Network science: Theory and real-world applications

Michele Bellingeri

Davide Cassi

Roberto Alfieri

Massimiliano Turchetto

Ho Chi Minh City, Vietnam, 7/12/2023

UNIVERSITÀ DI PARMA

ecosister

Ministero degli Affari Esteri e della Cooperazione Internazionale

UNIVERSITÀ DI PARMA NETWORKS UNIT

Robustness and important components of real-world social weighted networks

Progetto di Grande Rilevanza 2021-2023

Italia-Vietnam

Information and Communication Technologies

Ministero degli Affari Esteri
e della Cooperazione Internazionale

BỘ KHOA HỌC VÀ CÔNG NGHỆ

Italian group

Prof. Davide Cassi

Prof. Roberto Alfieri

Dr. Michele Bellingeri

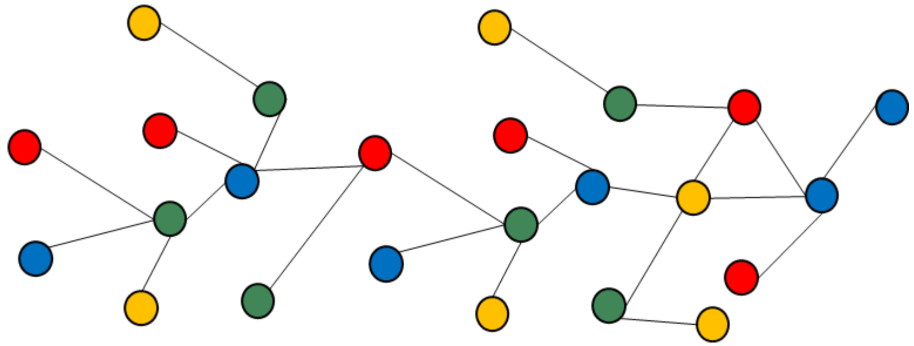
Dr. Massimiliano Turchetto

UNIVERSITÀ DI PARMA
NETWORKS UNIT

https://www.networks.unipr.it/wordpress/

Network Science

Network science investigates complex networks



A **network** is made by:

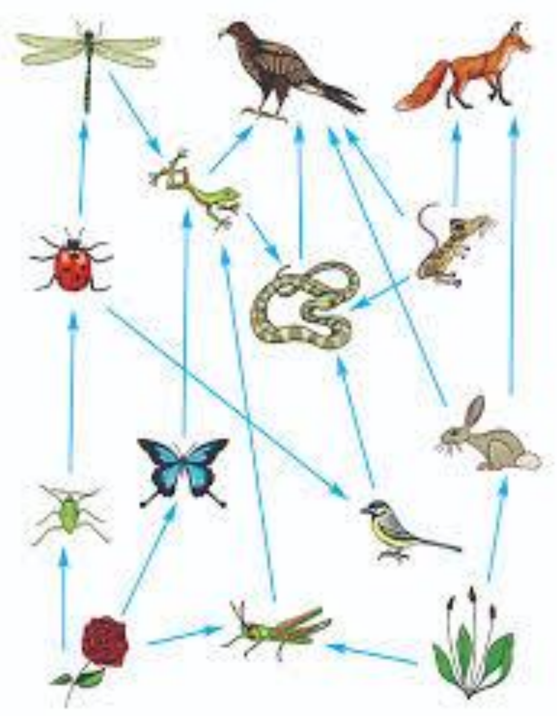
nodes (object, persons, places, species, computers…)

connected by **links** (relationships, contacts, cables, trophic flows, connections, …)

Real Networks

Ecological Trophic Networks

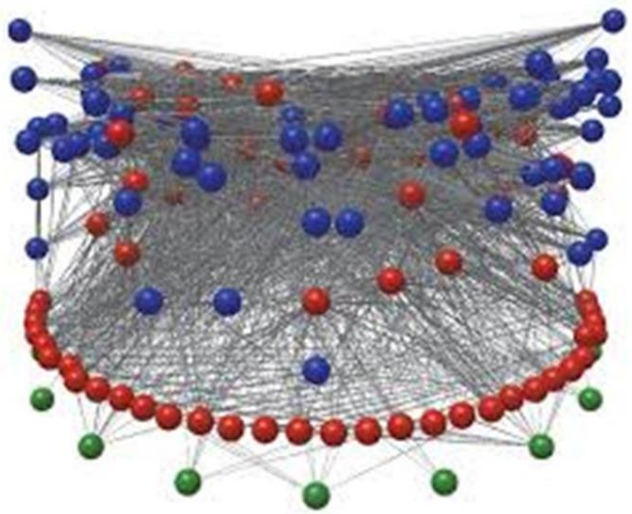
Food webs describe '*who eats whom*' in ecosystems



Complex Networks

Nodes → Species

Links → Trophic relations



Real Networks

Social Networks

Networks describing social interactions

Complex Networks

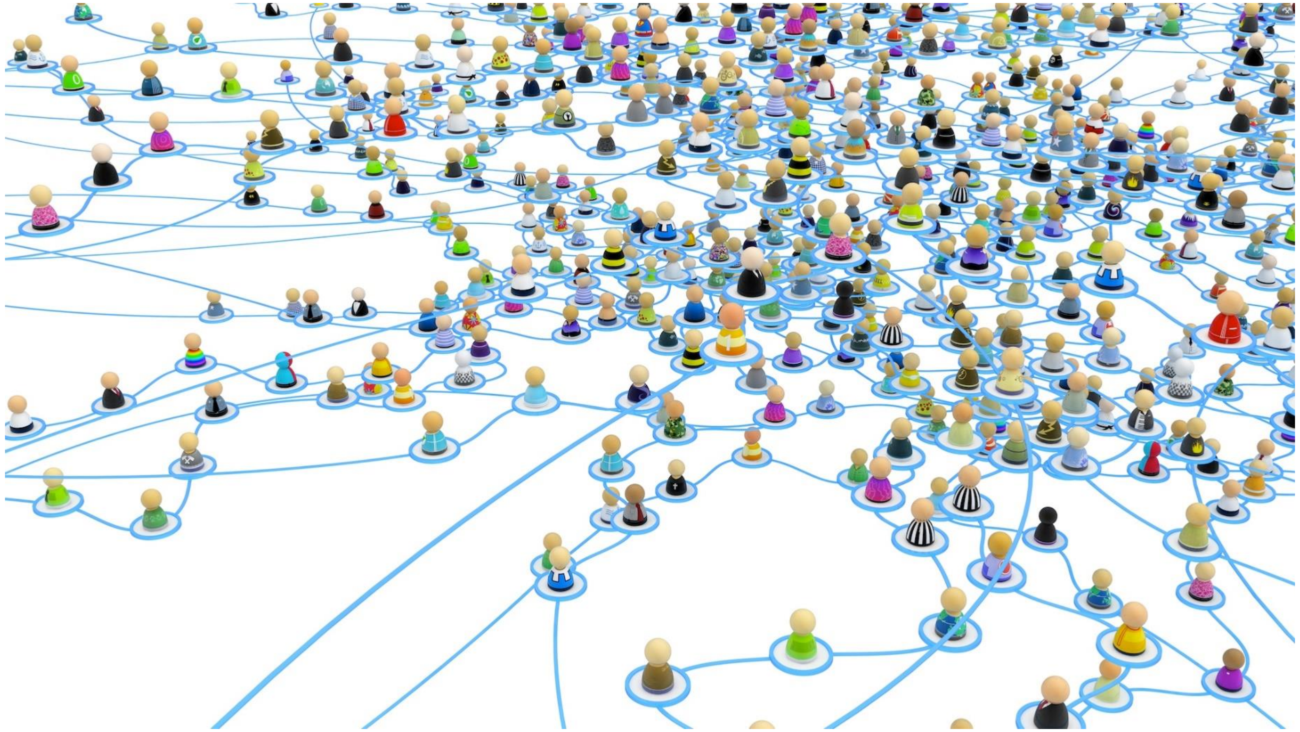
Nodes → Individuals

Links → Friendship

Contact

Working

Disease spreading



https://pmrpressrelease.com/global-location-based-social-networking-service-lbsns-market/

Real Networks

Social Networks

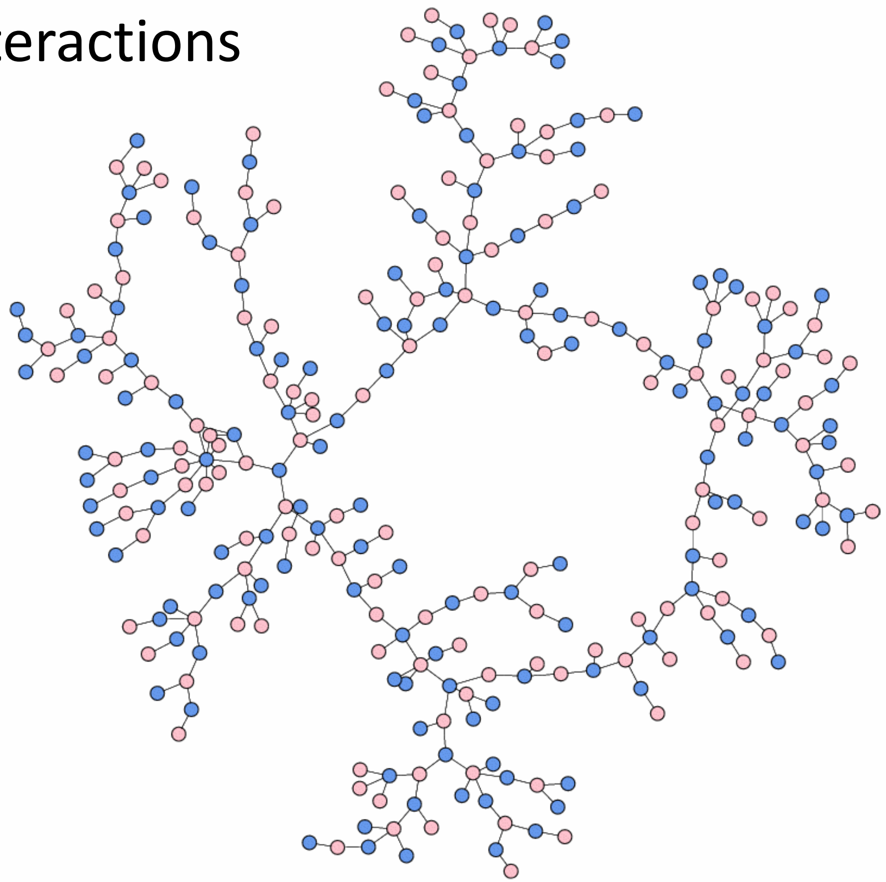
Online social networks account online social interactions

