Complex Systems & Network Thinking

Pierre-Alexandre Balland

What is the **New Science** of Cities?

This is the **new science** of cities

The new science of cities is about applying network thinking (and complex system thinking) to cities

Today’s objectives

- Introduction to network thinking
- Complex networks in natural sciences, social sciences, and business
- Elements of graph theory and matrix calculus
- Random, small worlds, and growing complex networks
- Key structural patterns of real-world complex systems
- Cities as complex systems?

- Linked to the computer labs in the QUEA class – bring your laptop!

On the side

- Discussion on big data
- Do we still need theory when we have big data?
- Data visualization techniques – art or science?

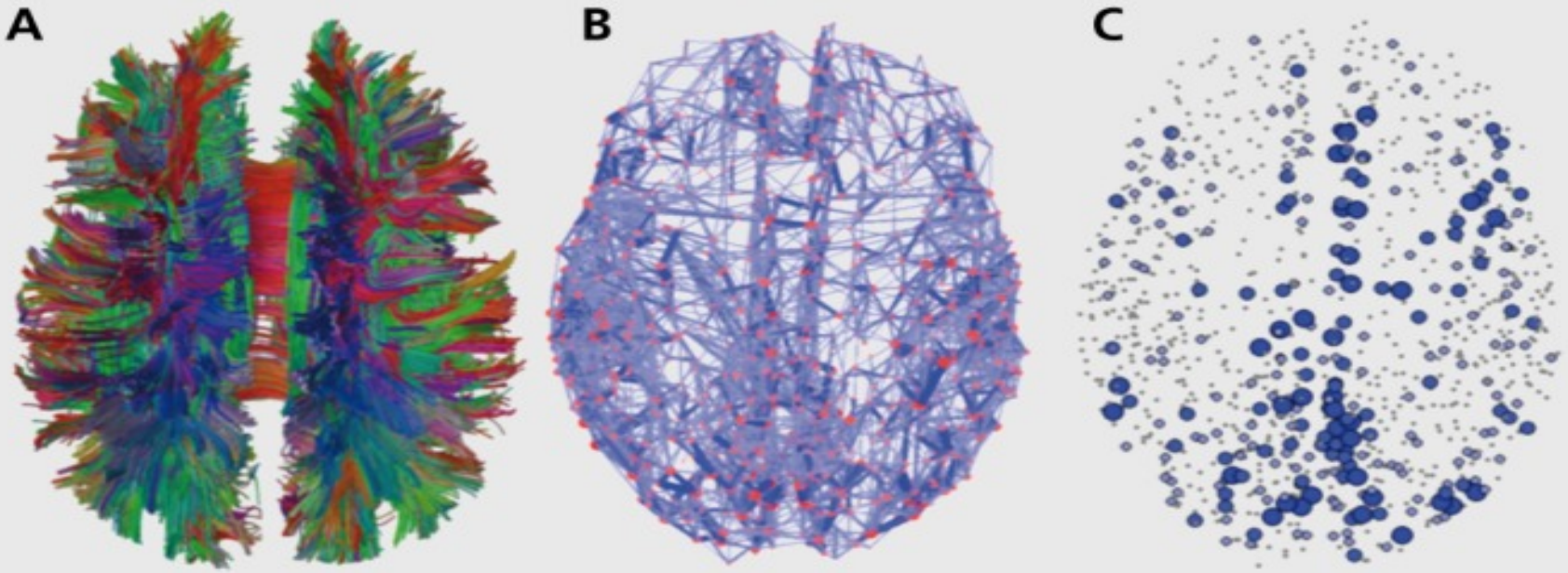
Class schedule & overview of the class

https://paballand.github.io/teaching/nsc.html

What is network thinking?

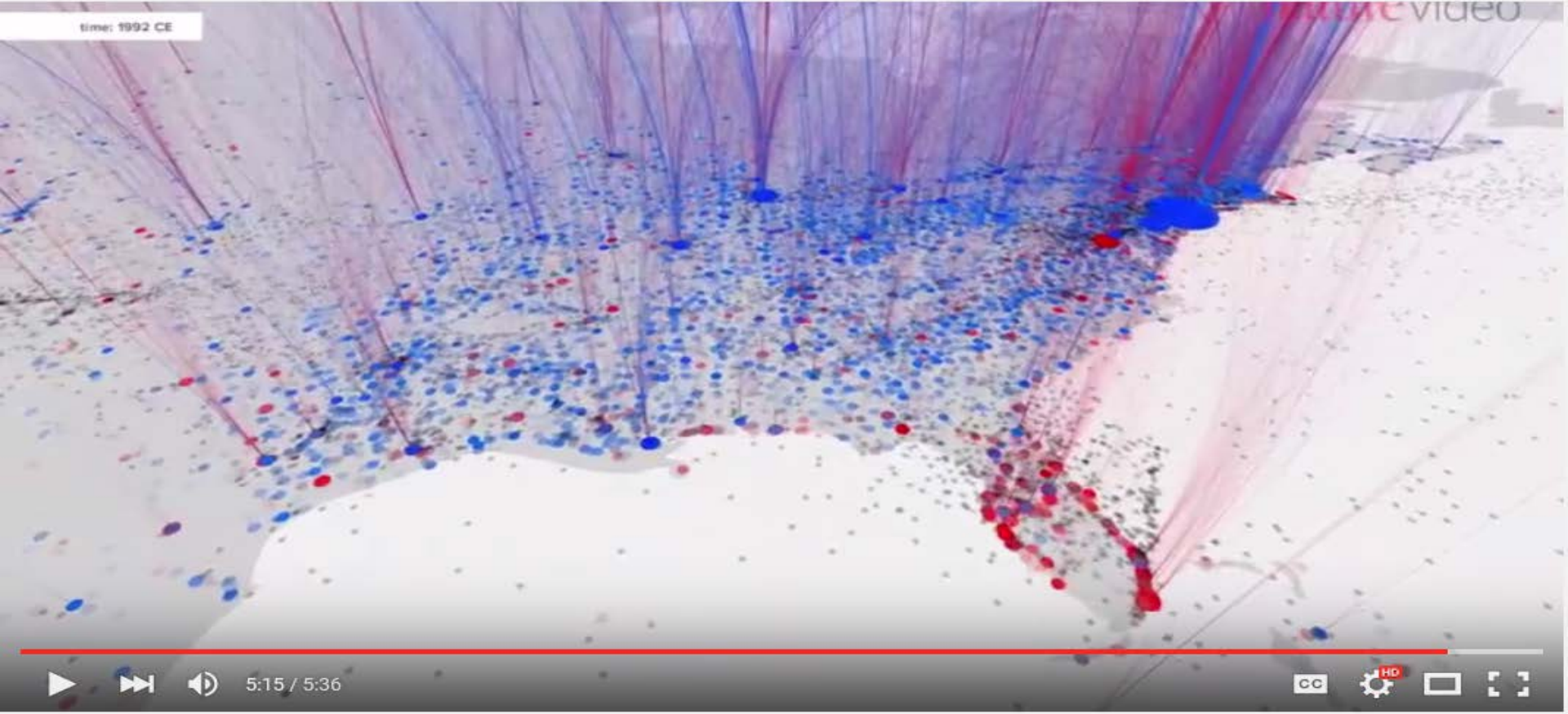
- A network-based paradigm is taking science by storm (Barabási, 2012)

