(Intro to)

Social Network Analysis

Letizia Milli
University of Pisa
letizia.milli@unipi.it
@LetiziaMilli

Katherine Abramski
University of Pisa
katherine.abramski@phd.unipi.it

Giulio Rossetti
CNR-ISTI
giulio.rossetti@isti.cnr.it
@GiulioRossetti

Slides available at: https://t.ly/_bWo

Chapter 0

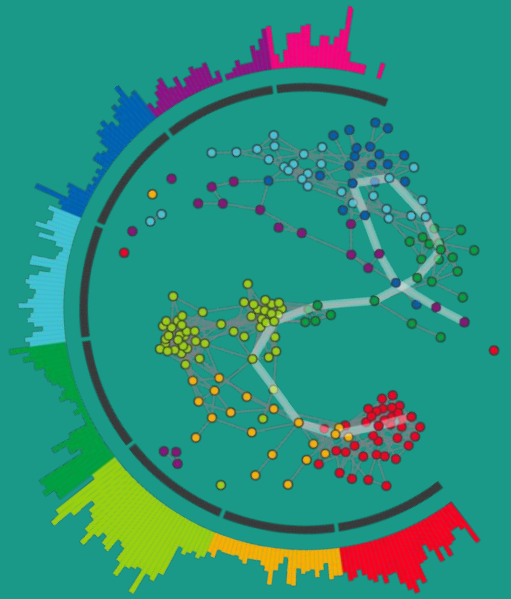
Why should we care about Complex Networks?

Behind each complex system
there is a network,
that defines the interactions
between the components.

Suggested Reading
Complexity Explained
https://complexityexplained.github.io/

The role of networks

Behind each system studied in complexity there is an intricate wiring diagram, or a **network**, that defines the interactions between the component.



We will never understand **complex system** unless we map out and understand the networks behind them.

Examples of

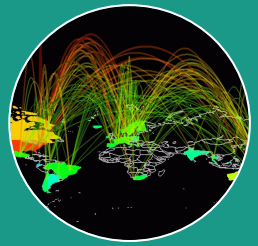
Real world Networks

Type: Social
Nodes: Individuals
Links: Social relationship

NetSci
network science society

Type: Scientific Collaborations
Nodes: Researchers
Links: Co-Authorships

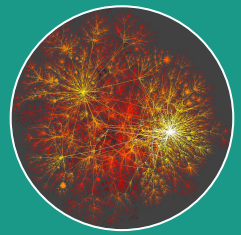
Type: Actor connectivity
Nodes: Actors
Links: Cast jointly



Type: Communication
Nodes: Phones, Airports..
Links: Phone calls, Flights..

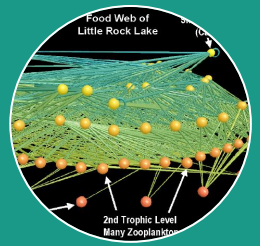
Examples of

Real world Networks (cont'd)



Type: Technological
Nodes: PC, Routers
Links: Physical lines

Type: Scientific Citation
Nodes: Papers
Links: Citations



Type: Biological
Nodes: Species
Links: Trophic interactions

Type: Mobility
Nodes: Individuals, Cars...
Links: Co-Location...

The Emergence of Network Science

The (urgent) need to understand complexity

Despite the challenges complex systems offer us, we cannot afford to not address their behavior, a view increasingly shared both by scientists and policy makers.

Networks are not only essential for this journey, but during the past decade some of the most important advances towards understanding complexity were provided in context of network theory.

Data Availability

1990 C. elegans neural wiring diagram
1998 - Movie Actor Network
1998 - Citation Networks
1999 -World Wide Web
2000 - Metabolic Networks
2001 - PPI network
2008 - OSNs

Universality

The architecture of networks emerging in various domains of science, nature, and technology are more similar to each other than one would have expected.