Facebook complex networks

Nodes → Pages

Links → Hyperlinks

↘ Mutual likes



https://www.cs.uoi.gr/~tsap/teaching/2014-cs-l14/references.html

Real Networks

Communication Networks

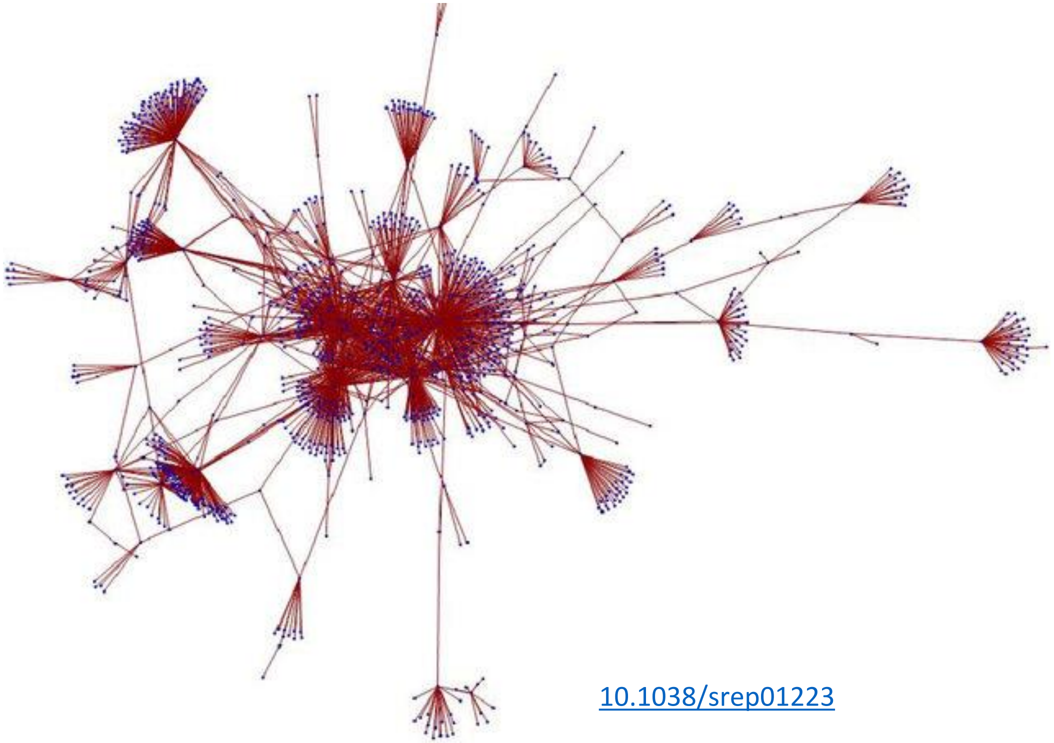
Modelling the communication activities among individuals or entities

Enron email network

Email complex networks

Nodes → Employees

Links → Email exchange



10.1038/srep01223

Real Networks

Computers Networks

Nodes → Routers

→ PC

Links → Cables

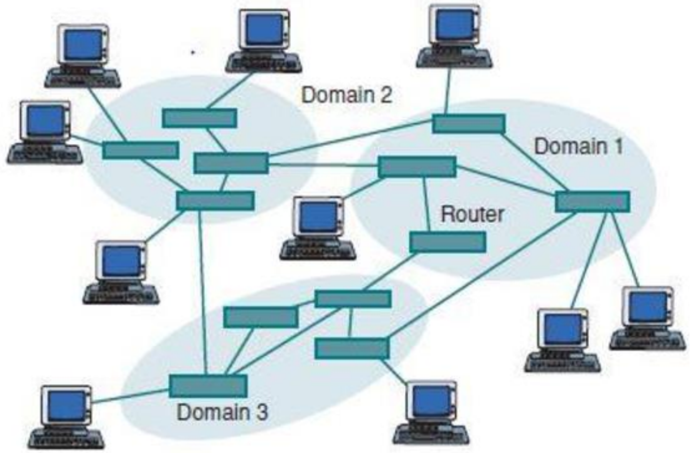
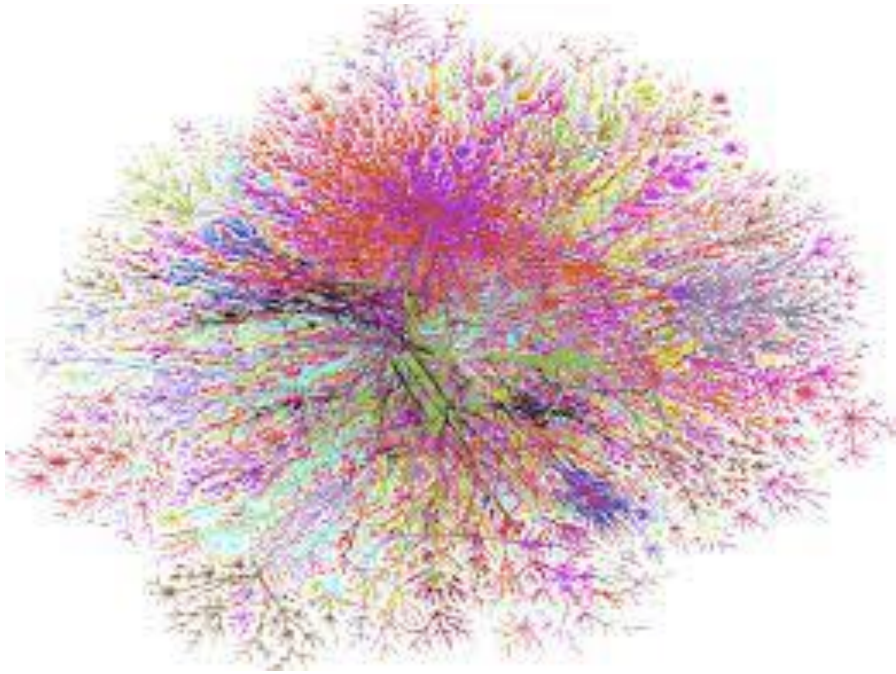


Fig. 1 – Network Structure of the Internet

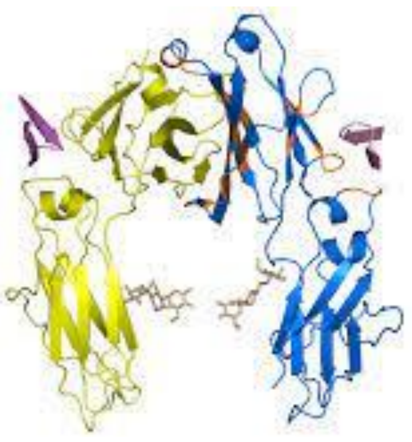
10.5120/17765-8882

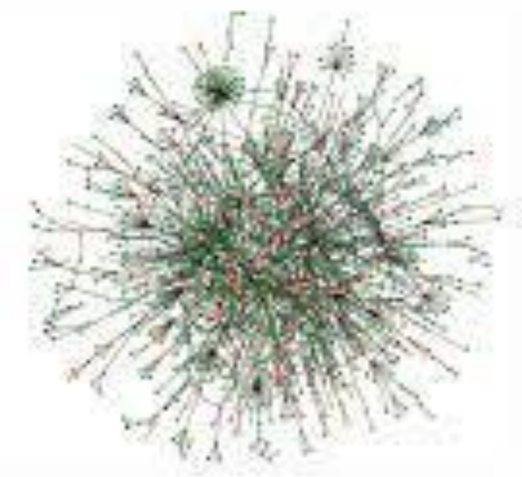


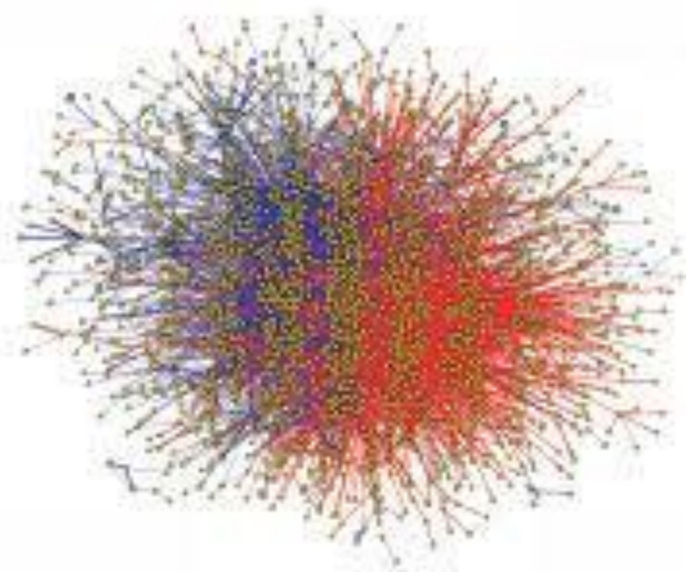
https://newmedialab.cuny.edu/project/complex-networks/

Real Networks

Proteins Networks



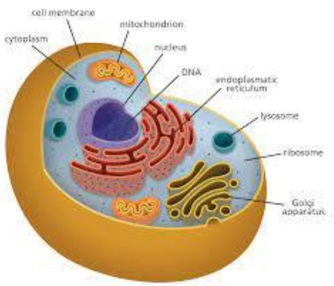




https://www.creative-proteomics.com/blog/index.php/brief-introduction-of-protein-protein-interaction-ppi/

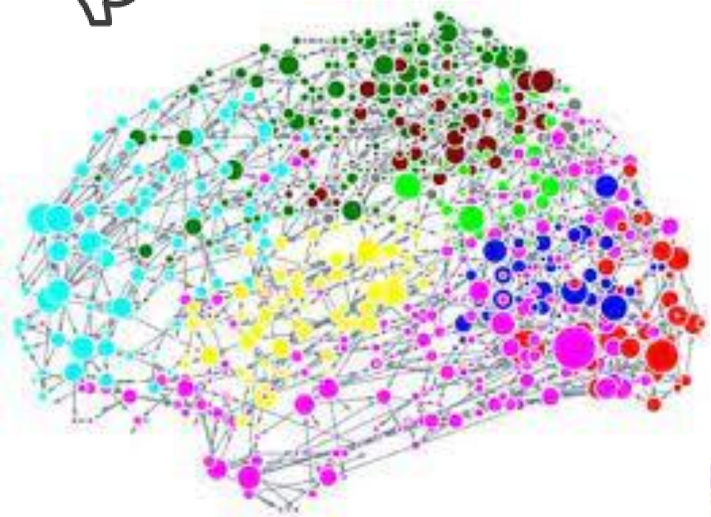
Nodes → Proteins

Links → Metabolic interactions

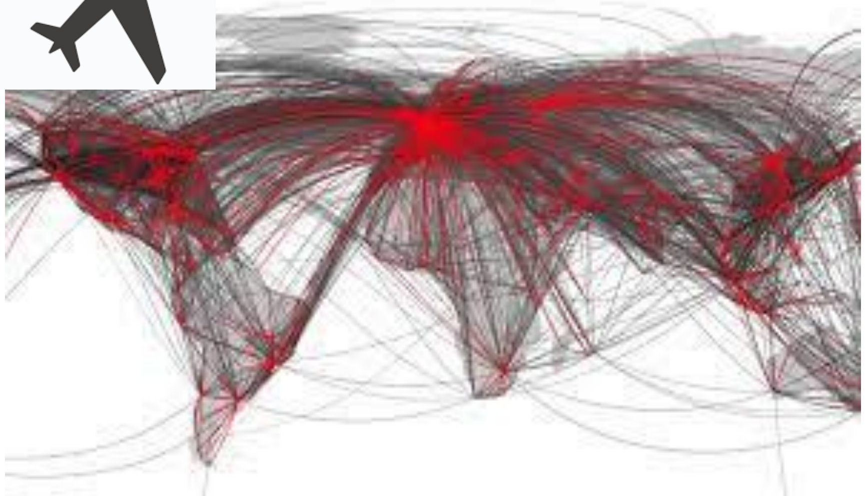


Many more

Brain networks



Airports networks



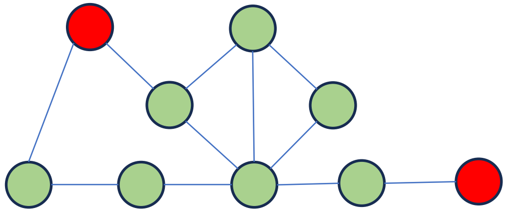
Road networks

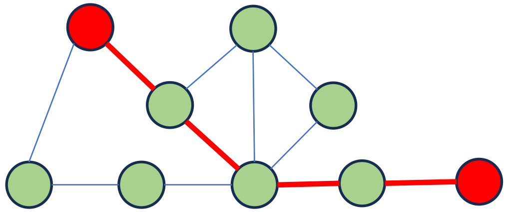
Notions

Shortest path

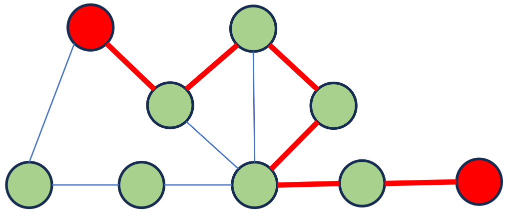
A ***"path"*** is a set of distinct and connected nodes.