Network structure of the brain



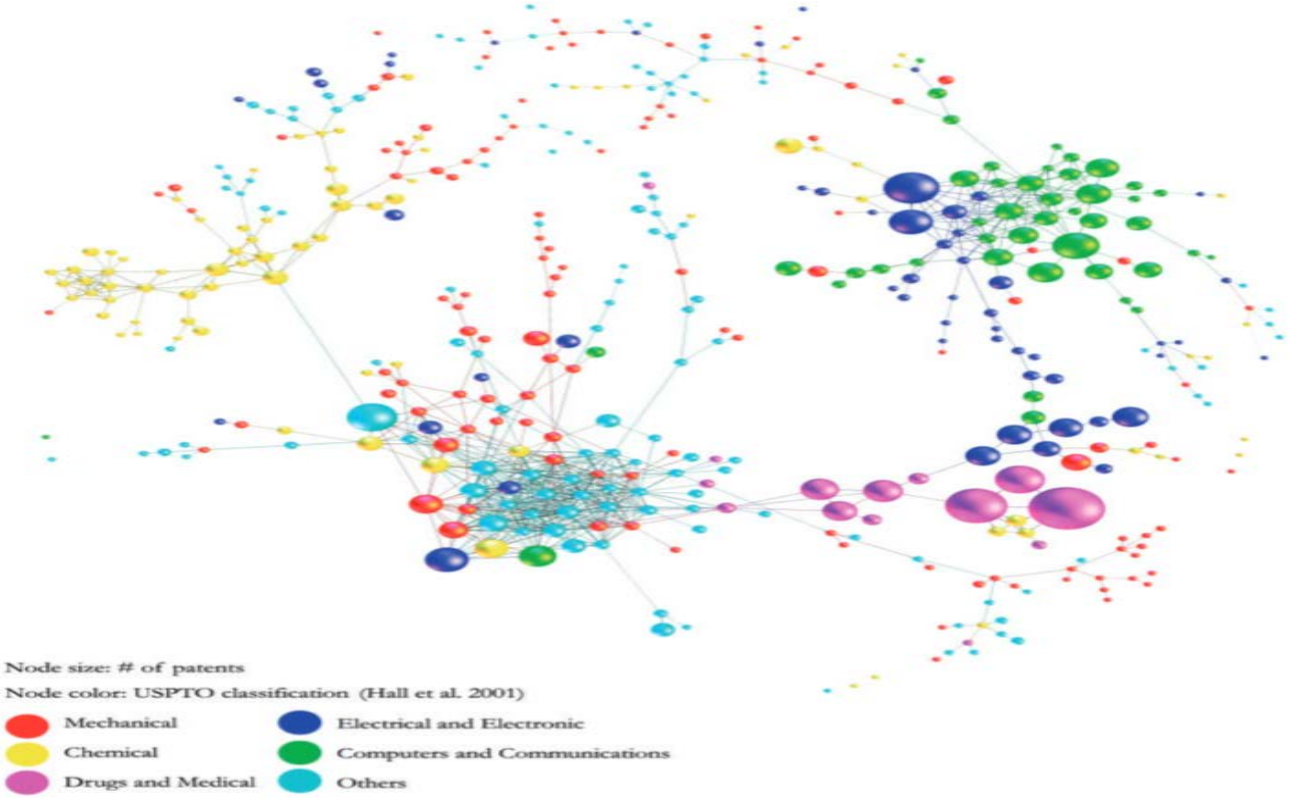
Sporns (2013)

Migration flows



Knowledge flows

Knowledge relatedness



Boschma, Balland and Kogler (2013)

What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)...but also business

An interesting patent

US006285999B1

(12) **United States Patent**
Page

(10) **Patent No.: US 6,285,999 B1**
(45) **Date of Patent: Sep. 4, 2001**

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/035,205, filed on Jan. 10, 1997.

(51) **Int. Cl.**[7] **G06F 17/30**
(52) **U.S. Cl.** **707/5**; 707/7; 707/501
(58) **Field of Search** 707/100, 5, 7, 707/513, 1–3, 10, 104, 501; 345/440; 382/226, 229, 230, 231

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106	*	8/1990	Gansner et al.	345/440
5,450,535	*	9/1995	North	395/140
5,748,954		5/1998	Mauldin	395/610
5,752,241	*	5/1998	Cohen	707/3
5,832,494	*	11/1998	Egger et al.	707/102
5,848,407	*	12/1998	Ishikawa et al.	707/2
6,014,678	*	1/2000	Inoue et al.	707/501

OTHER PUBLICATIONS

S. Jeromy Carriere et al, "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257–1267.*

Wang et al "Prefetching in Worl Wide Web", IEEE 1996, pp. 28–32.*

Ramer et al "Similarity, Probability and Database Organisation: Extended Abstract", 1996, pp. 272.276.*

Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221–231.*

L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometricka, vol. 18, pp. 39–43.

C.H. Hubbell, "An input–output approach to clique identification sociometry," 1965, pp. 377–399.

Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26–48.

E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471–479.

Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297–312.

N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93–95.

P. Doreian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45–56.

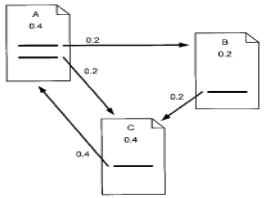
(List continued on next page.)

Primary Examiner—Thomas Black
Assistant Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder L.L.P.

