MADD! MASTER IN DATA & DESIGN

NETWORKS

Pablo Aragón @elaragon

ELISAVA
Escola Universitària de Barcelona
Disseny i Enginyeria

Hello, World!

———

Research scientist at the Big Data & Data Science group in Eurecat, Technology Centre of Catalonia.

Adjunct professor (former PhD fellow) at the Artificial Intelligence and Machine Learning group in Universitat Pompeu Fabra.

Researcher and board member at Decidim


WHO ARE YOU?

Sessions of the course

15.10 Introduction to Computational Social Science and Network Science: Gephi I

22.10 Community Detection Methods and Centrality Measures: Gephi II

29.10 Digital Methods Tools: DMI-TCAT, Netvizz, table2net, RAWGraphs

05.11 Technopolitical networks: The case of Decidim

20.11 Work on the project (hands-on session)

26.11 Work on the project (hands-on session)

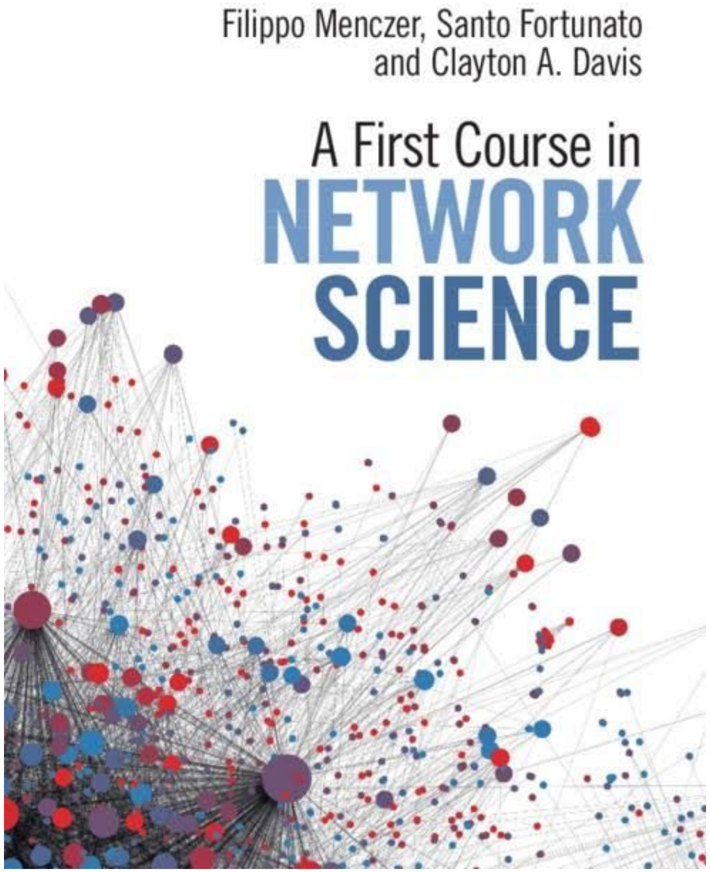
03.12 Work on the project (hands-on session)

10.12 Showcase and Evaluation

Course book

Outline

- Network Elements
- Small Worlds
- Hubs
- Directions and Weights
- Network Models
- Communities
- Dynamics
- Code examples



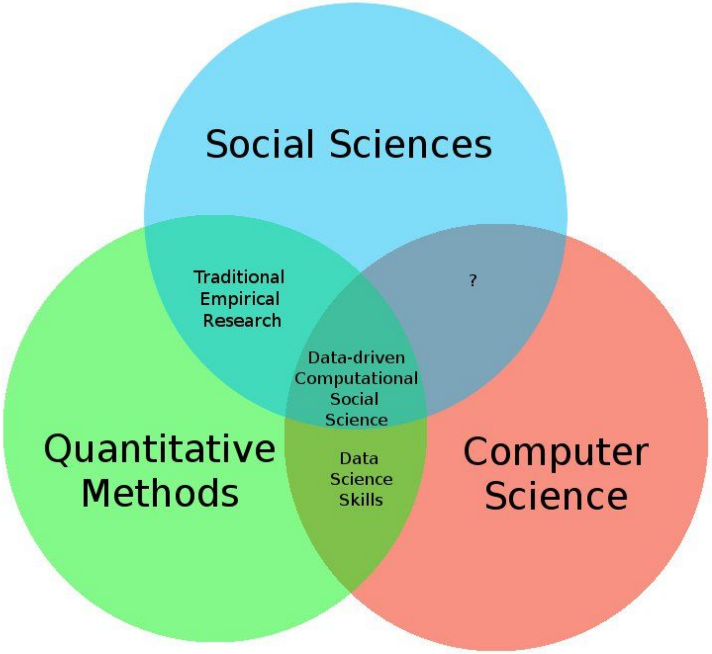
Session I

Computational Social Science and Network Science

Computational Social Science [Lazer 2009]

“The capacity to collect and analyze massive amounts of data has unambiguously transformed such fields as biology and physics.”

“Computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale.”



“These emerging data sets surely must offer some qualitatively new perspectives on collective human behavior.”

The End of Theory? [Anderson 2008]

———

“With enough data, the numbers speak for themselves”

“But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete”

“We can analyze the data without hypotheses about

what it might show”

CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

Illustration: Marian Bantjes "**All models are wrong**, but some are useful."

So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies like Google, which have grown up in an era of massively abundant data, don't have to settle for wrong models. Indeed, they don't have to settle for models at all.

The End of Theory? [Anderson 2008]

———

“With enough data, the numbers speak for themselves”

“But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete”

“We can analyze the data without hypotheses about

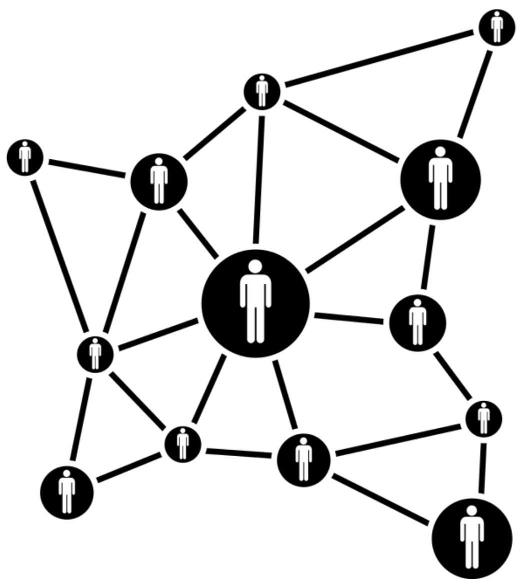
what it might show”

@elaragon

Scientific model

Abstract representation (a simplification of reality) to

- analyse
- explain
- simulate



Source: gpiberia.es

a complex system (e.g., a social network)

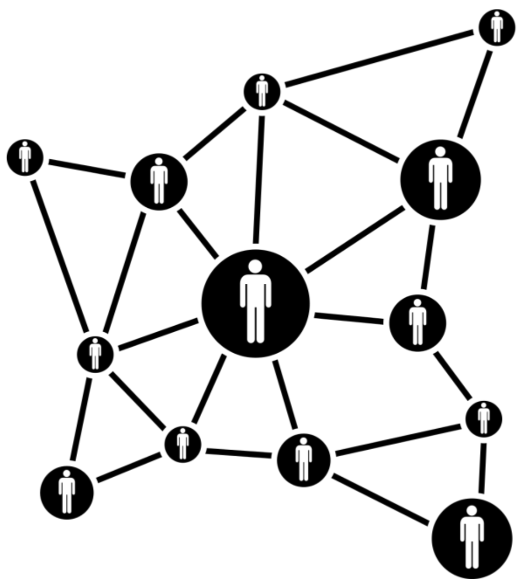
Social Network Analysis

What is social network analysis?

- Researching social relationships
- Mapping relationships between individuals with graph theory

Graph Theory

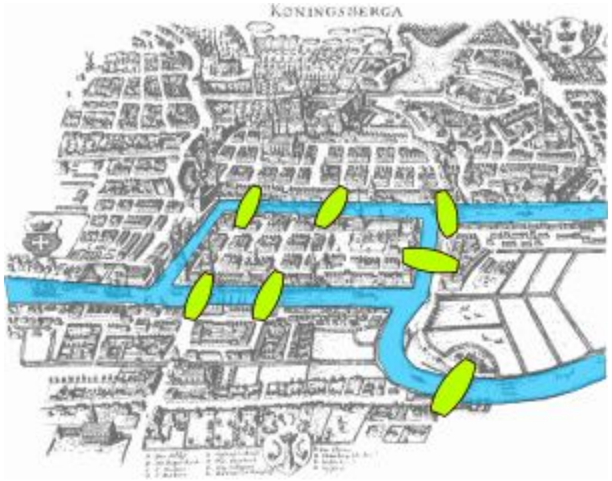
- Model network structure:
  - Individuals as nodes
  - Relationships as edges



Source: gpiberia.es

Origin of Graph Theory

The city of Königsberg in Prussia (now Kaliningrad, Russia) was set on both sides of the Pregel River, and included two large islands—Kneiphof and Lomse—which were connected to each other, or to the two mainland portions of the city, by seven bridges.. **is it possible to cross the seven bridges and return to the same point without going over the same bridge twice?**



Source: Wikipedia

Origin of Graph Theory

Leonard Euler determined that for this problem

- The intermediate points of a possible route must necessarily be connected to an even number of lines

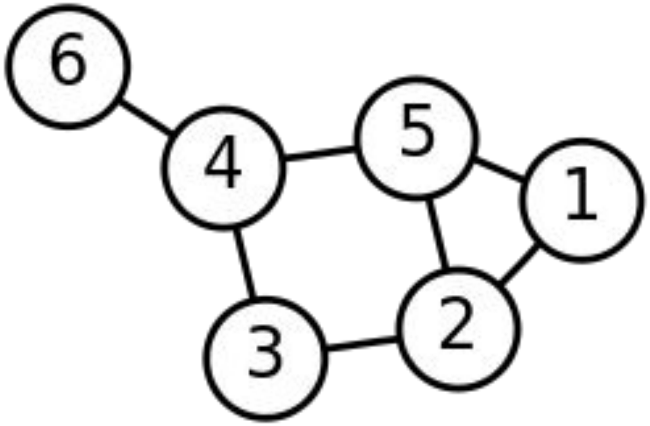
- If we get to a point from some line, then the only way to get out of that point is by a different line.

Source: Wikipedia

Graph Theory

A graph G is an ordered pair G=(V,A), where:

- V is a set of vertices or nodes, and
- A is a set of edges or links, which connect these nodes.

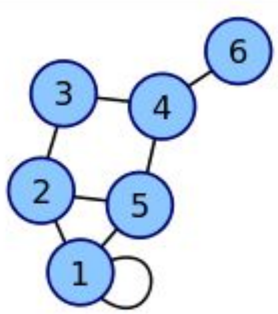


Source: Wikipedia

Graph representation

A graph G=(V,A), can be represented as:

- a list,
- a matrix.

Grafo G(V,A)	Conjuntos	Lista de Adyacencia	Matriz de adyacencia	Matriz de incidencia
	V = { 1, 2, 3, 4, 5, 6 } A = { {1,1}, {1,2}, {1,5}, {2,3}, {2,5}, {3,4}, {4,5}, {4,6} }	{ {1,2,5}, {3,5}, {4}, {5,6} }	$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$

Source: Wikipedia

Any network as a graph

Barcelona Metro: each station as a node and each connection as an edge.

Undirected vs. directed graphs

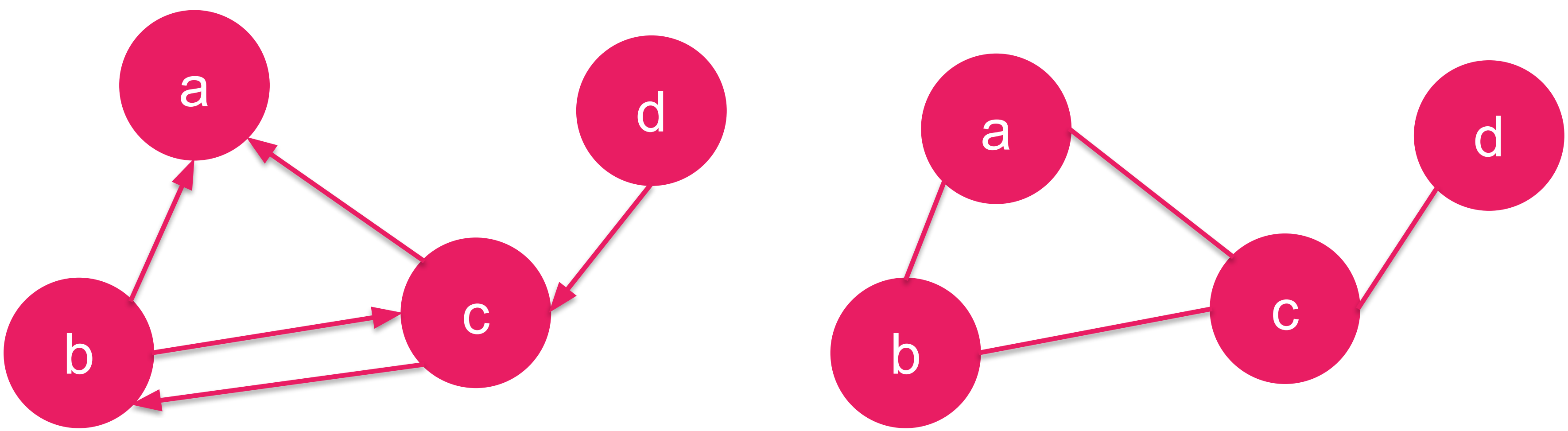
Undirected: the relationship is reciprocal

- Co-authors in scientific articles
- Friends on Facebook
- Sexual interactions

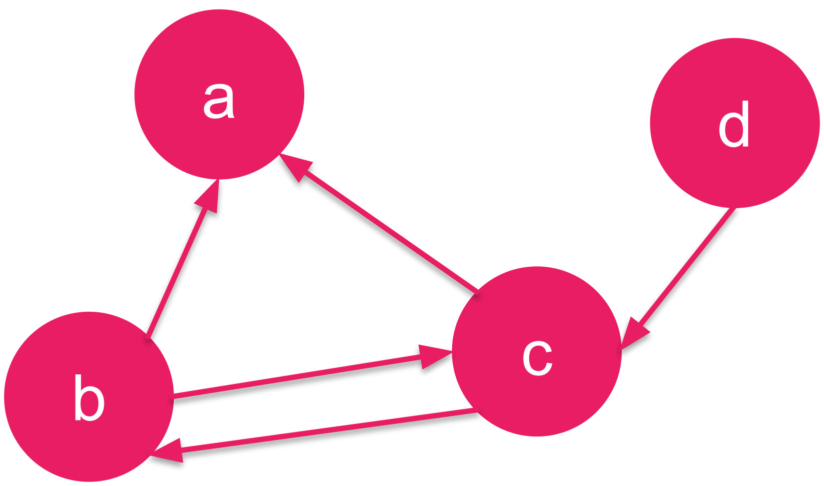
Directed: the relationship is not reciprocal

- Followers on Twitter
- Replies on Facebook
- Forum interactions

Directed vs. undirected graphs



Directed vs. undirected graphs



In directed graphs:

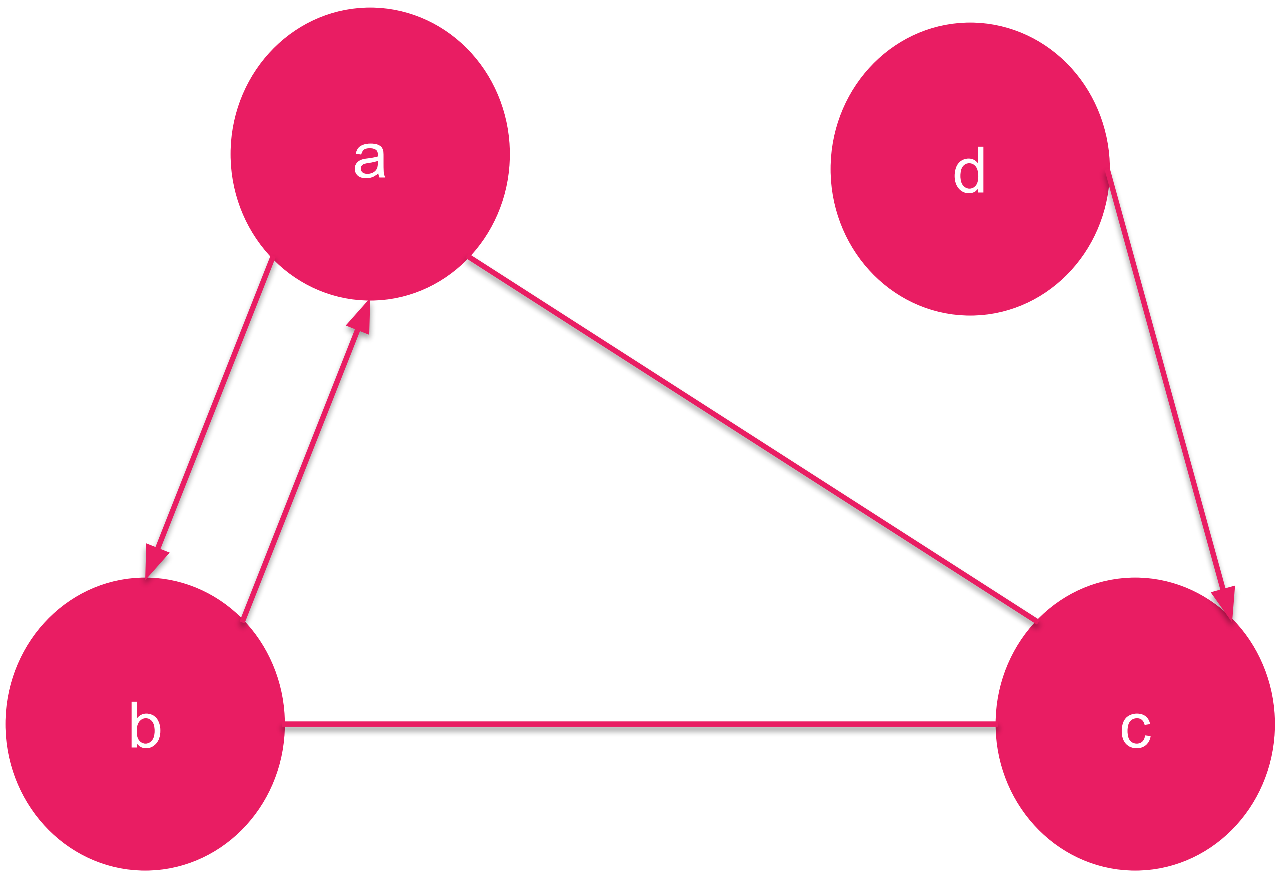
Sink

Node with incoming links only (A)

Source

Node with outgoing links only (D)