The **shortest path** between two nodes is the **minimum number of links** required to travel from one node to another.





4 Links



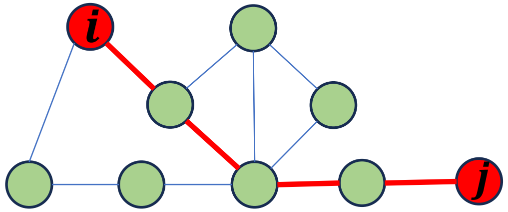
6 Links

shortest path

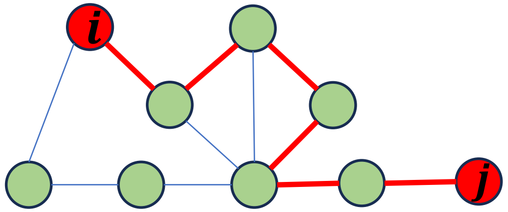
Notions

Distance between nodes

The distance d_{ij} is the shortest path length between nodes i and j



4 Links



6 Links

$$d_{ij} = 4$$

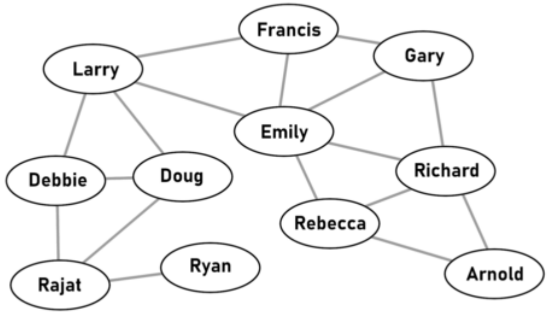
Notions

Distance between nodes - Applications

The node distance d_{ij} can find:

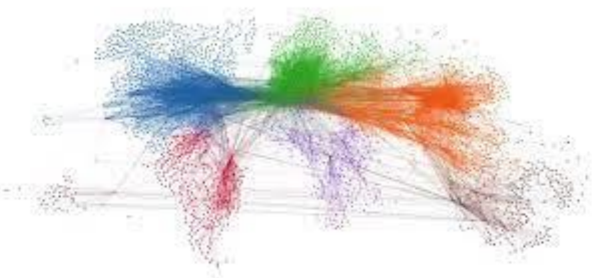
Best route between physical locations

Fastest information delivery among individuals



Six degrees of separation!

Minimum number of airports to travel in a journey



Node centrality

How to quantify the importance of a node?

Degree Centrality

Number of neighbors = Number of links

The degree of the node i

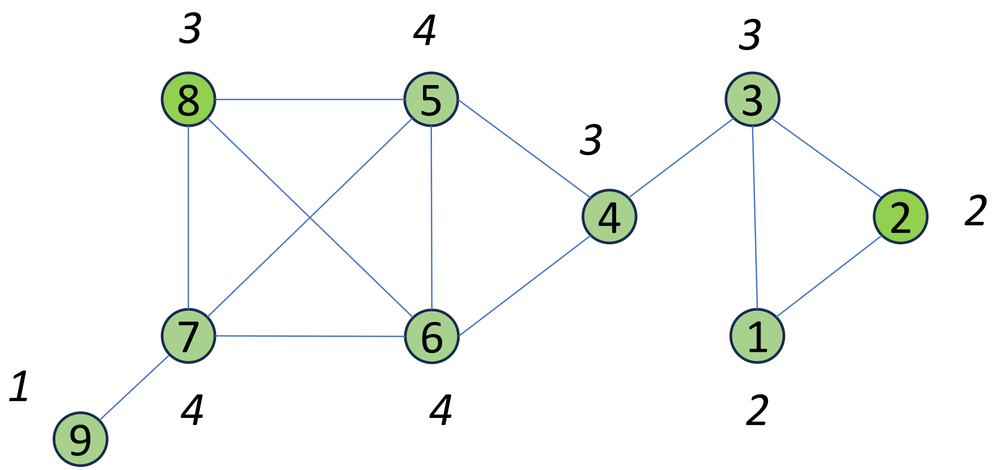
$$k_i = \sum_j a_{ij}$$

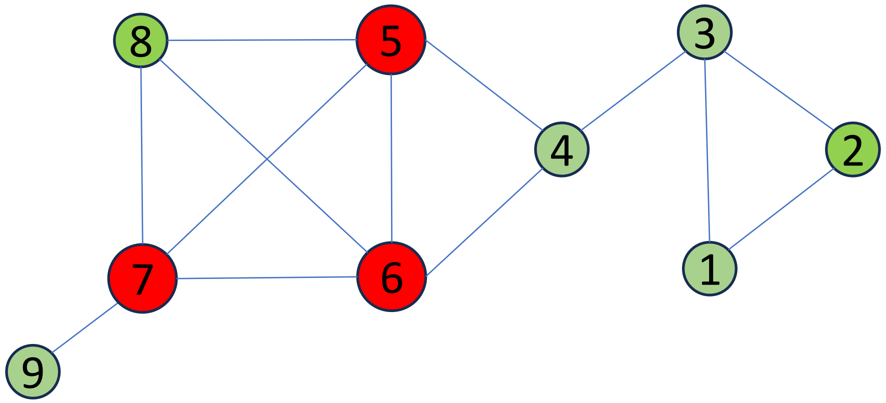
Where a_{ij} is the element of the adjacency matrix

Node centrality

Degree Centrality

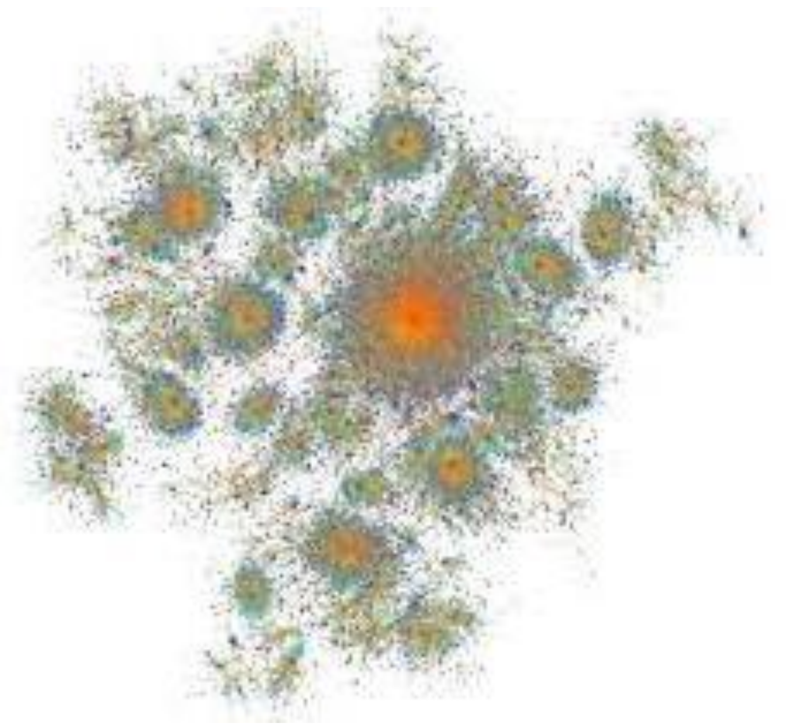
$$k_i = \sum_j a_{ij}$$



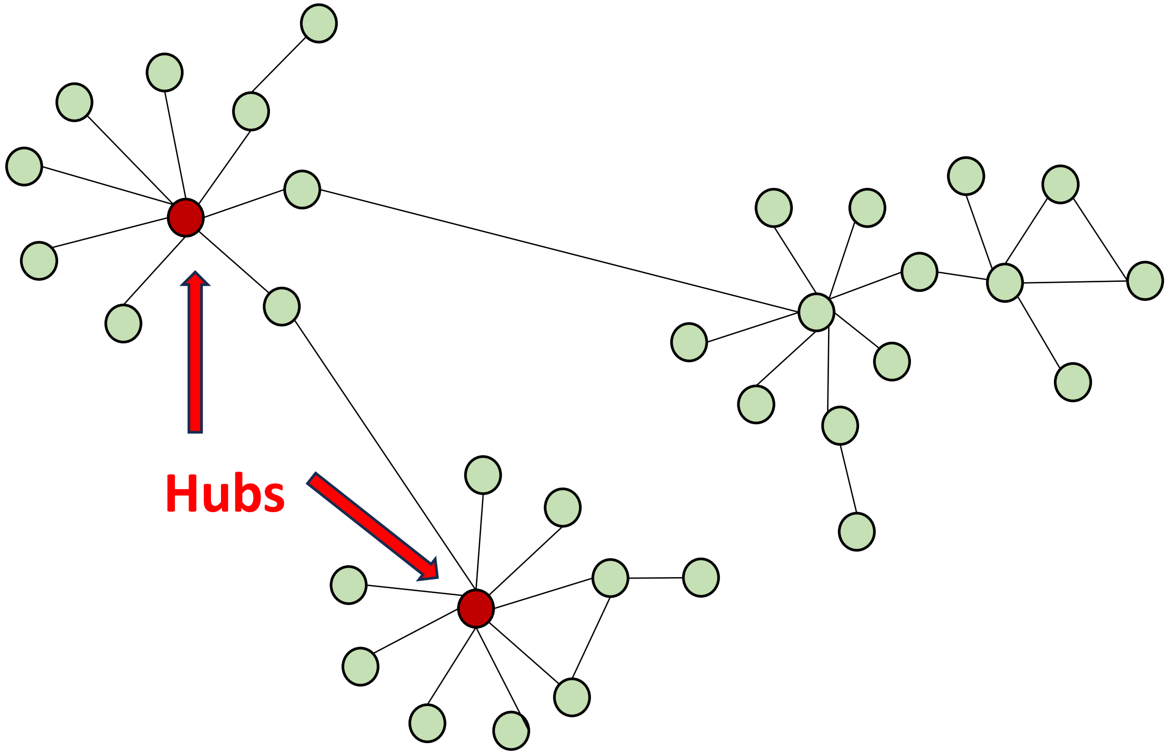


Node centrality

Degree Centrality



https://www.networkpages.nl/



Hyper connected nodes !!