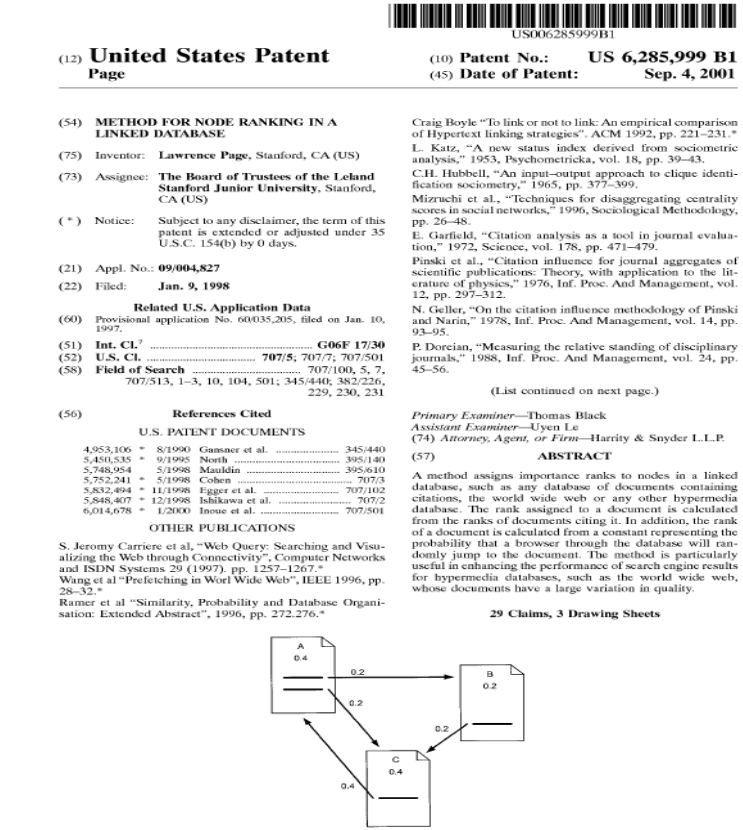
(57) **ABSTRACT**

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

29 Claims, 3 Drawing Sheets



The Google PageRank algorithm



US006285999B1

(12) **United States Patent**
Page

(10) **Patent No.: US 6,285,999 B1**
(45) **Date of Patent: Sep. 4, 2001**

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/035,205, filed on Jan. 10, 1997.

(51) **Int. Cl.**[7] **G06F 17/30**
(52) **U.S. Cl.** **707/5**; 707/7; 707/501
(58) **Field of Search** 707/100, 5, 7, 707/513, 1–3, 10, 104, 501; 345/440; 382/226, 229, 230, 231

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106	*	8/1990	Gansner et al.	345/440
5,450,535	*	9/1995	North	395/140
5,748,954		5/1998	Mauldin	395/610
5,752,241	*	5/1998	Cohen	707/3
5,832,494	*	11/1998	Egger et al.	707/102
5,848,407	*	12/1998	Ishikawa et al.	707/2
6,014,678	*	1/2000	Inoue et al.	707/501

OTHER PUBLICATIONS

S. Jeromy Carriere et al, "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257–1267.*
Wang et al "Prefeteching in Worl Wide Web", IEEE 1996, pp. 28–32.*
Ramer et al "Similarity, Probability and Database Organisation: Extended Abstract", 1996, pp. 272.276.*
Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221–231.*
L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometricka, vol. 18, pp. 39–43.
C.H. Hubbell, "An input–output approach to clique identification sociometry," 1965, pp. 377–399.
Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26–48.
E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471–479.
Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297–312.
N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93–95.
P. Doreian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45–56.

(List continued on next page.)

Primary Examiner—Thomas Black
Assistant Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder L.L.P.

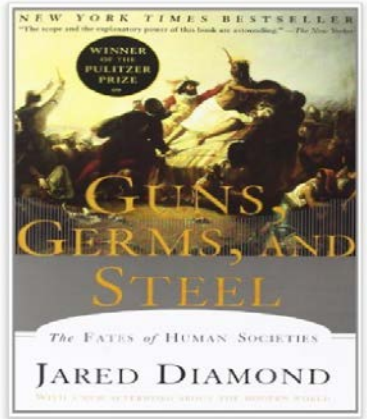
(57) **ABSTRACT**

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

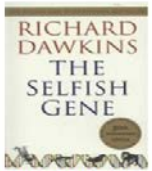
29 Claims, 3 Drawing Sheets

Google

Amazon

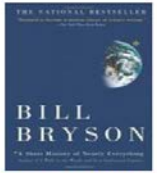


Customers Who Bought This Item Also Bought

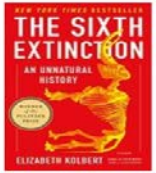


The Selfish Gene: 30th Anniversary Edition--with a new Introduction by the...
› Richard Dawkins
★★★★½ 707
#1 Best Seller in Genetics
Paperback

1493: Uncovering the New World Columbus Created
› Charles C. Mann
★★★★½ 513
Paperback
$10.08 Prime



A Short History of Nearly Everything
› Bill Bryson
★★★★½ 2,179
Paperback
$9.60 Prime



The Sixth Extinction: An Unnatural History
› Elizabeth Kolbert
★★★★½ 1,174
#1 Best Seller in Natural History
Paperback

Facebook recommendation

People You May Know Add Friend Remove

Social networks and population mapping

Paul Butler (Facebook - 2010)

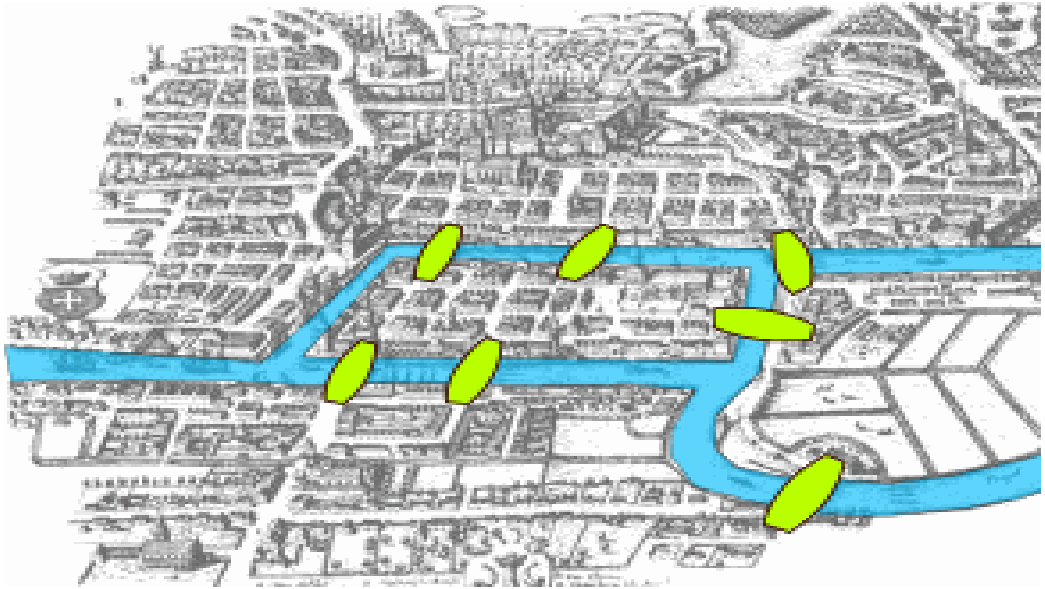
What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)…but also business

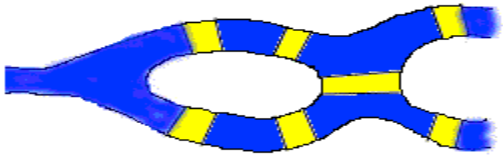
- Network analysis is a broad intellectual approach instead of a narrow set of methods (Wellman, 1983)

