SSAAD 2023

Université de Lille

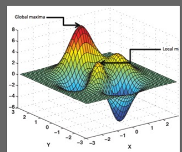
SEARCH LANDSCAPE ANALYSIS

PART 2 – INTRODUCTION TO COMPLEX NETWORKS

Gabriela Ochoa

UNIVERSITY of STIRLING

OUTLINE

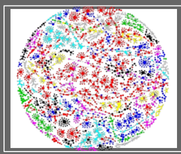


1932, ~1990

Landscape Analysis

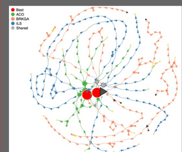
~2000

Complex Networks



2008

Local Optima Networks (LONs)



2020

Search Trajectory Networks (STNs)

1

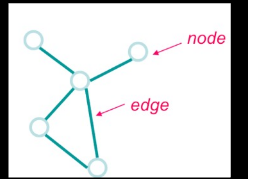
WHAT IS A NETWORK?

A collection *points* that are connected with *lines*

A Graph is an ordered pair $G = (V, E)$, where V set of vertices, E set of edges

Graph mathematical abstraction,

Network real-world data-driven instantiation



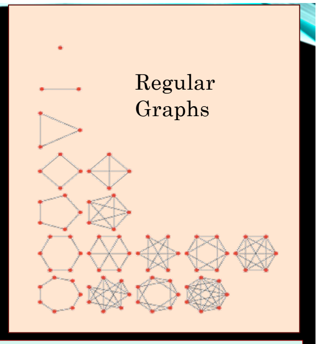
Points	Lines	Discipline
vertices	edges, arcs	Math (Graph Theory)
nodes	links	Computer Science
sites	bonds	Physics
actors	ties, relations	Sociology

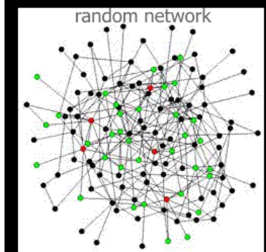
Terminology according to the discipline

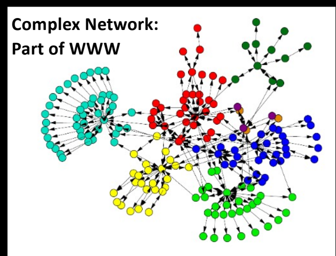
2

WHAT IS A COMPLEX NETWORK?

- Nontrivial topological features
  - Structure is irregular not regular/simple.
  - Not random either
- Can evolve over time



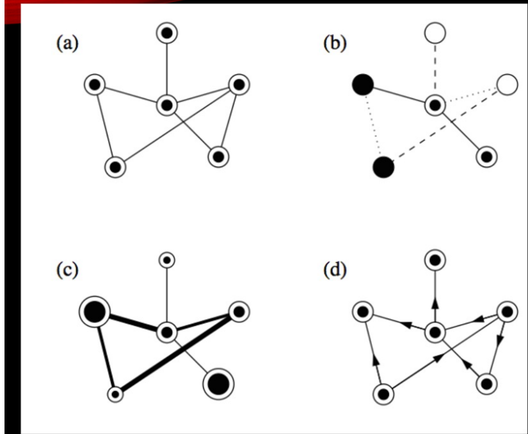




"Behind each complex system, there is an intricate network that encodes the interactions between the system's components."

Albert-László Barabási,
Network Science

TYPES OF NETWORKS



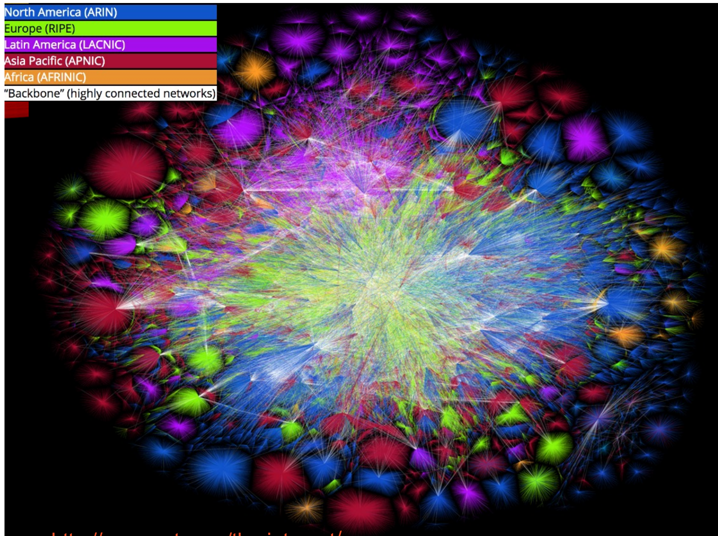
- (a) un-weighted, undirected
- (b) discrete vertex and edge types, undirected
- (c) varying vertex and edge weights, undirected
- (d) Directed (also called arcs)