Node centrality

Betweenness Centrality

The betweenness $g(i)$ of the node i is

$$g(i) = \sum_{s,t=1}^{N} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

$\sigma_{st}(i)$ is the number of the shortest paths passing through the node i

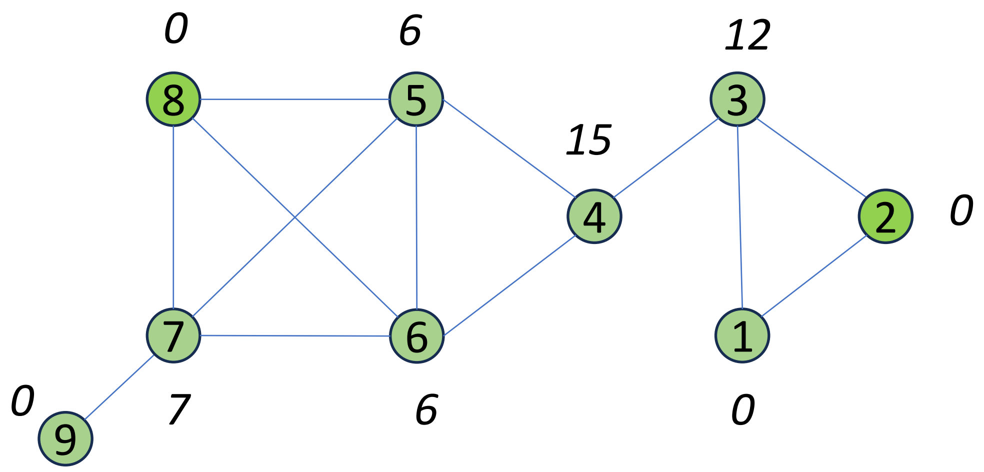
σ_{st} is the total number of shortest paths between nodes s and t

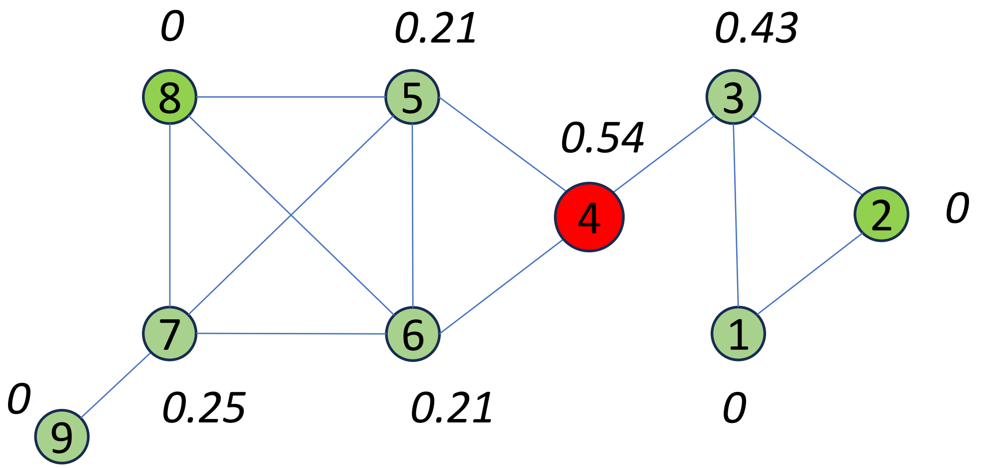
N the number of nodes.

Node centrality

Betweenness Centrality

$$g(i) = \sum_{s,t=1}^{N} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

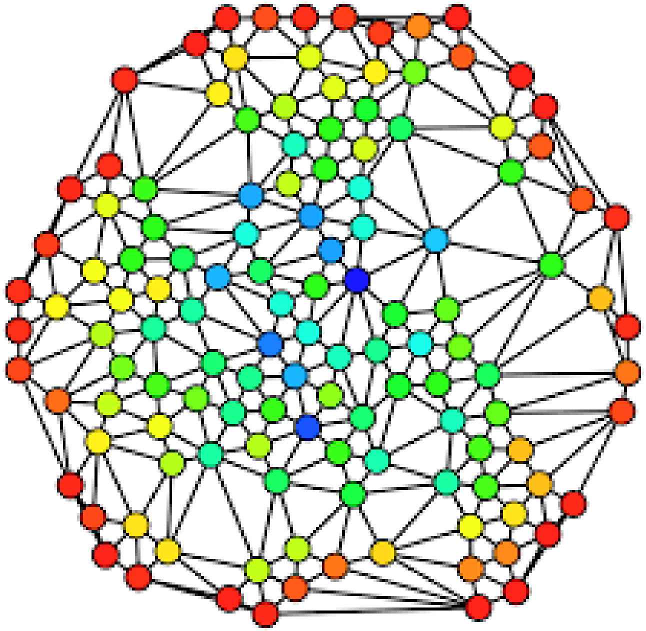




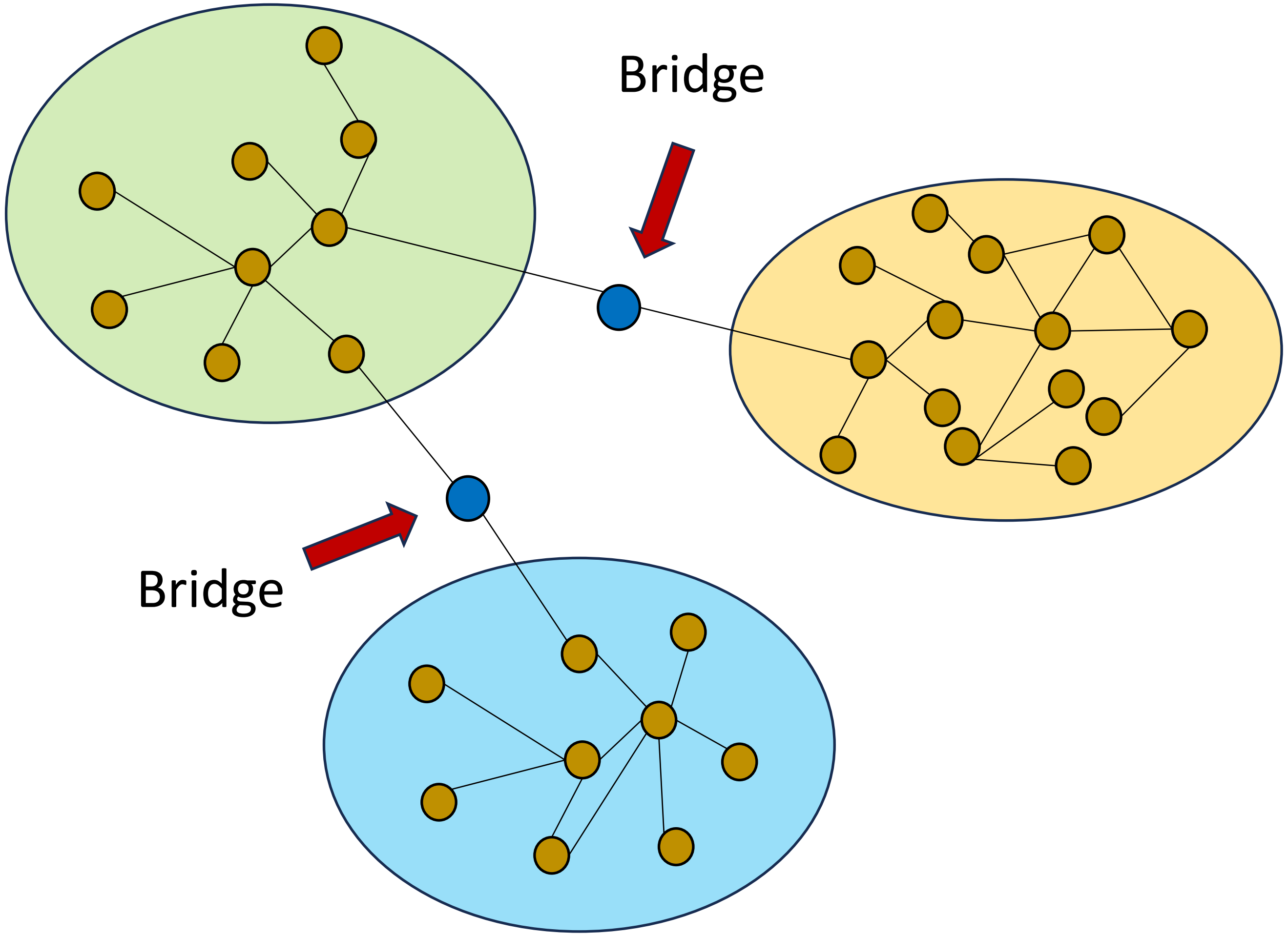
Node centrality

Betweenness Centrality

From least (**red**) to greatest (**blue**).



https://en.wikipedia.org/wiki/Betweenness_centrality



Node centrality

Closeness Centrality

The closeness is the inverse of the sum of distances to all nodes.

The closeness $c(i)$ of the node i is

$$c(i) = \frac{1}{\sum_{j\neq i, j=1}^{N} d_{ij}}$$

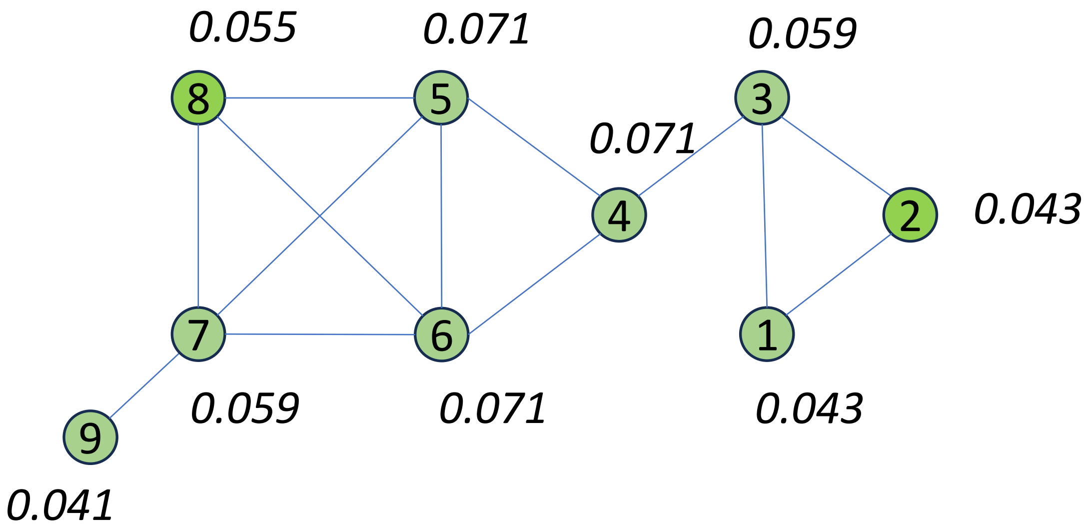
d_{ij} is the distance between nodes i and j

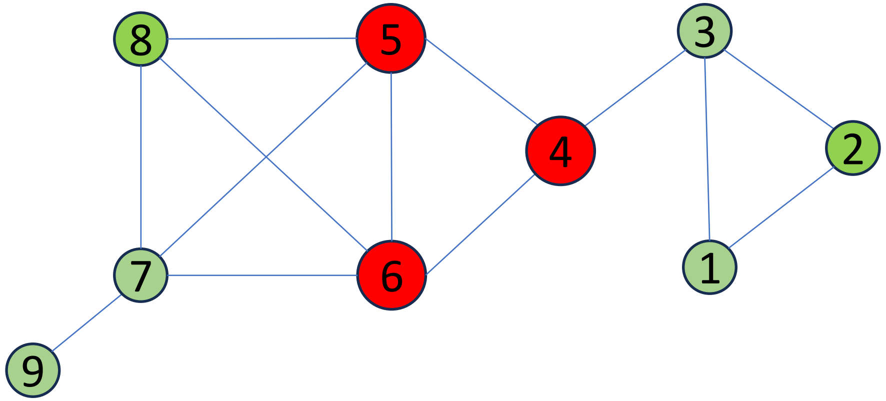
N the number of nodes.