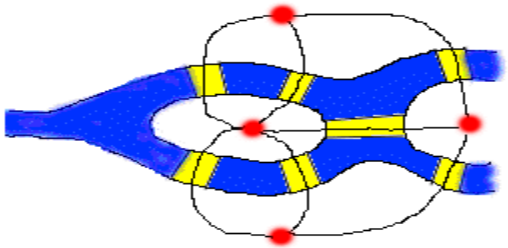
The seven bridges of Königsberg

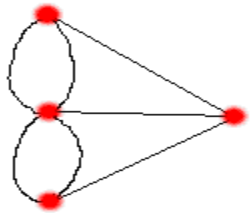
Fundamental problem in the history of mathematics : find a walk through the city that would cross each bridge once and only once



Leonhard Euler solution (1735)







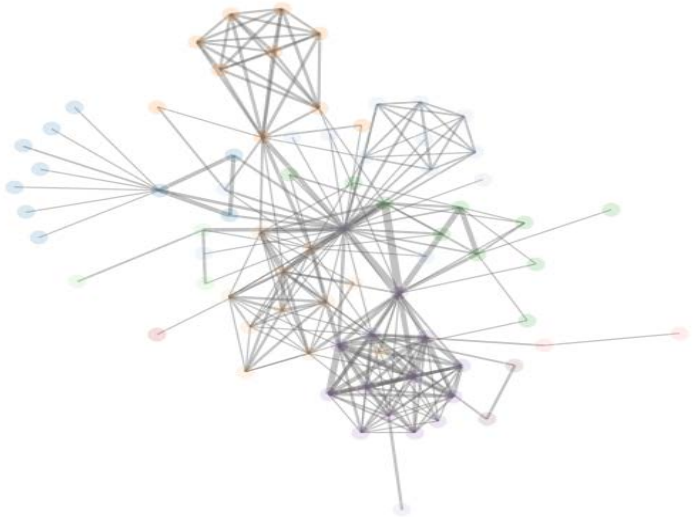
- The route inside each land mass is irrelevant, only the patterns of connection are important
- Abstract reformulation: collapse areas of land separated by the river into points (nodes) connected by the 7 bridges (edges)
- Euler used a network (graph-based) approach to prove that there is no path that would cross each bridge once and only once
- Foundation of graph theory and mathematical topology

What is network thinking?

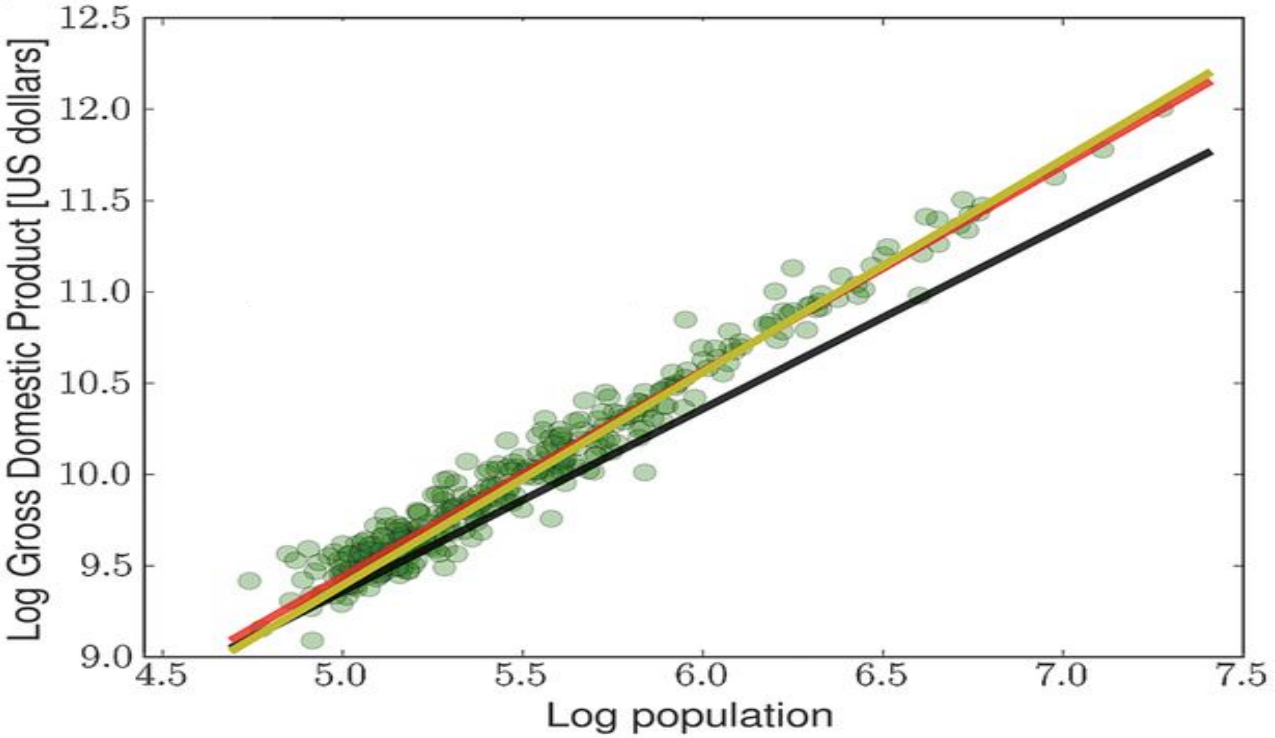
- A network-based paradigm is taking science by storm (Barabási, 2012)…but also business
- Network analysis is a broad intellectual approach instead of a narrow set of methods (Wellman, 1983)
- A network-based paradigm shifts the unit of analysis from **individuals** and their attributes to (the structure of) their **relationships**

Network metrics & visualization

- Network centrality
- Brokerage
- Network density
- Core-periphery structure
- Average path length
- Clustering coefficient
- Communities
- Degree distribution
- Statistical model of network dyna
- ...



Scaling in cities



Bettencourt (2013)

Cities as (complex) networks

Network thinking in music

Graph or networks?