Mixed graphs



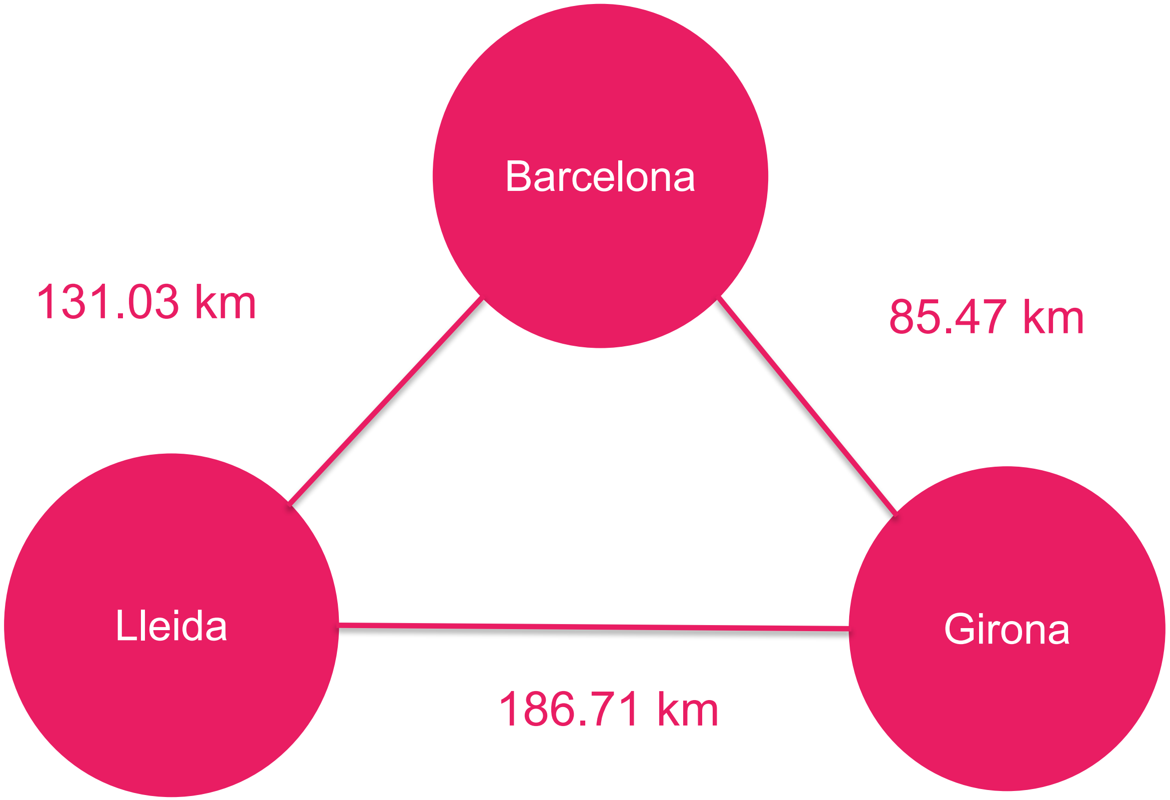
Weighted graphs

The edges are weighted to express some attribute of intensity such as probability, distance, time, e.g:

- Distance between two cities
- Time between metro stations

A network with weighted edges is denoted as weighted network

Weighted graphs



@elaragon

Signed graphs

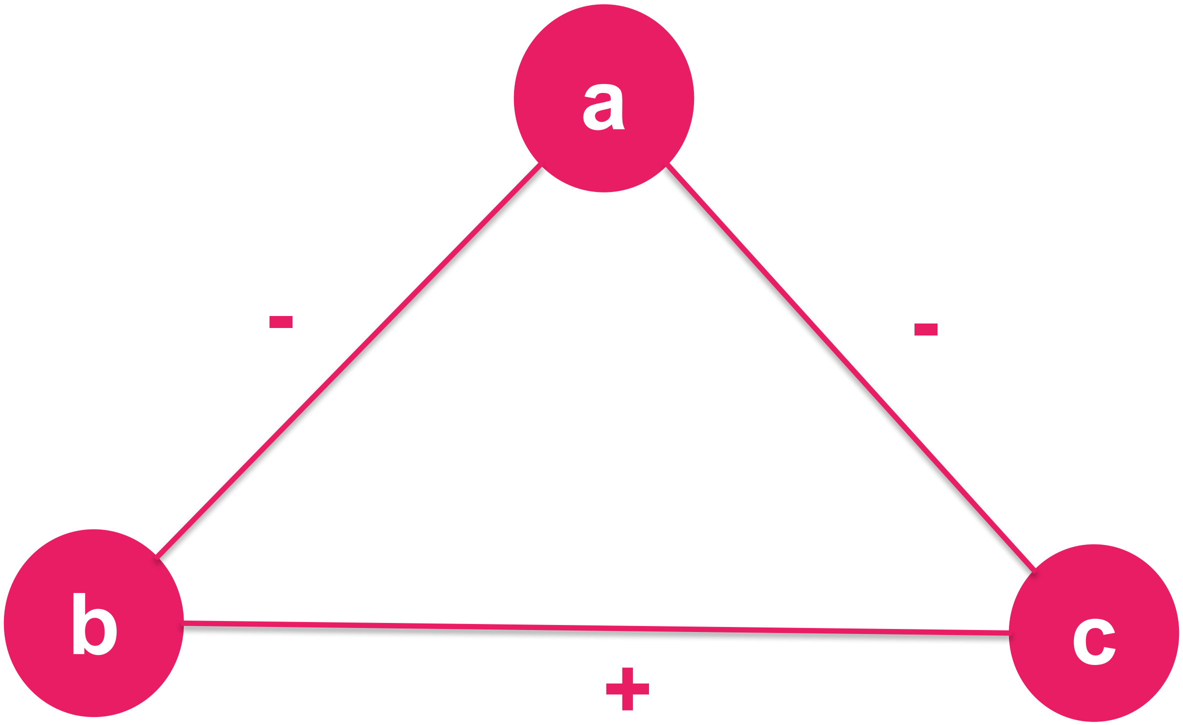
Definition

- Network in which the edges can be positively charged (+) or negative charged (-)
- Some allow zero charge

Examples

- Networks to represent political allies/adversaries
- Networks to represent a belief system

Signed graphs

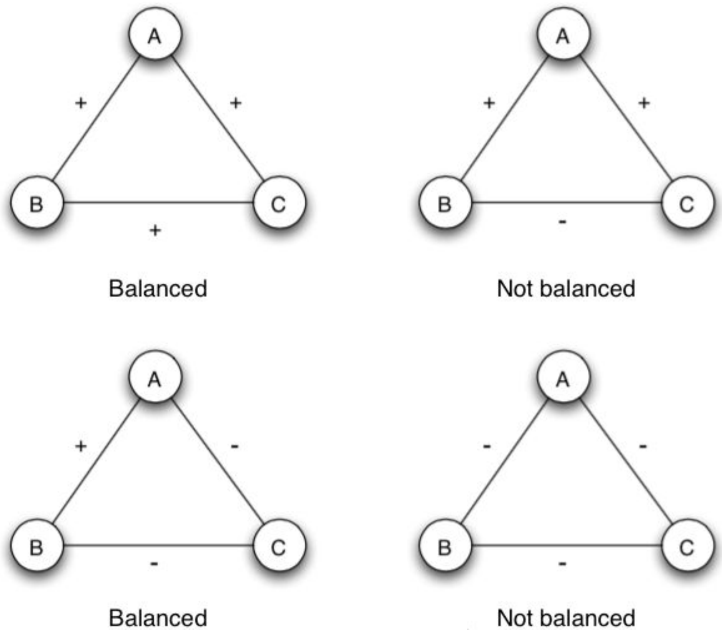


Balance Theory [Heider 1946]

“People tend to maintain consistency in patterns of their liking and disliking of one another and of inanimate objects.

When patterns of liking and disliking are balanced, structures are stable.

When they are imbalanced, structures are unstable and there is pressure to change in the direction that makes them balanced”.

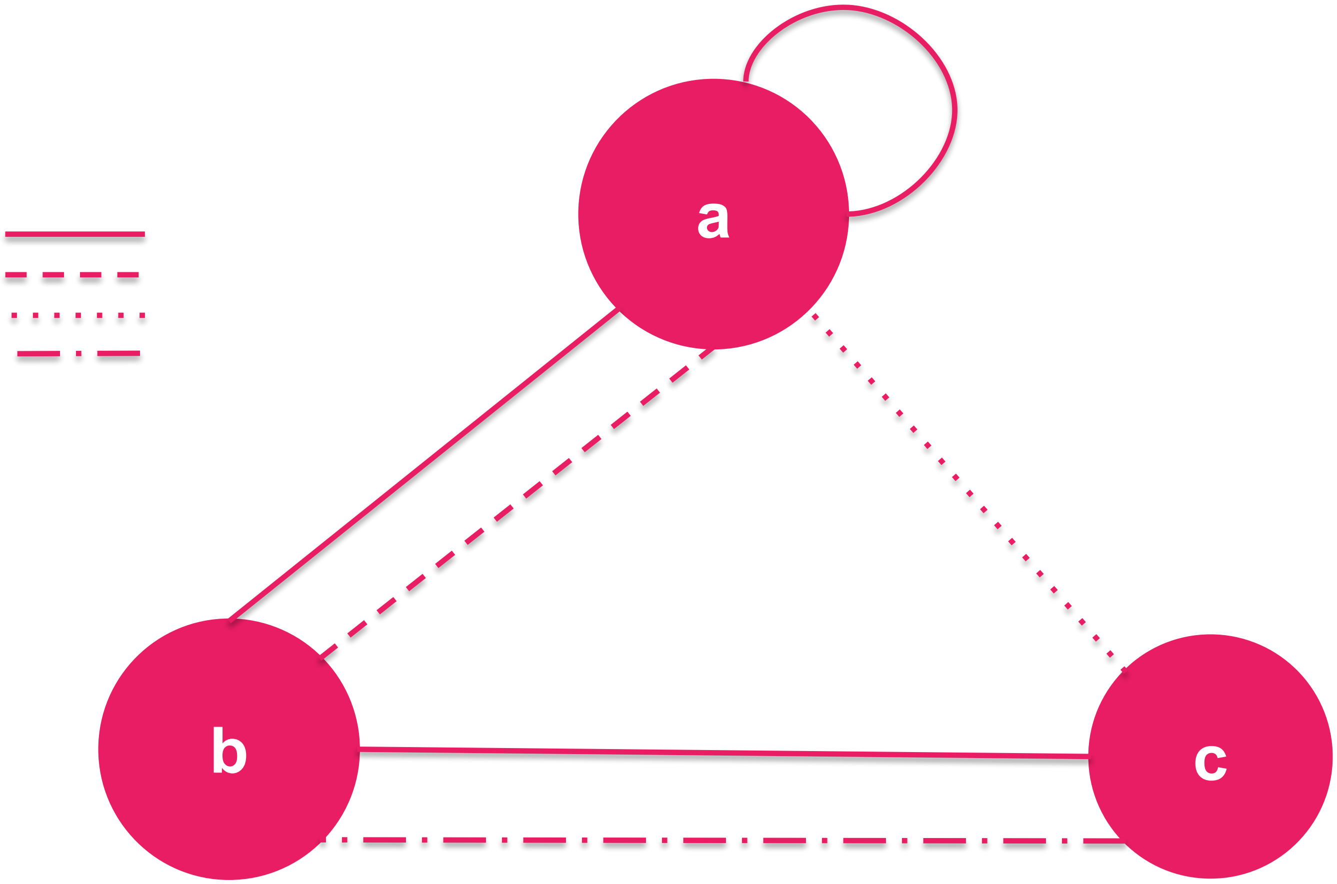


Multigraphs

Definition

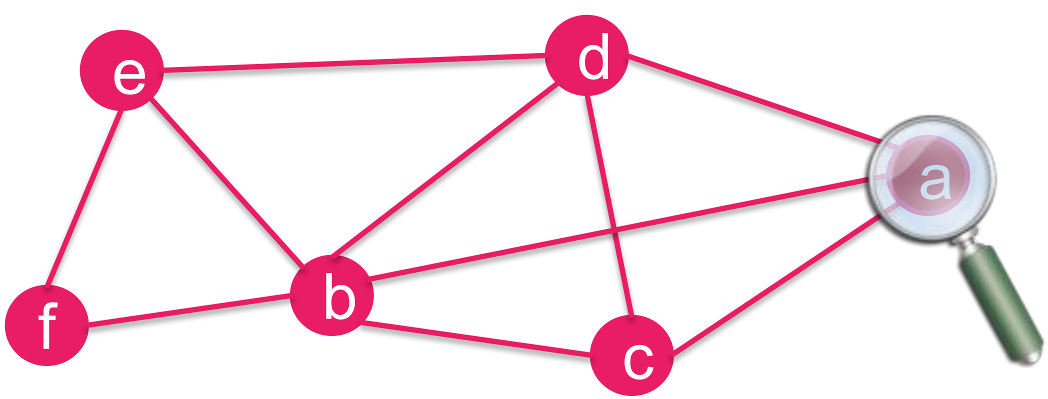
- Graph in which two nodes can be connected with multiple and usually different types of edges.
- The different edges between nodes model different types of relationships.
- They contain loops depending on the domain
- Many "real world" networks are multi-graph.

Multigraphs



Walks

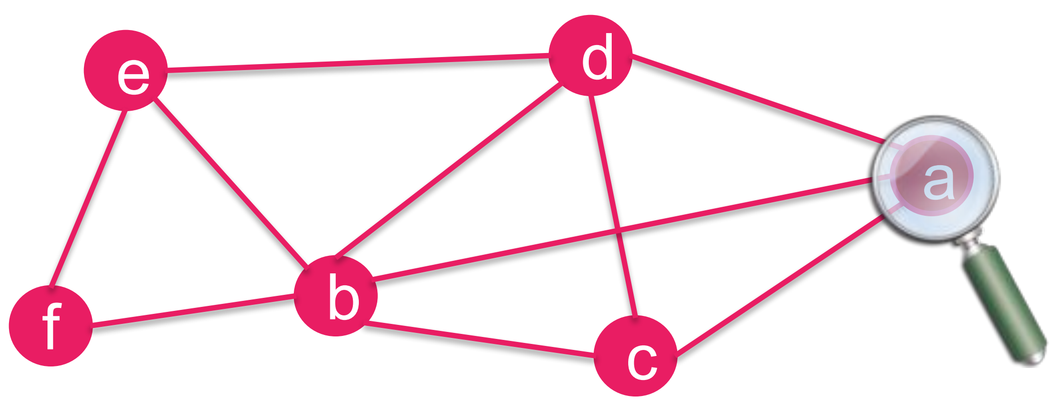
- **Walk:** Sequence of alternating vertices and edges such as $v_0,e_1,v_1,e_2,...,e_k,v_k$ where each edge $e_i=\{v_{i-1},v_i\}$
- **Closed walk:** Walk where $v_0=v_k$
- **Open walk:** walk where $v_0\neq v_k$



@elaragon

Trails, paths, cycles and circuits

- **Trail**: Walk with no repeated edges
- **Path**: Open trail with no repeated vertices
- **Cycle**: Closed trail where no other vertices are repeated apart from the start/end vertex
- **Circuit**: Closed trail with no repeated edges (may have repeated vertices)



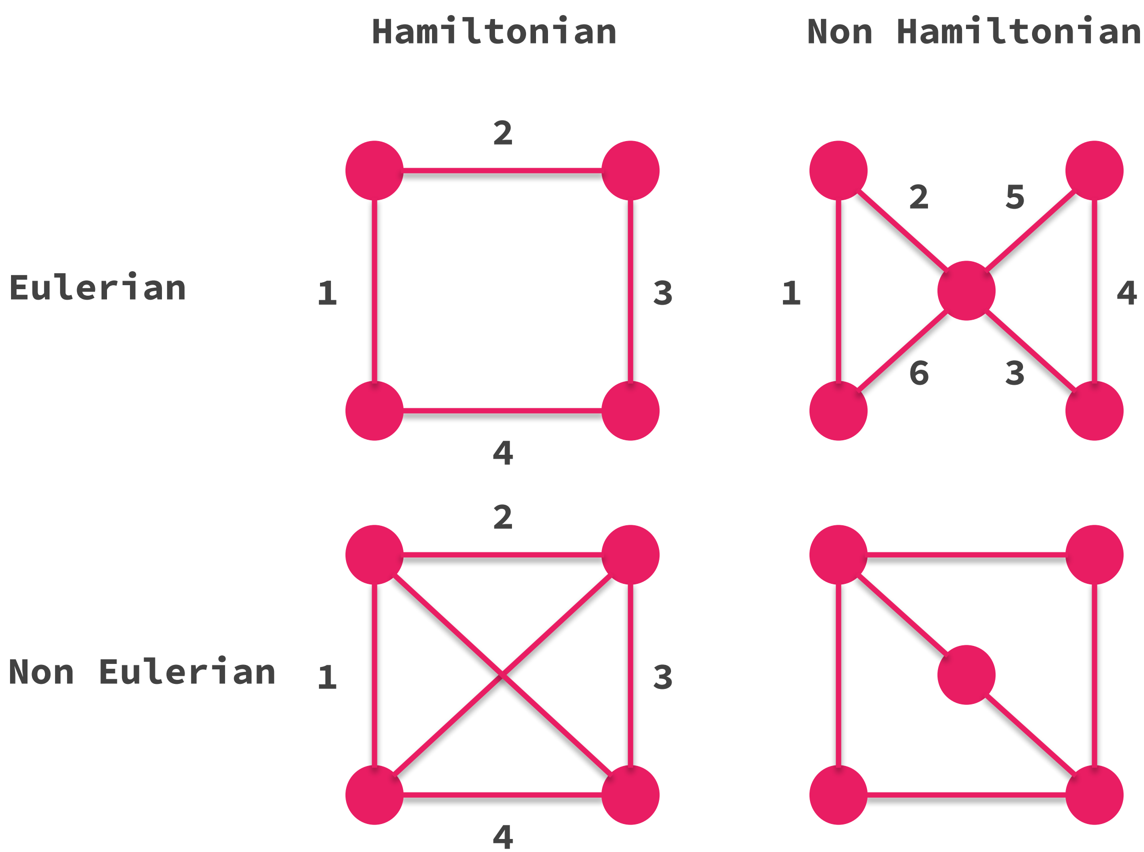
Type of cycles

Eulerian

A cycle that travels exactly once over each edge in a graph.

