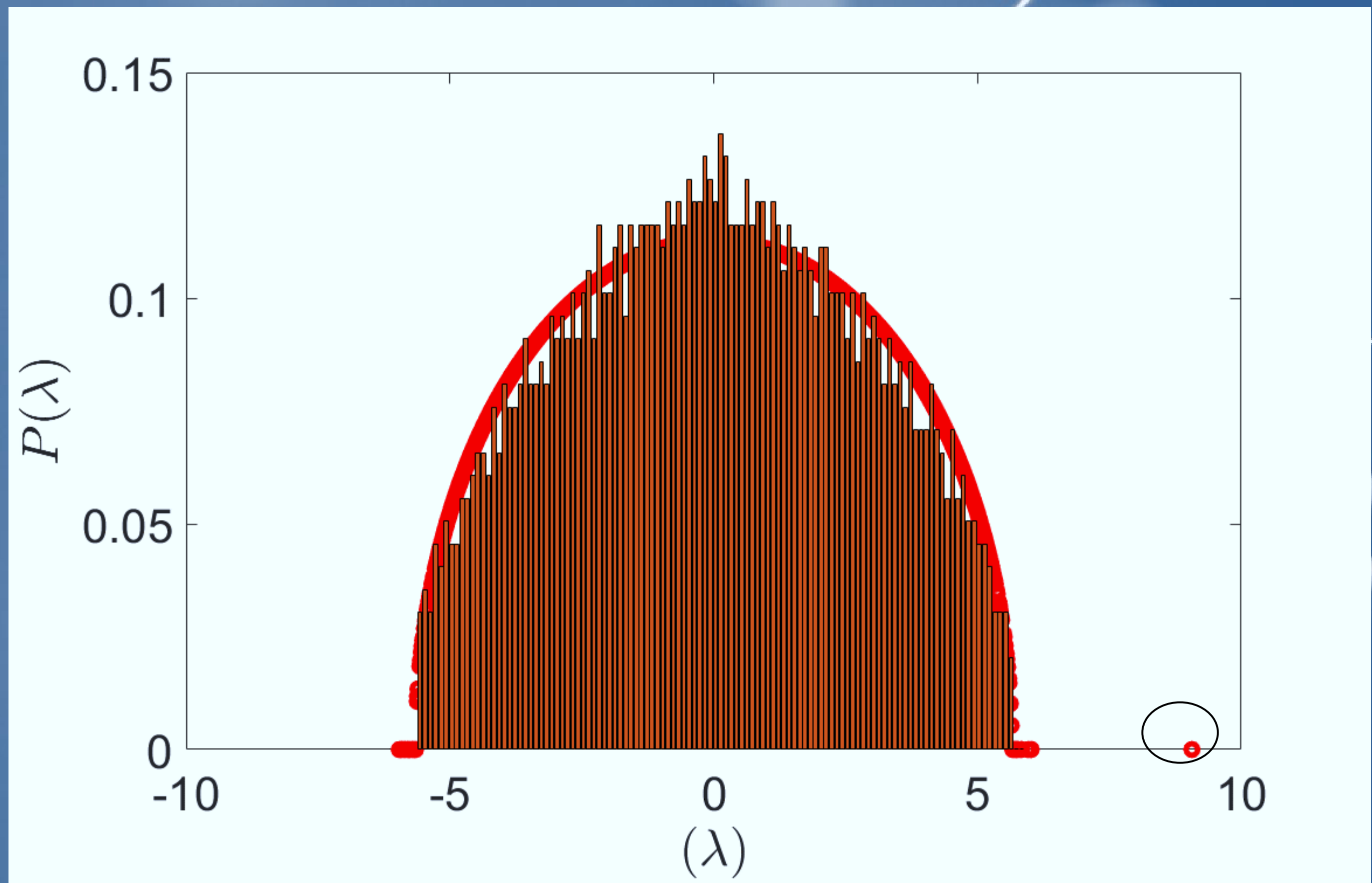


By Kenji Hiranabe with the kindest help of
Prof. Gilbert Strang


Q
Orthogonal
QQ^T = I
1
∀∨λ∨¿1
Ortho=[0 0 1; 1 0 0; 0 1 0];
I2=eig(A_ortho);
%% Large Orthogonal %%%
sigma=1;N=200;
AA=normrnd(0,sigma,N);
[Q R]=qr(AA);
EIg_ortho=(eig(Q));
plot(real(EIg_ortho), imag(EIg_ortho),'o');
figure; plot(abs(EIg_ortho));
S
Symmetric
S=S^T
∀λ∈ℝ
sigma=1;N=2000;
B_May=normrnd(0,sigma,N);
B_May = B_May/max(abs(eigs(B_May)));
Jac_M_May=triu(B_May);
% % Jac_M_May(1:length(B_May)+1:end)=0;
% % % %eigen_all_M=eig(Jac_M_L);
B_May_symm= Jac_M_May+Jac_M_May'; %%% Symmetric matrix
histogram((eigen_all_M),20);

Spectral Density

$$\rho(\lambda) = \begin{cases} (2\pi\sigma^2)^{-1}\sqrt{4\sigma^2 - \lambda^2} & \text{if } |\lambda| < 2\sigma \\ 0 & \text{otherwise} \end{cases}$$