The Tools of

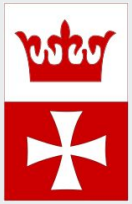
Modern Network Theory

1. Graph Theory
2. Social Network Theory
3. Statistical Physics
4. Computer Science
5. Biology
6. Statistics

Chapter 1

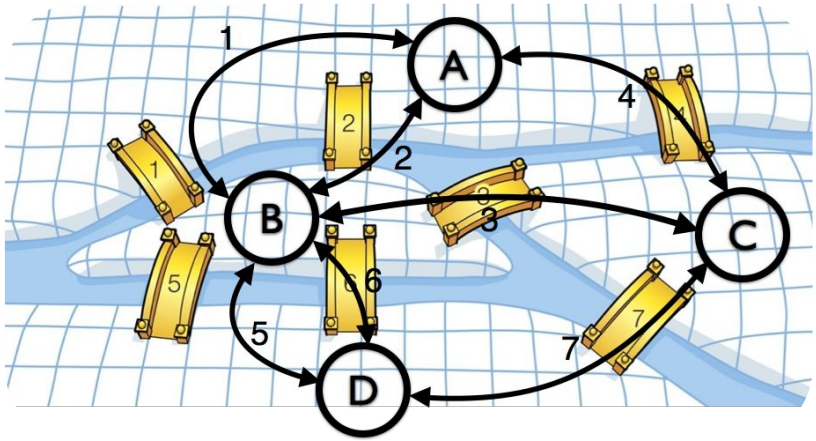
Networks: Basic Measures

The Bridges of Konigsberg



Can one walk across the seven bridges and never cross the same bridge twice?

Famous Konigsberg Citizens
Immanuel Kant (philosopher, 1724-1804)



Euler's theorem (1735)

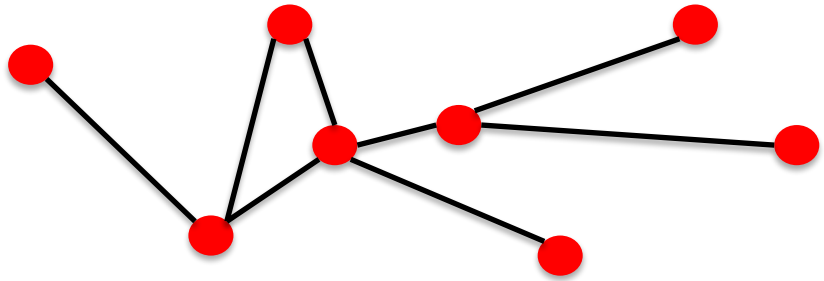
- If a graph has more than two nodes of odd degree, there is no path/cycle that crosses each bridge exactly once.
- If a graph is connected and has no odd degree nodes, it has at least one path.

Components of a Complex System

Networks or Graphs?

Network *<nodes, links>*
refers to real systems
(www, social network, metabolic network)

Graph *<vertices, edges>*
mathematical representation of a network
(web graph, social graph)



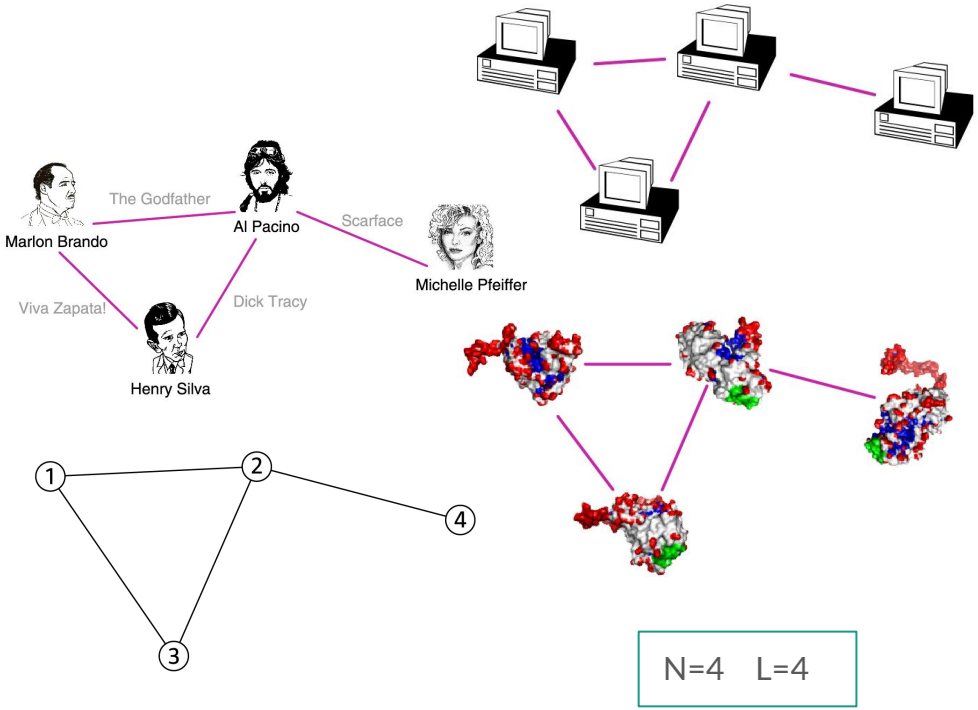
		Symbol
Components	nodes, vertices	N
Interactions	edges, links	L
System	network, graph	(N,L)

A Common Language

The choice of the proper network representation determines our ability to use network theory successfully.