Hamiltonian

A cycle that travels exactly once over each vertex in a graph



Tree graphs

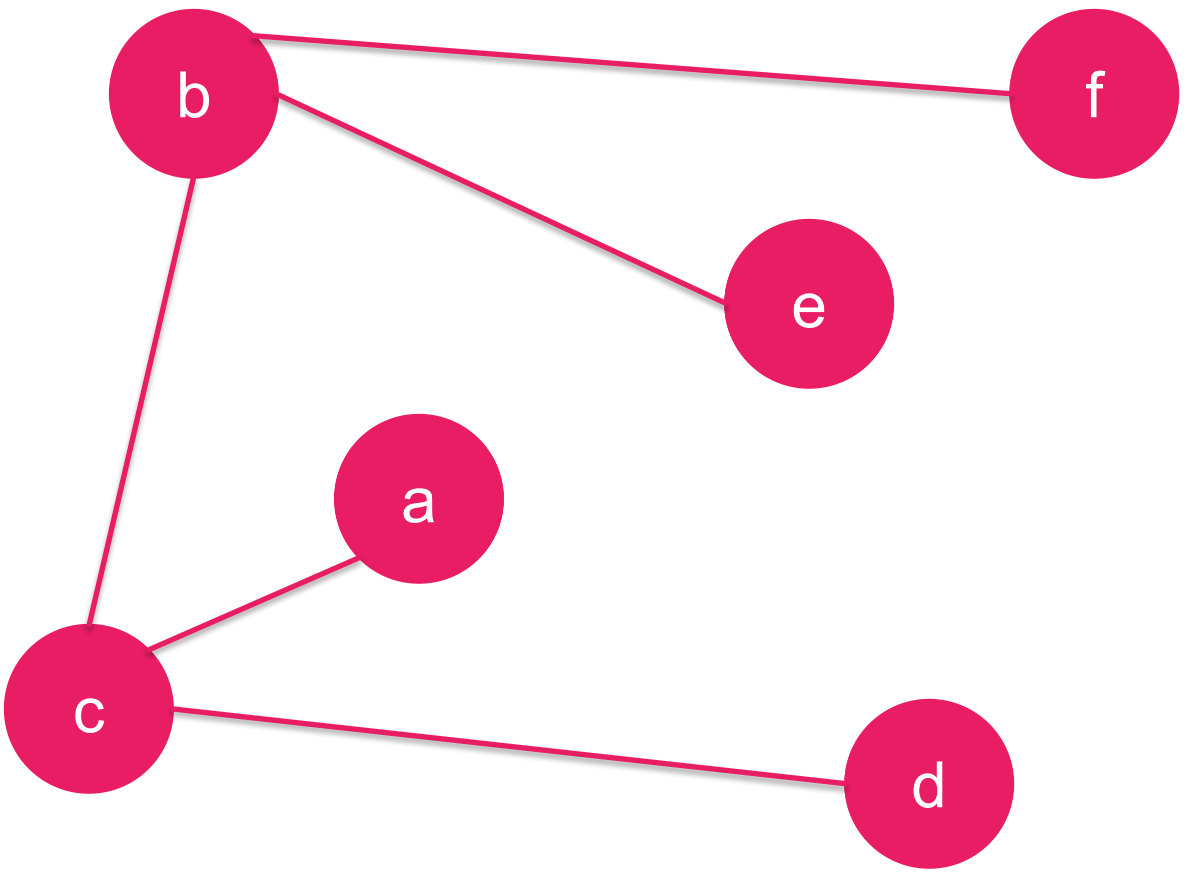
Definition

A graph in which any two vertices are connected by exactly one path.

Properties

- It is connected
- It does not contain cycles
- The number of edges is one unit less than the number of vertices

Tree graphs



Rooted tree graphs

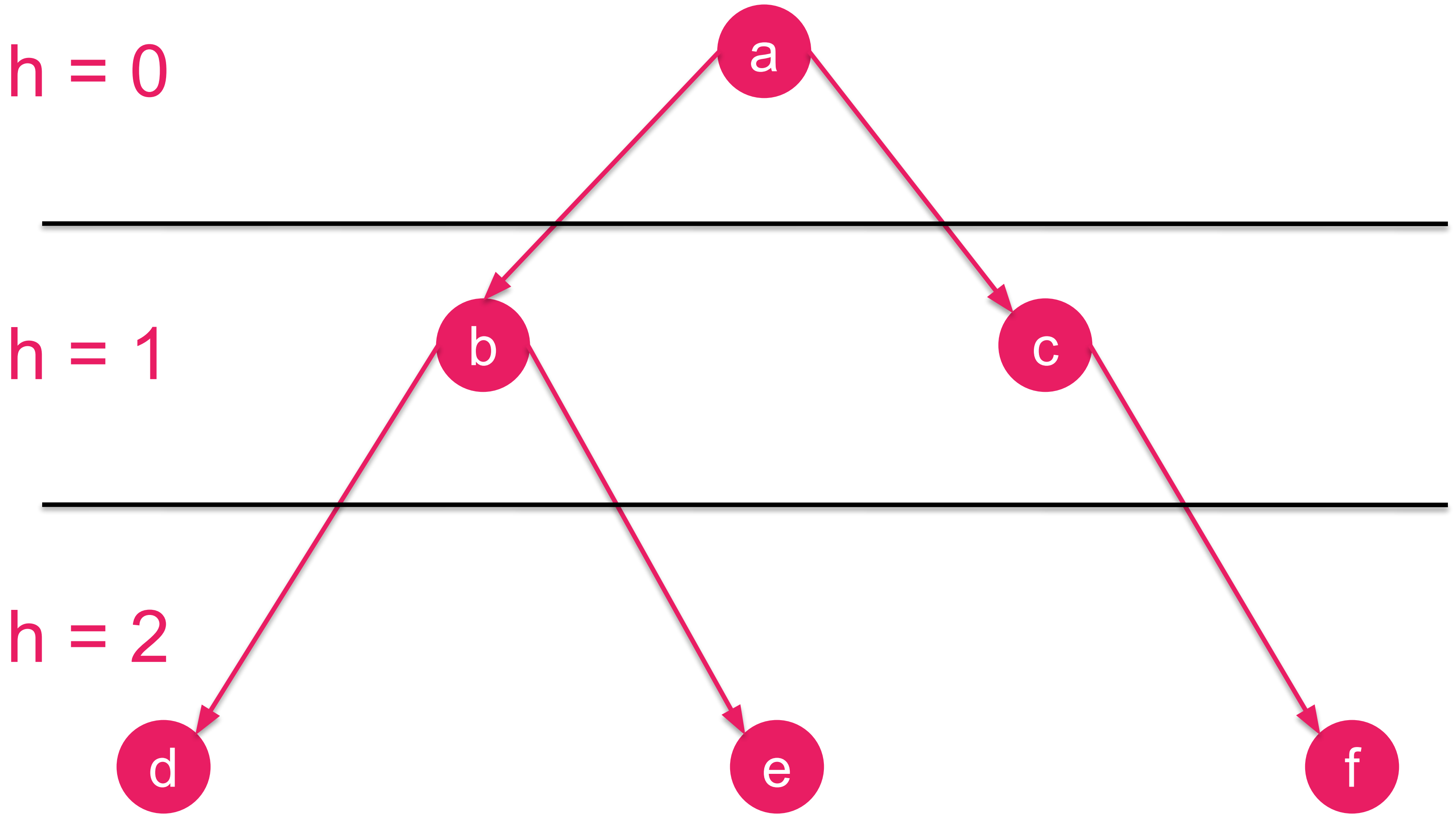
Definition

Directed tree with a node designated as root and all edges are oriented to/from the root.

Properties

- Each node has a parent and 0-n children
- All nodes at a distance L from the root are at the level L
  - Depth: Maximum tree level
- They model threads, taxonomies, organization charts, etc.

Rooted tree graphs



Forest graphs

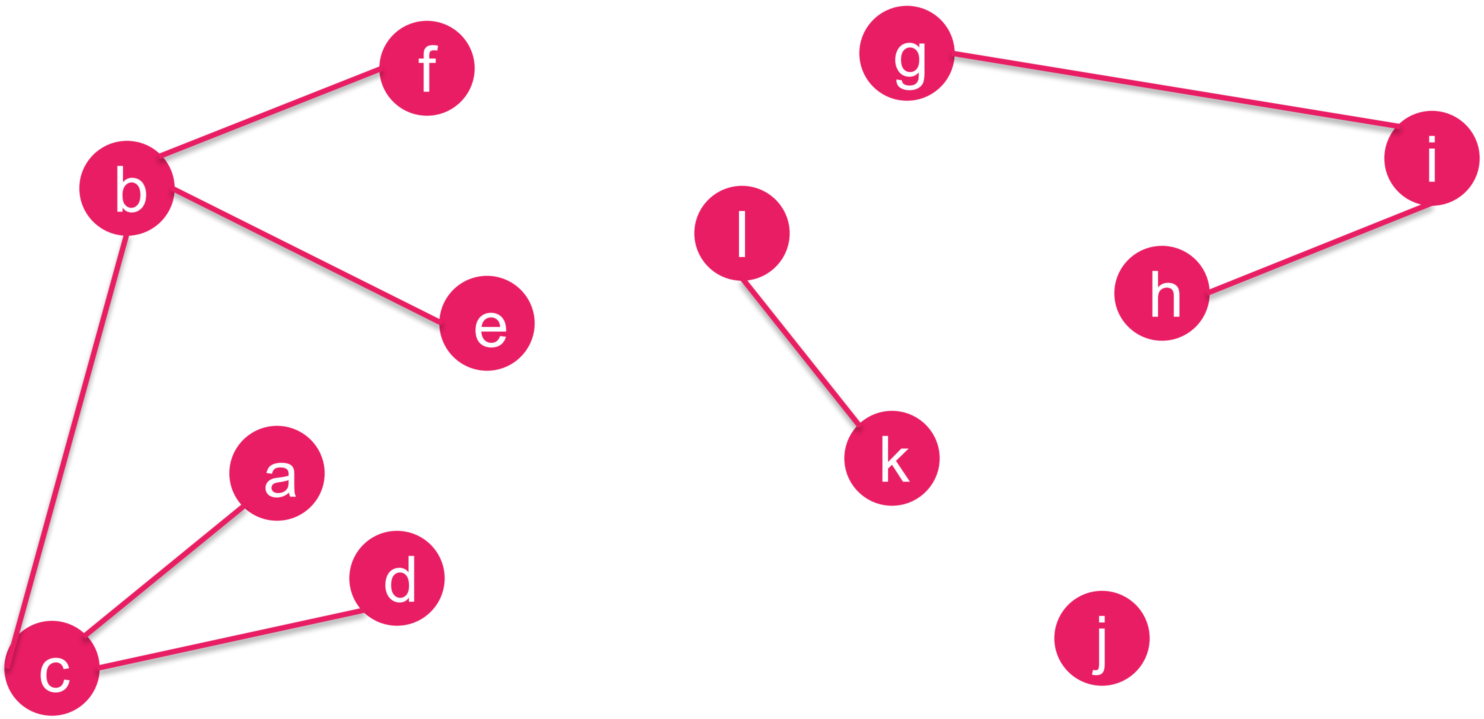
Definition

Graph with no cycles

Properties

- It doesn't have to be connected
- If it's connected, it's a tree
    - Every tree is a forest but not every forest is a tree

Forest graphs



@elaragon

Complete graphs

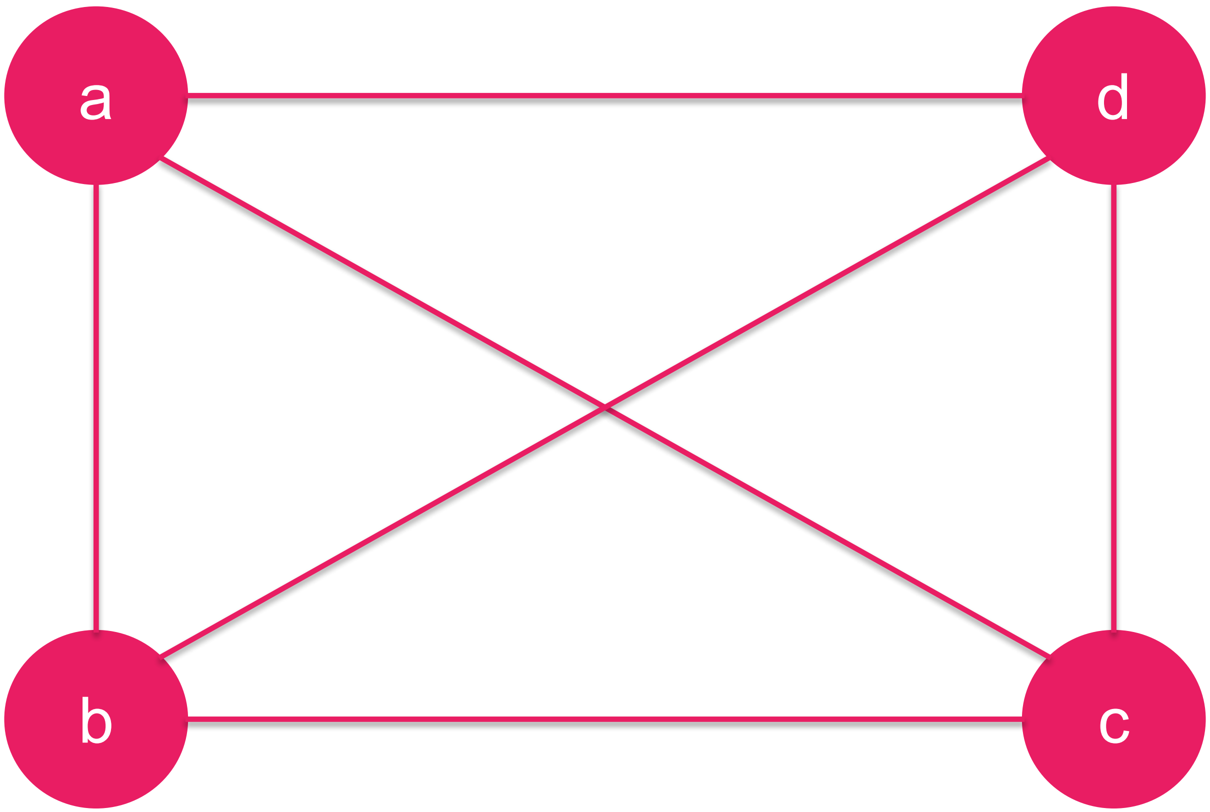
Definition

Undirected graph in which every pair of distinct nodes is connected by a unique edge (aka. clique)

Properties

- If it has n nodes, it denoted K_n (after Kuratowski)
- The number of edges is n(n-1)/2

Complete graphs



Bipartite graphs

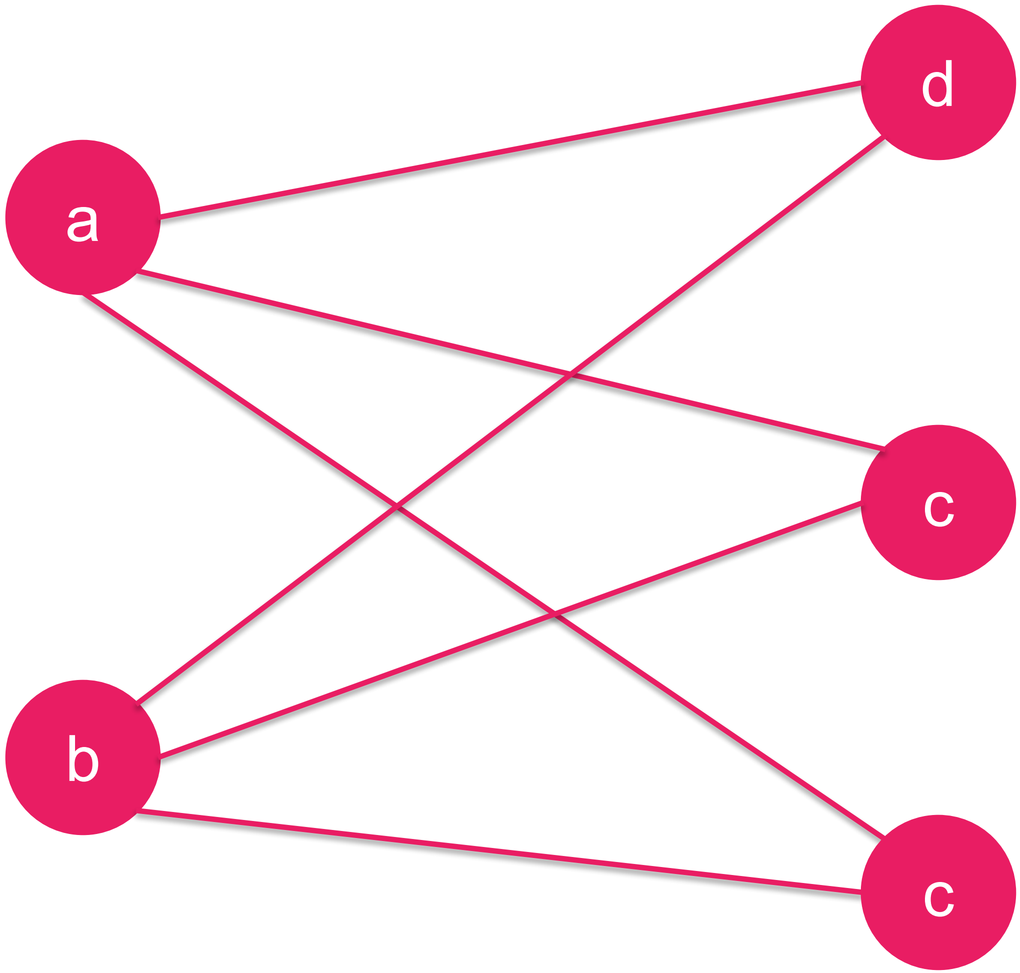
Definition

Graph whose nodes can be divided into two disjoint and independent partitions such that every edge connects two vertices from the two different partitions.