4

EXAMPLES OF REAL-WORLD COMPLEX NETWORKS

Technological networks	• Internet, Telephone, Railway, Airlines, …
Information (knowledge) networks	• WWW, citation, semantic word, …
Biological networks	• Neural networks, metabolic networks, food webs, protein interaction networks, …
Social networks	• Friendships, scientific collaborations, …

5



The Internet

Map of routing paths: paths through which information flows from router to router all across the world.

Nodes IP addresses (devices such as routers)

Edges links between nodes. The length of the lines are indicative of the delay between those two nodes.

Colour different regions of the world

http://www.opte.org/the-internet/

6

BIOLOGICAL NETWORK

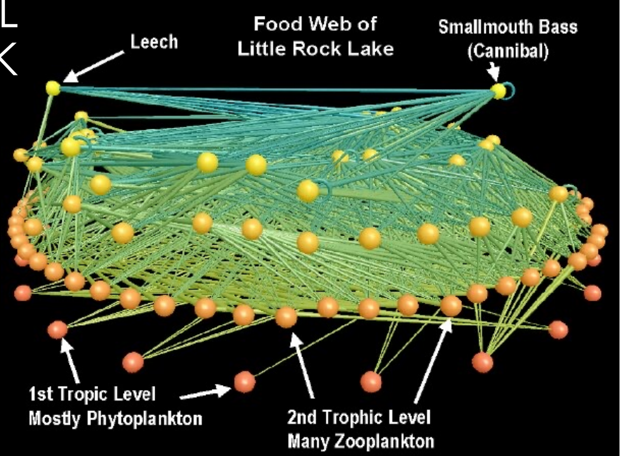
Food webs describe who eats whom among species within a habitat.

Nodes species

Edges feeding interactions *trophic links.*

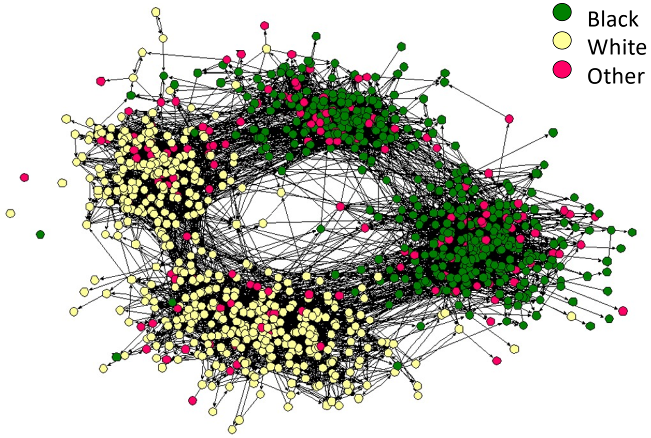
3D Visualisation of the food web of Little Rock Lake, Wisconsin (181 nodes and 2375 links). The network shows strong connectivity. Average links per node (in-coming & out-going) is 26.

https://doi.org/10.1145/1066677.1066950



7

A Social Network (Friendships in a US School)



A network with **community structure**, divides into groups of nodes with dense connections within groups and sparser connections between groups.

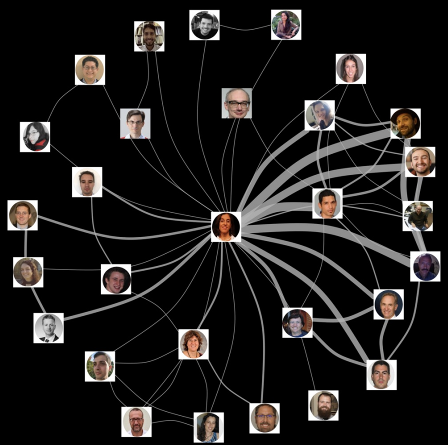
Nodes students

Edges reported friendship

Colour race

Moody, J., *Race, school integration, and friendship segregation in America*, *Am. J. Sociol.* 107, 679–716 (2001