Node centrality

Closeness Centrality

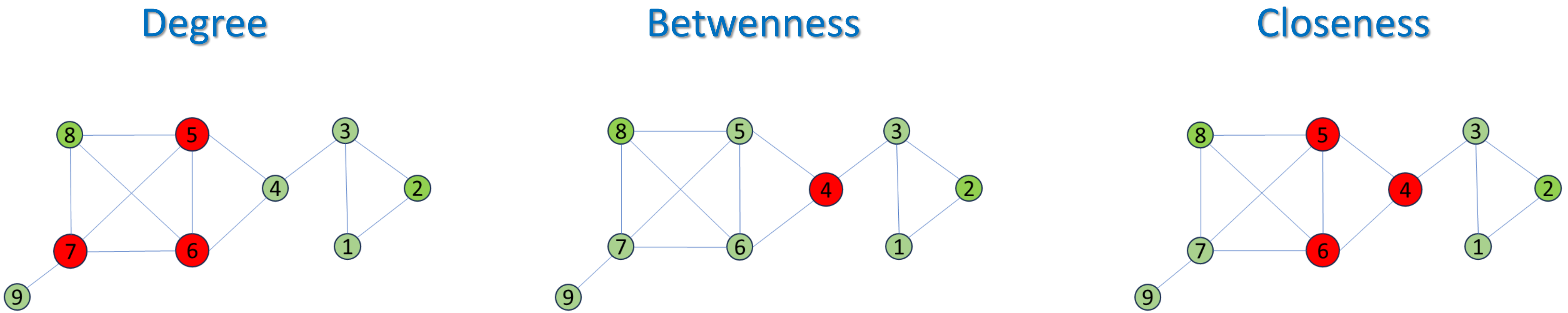
$$c(i) = \frac{1}{\sum_{j \neq i, j=1}^{N} d_{ij}}$$





The **more central** a node is ⟹ **the closer** it is to all other nodes

Node centrality



Different centrality indicators → Different node importance

Network robustness

ROBUSTNESS

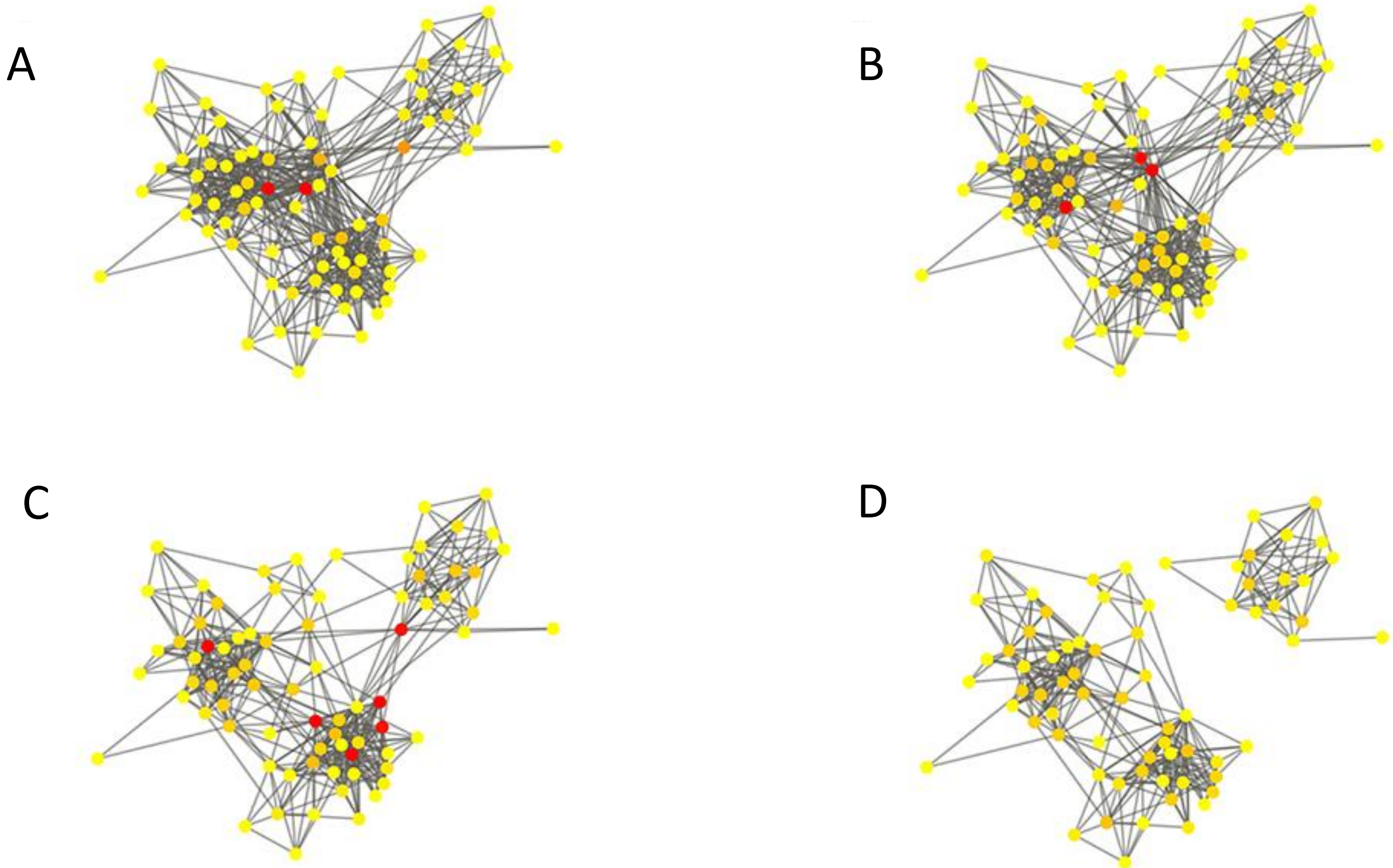
↓

Ability to withstand failures and perturbations

Robustness is a critical attribute of many complex systems and complex networks.

Network robustness

Predict network robustness to node removal



Network robustness

Robustness Measure

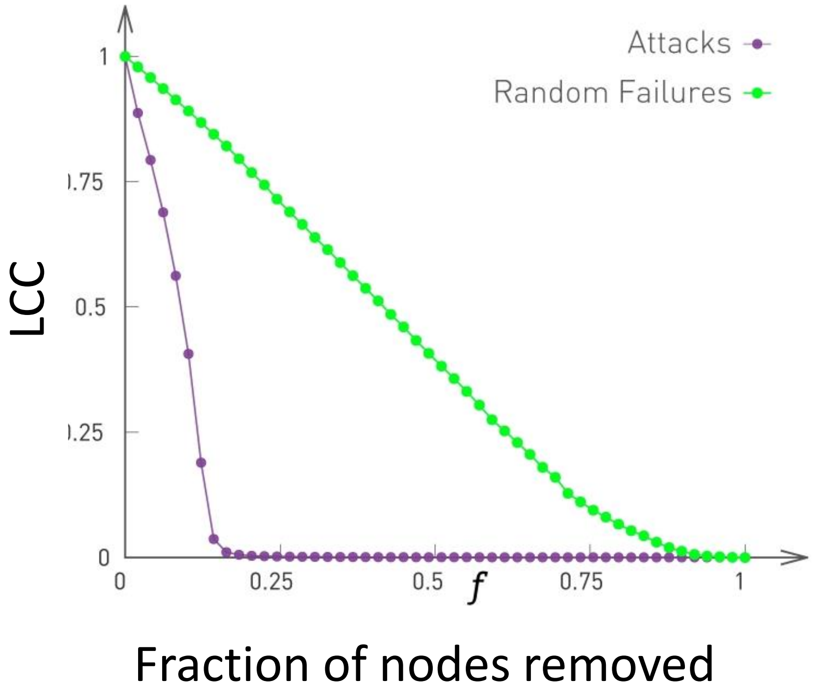
Largest Connected Component LCC

↓

Maximum number of connected nodes

Real networks ➡ **resistant** to random removal

Real networks ➡ **vulnerable** to targeted attack



Albert and Barabási 2002, Statistical mechanics of complex networks Rev. Mod. Phys. 74, 47.

Best attack strategy

The quickest network dismantle

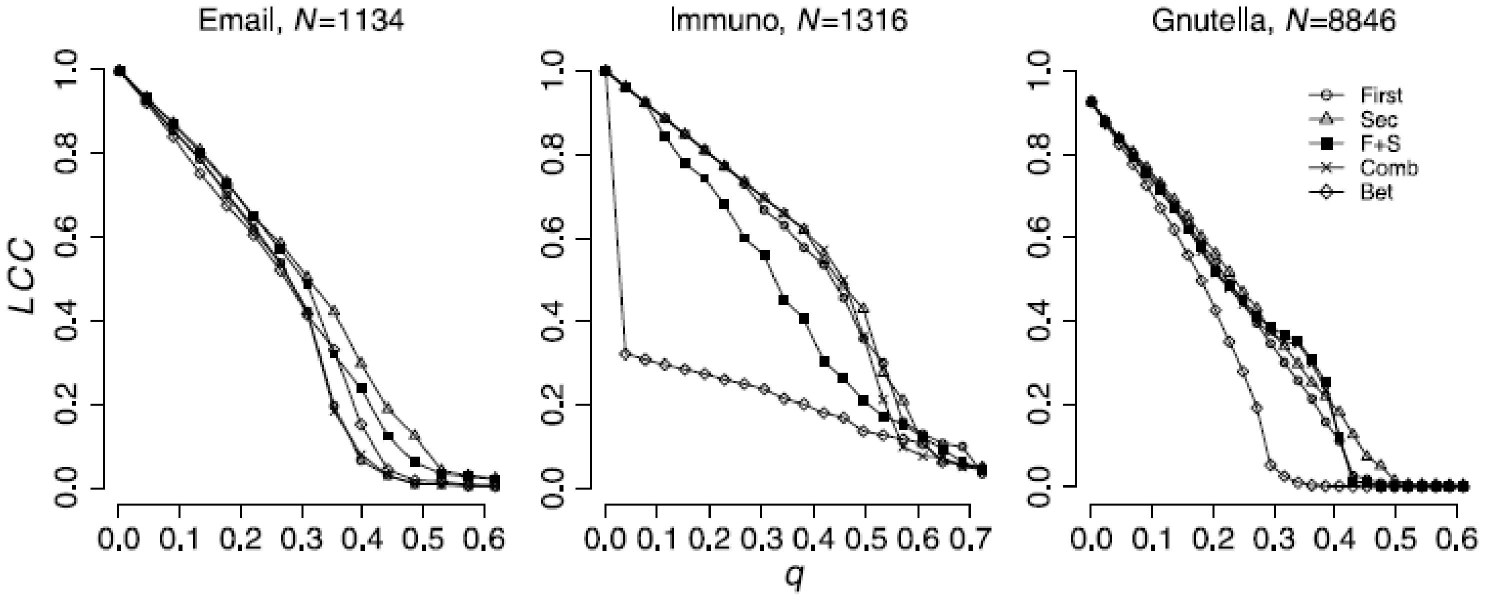


Fig. 3. Size of normalized *LCC* and the fraction *q* of nodes removed for recalculated targeted attacks to model networks. Points are plotted every 20 nodes removed for networks with $N = 500$ and $N = 1000$, and every 200 nodes removed for $N = 10\,000$.