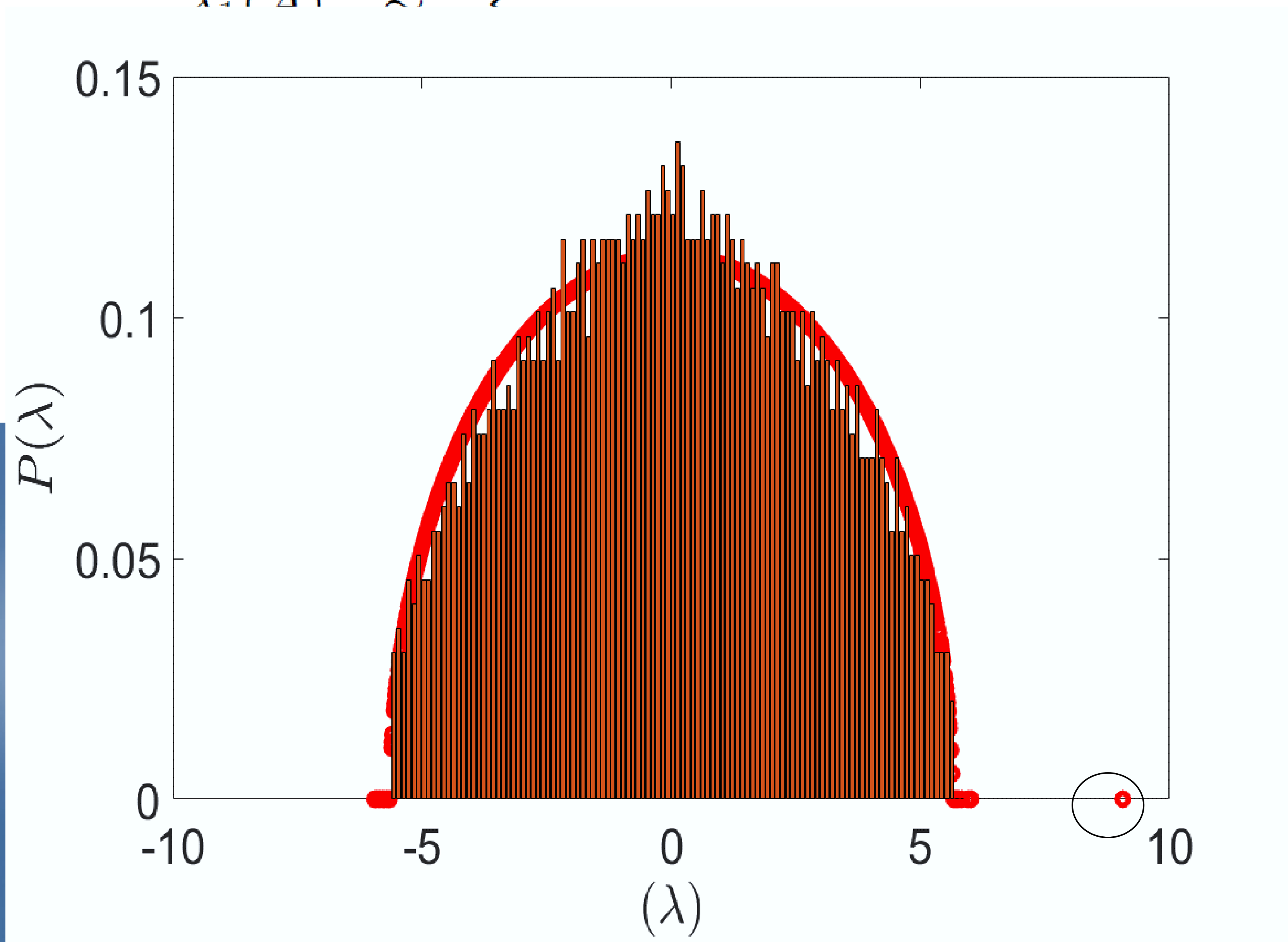


Spectral properties of complex networks, Camellia Sarkar1 & Sarika Jalan;

Spectral Density of a random graph

Theorem [Krivelevich, Sudakov ; 2003 + Vu ; 2007].

$$\begin{cases} (\log n)^{1/2} & \text{if } d \ll (\log n)^{1/2} \end{cases}$$



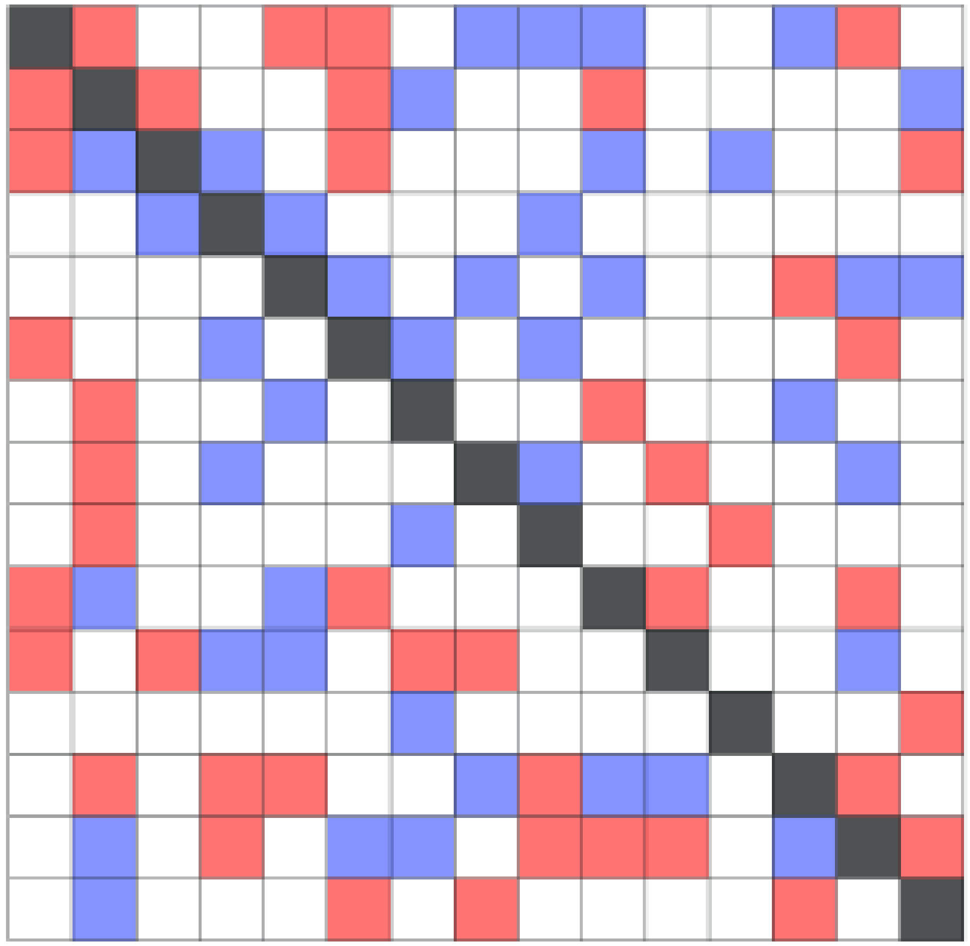
e) ~d

Spectral properties of complex networks, Camellia Sarkar1 & Sarika Jalan;

Map of Eigenvalues for real $n \times n$ square matrices

For a species to be self-regulating, we need $M_{ii} < 0$. This self-regulation is equivalent to setting a carrying capacity (or other similar density-dependent mechanism) for the population. May set all the diagonal elements $M_{ii} = -1$. He then set the off-diagonal elements to 0 with probability $1 - C$, and with probability C, he drew them independently from a distribution with mean 0 and variance σ^2. Note that, although in the subsequent literature