Properties

- If each node in a partition is connected to all nodes in the other partition, it is called $K_{n,m}$ (after Kuratowski)
- The number of edges is $n \cdot m$

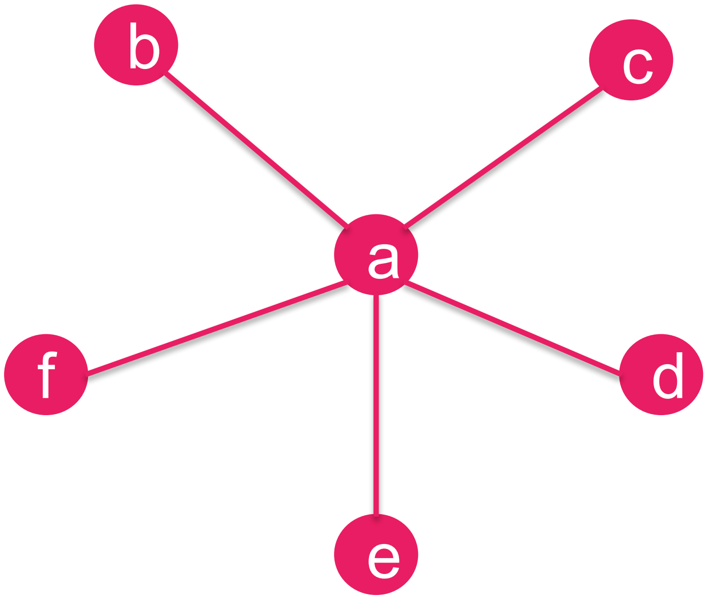
Bipartite graphs



Star graphs

Definition

Graph in which one node is connected to all the nodes and all the other nodes are only connected to it



[Ugarte 2011]

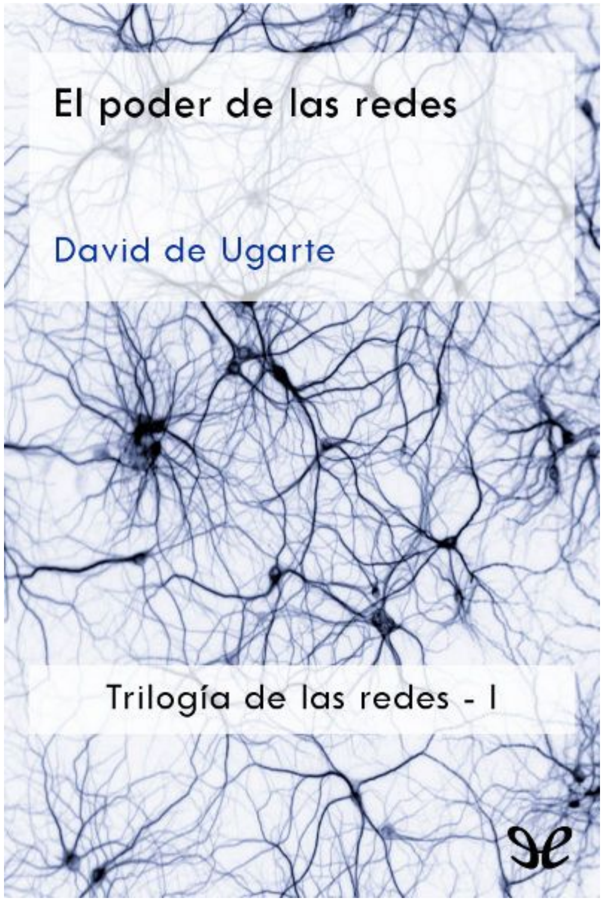
Illustrated manual for

- citizens,
- civic organizations
- enterprises

involved in cyberactivism

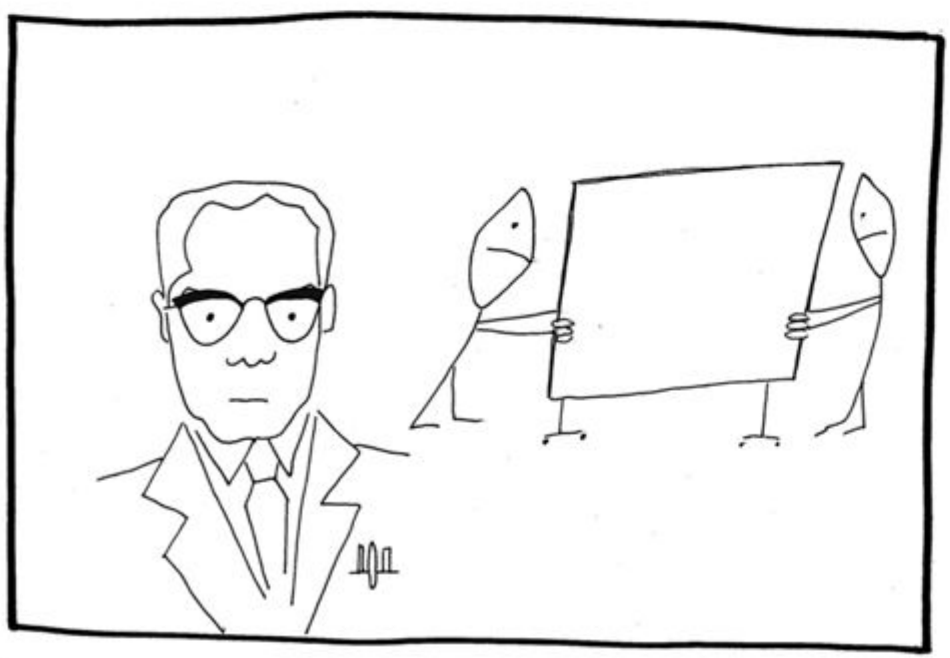
Available at:

http://www.pensamientocritico.org/davuga0313.pdf



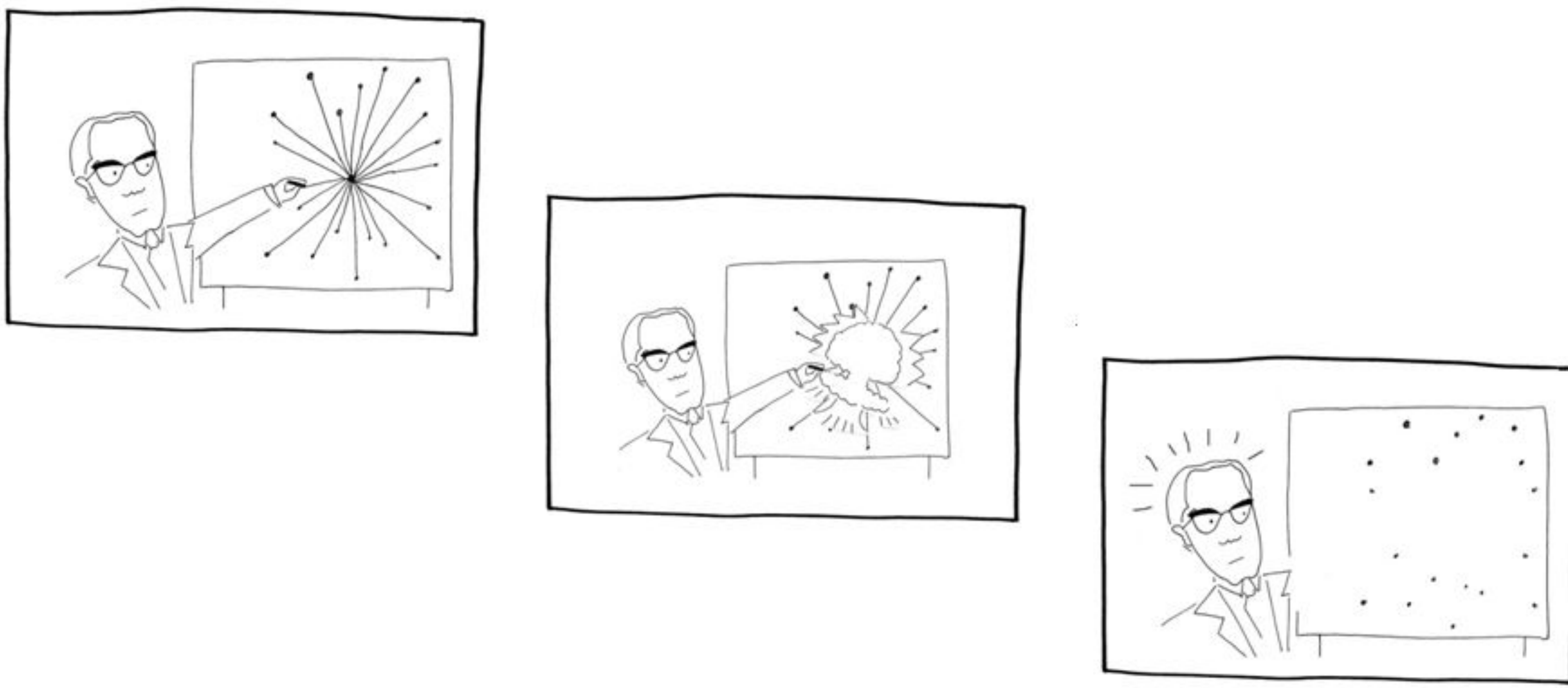
[Ugarte 2011]

In 1964 Paul Baran was commissioned by the RAND Corporation, the American defence scientist think tank to protect American networks from a possible Soviet attack.



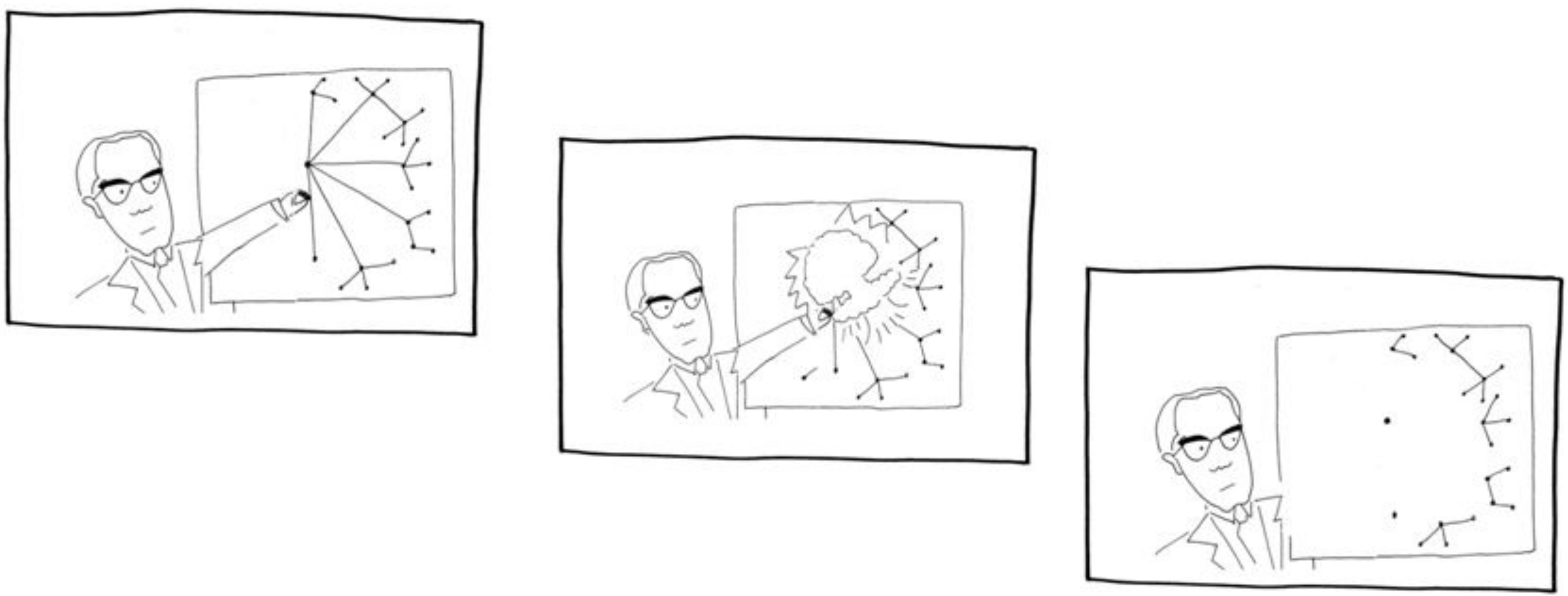
[Ugarte 2011]

The disconnection of the **central node** immediately destroys the entire network.



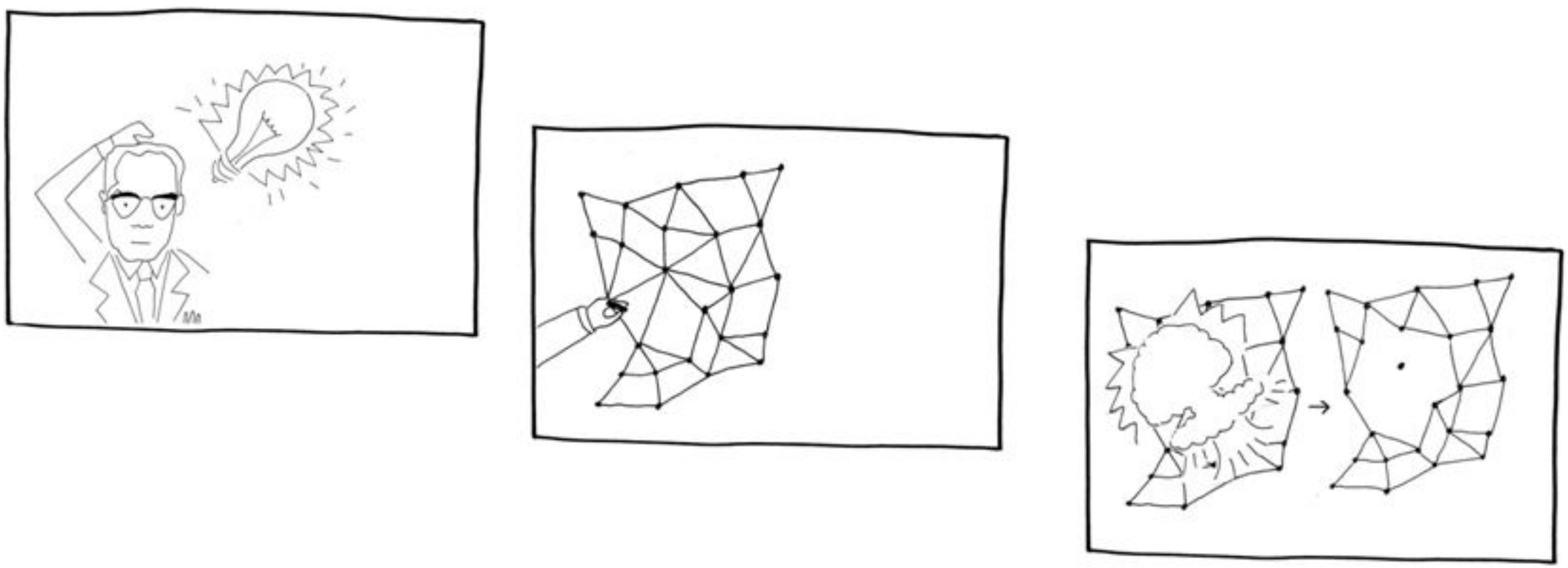
[Ugarte 2011]

A **decentralized** network was more robust: by eliminating one of the locally centralized nodes the network did not completely disappear even if some nodes were disconnected.



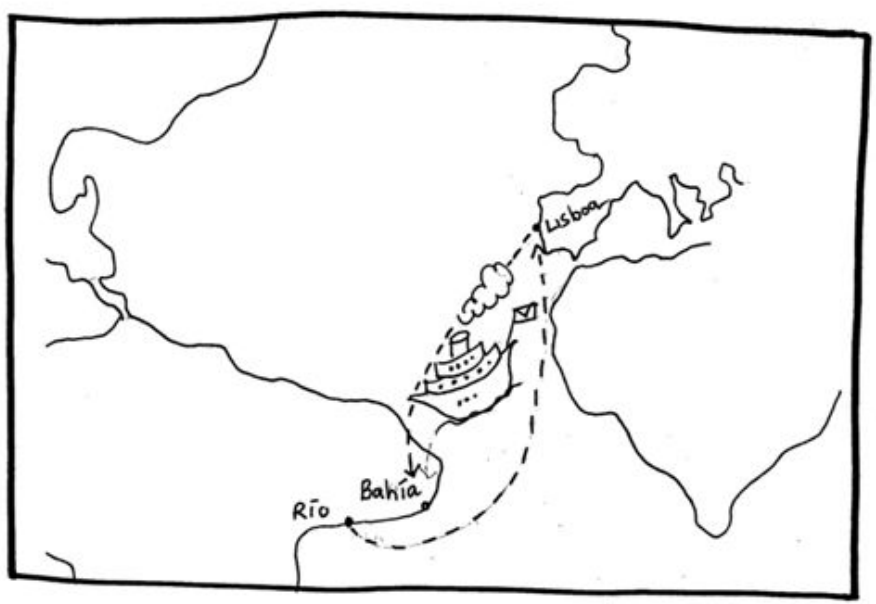
[Ugarte 2011]

Distributed networks to connect computers from major universities that had received defense research funds. That network (DARPANet) would later be known as the Internet



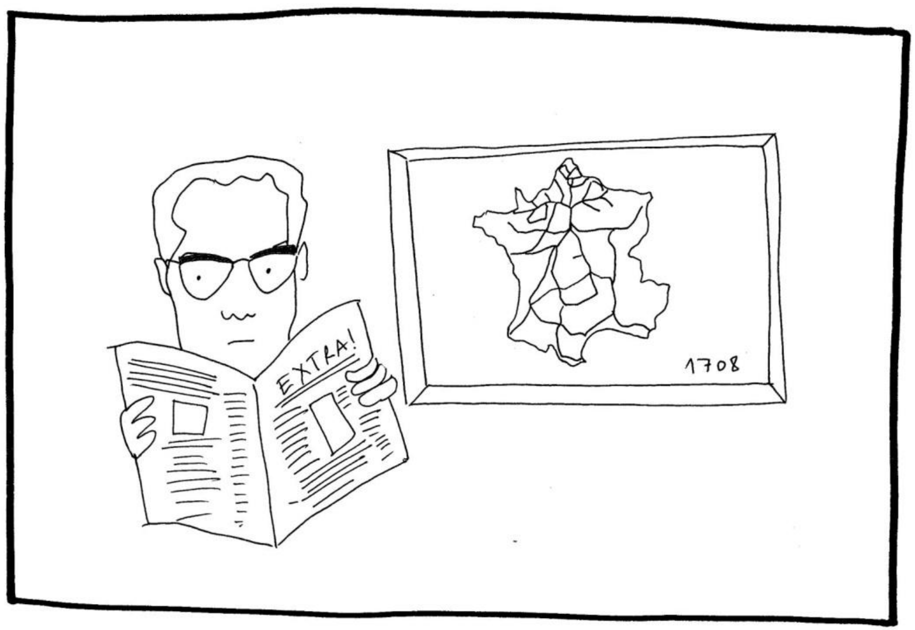
[Ugarte 2011]

The **centralized** network would correspond almost perfectly to the structure of the post networks that were the main system of communication three centuries ago. Every letter sent from one point to another necessarily had to pass through the capital, the central city where the power is located.



[Ugarte 2011]

The first newspapers were in the ports and their main news was precisely the arrivals and departures of ships.



[Ugarte 2011]

Newspapers with political news and opinion will be in the capitals, the center of the network.

Even today those who write outside the headquarters are called "correspondents".

@elaragon

[Ugarte 2011]

When the first form of modern political parties appeared during the French Revolution, clubs such as Robespierre's "Jacobin Club" would function exactly the same, with a centre in Paris.

[Ugarte 2011]

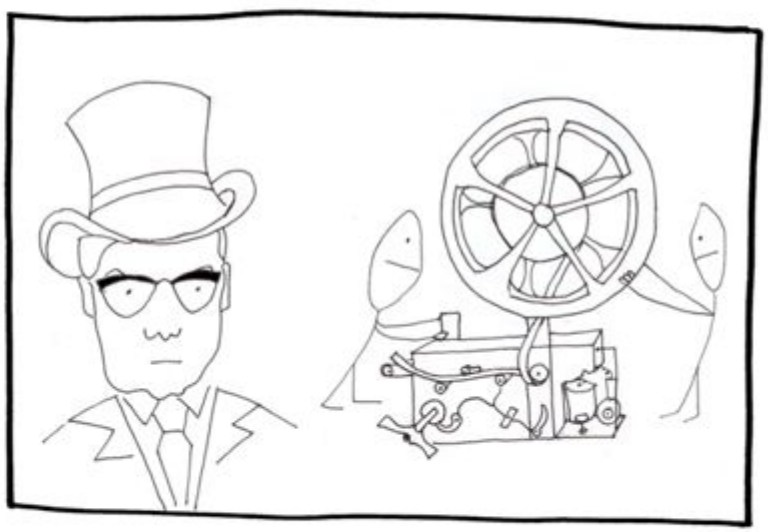
Centralized system of communications, the model of state to which the parties of the French revolution will give rise will be as centralist as that of the absolute monarchy.