8



COLLABORATION NETWORK

50 Articles on LONs (45) and STNs (5) 2008 - 2022 where I am a co-author

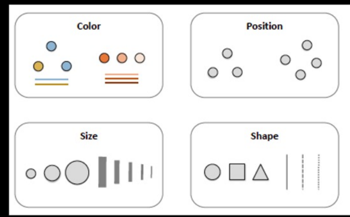
Nodes Researchers

Edges Co-authorship, width proportional to number of of papers

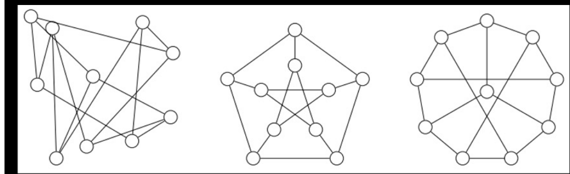
9

NETWORK VISUALISATION

Art of choosing an appropriate representation that is *aesthetically pleasing* and highlights important *structural properties*



https://kateto.net/network-visualization

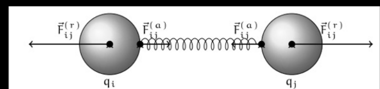


Petersen graph

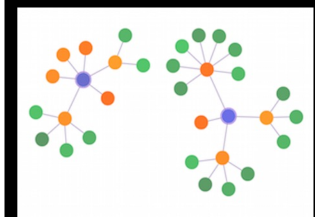
Node-edge diagram graph layouts

10

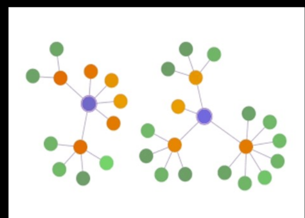
FORCE-DIRECTED GRAPH LAYOUT ALGORITHMS



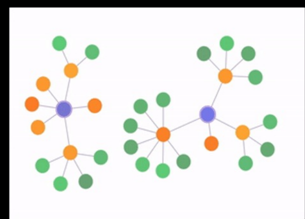
Aesthetic criteria
- Vertices are evenly distributed
- The number of edge crossings is minimised
- The lengths of edges are approximately uniform
- Inherent symmetries in the graph are respected



Fruchterman-Reingold

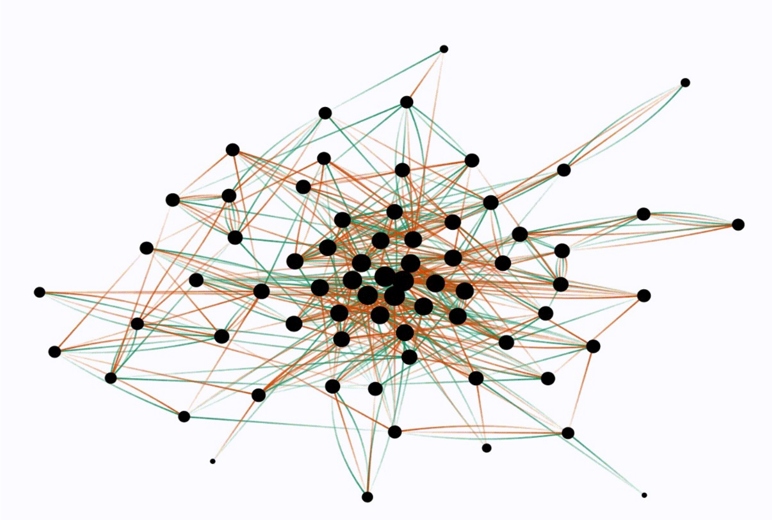


Kamada–Kawai (*organic*)



Reingold-Tilford (trees)

11



Graph Layout

Many Algorithms

Some are simple: circular, star, grid, random.

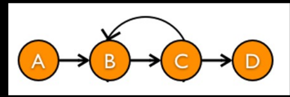
Others optimise the position of nodes based on different characteristics of the graph.

12

Source https://www.data-imaginist.com/2017/ggraph-introduction-layouts/

NETWORK METRICS

DIAMETER AND SHORTEST PATH



- Shortest path: the path between two nodes that visits the fewest intermediate nodes (path with less edges)
  - In the graph above, A->B->C->D is shorter than A->B->C->B->D (disallowing loops)
- Let $d(v_i, v_j)$ be the shortest-path distance between nodes *i* and *j*

Diameter**:** length of the longest shortest path between two vertices of the graph

Average shortest path distance

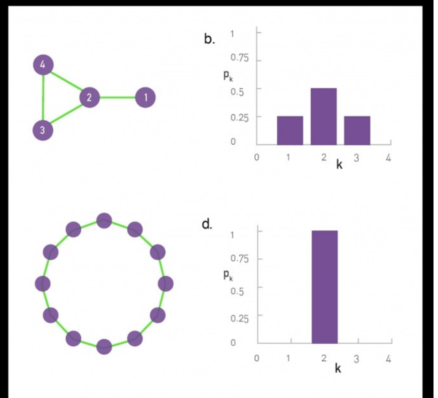
$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j)$$