Random



With probability C
M_{ij} sampled from $\mathcal{N}(0, \sigma^2)$

With probability $(1 - C)$
$M_{ij} = 0$

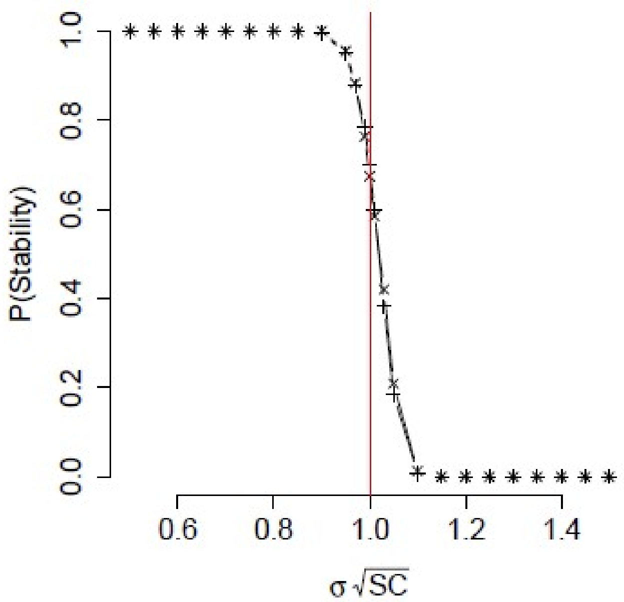
Diagonal entries
$M_{ii} = -d$

May's theorem

The probability of stability is close to 1 whenever:

$$\sigma\sqrt{SC} < d$$

and is close to 0 otherwise.



Following the circular law

Allesina,Tang, Nature 2012

Spectral Density/Eigenvalue Distribution

- *What will happen (eigenvalue distribution) if the elements are randomly drawn from Gaussian distribution?*

(Robert May-Nature, Stefano Alesina-Nature 2012)

- *What will happen (eigenvalue distribution) if the elements are correlated?*

(Stefano Alesina-Nature 2012, Somplonsky-PRL)

- *What will happen if the underlying matrix is sparse: a diluted graph with known degree distribution?*

(Note: If the degree distribution is Poisson/Binomial, it can capture the notion of May/Alesina work: Next slide.)

- *State of the art: Given:*

❑ *Degree distribution, Underlying connectivity,* **BUT** *elements are not random rather coming from particular distributions.*

What will be the eigen value distribution? How the distribution parameters control the largest eigen value?