Graph and networks

- Networks are found in nature and society

Graph and networks

- Networks are found in nature and society
- Graphs are the mathematical representation of these networks (a map)

Graph and networks

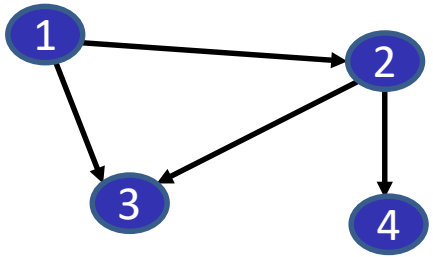
- Networks are found in nature and society
- Graphs are the mathematical representation of these networks (a map)
- In the literature, both are used interchangeably

An advice network

- Emma (1) helps Mason (2)
- Emma (1) helps William (3)
- Mason (2) helps William (3)
- Mason (2) helps Sophia (4)

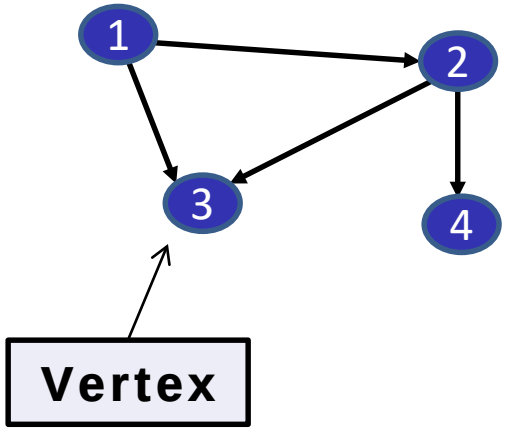
A graph

- Emma (1)
- Mason (2)
- William (3)
- Sophia (4)



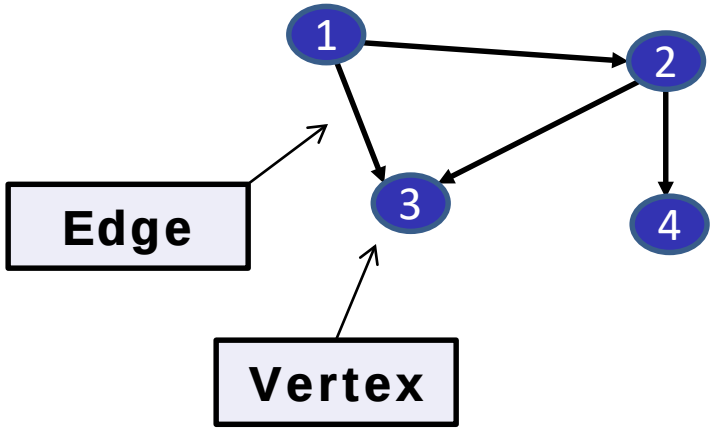
A graph

- Emma (1)
- Mason (2)
- William (3)
- Sophia (4)



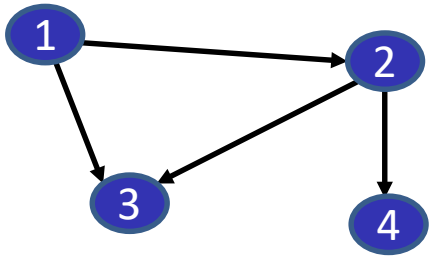
A graph

- Emma (1)
- Mason (2)
- William (3)
- Sophia (4)



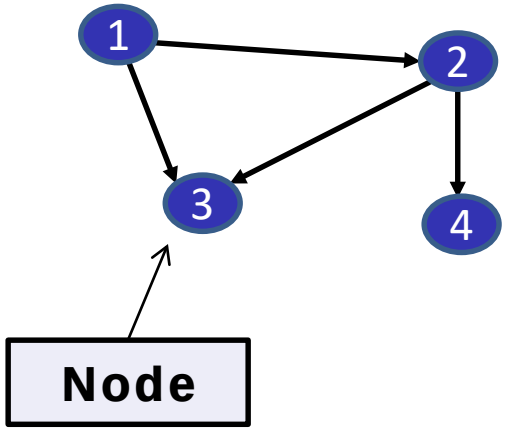
A network

- Emma (1)
- Mason (2)
- William (3)
- Sophia (4)



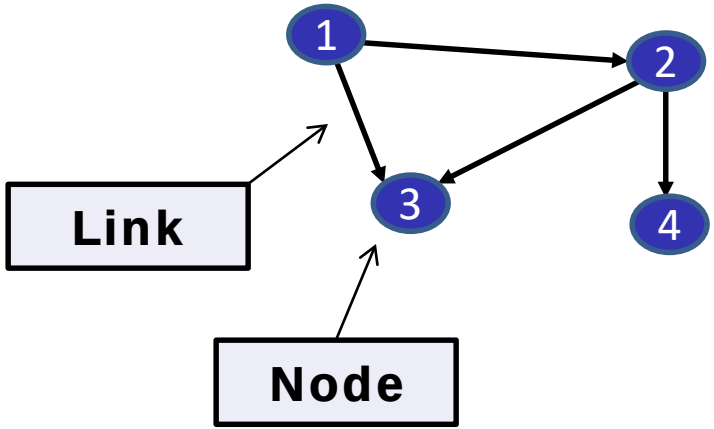
A network

- Emma (1)
- Mason (2)
- William (3)
- Sophia (4)



A network

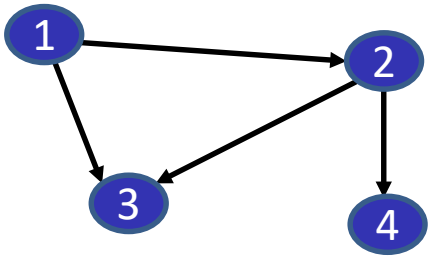
- Emma (1)
- Mason (2)
- William (3)
- Sophia (4)



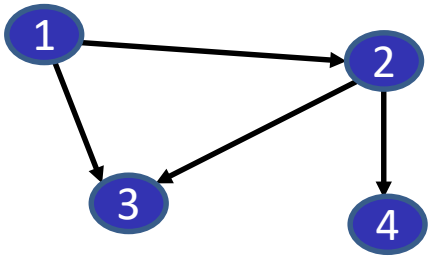
Network terms

- *N* = number of nodes (size of the network)
- *N* = 4
- The network is composed by the nodes *i* = 1, 2, …, N
- *L* = number of links
- *L* = 4
- The connection between Mason and William [*Mason (2) helps William (3)*] is denoted as (2,3)
- A graph might be denoted as G, its vertex set as V(G), and its edge set as E(G)