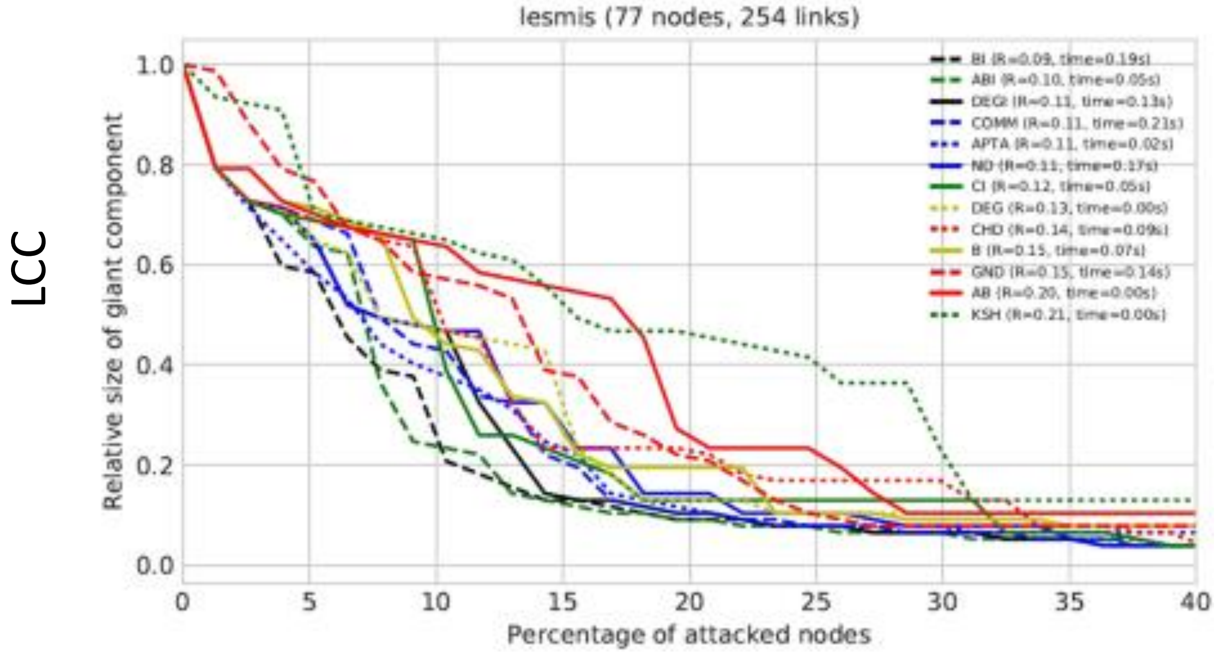
Betweenness

Highly effective strategies to dismantle real networks !!

Bellingeri M., Cassi D., Vincenzi S., 2014.
Efficiency of attack strategies on complex model and real-world networks, Physica A, 414.

Best attack strategy

Compare node attack strategies



Wandelt et al. 2018. A comparative analysis of approaches to network-dismantling. Sci Rep 8, 13513.

The best algorithm is betweenness!

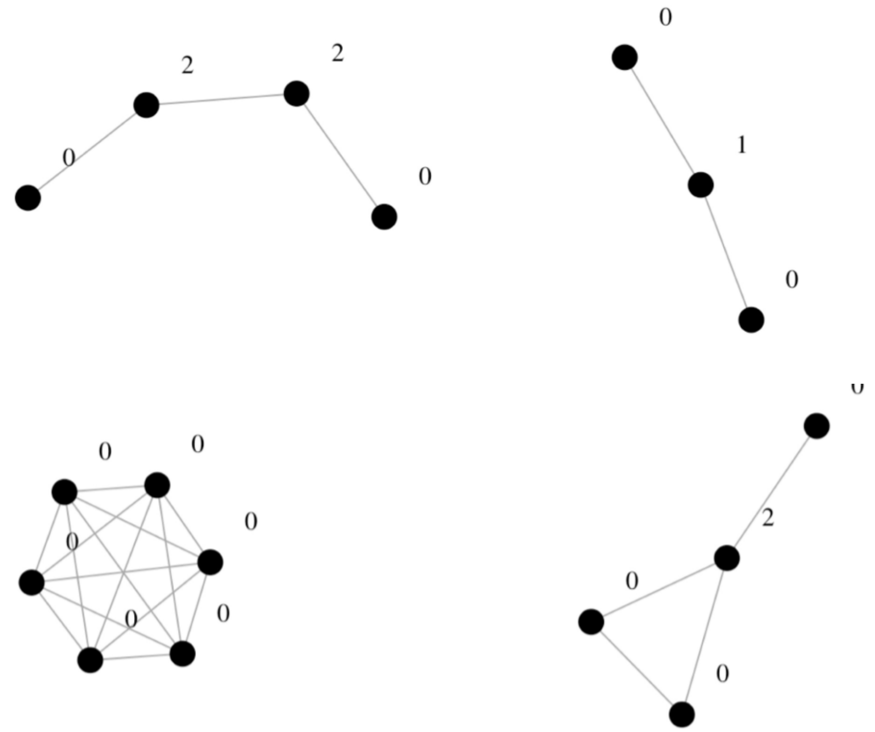
The definition is extremely well aligned with the problem

↓

disrupt the main communication paths of the network.

Best attack strategy

Betweeneess drawbacks



LCC is a complete graph

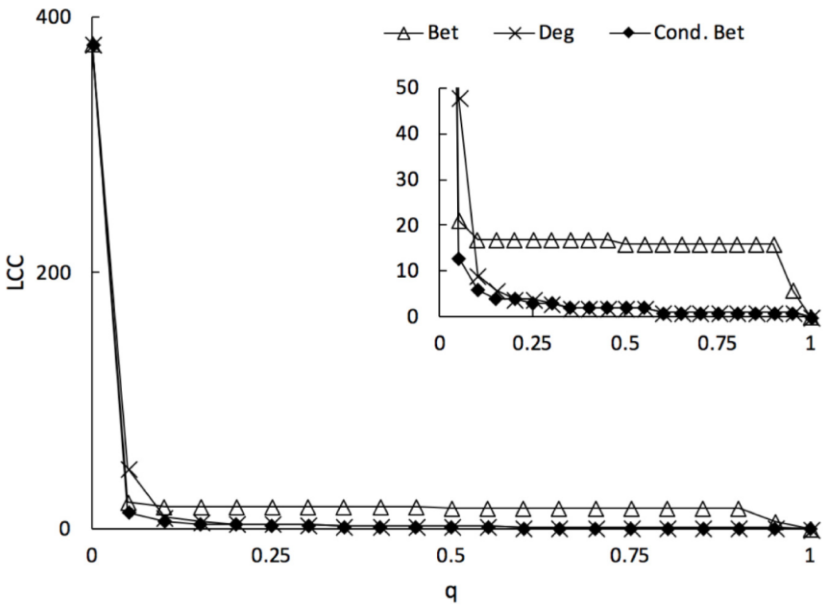
Nodes has zero betweenness

Betweenness does not select nodes inside a LCC complete graph

Nguyen, Pham, Cassi, Bellingeri, 2019. Conditional attack strategy for real-world complex networks, Physica A, 530, 121561

Best attack strategy

Betweeneess drawbacks



LCC is a complete graph

Betweenness ineffective at the end

Propose → **Conditional betweenness**

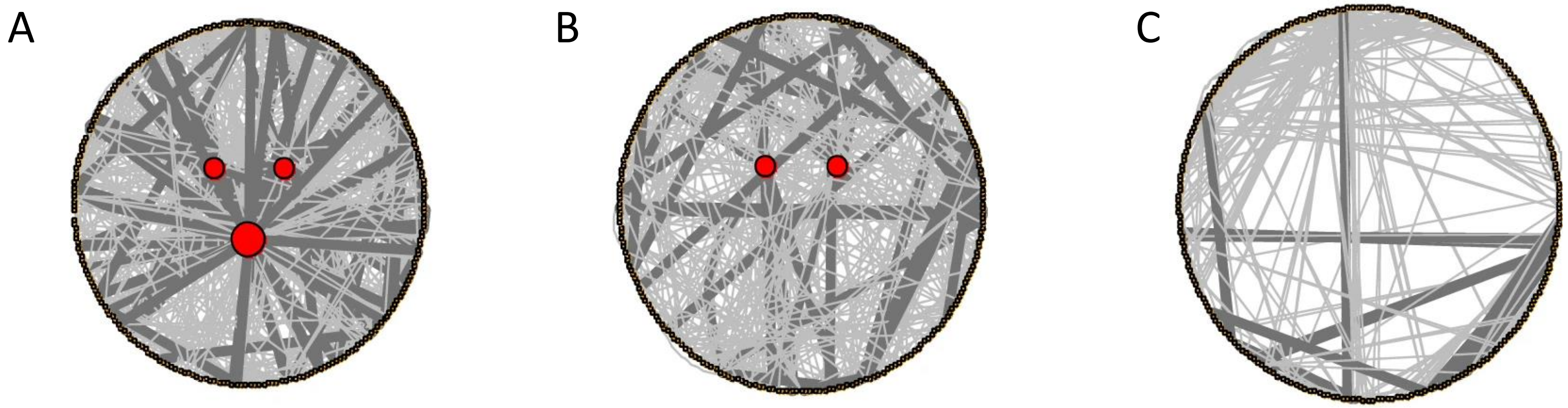
Remove the highest betweenness node only inside LCC.

Nguyen, Pham, Cassi, Bellingeri, 2019. Conditional attack strategy for real-world complex networks, Physica A, 530, 121561

Network Robustness

LCC drawbacks

LCC accounts the max number of connected nodes **only**



‘**connected but inefficient**’ network

Bellingeri et al. 2019. The heterogeneity in link weights may decrease the robustness of real-world complex weighted networks. Sci Rep 9, 10692.

Network Robustness

Network efficiency Eff

The efficiency measures how **efficiently networks exchange information**

$$Eff = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$$

d_{ij} is the distance between nodes i and j

N the number of nodes.

Also called communication efficiency.