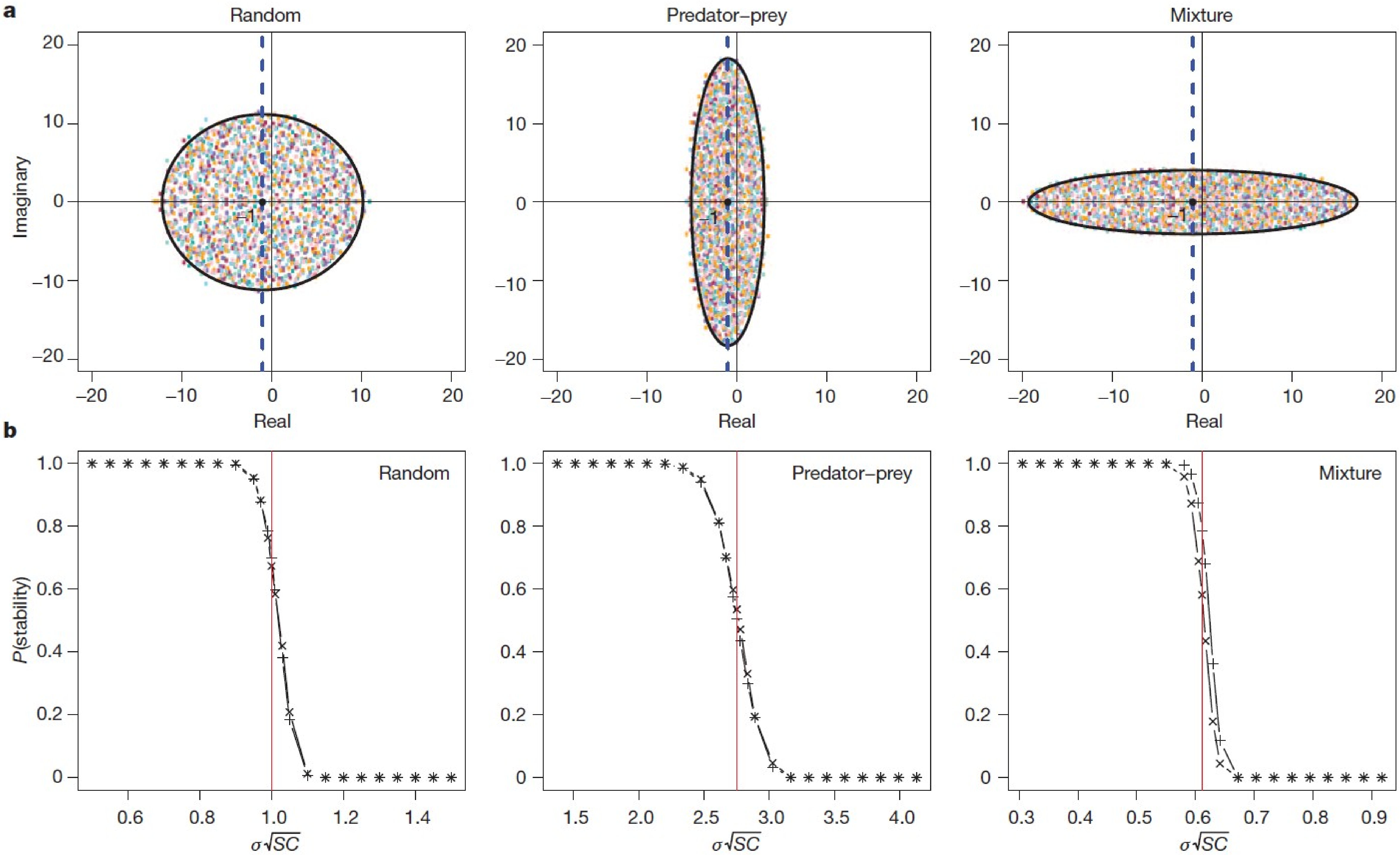


Figure 1 | Distributions of the eigenvalues and corresponding stability profiles. **a**, For $X \sim N(0, \sigma^2)$, $S = 250$, $C = 0.25$ and $\sigma = 1$, we plot the eigenvalues of 10 matrices (colours) with $-d = -1$ on the diagonal and off-diagonal elements, following the random, predator–prey or mixture

Allesina,Tang, Nature 2012

Graph Laplacian

Definition

The graph Laplacian $\mathbf{L}$ of a simple undirected graph is the matrix with elements L_{ij} such that

$$L_{ij} = \begin{cases} k_i \text{ if } i = j \\ -1 \text{ if there is a link between nodes } i \text{ and } j \text{ and } i \neq j \\ 0 \text{ otherwise.} \end{cases} \tag{50}$$

Graph Laplacian

- Alternatively, we can write
- δ_{ij} is the Kronecker delta, which is 1 for $i = j$ and 0 otherwise

$$L_{ij} = \delta_{ij} k_i - A_{ij} \tag{51}$$

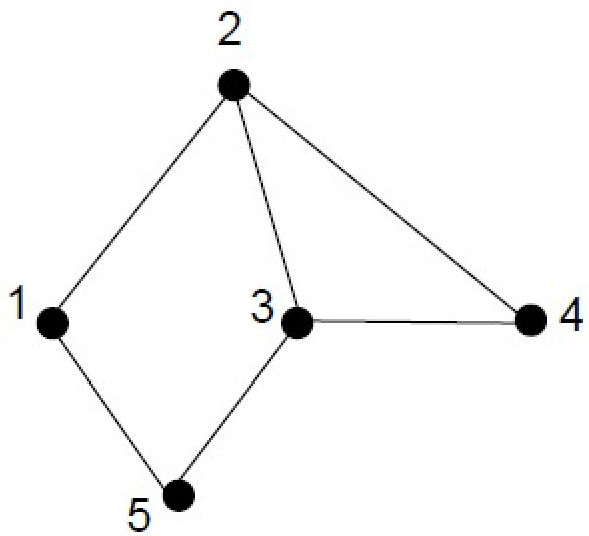


Figure: $L = D - A$

$$\mathbf{L} = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & -1 & -1 & 2 & 0 \\ -1 & 0 & -1 & 0 & 2 \end{pmatrix}$$