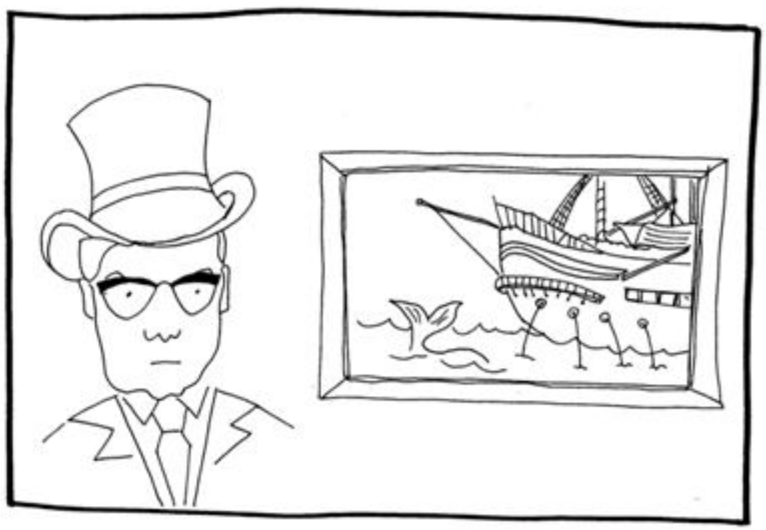
Even today extreme centralism is known as the **Jacobin state**.

[Ugarte 2011]

In 1844 Samuel Morse first sent out news via his **electric telegraph**.

In 1852, Paul Julius Reuter, connects for the first time Paris and London through a submarine cable.



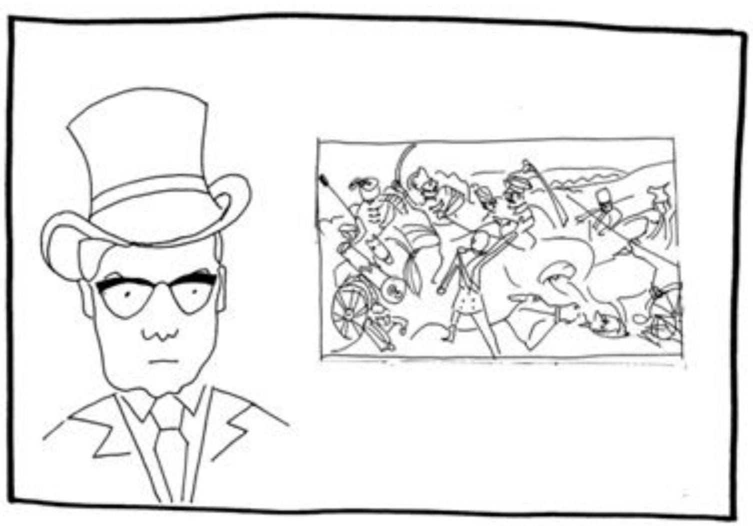


@elaragon

[Ugarte 2011]

Initially, for sending the quotations of the European stock exchanges. Then, news of all kinds, **Reuters** was born.

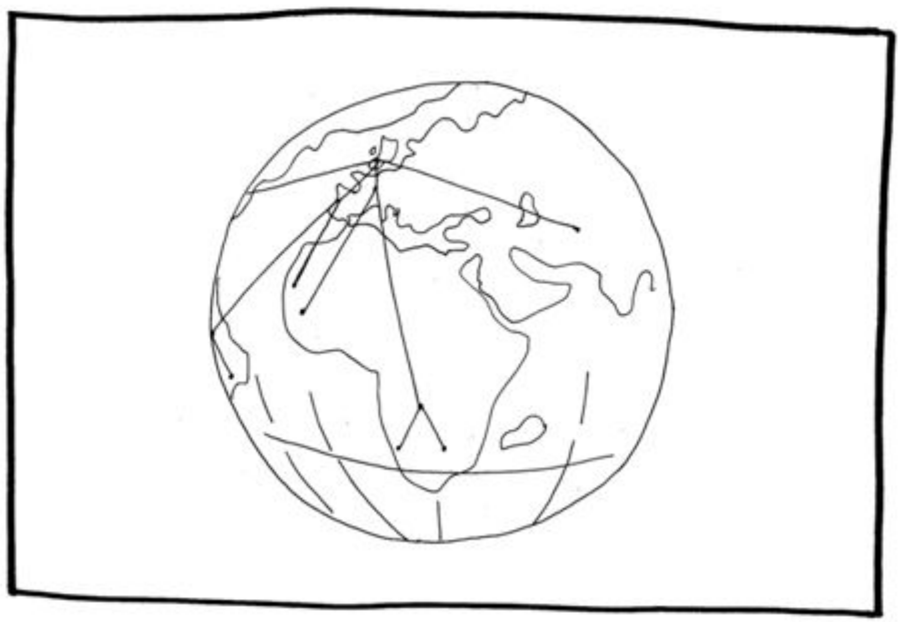
The information that allowed decisions to be taken on the state of peace and war became available to everyone.



[Ugarte 2011]

Public opinion now presses governments in a new manner.

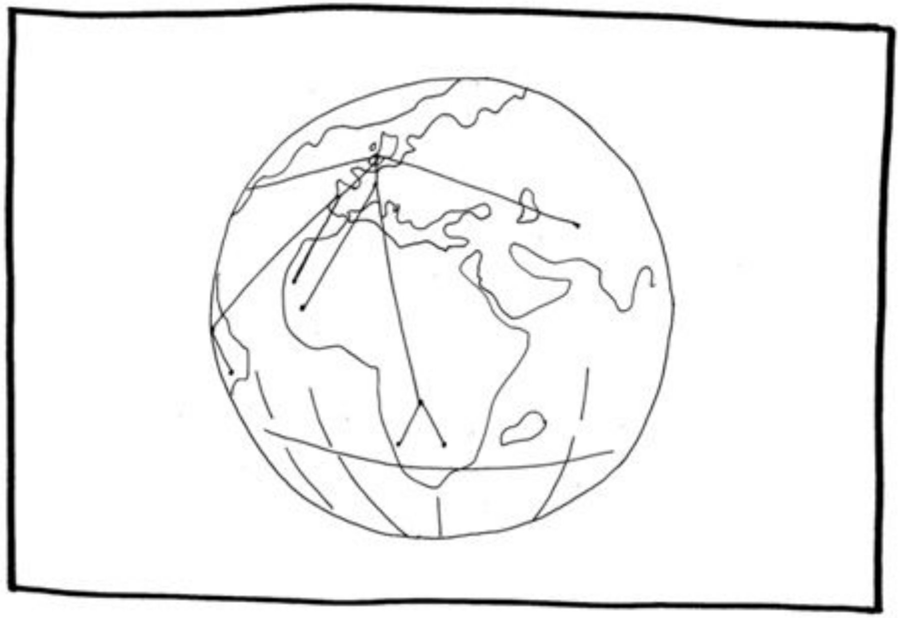
British businessmen realise that in the face of a strike they can divert production to their colleagues in France by contacting them by telegraph.



[Ugarte 2011]

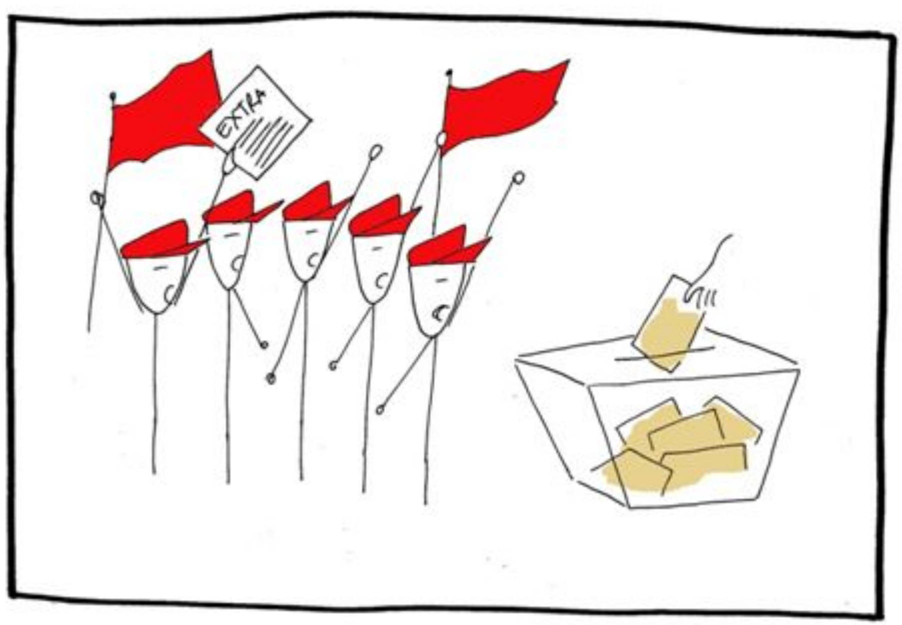
Workers responded by convening in London in 1863 all the workers' and radical movements of the continent with the aim of coordinating, in the same way, the workers of the whole continent by **telegraph**.

The **International Workingmen's Association** was born.



[Ugarte 2011]

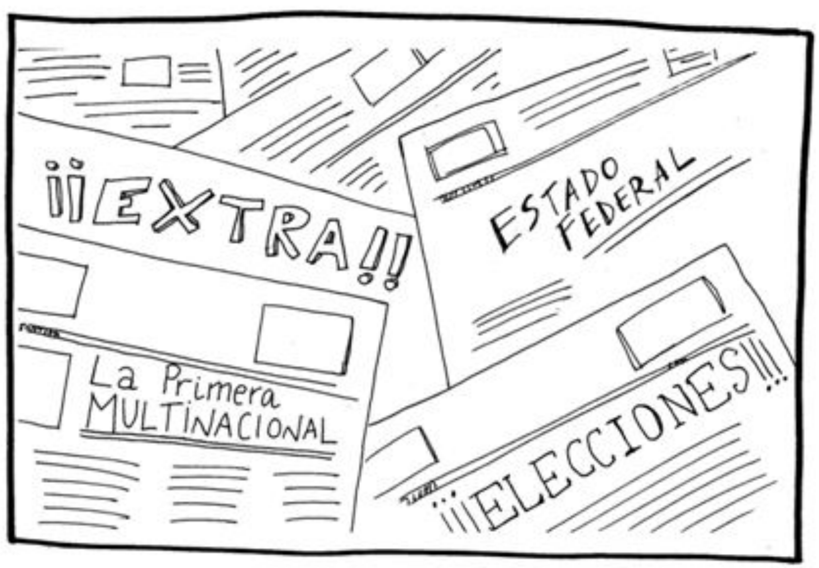
In 1889 the Second International was founded. Socialist parties follow the structure of the **decentralised** telegraph, organising themselves into territorial groups which in turn delegate to local nodes, these in turn being regional and these national nodes forming the International.



[Ugarte 2011]

Mass political participation, popular newspapers and universal suffrage The state is transformed, it becomes **decentralized** following the structure of the same communication network.

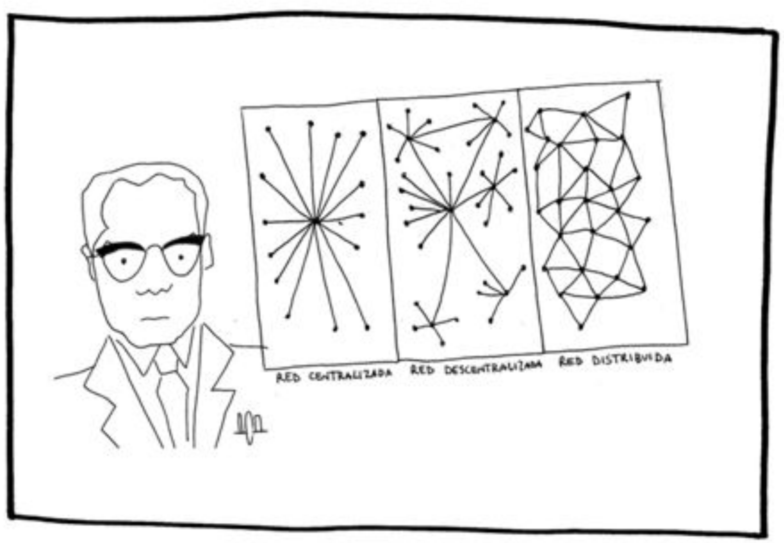
States are federalized and the companies are divided into regional divisions and delegations on the one hand and grouped as multinationals on the other.



[Ugarte 2011]

The transition from the post (centralized network) to the telegraph (decentralized network) meant the end of the power of a single agent: representative democracy, news agencies, multinationals and the federal state, etc.

What about the transition from decentralized networks to fully distributed networks like the Internet?



[Ugarte 2011]

What about the transition from decentralized networks to fully distributed networks like the Internet?

Conclusions

- Social network analysis metrics are based on graph theory.

- In addition to the graphs presented there are many others, for example:
  - Hypergraphs (edges with more than two vertices)
  - Multilayer graphs

- The history of networks helps to understand the rise of socio-political issues.

Structural metrics of networks

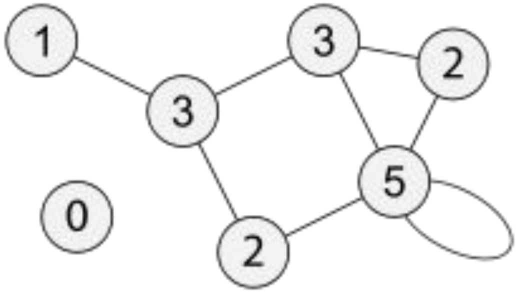
Degree

Definition

The number of edges connected to the node

Types

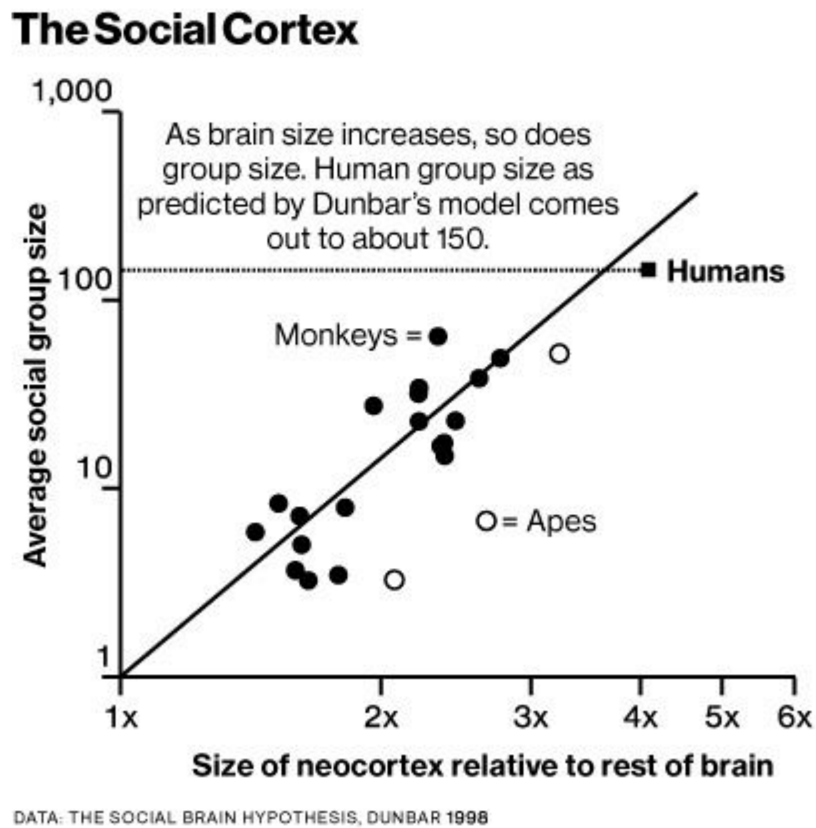
- Undirected networks: degree
- Directed networks: indegree / outdegree
- Non-weighted networks: degree
- Weighted networks: sum of the weights of each edge



Source: Wikipedia

Dunbar's number

- How many social contacts can we maintain?
- What are the cognitive limits to social interactions?
- One person is able to relate fully with a capacity of 150 people [Dunbar 1992]
- Studies indicate that conflicts often occur in organizations larger than this number



Dunbar's number

Examples

- Neolithic villages 6500 BC 150-200
- Modern armies (company) 180
- Hutterite communities 107
- Nebraska Amish parishes 113
- Business organisation <200
- Ideal church congregations <200
- Domesday Book villages (1087 AD) 150
- C18th English villages 160
- Research sub-disciplines 100-200
- Small world experiments 134
- Hunter-Gatherer communities 148
- Xmas card networks 154

Dunbar's number

Examples

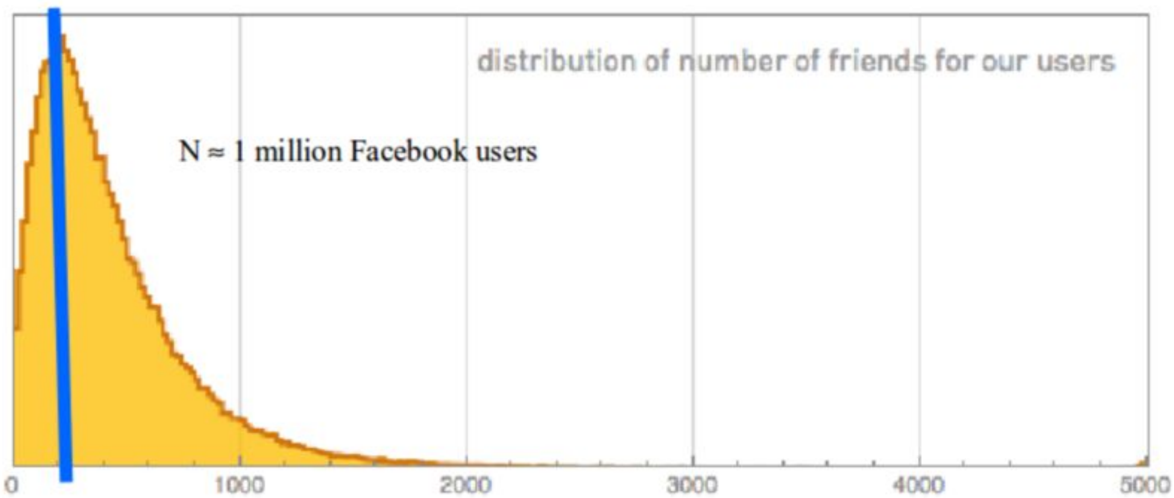
- The Swedish tax offices were reorganised in 2007 into units of 150 persons.

- Each Gore-Tex factory is also limited to 150 people.