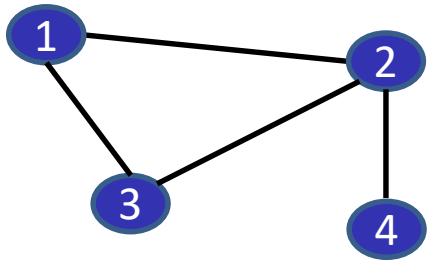
Directed network



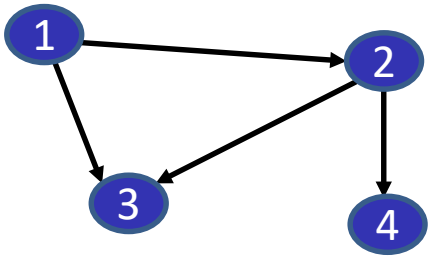
Directed network



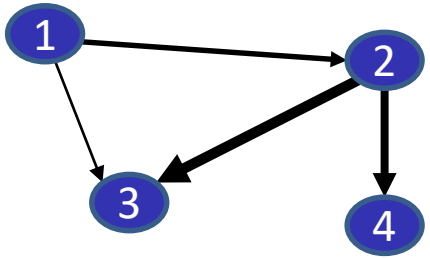
Undirected network



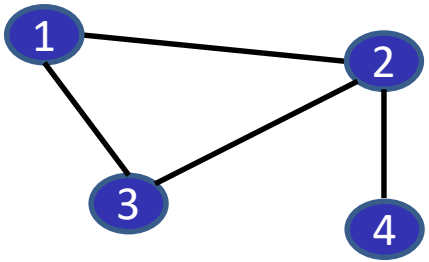
Directed network



Weighted network

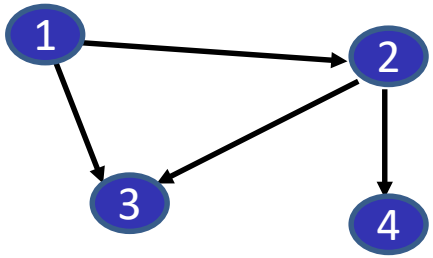


Undirected network

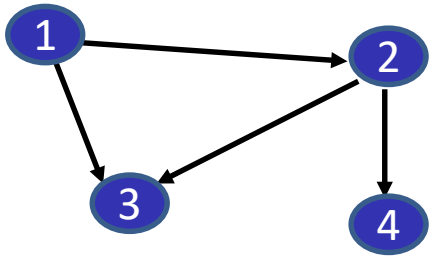


Network name	Nodes	Links	Direction of ties	N
Internet	Routers	Internet connections	No	200,000
WWW	WebPages	Hyperlinks	Yes	500,000
Friendship network	Individuals	Friendship	No	200
Actor network	Actors	Co-acting	No	200,000
Patent citations network	Patent documents	Citations	Yes	7,000,000
Co-invention network	Inventors	Co-patenting	No	200,000

Directed graph (digraph)



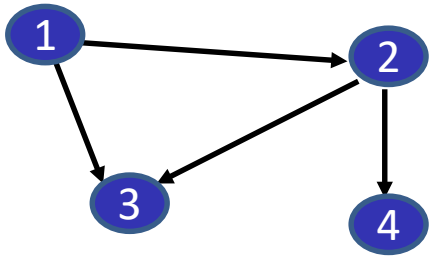
Directed graph (digraph)



Edge list

Vertex	Vertex
1	2
1	3
2	3
2	4

Directed graph (digraph)

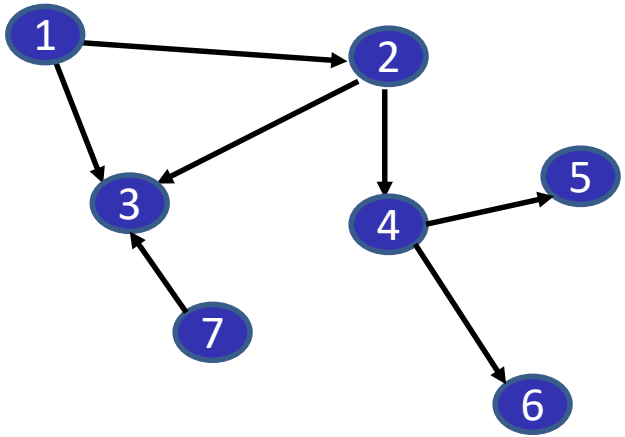


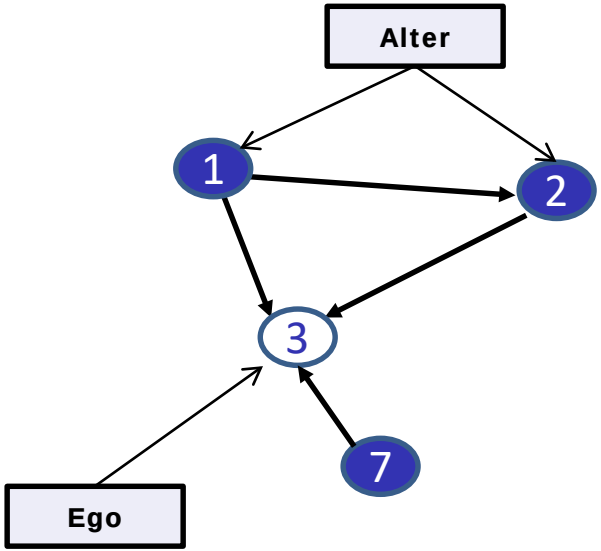
Edge list

Vertex	Vertex
1	2
1	3
2	3
2	4

Adjacency matrix

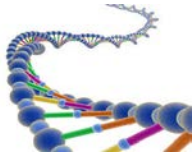
Vertex	1	2	3	4
1	-	1	1	0
2	0	-	1	1
3	0	0	-	0
4	0	0	0	-


1
2
3
4
5
6
7


Alter
1
2
3
7
Ego

Bipartite network (2-mode)



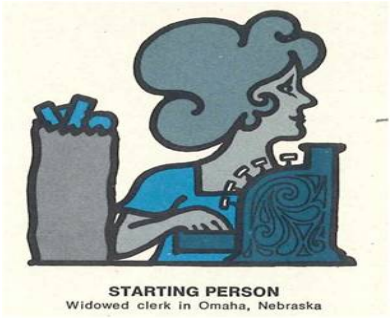
k knowledge domains



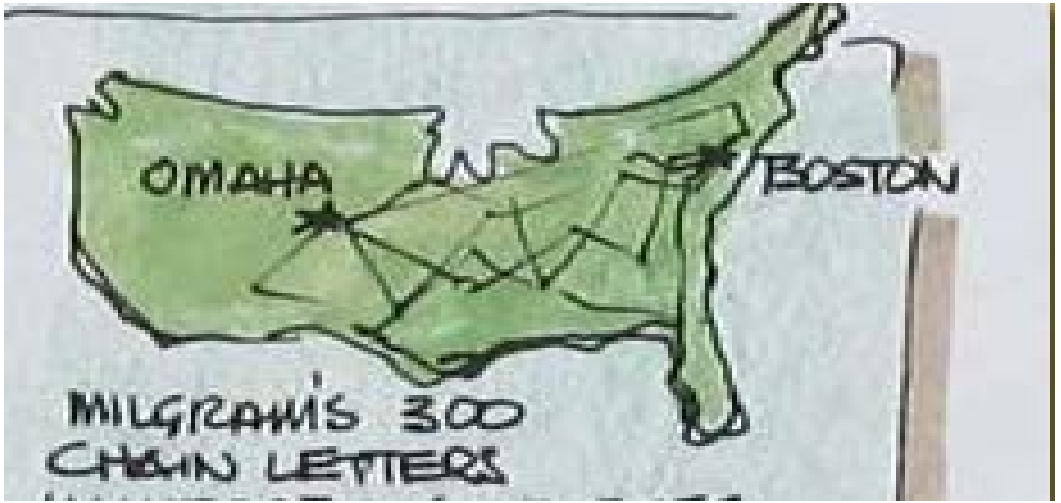

n cities

It is possible to formalize the data that connect **cities** to the **knowledge** they produce as a *n* by *k* bipartite **network** (two different sets of nodes)
A **link** between a city *i* and a knowledge domain *j* means that *i* produces knowledge in category *j*

The small world experiment of Milgram



?