Graph Laplacian

- Alternatively, we can write
- δ_{ij} is the Kronecker delta, which is 1 for $i = j$ and 0 otherwise

$$L_{ij} = \delta_{ij} k_i - A_{ij} \tag{51}$$

The eigenvalues of the graph Laplacian are its most interesting property

The Laplacian is a symmetric matrix → it has real eigenvalues

We can even show that all of its eigenvalues are non-negative

Also, we can show that its smallest eigenvalue $\lambda_1 = 0$

Graph Laplacian

- Alternatively, we can write
- δ_{ij} is the Kronecker delta, which is 1 for $i = j$ and 0 otherwise

$$L_{ij} = \delta_{ij} k_i - A_{ij} \tag{51}$$

$$\mathbf{L}\begin{pmatrix}1\\ \vdots\\ 1\\ \vdots\end{pmatrix} = \begin{pmatrix}\sum_j(\delta_{1j}k_{1j} - A_{1j})\\ \vdots\\ \sum_j(\delta_{ij}k_{ij} - A_{ij})\\ \vdots\end{pmatrix} = \begin{pmatrix}k_1 - \sum_j A_{1j}\\ \vdots\\ k_i - \sum_j A_{ij}\\ \vdots\end{pmatrix} = \begin{pmatrix}k_1 - k_1\\ \vdots\\ k_i - k_i\\ \vdots\end{pmatrix} = \begin{pmatrix}0\\ \vdots\\ 0\\ \vdots\end{pmatrix} = 0\begin{pmatrix}1\\ \vdots\\ 1\\ \vdots\end{pmatrix}$$

Connected Components and the algebraic connectivity

Suppose we have a network with different components

The components have sizes

$$\mathbf{L} = \begin{pmatrix} \boxed{} & 0 & \cdots \\ 0 & \boxed{} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Connected Components and the algebraic connectivity

Suppose we have a network with different components

The components have sizes

$$\mathbf{L} = \begin{pmatrix} \boxed{} & 0 & \cdots \\ 0 & \boxed{} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

- We have n_1 ones and this is an eigenvector with eigenvalue 0
- We have c such eigenvectors

Connected Components and the algebraic connectivity

- In a network with c components c eigenvalues are equal to 0
- The second eigenvalue λ_2 of the graph Laplacian is non-zero iff the network is connected
- The second eigenvalue of the Laplacian is called *algebraic connectivity*
- It is a measure of how connected is a network, i.e. how difficult is to divide that network

Ring : $k=4$ neighbours.

$$P(k) = \delta(k-4)$$

Now, choose a random node, i.

In the first hop ($l=1$) it has 4 nodes

In the second hop ($l=2$) it has connects to next 4 nodes. Let's assume there is max hop: l_{max} hop (highest distance)

So, $$1 + \sum_{l=1}^{l_{max}} 4 \approx N \quad (\text{why approx.}) \longrightarrow (1)$$

$$\Rightarrow 1 + 4 l_{max} = N$$

$$l_{max} \approx \frac{N}{4} \longrightarrow (2)$$

For one particular node (Sum of all distances for $N \to \infty$)

$$4\left(1+2+\cdots+\frac{N}{4}\right) = 4 \times \frac{\frac{N}{4}\left(\frac{N}{4}+1\right)}{2}$$

$$\therefore \langle l \rangle = \frac{\frac{N}{2} \; \frac{4 \times \frac{N}{4}\left(\frac{N}{4}+1\right)}{2 \times \frac{N(N-1)}{2}}}{} = \frac{1}{4} \frac{N(N+4)}{N(N-1)} \times \frac{N}{2} \approx \frac{N}{2 \times 4}$$

For N nodes

For double counts

(See pages 2 to 4 for ring)

~~For~~ For a ring (degree 2)

For a chosen node 1

	Dist
To node 2 :	1
To note 3 :	2

⋮

$$\begin{cases} \text{To node } \frac{N}{2}+1 : & \frac{N}{2} \quad \text{if } N \text{ is even} \\ \text{To node } \frac{N+1}{2} : & \frac{N-1}{2} \quad \text{if } N \text{ is odd} \end{cases}$$