Dunbar’s number

Examples in social media

- The number of "friends" on Facebook ≈ 150-250
- One might have hundreds of friends, while only talking to a few



- Similar results on Twitter [Gonçalves 2011]

@elaragon

Dunbar’s circles

Hierarchical Structures

- Social relationships form an inclusive, hierarchical series of circles: **when size increases, intensity decreases**

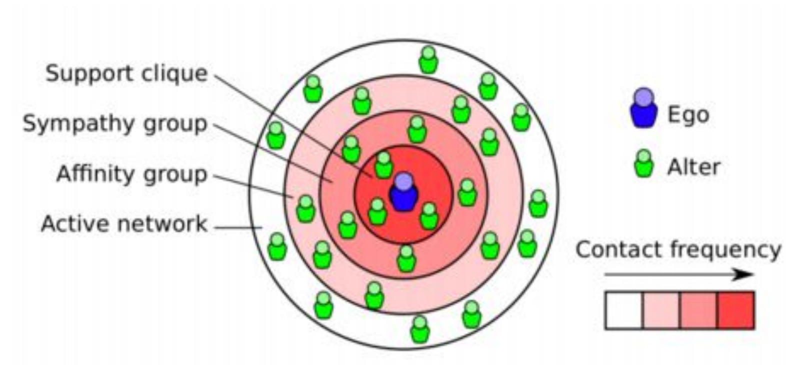
- Ratio ≈ 3 [Zhou 2005]

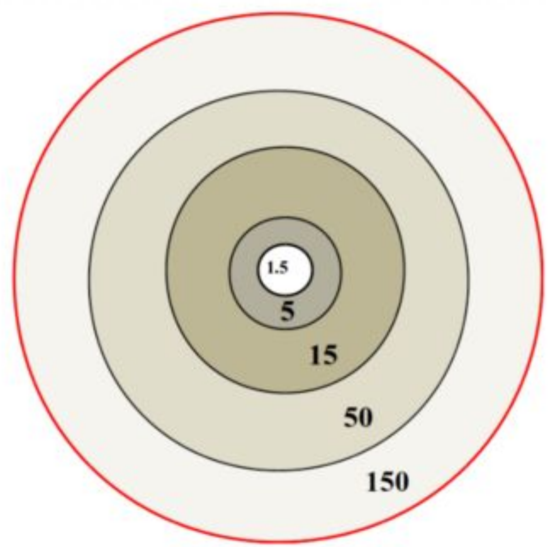
Examples

- Military groups [Dunbar 2011]

- Facebook [Arnaboldi 2012]

- Twitter [Arnaboldi 2013]

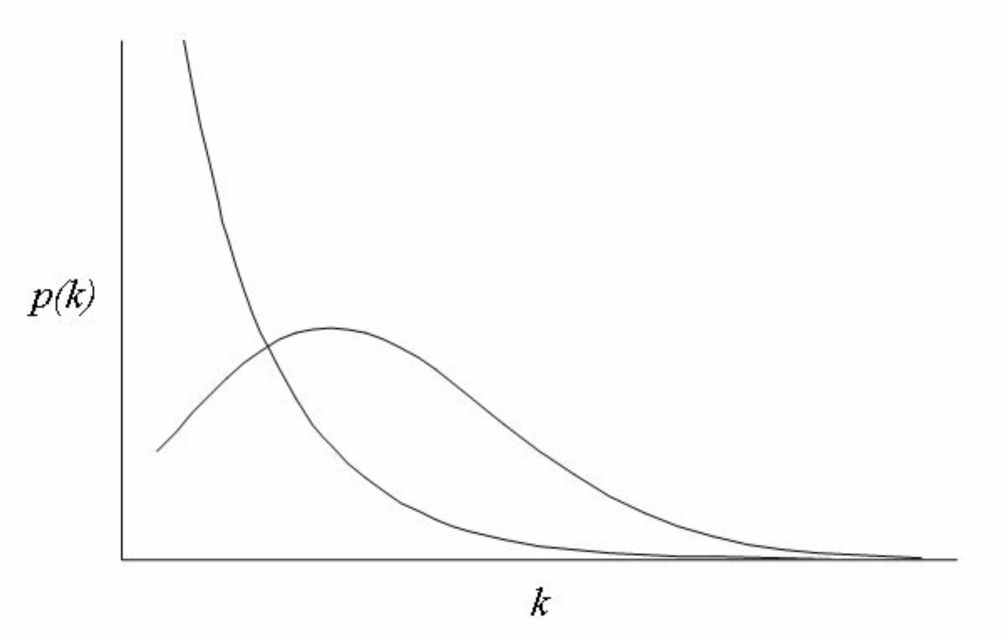




Degree distribution

Definition

- Density of nodes in the network with a certain degree k
- The degree of a node in a network is the number of connections of a vertex or node to other nodes



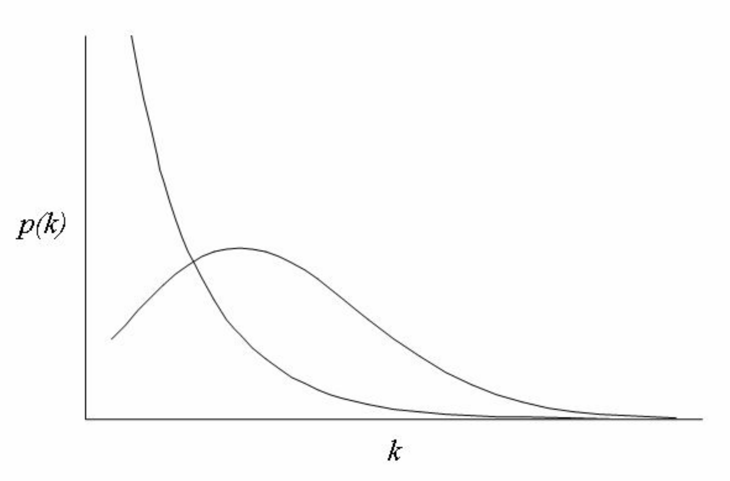
Source: Wikipedia

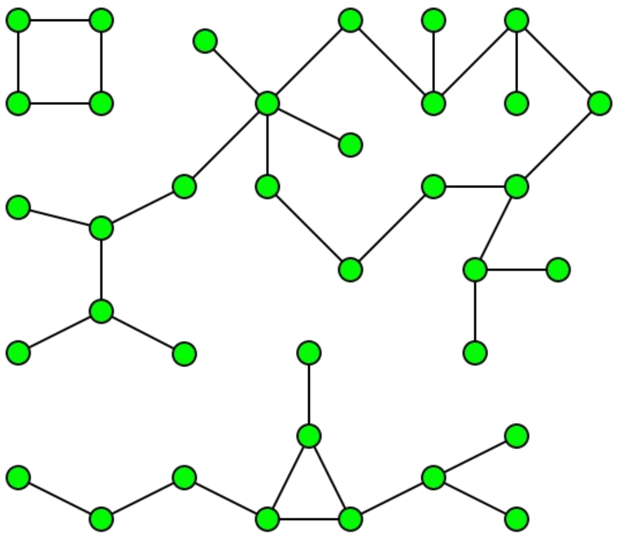
Degree distribution

Random networks

A network is generated by some kind of random process, e.g., [Erdös 1959]

- Nodes are randomly connected
- Undirected network





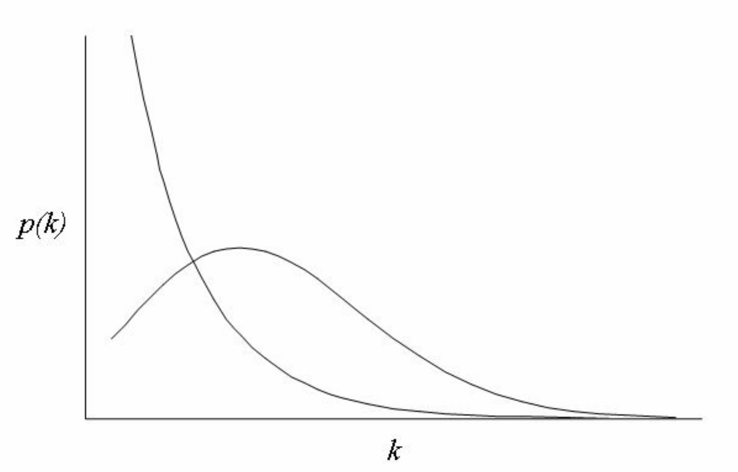
Source: Wikipedia

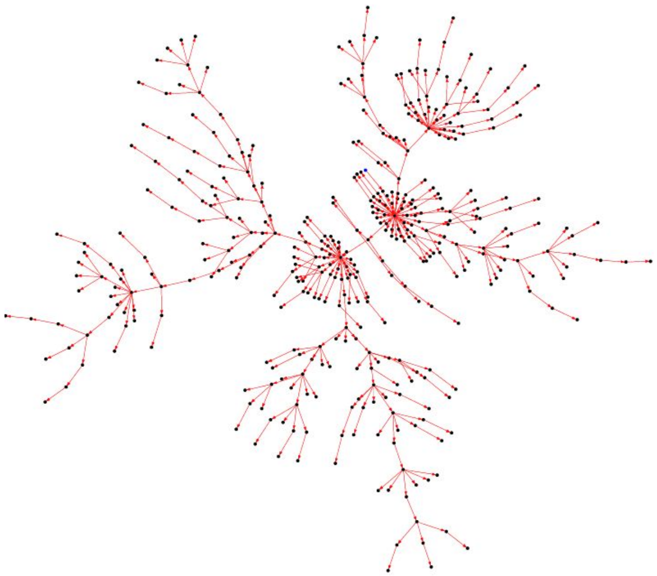
Degree distribution

Scale-free networks

A small group of nodes with a high degree and a large tail of nodes with a low degree [Barabási 1999]

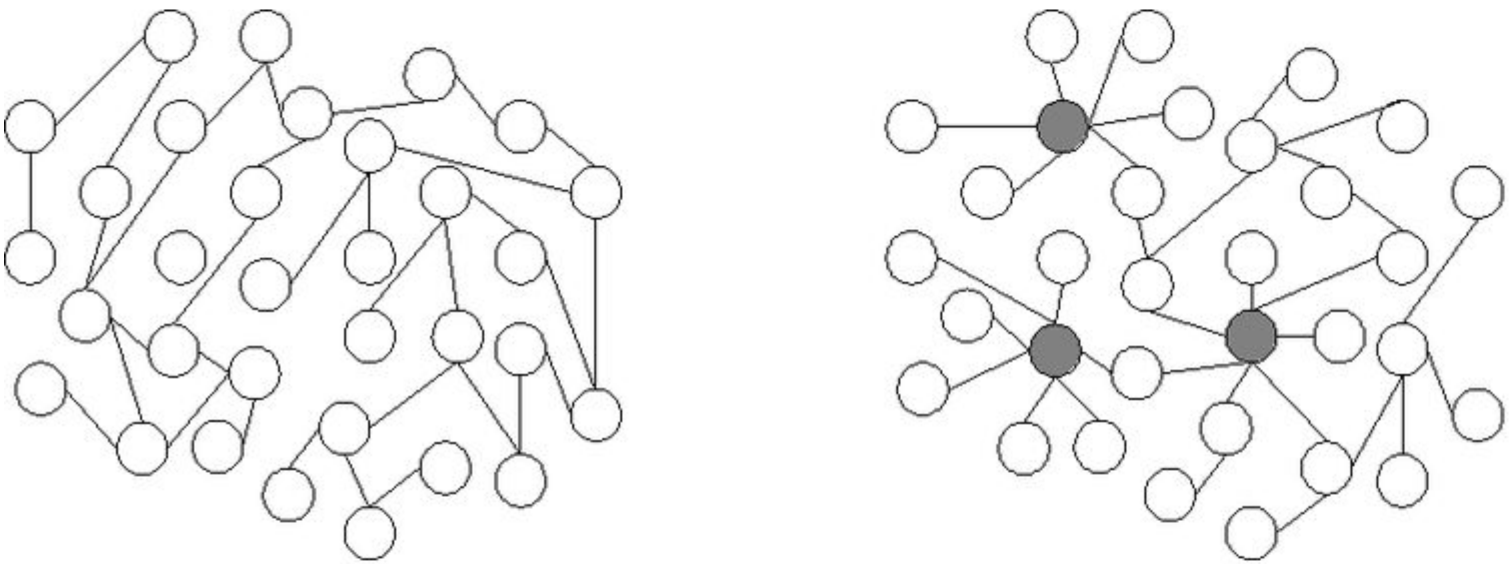
- Telephone call networks
- International trade networks
- Airport Routes





Source: convetry.ac.uk

Degree distribution



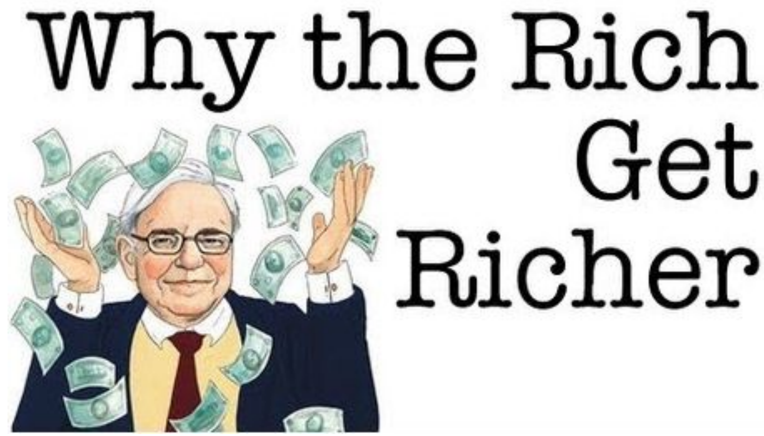
(a) Random network (b) Scale-free network

Source: Wikipedia

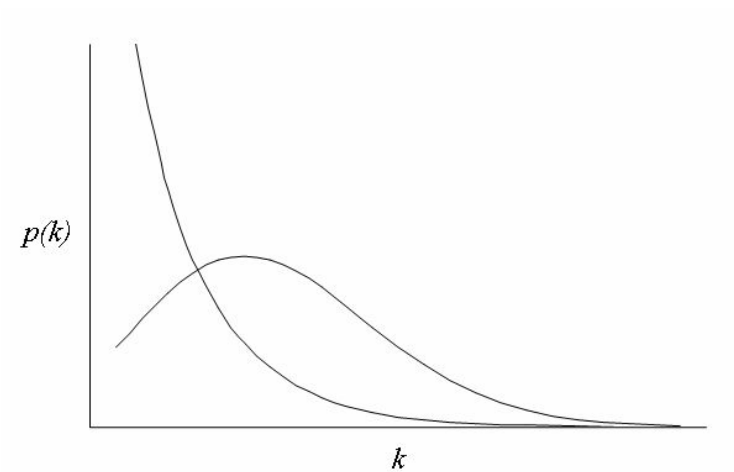
Preferential attachment

Fundamental model to explain why the degree distribution responds to a power law (long tail):

- Links to web pages
- Citations to scientific articles



Source: YouTube



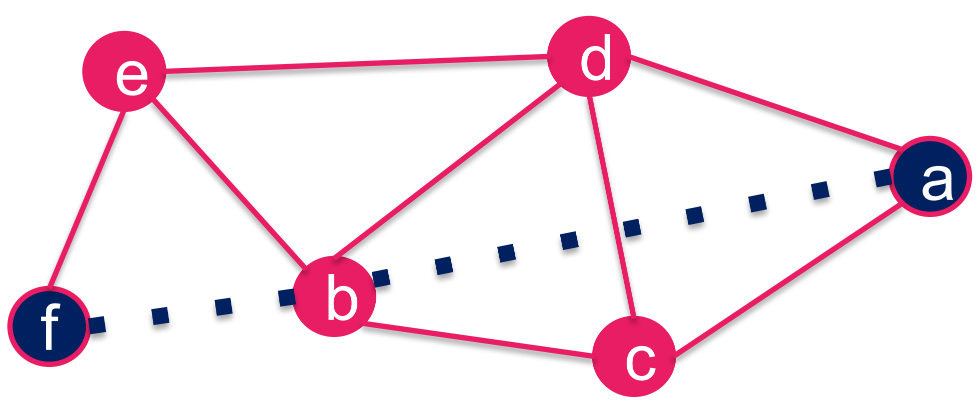
Distance

Definition

- The shortest path between two nodes the path that joins them with the least number of edges
- This number of edges is called the distance between nodes

Application

- Short paths are important since they require a lower cost in diffusion dynamics, propagation, etc.



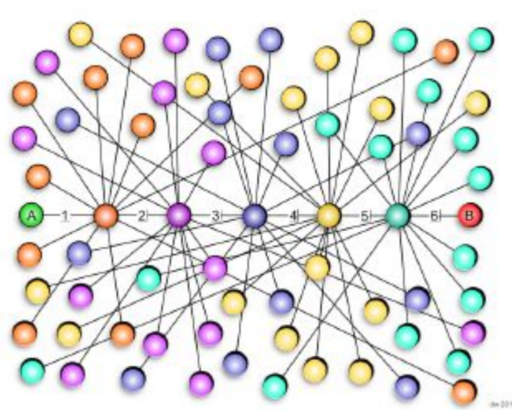
Average path length

Definition

Average of possible distances between all possible nodes

Six degrees of separation

- Milgram's Small-world experiment [Milgram 1967]
- Average distance = 6



Source: Wikipedia

Average path length

Definition

Average of possible distances between all possible nodes

Distances are usually short

- d = 6.6 MSN Messenger (N = 220M users) [Leskovec 2008]
- d = 3.48 movie actors (N = 450 000)
- d = 6.19 co-authors of articles in Physics (N = 53 000)
- d = 4.95 e-mails (N = 60 000) [Newman 2003a]

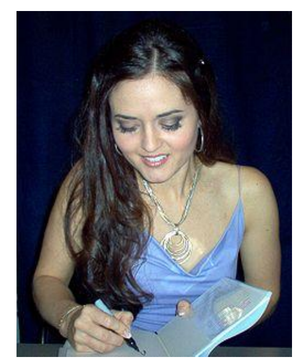
Six degrees of separation

Six degrees of Kevin Bacon

Only 17 of the 700K actors in IMDB's network are at distance of 8 to Kevin Bacon http://oracleofbacon.org/ [Ruthven 1994]

Other domains

- Mathematics: Paul Erdös
- Physics: Albert Einstein
- Linguistics: Noam Chomsky
- Economics: Joseph Stiglitz
- Math+Acting: Paul Erdös + Kevin Bacon



Source: Wikipedia

Diameter

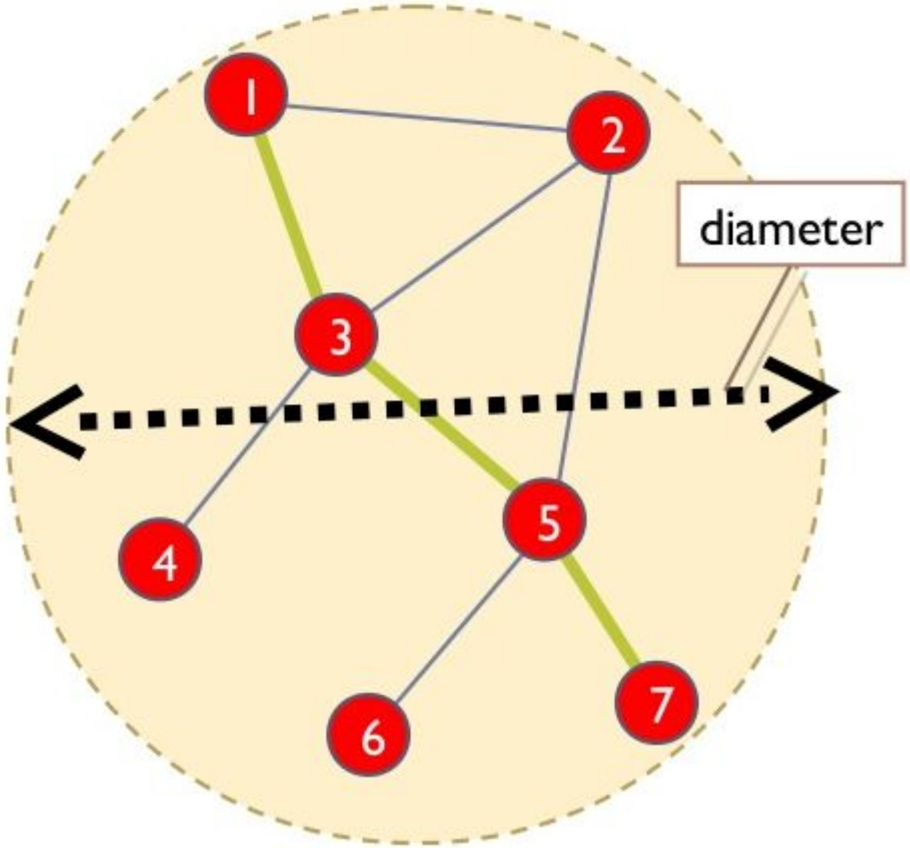
Definition

- Maximum distance between all possible user pairs
- Effective diameter (90% percentile of distances)