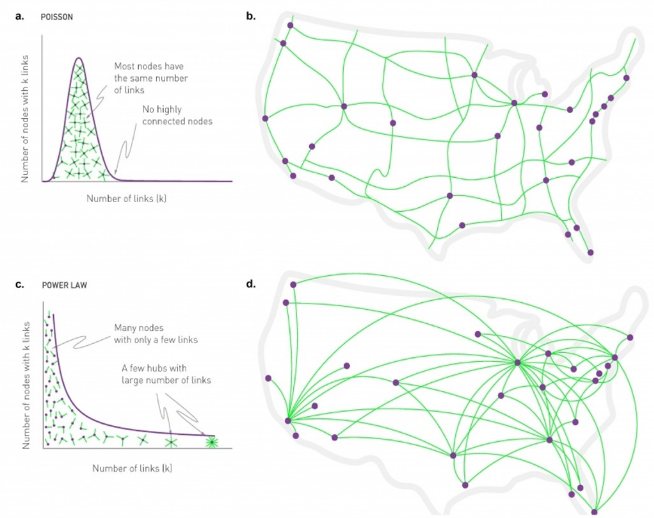
14

DEGREE AND DEGREE DISTRIBUTION

Degree *δ(i)* of vertex *i*
number of edges incident on *i*

Degree distribution
probability distribution of these degrees over the whole network.





POISSON VS. POWER-LAW DISTRIBUTIONS

Random Network

- Poisson distribution
- Most nodes have comparable degrees
- No nodes with large number of links
- Motorway network

Many Real-world Networks

- Power-law distribution
- Most nodes have only a few links
- Hubs: a few highly connected nodes
- Air-traffic network

16

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

(Newman, 2003)

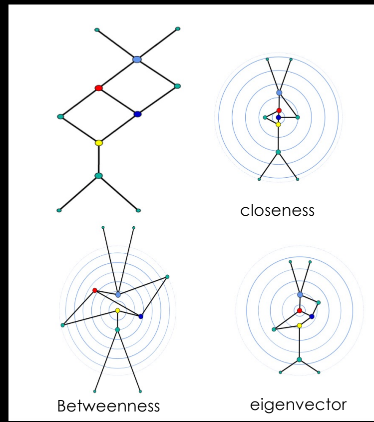
NETWORK METRICS

number of vertices n
number of edges m
mean degree z
mean vertex–vertex distance l
exponent power law dist. α
clustering coefficient C
degree correlation coefficient r

17

CENTRALITY METRICS

Degree	• importance score based on the number of links held by each node
Betweenness	• measures the number of times a node lies on the shortest path between other nodes
Closeness	• scores each node based on their 'closeness' to all other nodes
Eigenvector	• measures how well connected a node is, and how many links their connections have.
PageRank	• version of Eigenvector that considers edges direction and weight

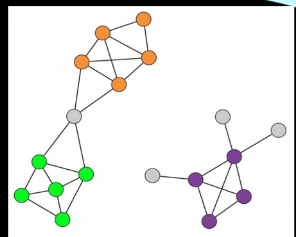


18

COMMUNITY STRUCTURE

Communities are locally dense connected subgraphs in a network. This expectation relies on two distinct hypotheses:

- Connectedness Hypothesis Each community corresponds to a connected subgraph
- Density Hypothesis Nodes in a community are more likely to connect to other members of the same community than to nodes in other communities.



The orange, the green and the purple nodes form communities.

They satisfy the connected and density hypothesis expectation.

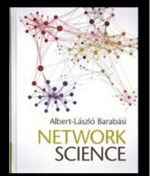
19

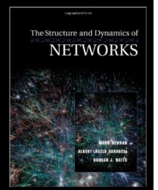
SOFTWARE

R	Python	C++	Interactive Apps
• igraph ggraph visNetwork	• NetworkX igraph SNAP	• SNAP	• Gephi Pajek Cytoscape

20

RESOURCES





Books

- L Barabasi Network Science http://barabasi.com/networksciencebook/
- M. E. J. Newman (2010) Networks: An Introduction, Oxford University Press
- E. D. Kolaczyk, G. Csárdi (2014) Statistical Analysis of Network Data with R, Springer

Articles

- Newman, M. E. (2003) The structure and function of complex networks. *SIAM review*, 45(2):167–256
- Newman, M. E. (2001) The structure of scientific collaboration networks.

21