Rationale

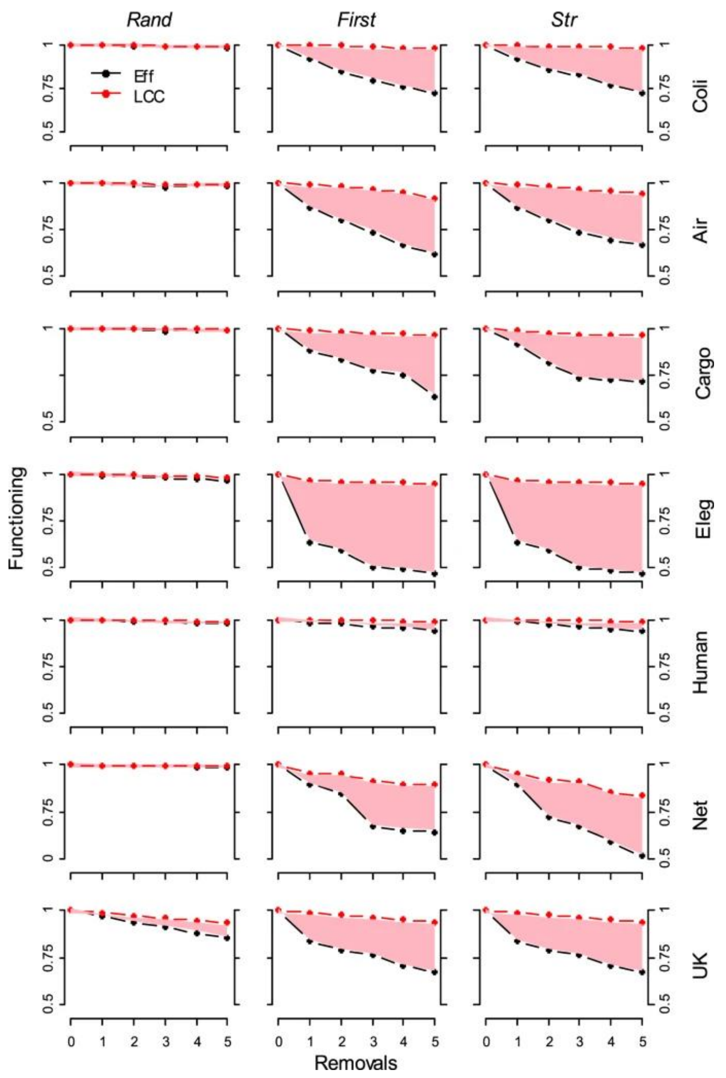
More distant nodes

↓

the less efficient their communication

Network Robustness

LCC *vs* Eff



Handful node removals

Quick Eff decrease

LCC is roughly constant

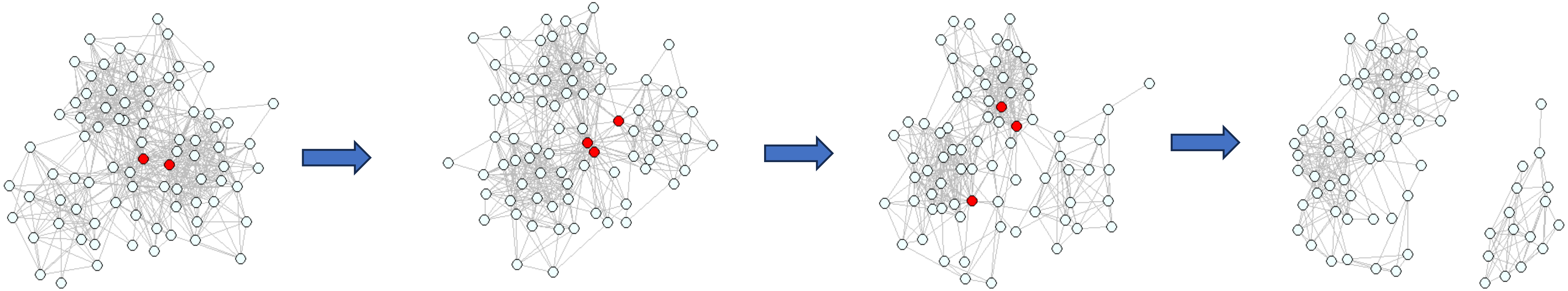
Network ➡ '**connected but inefficient**'

Bellingeri et al. 2019. The heterogeneity in link weights may decrease the robustness of real-world complex weighted networks. Sci Rep 9, 10692.

Modelling Vaccination

Node removal model vaccination process

Social contact networks Animal networks Computers networks



Node vaccination → Disrupt epidemic paths → Curb disease spreading

Bellingeri M, Bevacqua D, Scotognella F, Alfieri R, Nguyen Q, Montepietra D, Cassi D (2020), Link and Node Removal in Real Social Networks: A Review, Frontiers in Physics, 8,228.

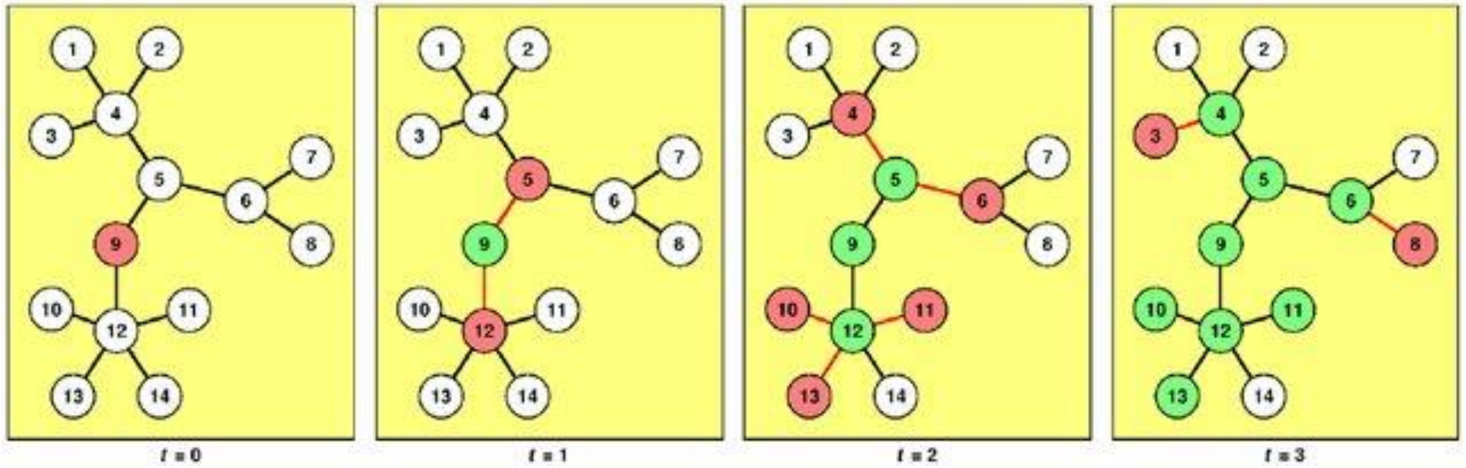
SIR Epidemic spreading

The population is assigned to compartments:

S ➡ Susceptible ○

I ➡ Infectious ●

R ➡ Recovered ●

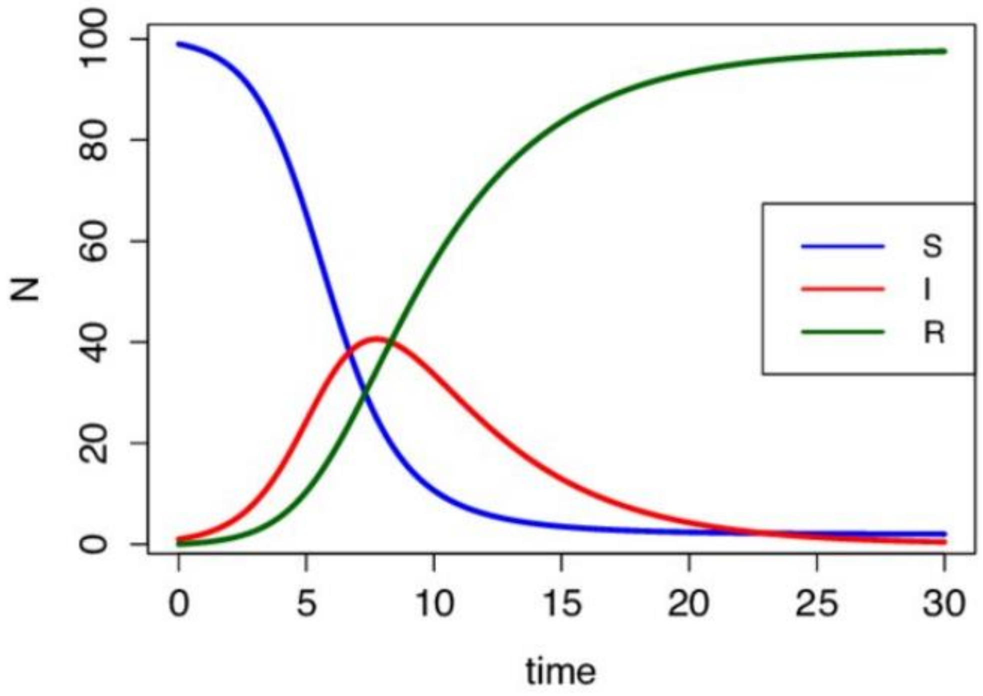


https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0022124.g001

epipack 0.1.5

SIR Epidemic spreading

Evaluates the epidemic spreading pace



10.13140/RG.2.2.28517.19689

τ_{15} time steps needed for the disease to strike 15% of nodes

TI overall number of nodes affected by the disease

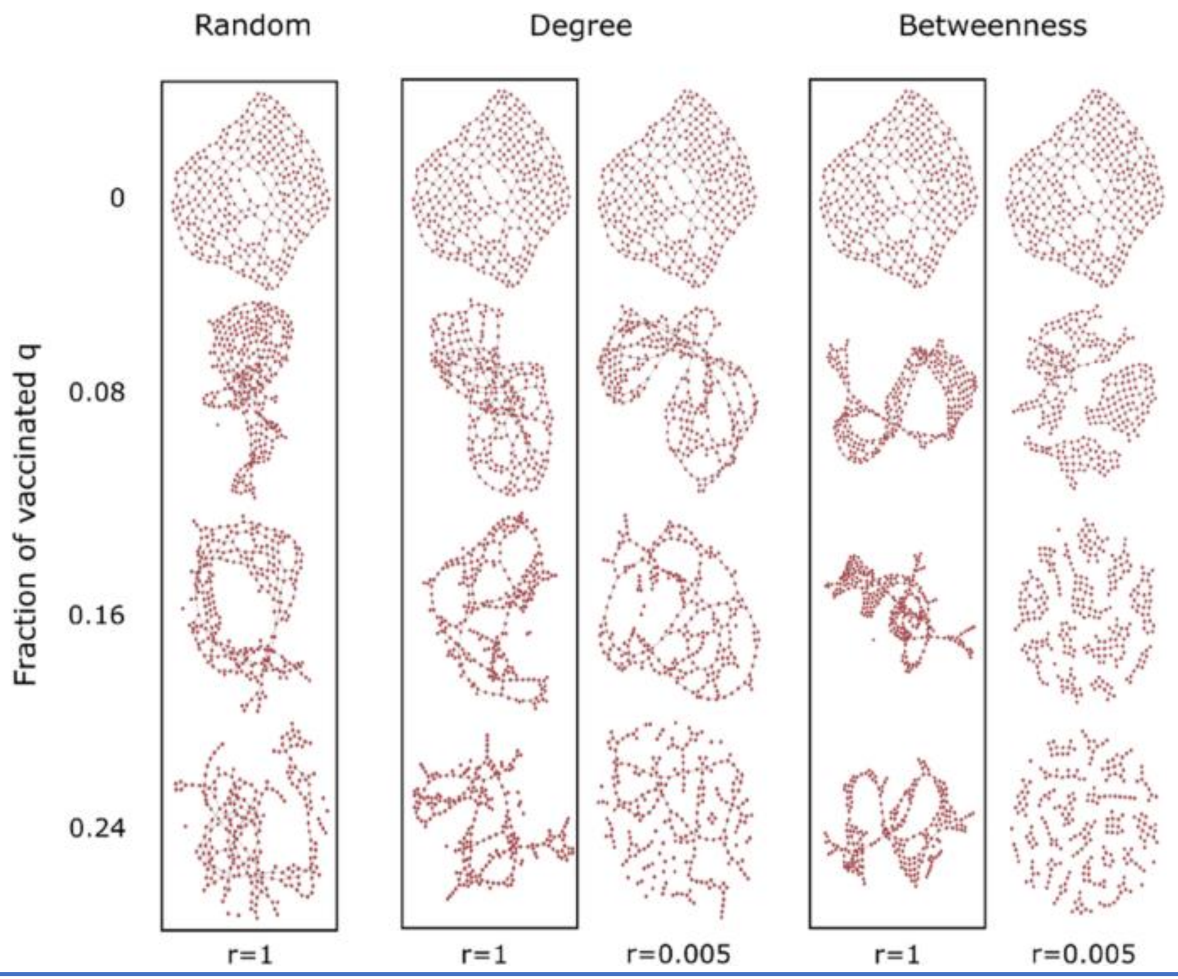
ζ the maximum number of nodes concurrently infected

↓

the maximum disease prevalence

Modelling Vaccination

Compare node vaccination strategies to halt SIR



Non-adaptive approach

According to the initial network structure

Semi-adaptive approach.