Paradox

- The diameter decreases when the network grows [Leskovec 2007]
- The more users, the closer they are.
- More users means more relationships

Diameter



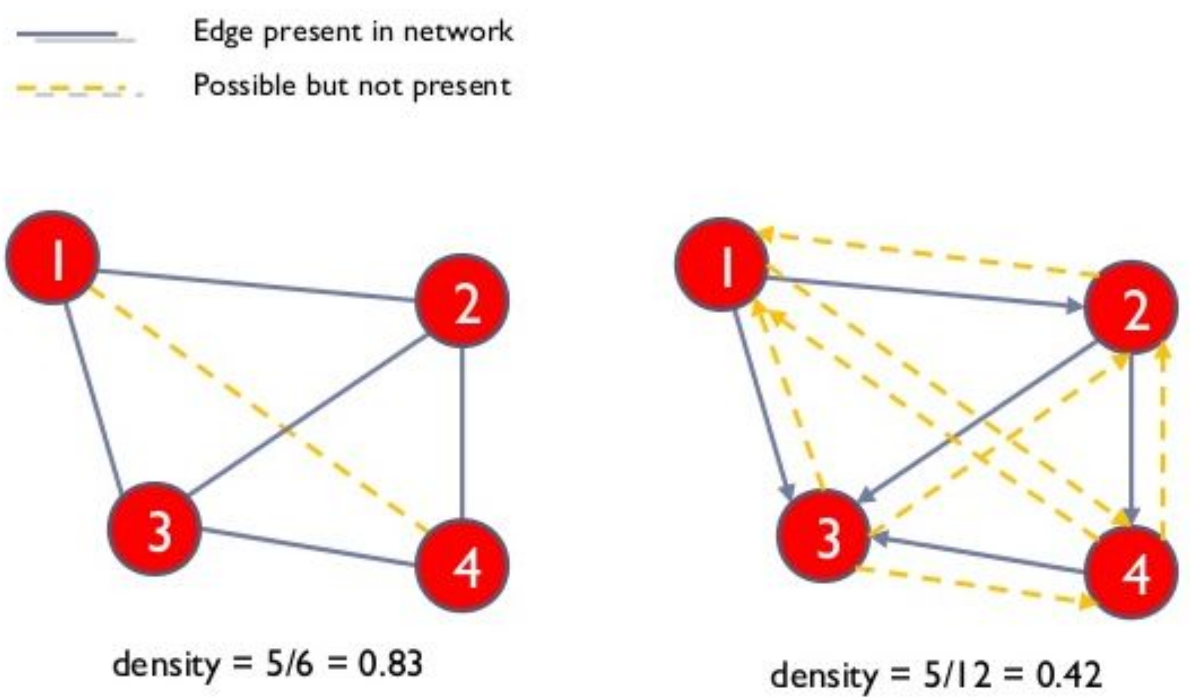
Fuente: slideshare.net/gcheliotis/

Density

Definition

- Density is used to compare networks of similar size. Ratio between the number of network edges over the total number of possible edges between all pairs of nodes ($n \cdot (n-1)/2$, where n is the number of vertices, for an undirected network)
- Common measure of how well connected a network is
- A network with density=1 is called a clique
- A directed network will have half the density of its undirected equivalent, because there are twice as many edges possible, i.e. $n \cdot (n-1)$

Density



Fuente: slideshare.net/gcheliotis/

Density

Properties

- Social networks tend to be sparse (non dense).
- Density depends on the size of the network.
- The more users, the lower the density.
- Density should only be used to compare networks of similar size.

Density

Table 1.1 Basic statistics of network examples. Network types can be (D)irected and/or (W)eighted. When there is no label the network is undirected and unweighted. For directed networks, we provide the average in-degree (which coincides with the average out-degree).

Network	Type	Nodes (N)	Links (L)	Density (d)	Average degree ($\langle k \rangle$)
Facebook Northwestern Univ.		10,567	488,337	0.009	92.4
IMDB movies and stars		563,443	921,160	0.000006	3.3
IMDB co-stars	W	252,999	1,015,187	0.00003	8.0
Twitter US politics	DW	18,470	48,365	0.0001	2.6
Enron Email	DW	87,273	321,918	0.00004	3.7
Wikipedia math	D	15,220	194,103	0.0008	12.8
Internet routers		190,914	607,610	0.00003	6.4
US air transportation		546	2,781	0.02	10.2
World air transportation		3,179	18,617	0.004	11.7
Yeast protein interactions		1,870	2,277	0.001	2.4
C. elegans brain	DW	297	2,345	0.03	7.9
Everglades ecological food web	DW	69	916	0.2	13.3

Comparison of networks

Jaccard index

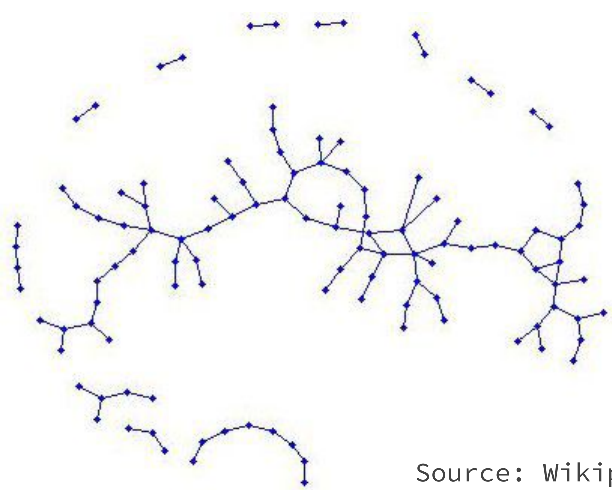
- Number of times two nodes have the same neighbours as a percentage of the total number of links.
- This is relevant because two networks may have similar topology, but differ in the connections.

$$score(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

Giant component

Definition

- Largest connected component of the network.
- In a connected component there is a path between any pair of nodes belonging to this component, i.e., there are no isolated users in any component.

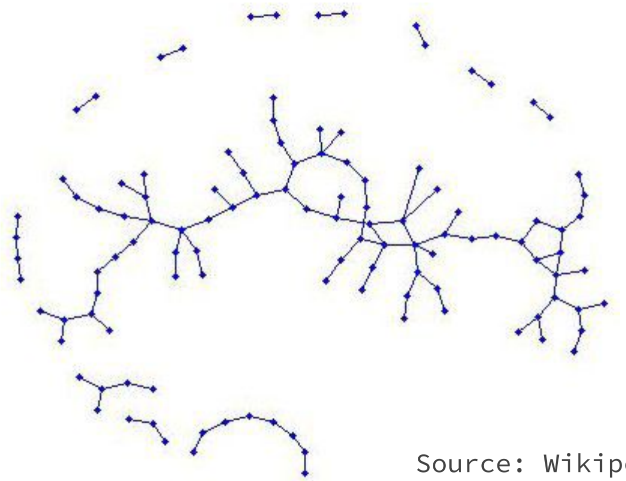


Source: Wikipedia

Giant component

Properties in social networks

- Usually very big, the vast majority of the nodes belong to it
- There are only very small components apart from this one
- If an isolated group starts to have around 10 members it eventually joins the giant component.



Source: Wikipedia

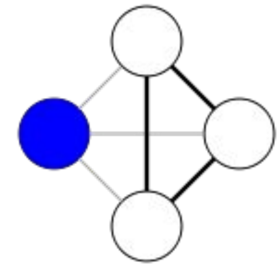
Clustering coefficient

Definition

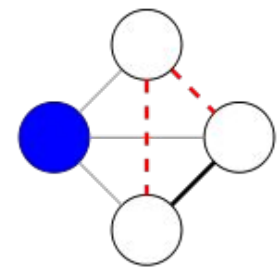
- Density of connections between a node's direct neighbours.
- C is the average of local coefficients C_i
- C_i was calculated for each node

$$C_i = |E_i| / (k_i \cdot (k_i - 1))$$

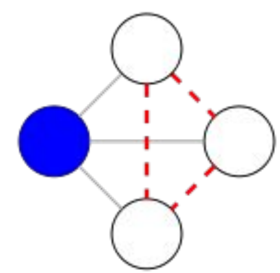
- k_i = total degree of the node
- E_i = set of edges between the direct neighbours of the node i



$c = 1$



$c = 1/3$



$c = 0$

Source: Wikipedia

Clustering coefficient

- Some networks, e.g., social networks, tend to have high clustering coefficients because of **triadic closure**: we meet through common friends

- Other networks, e.g., bipartite and tree-like networks, have low clustering coefficient

Table 2.1 Average path length and clustering coefficient of various network examples. The networks are the same as in Table 1.1, their numbers of nodes and links are listed as well. Link weights are ignored. The average path length is measured only on the giant component; for directed networks we consider directed paths in the giant strongly connected component. To measure the clustering coefficient in directed networks, we ignore link directions.

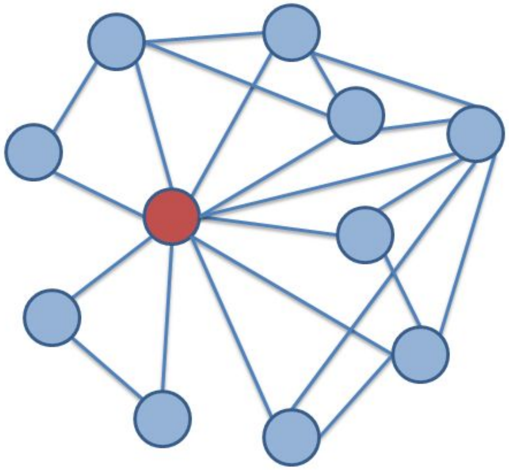
Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron Email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

Ego-network

Definition

Network of neighbours of a node:

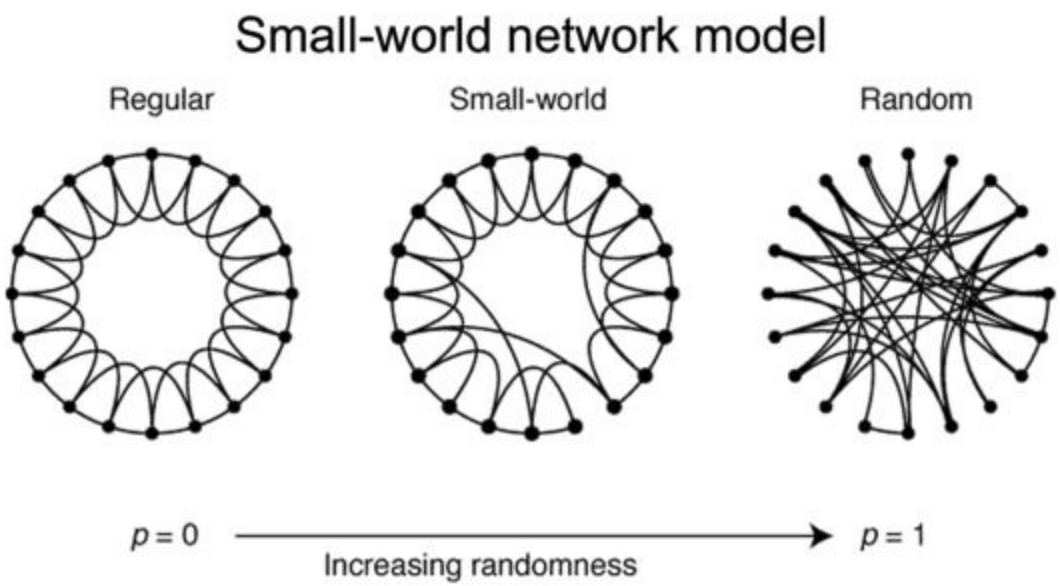
- All nodes are linked to the main node node
- Edges between neighbours are also included



Small-world networks

Properties [Watts 1998]

- Low average distance
- High clustering coefficient

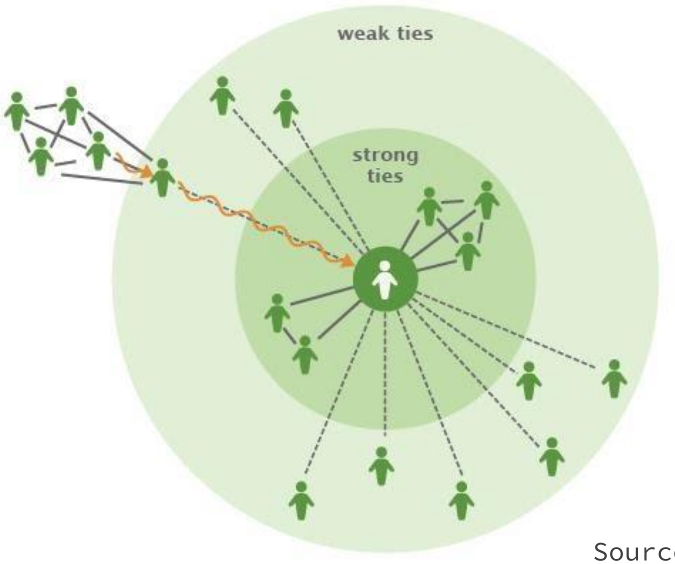


Weak ties

Properties [Granovetter 1973]

- The dynamics of dissemination and coordination are influenced by links established with nodes belonging to other clusters.

- This phenomenon has become very relevant with the rise of online social networks networks.

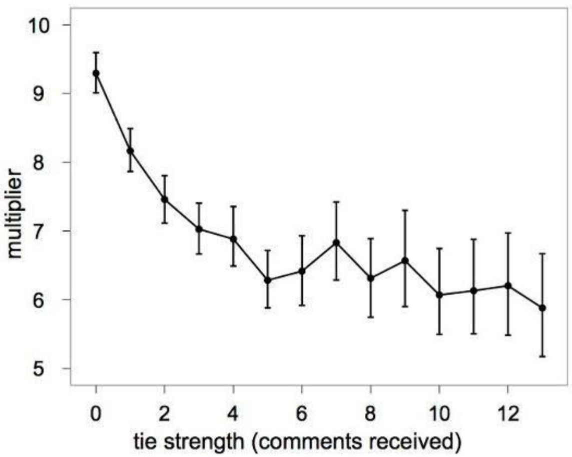


Source: Facebook

Weak ties

Properties [Bakshy 2012]

- Weak links have the greatest potential to expose a link to friend that you would not otherwise have discovered it



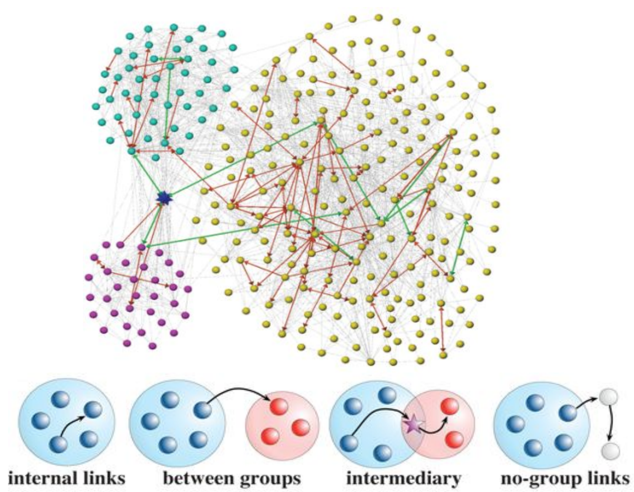
Source: Facebook

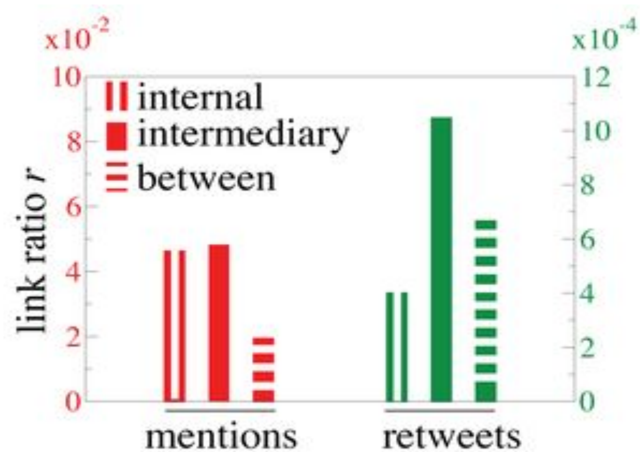
@elaragon

Weak ties

Properties [Grabowicz 2012]

- Personal interactions are more likely to occur in internal links to groups (strong links)
- The propagation of events or new information is spread by intermediate links (weak links)





Reciprocity

Different approaches

- Standard Reciprocity

$$r = \frac{E^{\leftrightarrow}}{E}$$

- Corrected reciprocity with density [Garlaschelli 2004]

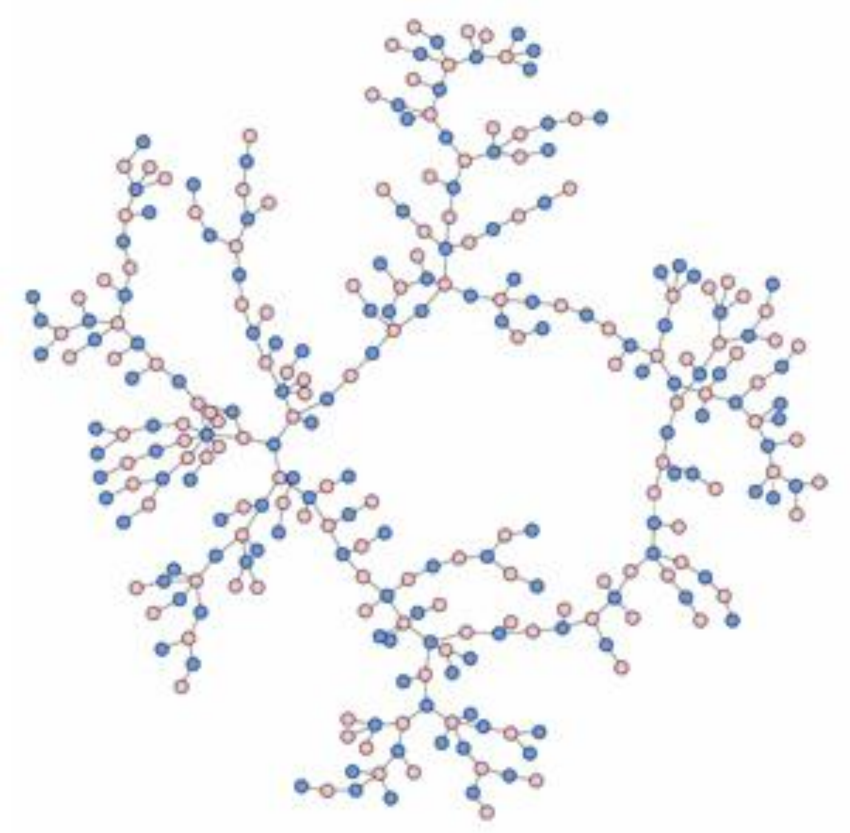
$$\rho = \frac{r - \bar{a}}{1 - \bar{a}}$$

- Reciprocity in weighted networks [Squartini 2013]

$$r_w = \frac{W^{\leftrightarrow}}{W} = \frac{\sum_u \sum_{v \neq u} w_{uv}^{\leftrightarrow}}{\sum_u \sum_{v \neq u} w_{uv}}$$

Reciprocity

Structure of adolescent romantic and sexual networks



Assortativity

Definition

- Preference for relationships between nodes with the same or different attributes:
  - Degree
  - Sex
  - Age
  - Race
  - Weight
  - Mother tongue

Assortativity

Interpretation

- r > 0: assortative mixing
  - There is a preference for connections between similar nodes.
- r = 0: neutral mixing
  - No relationship preference.
- r < 0: disassortative mixing
  - There is a preference for connections between different nodes.