Structural features of networks

- Real-world networks are characterized by:

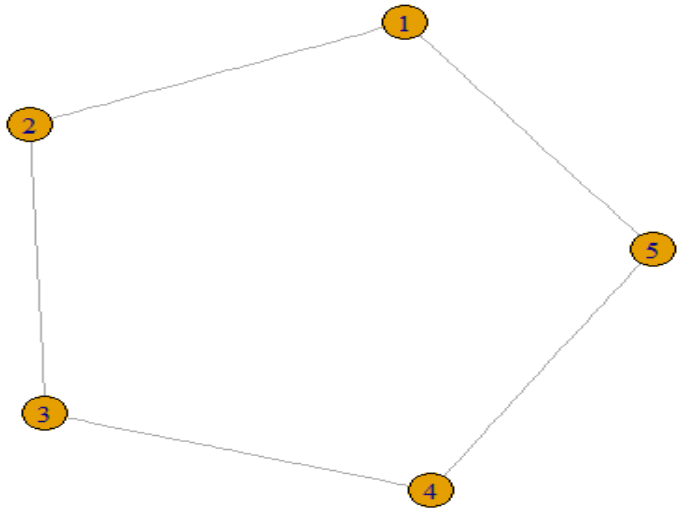
  (1) Small average path length

Structural features of networks

- Real-world networks are characterized by:

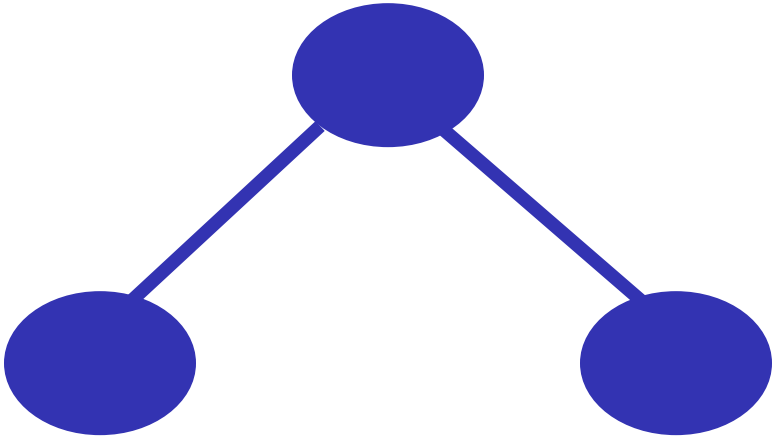
  (1) Small average path length

$$l_g = \frac{1}{n.(n-1)} . \sum_{i \# j} d\ (v_i, v_j)$$



Average path length = (1+2+2+1)/4 = 1.5

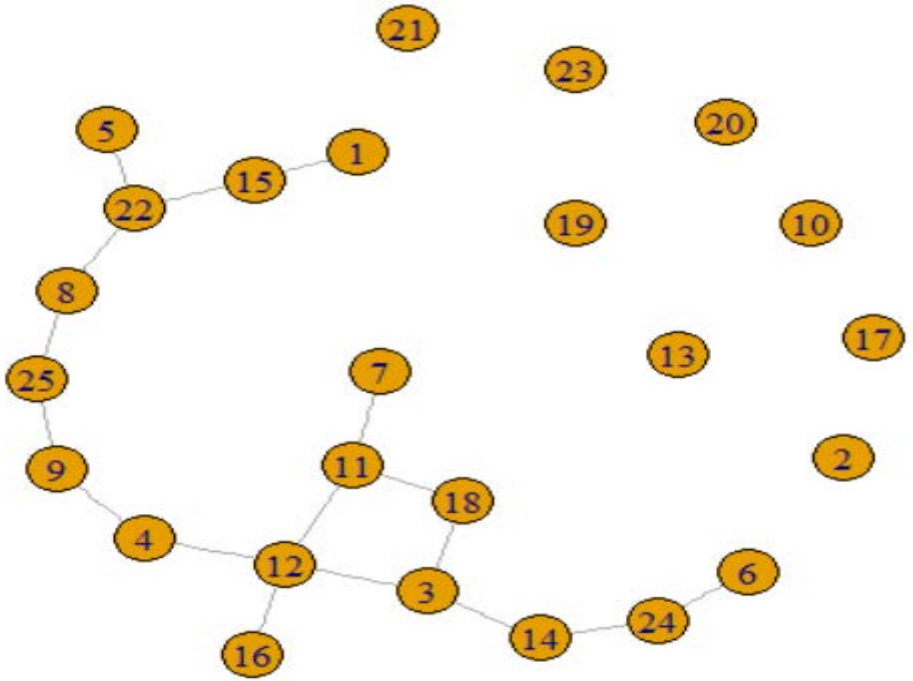
Clustering in networks



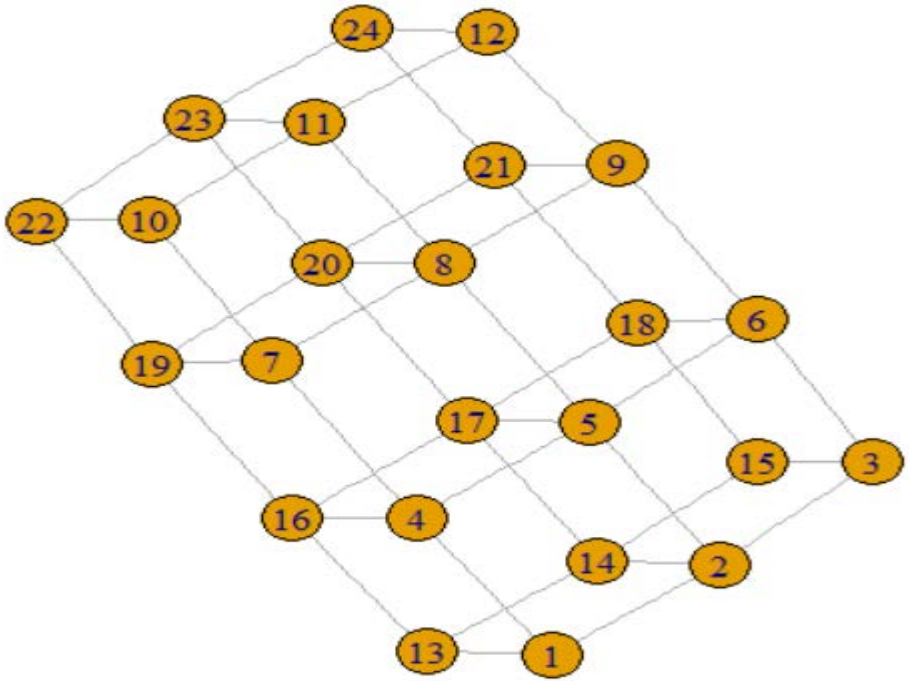
Structural features of networks

- Real-world networks are characterized by:

  (1) Small average path length
  (2) High clustering coefficient [equation]

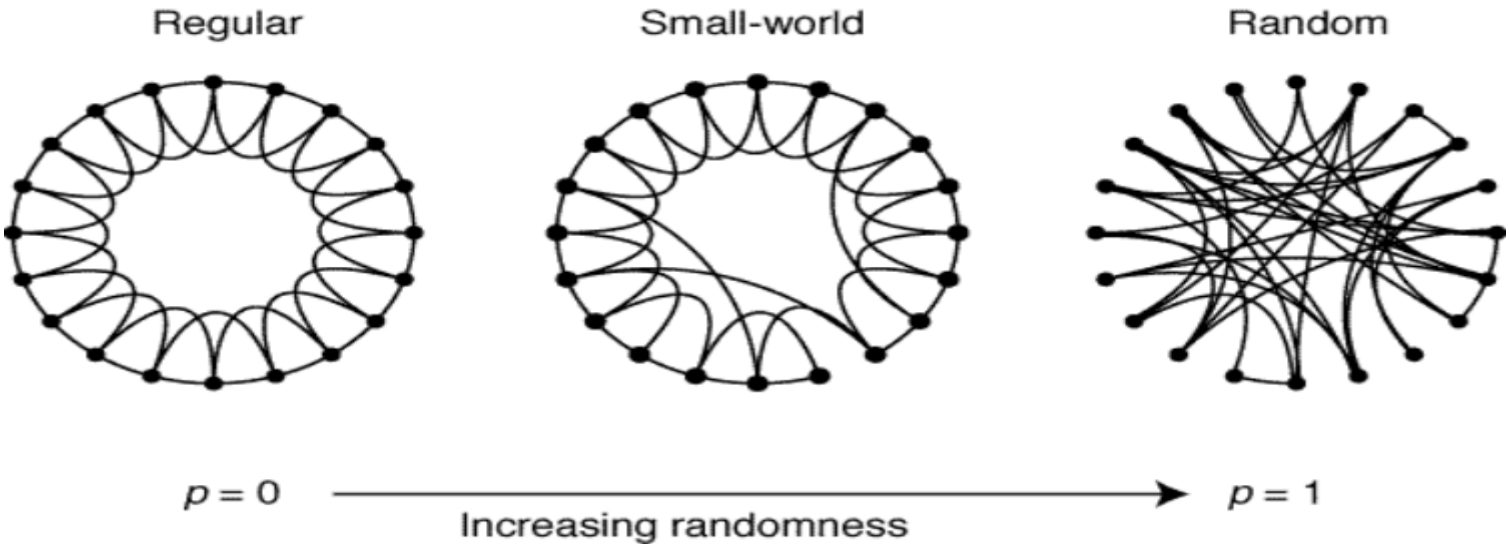


(1) Small average path length



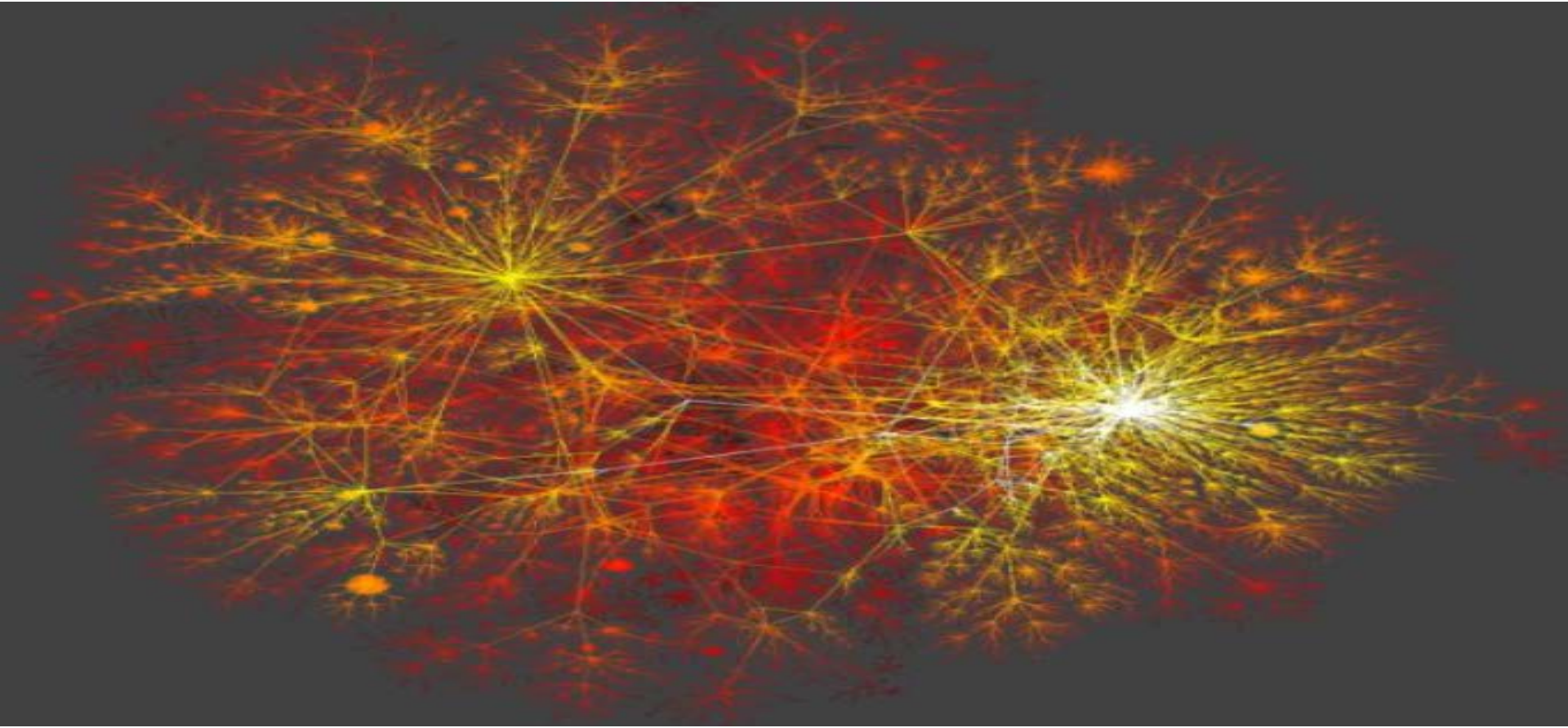
(2) High clustering coefficient

Small world model



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

Structural unequality



Structural features of networks

- Real-world networks are characterized by:

  (1) Small average path length
  (2) High clustering coefficient
  (3) Unequal degree distribution

Complex networks

- Non-trivial topology
- Inherently dynamical properties – emergent behavior
- High structural heterogeneity

Thanks!

paballand.com

github.com/PABalland/EconGeo