Performing partial node rank recalculation

Sartori et al. A comparison of node vaccination strategies to halt SIR epidemic spreading in real-world complex networks. Sci Rep. (2022).

Modelling Vaccination

Compare node vaccination strategies to halt SIR

Results

- Different best strategy in the non-adaptive and semi-adaptive approaches

- Best strategy depends on the number of available vaccines.

-Partial rank recalculation increases the efficacy by up to 80%.

Sartori et al. A comparison of node vaccination strategies to halt SIR epidemic spreading in real-world complex networks. Sci Rep. (2022).

Forecasting SIR epidemic spreading

Correlate

epidemic spreading pace

↕

40 indicators of **network structure**

Salient structural features

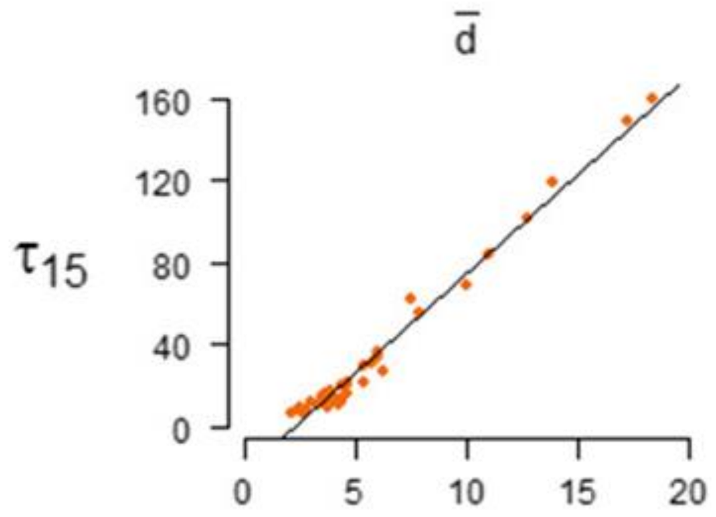
- Degree heterogeneity
- Nodes Communities
- Connectivity Level
- Efficiency
- Node Distance

40 Network Structural Indicators (NSI)

ID	Key	Full name	Formula	Definition	Reference
1	N	Node number		N is the number of nodes in the network	
2	L	Link number		L is the number of links in the network	
3	C	Connectance	$C = \frac{2L}{N(N-1)}$	L is the number of links, and N is the number of nodes	[15]
4	$\bar{k}$	Average node degree	$\bar{k} = \frac{1}{N}\sum_{i=1}^{i=N} k_i$	k_i is the degree of the node i, and N is the nodes' number	[1]
5	σ_k	Node degree standard deviation	$\sigma_k = \sqrt{\frac{\sum_{i=1}^{i=N}(k_i-\bar{k})^2}{N-1}}$	k_i is the degree of the node i, $\bar{k}$ is the average node degree, and N is the nodes' number	[49]
6	AH	Albertson index	$AH = \sum_{i,j\in L} \lvert k_i - k_j \rvert$	i, j is the link connecting nodes i and j, k_i is the degree of the node i, k_j is the degree of the node j, and L is the network link set.	[46]
7	nAH	Normalized Albertson index	$nAH = \frac{AH}{L}$	AH is the Albertson index, and L is the number of links	[49]
8	EH	Estrada heterogeneity index	$EH = \frac{\sum_{i,j\in L}(k_i^{-1/2}-k_j^{-1/2})^2}{N-2\sqrt{N-1}}$	i, j is the link connecting nodes i and j, k_i is the degree of the node i and k_j is the degree of the node j, L is the network link set, and N is the node number	[30]
9	A	Network assortativity	$A = \frac{1}{\sigma_q^2}\sum_{j,k\in N} jk(e_{jk} - q_j q_k)$	σ_q is the standard deviation of the excess degree distribution, e_{jk} is the fraction of links connecting nodes of degree j and k, and q_j and q_k are the excess degree of nodes of degrees j and k, respectively	[39]
10	$\bar{d}$	Average node distance	$\bar{d} = \frac{1}{N(N-1)}\sum_{i,j\in N, i\neq j} d_{ij}$	d_{ij} is the distance between nodes i and j, and N is the node number	[41]
11	Φ	Network eccentricity	$\Phi = \frac{1}{N}\sum_{i=1}^{i=N}\varepsilon(i)$	$\varepsilon(i)$ is the eccentricity of the node i, and N is the node number	[41]
12	D	Network diameter	$D = \max_{i,j\in N, i\neq j}(d_{ij})$	d_{ij} is the distance between i and j, and N the node number	[41]
13	π	Network radius	$\pi = \min_{i\in G}(\varepsilon(i))$	$\varepsilon(i)$ is the eccentricity of the node i	[41]
14	Eff	Network efficiency	$Eff = \frac{1}{N(N-1)}\sum_{i\neq j}\frac{1}{d_{ij}}$	d_{ij} is the distance between node i and node j, and N is the node number	[52]
15	T	Average node transitivity	$T = \frac{1}{N}\sum_{i=1}^{i=N}\tau_i$	τ_i is the transitivity of the node i, and N is the node number	[3]
16	B	Average node betweenness	$B = \frac{1}{N}\sum_{i=1}^{i=N} g(i)$	N is the number of nodes and $g(i)$ the betweenness of the node i	[53]
17	nB	Average normalized node betweenness	$nB = \frac{1}{N}\sum_{i=1}^{i=N} {}^{n}g(i)$	N is the number of nodes, and ${}^{n}g(i)$ is the normalized betweenness of the node i	[53]
18	Clo	Average node closeness	$Clo = \frac{1}{N}\sum_{i=1}^{i=N} C_i$	C_i is the closeness of the node i, and N is the node number	[54]

Forecasting SIR epidemic spreading

τ_{15} Time steps for 15% of infected nodes



Best Structural Predictor

The average node distance $\bar{d}$

The mean number of links among node pairs

τ_{15} strong positive linear relationship $\bar{d} \sim \tau_{15}$

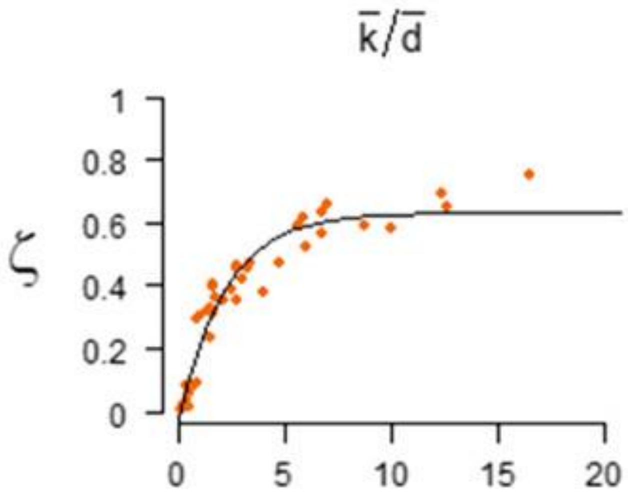
Higher $\bar{d}$ → higher the time to infect the 15% of nodes

Bellingeri, Bevacqua, Turchetto, Scotognella, Alfieri, Nguyen, Le, Nguyen , Cassi (2022)
Network structure indexes to forecast epidemic spreading in real-world complex networks. Front. Phys.

Forecasting SIR epidemic spreading

ζ Infected peak ⟶ maximum nodes number concurrently infected

Best Structural Predictor



$\frac{\bar{k}}{\bar{d}}$ ratio

the average node degree $\bar{k}$ / the average node distance $\bar{d}$

ζ follows a saturating function of $\frac{\bar{k}}{\bar{d}}$

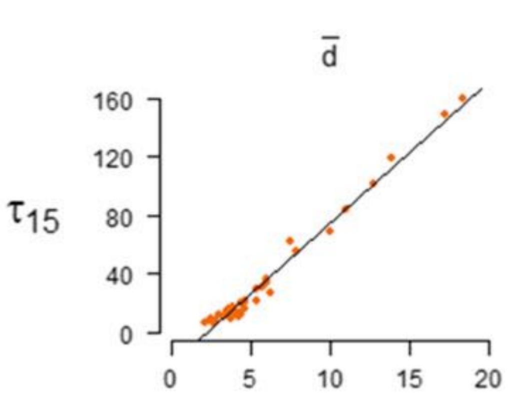
Higher $\bar{k}$ ⟶ Higher infected peak

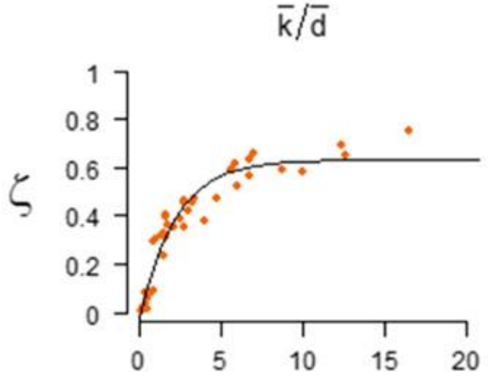
Higher $\bar{d}$ ⟶ Lower infected peak

Bellingeri, Bevacqua, Turchetto, Scotognella, Alfieri, Nguyen, Le, Nguyen , Cassi (2022)
Network structure indexes to forecast epidemic spreading in real-world complex networks. Front. Phys.

UNIVERSITÀ DI PARMA ecosister Ministero degli Affari Esteri e della Cooperazione Internazionale NETWORKS UNIT

Forecasting SIR epidemic spreading





Results

To predict spreading

The **distance among nodes** is more important than focusing on their connectivity level

Non-pharmaceutical interventions (NPIs) to halt Covid reduce social interactions → to decrease the number of the network links

More effective NPIs → space out the nodes → increasing node distance in the network

Bellingeri, Bevacqua, Turchetto, Scotognella, Alfieri, Nguyen, Le, Nguyen , Cassi (2022)
Network structure indexes to forecast epidemic spreading in real-world complex networks. Front. Phys.

Thanks

michele.bellingeri@unipr.it

https://www.networks.unipr.it