Assortativity

Examples

- movie actors: r = 0,208
- co-authors of physical articles: r = 0,36
- e-mails (address books): r = 0,092

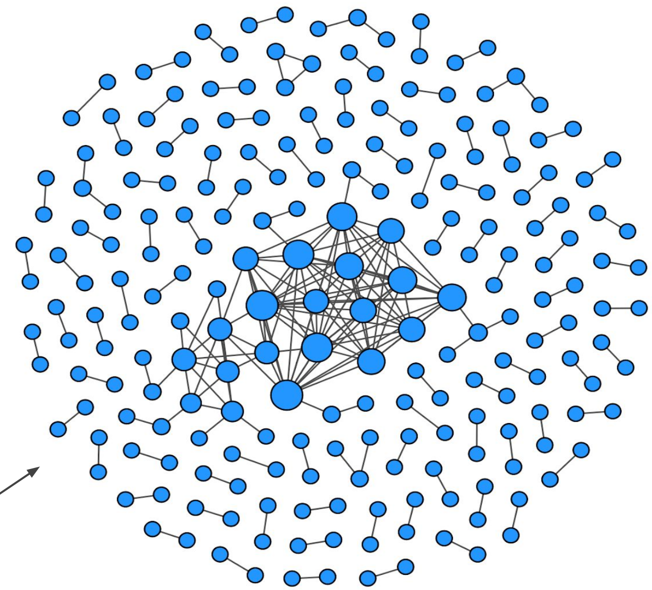
Results

- By degree (number of connections):
  - Many social networks are assortative: the most active nodes connect more with each other
    - In Wikipedia, it's the other way around (disruptive network): the more active ones interact mostly with the less expert ones [Laniado, 2011]
- By gender:
  - In Wikipedia, women talk more to each other [Laniado, 2012]
  - In Tuenti, there are no significant preferences

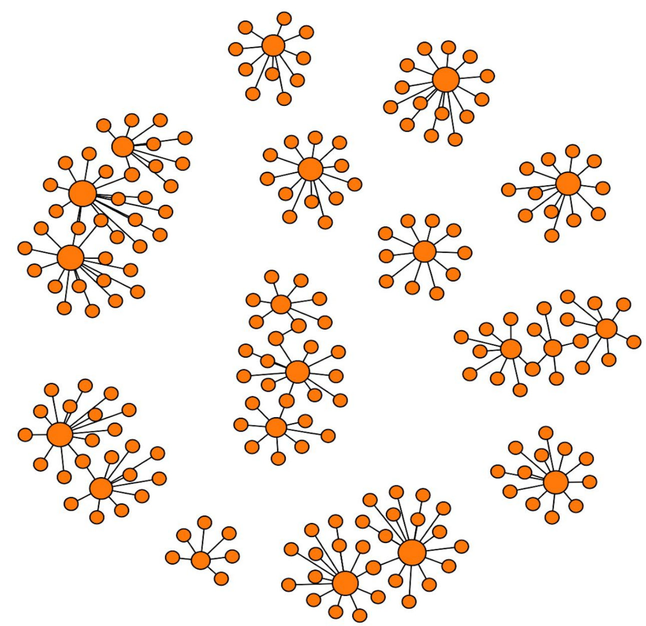
Degree assortativity

- A.k.a. degree correlation

- Assortative networks have a **core-periphery** structure with hubs in the core
    - Ex: social networks



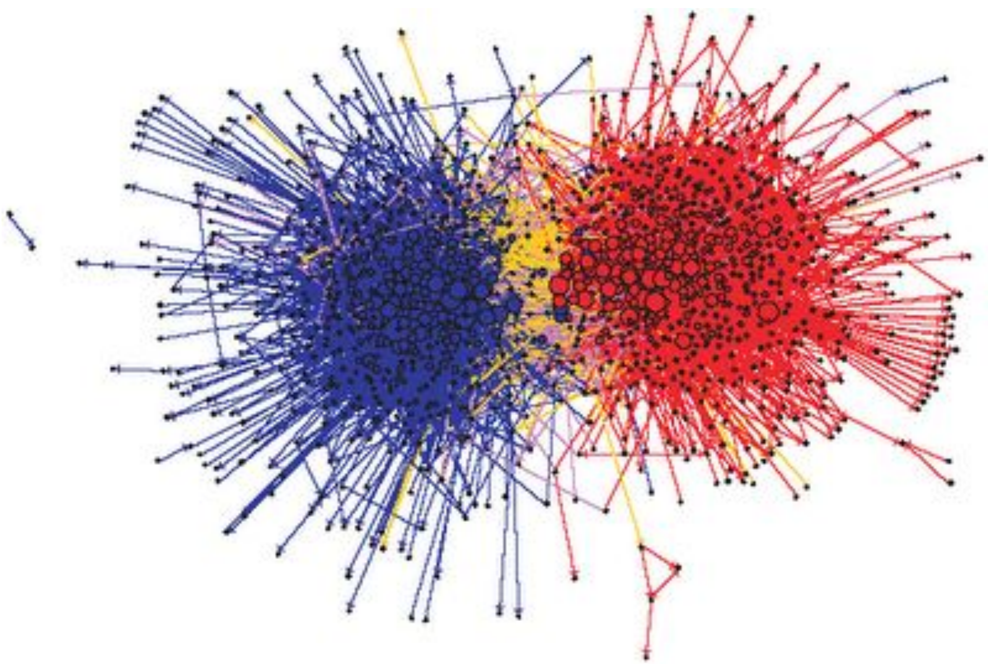
- Disassortative networks have **hub-and-spoke** (or star) structure
    - Ex: Web, Internet, food webs, bio networks



Political assortativity

Example [Adamic 2005]

Political blogs in USA

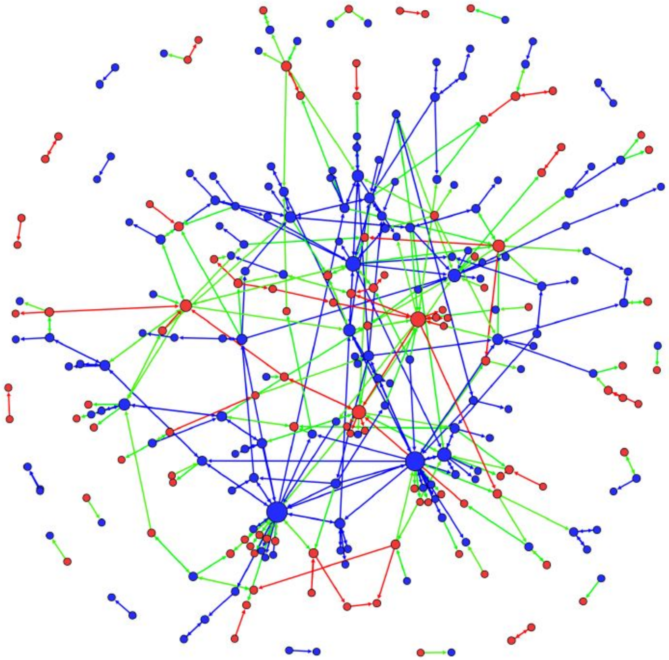


- Republicans
- Democrats
- Mixed

Political assortativity

Example [Neff 2013]

Political Wikipedia in USA



- Republicans
- Democrats
- Mixed

Network temporality

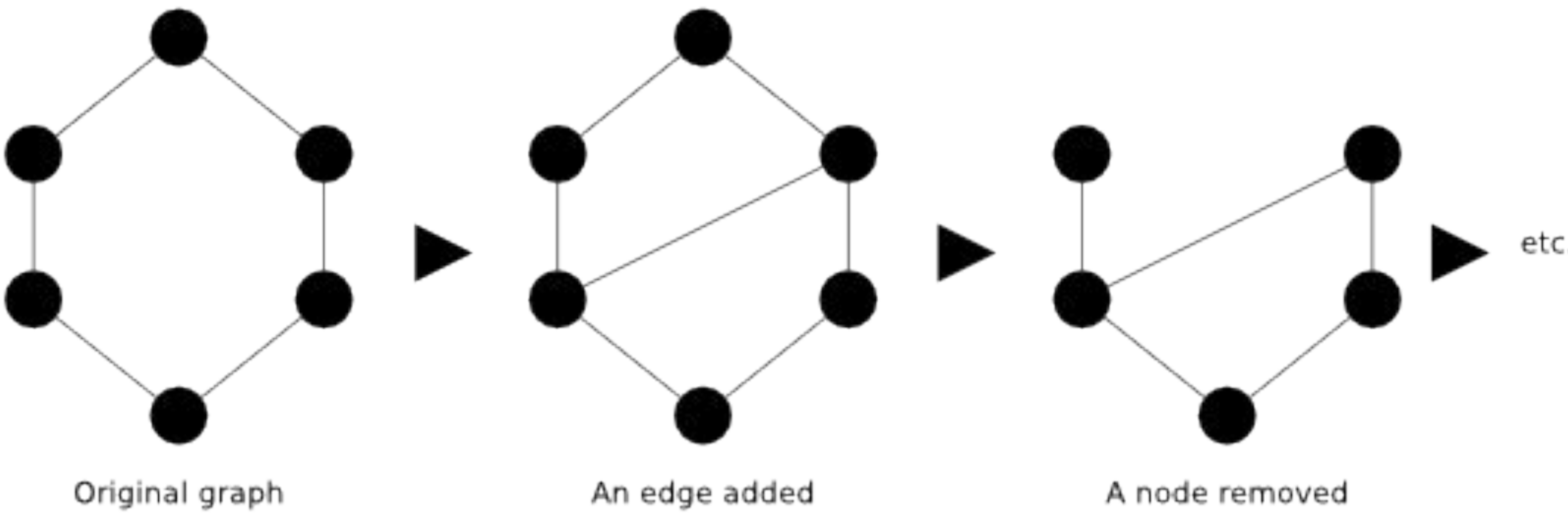
Definition

A dynamic network is a G(t) network that changes state according to a time unit t

Properties

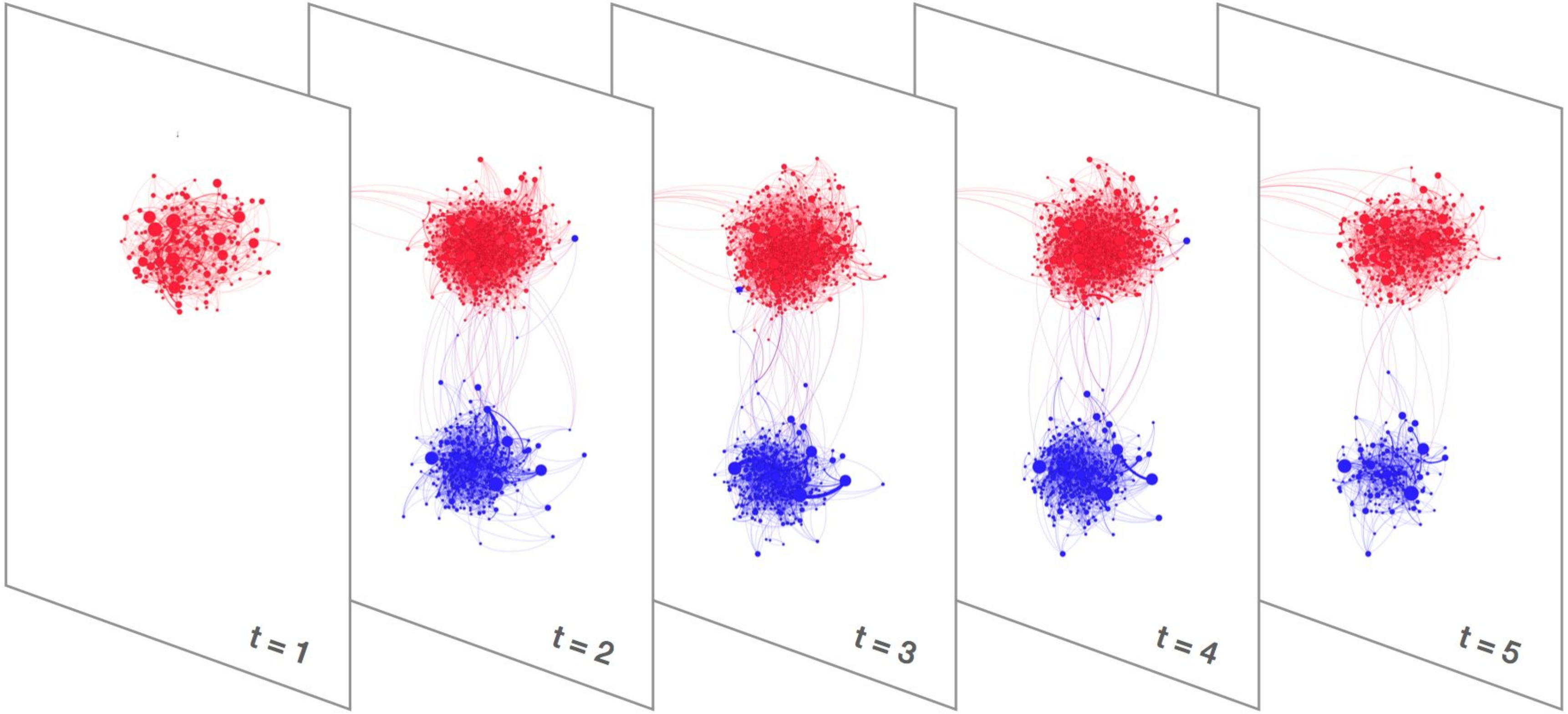
- Social networks typically change over time
- Transitions can represent different phenomena:
  - Evolution
  - Growth
  - Transformation
  - Decay

Network temporality



Source: Wikipedia

Network temporality



Conclusions

- Social relationships can be modeled through networks

- Structural metrics allow characterization in different types of networks:
  - Small world
  - Erdös - Renyi
  - Scale-free
  - and many more!

- Network visualization makes it easier to observe the structural metrics of a network

References

- Adamic, L. A., & Glance, N. (2005, August). The political blogosphere and the 2004 US election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery (pp. 36-43). ACM.
- Arnaboldi, V., Conti, M., Passarella, A., & Dunbar, R. (2013, October). Dynamics of personal social relationships in online social networks: a study on twitter. In Proceedings of the first ACM conference on Online social networks (pp. 15-26). ACM.
- Arnaboldi, V., Conti, M., Passarella, A., & Pezzoni, F. (2012, September). Analysis of ego network structure in online social networks. In 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing (pp. 31-40). IEEE.
- Bakshy, E. (2012). Rethinking information diversity in networks. Facebook Data Science Notes.

References

- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. Wired magazine, 16(7), 16-07.
- Barabási, Albert-László; Albert, Réka. (October 15, 1999). "Emergence of scaling in random networks". Science 286 (5439): 509–512. arXiv:cond-mat/9910332. Bibcode 1999Sci...286..509B. doi:10.1126/science.286.5439.509. MR 2091634. PMID 10521342.
- Dunbar, R.I.M. (1992). "Neocortex size as a constraint on group size in primates". Journal of Human Evolution 20: 469–493. Erdős, P.; Rényi, A. (1959). "On Random Graphs. I" (PDF). Publicationes Mathematicae. 6: 290–297.
- Garlaschelli, D., & Loffredo, M. I. (2004). Patterns of link reciprocity in directed networks. Physical review letters, 93(26), 268701.

References

- Gonçalves, B., Perra, N., & Vespignani, A. (2011). Modeling users' activity on twitter networks: Validation of dunbar's number. PloS one, 6(8), e22656.
- Granovetter, M. S. (1973). The strength of weak ties. In Social networks (pp. 347-367). Academic Press.
- Heider, F. (1946). Attitudes and cognitive organization. The Journal of psychology, 21(1), 107-112.
- Laniado, D., Castillo, C., Kaltenbrunner, A., and Fuster-Morell, M. (2012). Emotions and dialogue in a peer-production community: the case of Wikipedia. In Proceedings of WikiSym'12.
- Laniado, D., Tasso, R., Volkovich, Y., & Kaltenbrunner, A. (2011). When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages. Proceedings of ICWSM.

References

- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... & Jebara, T. (2009). Computational social science. Science, 323(5915), 721-723. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2745217/
- Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), 2.
- Leskovec, J., & Horvitz, E. (2008, April). Planetary-scale views on a large instant-messaging network. In Proceedings of the 17th international conference on World Wide Web (pp. 915-924). ACM.
- Newman, M. E. (2003). Mixing patterns in networks. Physical Review E, 67(2), 026126.
- Milgram, S. (1967). The small world problem. Psychology today, 2(1), 60-67.

References

- Neff, J. J., Laniado, D., Kappler, K. E., Volkovich, Y., Aragón, P., & Kaltenbrunner, A. (2013). Jointly they edit: Examining the impact of community identification on political interaction in wikipedia. PloS one, 8(4), e60584.
- Grabowicz, P. A., Ramasco, J. J., Moro, E., Pujol, J. M., & Eguiluz, V. M. (2012). Social features of online networks: The strength of intermediary ties in online social media. PloS one, 7(1), e29358.
- Ruthven. A. Kevin Bacon is the Center of the Universe rec.arts.movies. Google groups. Retrieved 2009-07-19.
- Squartini, T., Picciolo, F., Ruzzenenti, F., & Garlaschelli, D. (2013). Reciprocity of weighted networks. Scientific reports, 3, 2729.

References

- Ugarte, D. D. (2011). El poder de las redes. Manual ilustrado para ciberactivistas. Colección Biblioteca de las Indias.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’networks. nature, 393(6684), 440.
- Zhou, W. X., Sornette, D., Hill, R. A., & Dunbar, R. I. (2005). Discrete hierarchical organization of social group sizes. Proceedings of the Royal Society B: Biological Sciences, 272(1561), 439-444.

Homework

Watch the following documentary:

Jacques, A. (2008)

Connected: The Power of Six Degrees
How Kevin Bacon cured cancer

London: BBC Four

http://y2u.be/2rzxAyY7D7k