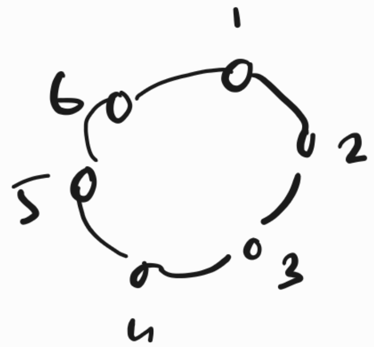


Fig. 1

The distance then decrease due to the periodic boundary condition.
For example, the distance to node N is 1

Calculate the sum of distances from a single node:

if N is even

For 4th move (it will not be multiplied by 2)

$$2 \times \left(1 + 2 + 3 + \dots + \frac{N}{2} - 1\right) + \frac{N}{2}$$

1-2 or 1-6   1-3 or 1-5   1 to 4

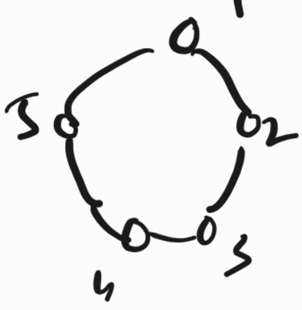
(CFIS 1)

$$= 2 \times \frac{\left(\frac{N}{2} - 1\right)\left(\frac{N}{2} - 1 + 1\right)}{2} + \frac{N}{2}$$

$$= \frac{N(N-2)}{4} + \frac{N}{2}$$

$$= \frac{N^2 - 2N + 2N}{4} = \frac{N^2}{4}$$

If N is odd



$$2 \times \left(1 + 2 + 3 + \dots + \frac{N-1}{2}\right)$$

$$= \frac{2 \times \left(\frac{N-1}{2}\right)\left(\frac{N-1}{2} + 1\right)}{2} = \frac{(N-1)(N+1)}{4} = \frac{N-1}{4}$$

Calculate the total sum of all shortest paths

Therefore, the total sum of shortest paths is N times the sum for a single node. However, each path is counted twice (once for i to j and the other j to i), so the sum must divided by 2:

If N is even: $\dfrac{N\left(\frac{N^2}{4}\right)}{2} = \dfrac{N^3}{8}$

If N is odd: $\dfrac{N\left(\frac{N^2-1}{4}\right)}{2} = \dfrac{N(N^2-1)}{8}$

Therefore the avg. path length

For even: $\dfrac{N^3}{8} \Big/ {}^{N}C_2 = \dfrac{N^3 \times 2}{8 \times N(N-1)}$

$= \dfrac{N^2}{4(N-1)} \sim \dfrac{N}{4}$

For odd $\quad \dfrac{N(N^2-1) \times 2}{8 \times N(N-1)} = \dfrac{N+1}{4}$

$$\sim \frac{N}{4}$$

What about degree : k

$$\langle l \rangle \approx \frac{N}{2K}$$

Avg. pathlength of chain

$N=5$

pathlength of 1 : $\quad AB \,/\, BC \,/\, CD \,/\, D\text{-}E$

$L_1 : \quad N-1$

Pathlength of 2 : $\quad \overline{AC} \,/\, \overline{BD} \,/\, \overline{CD}$

$L_2 : \quad N-2$

Path length of 3 : $\tilde{A}D$ / $\tilde{B}E$

L_3 : $N-3$

Path length of 4 : $L_4 = N-4$

. . . .

Thus avg. path length :

$$\langle l \rangle = \frac{1\times(N-1) + 2(N-2) + 3(N-3) + \cdots + (N-1)\times(N-(N-1))}{{}^{N}C_2}$$

$$\langle l \rangle = \frac{N(1+2+\cdots+N-1) - (1^2+2^2+3^2 \cdots + (N-1)^2)}{{}^{N}C_2}$$

$$= \frac{1}{{}^{N}C_2}\left[\frac{N \times N(N-1)}{2} - \frac{(N-1)N(2N-2+1)}{6}\right]$$

$$\langle l \rangle = \frac{2}{2}\frac{N(N-1)}{N(N-1)}\left[N - \frac{2N-1}{3}\right]$$

$$\langle l \rangle = \frac{3N-2N+1}{3} = \frac{N+1}{3} \approx \frac{N}{3}$$