In some cases there is a unique, unambiguous representation. In other cases, the representation is by no means unique.

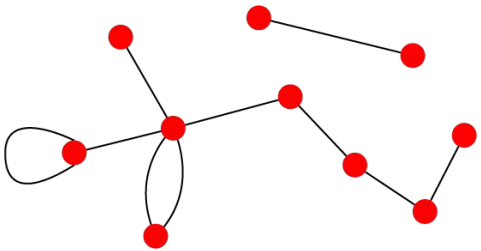
The way we assign the links between a group of individuals will determine the nature of the question we can study.



Directedness

Undirected graphs

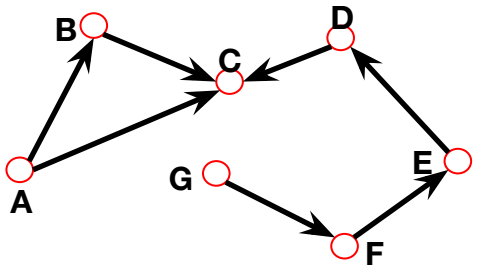
Links: undirected (*symmetrical*)



Examples of Undirected links
- Co-authorship links
- Actor network
- Protein interactions

Directed graphs (DiGraphs)

Links: directed (*arcs*).

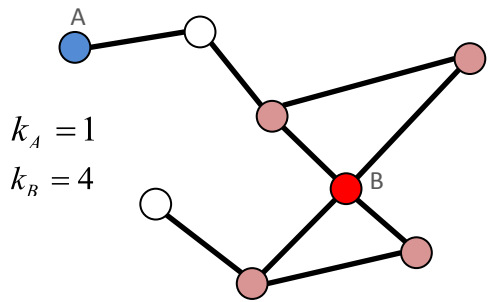


Example of Directed links
- URLs on the www
- Phone calls
- Metabolic reactions

Node Degree

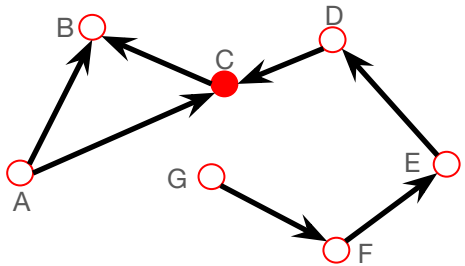
Undirected graphs

the number of links connected to the node



Directed graphs (DiGraphs)

we can define an in-degree and out-degree.
The (total) degree is the sum of in- and out-degree.



Source: a node with $k^{in} = 0$;

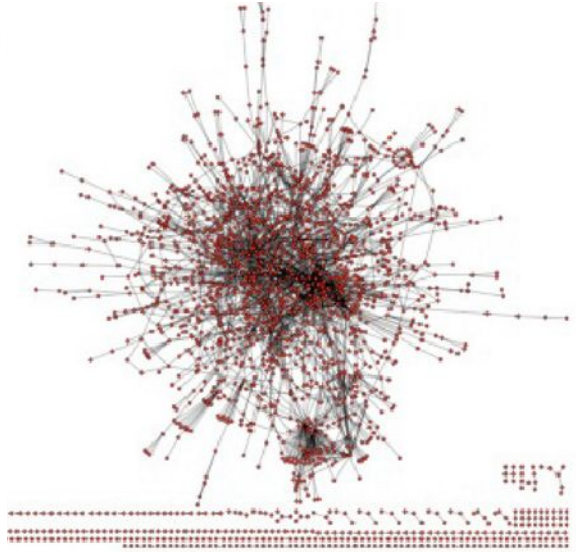
Sink: a node with $k^{out} = 0$.

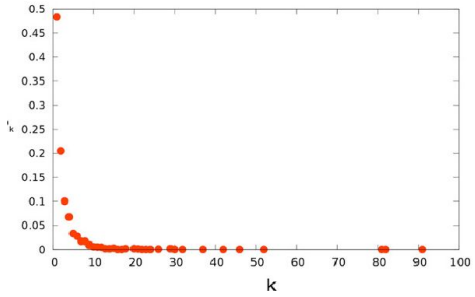
Degree Distribution

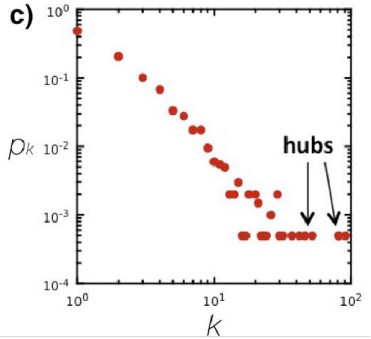
P(k): probability that a randomly chosen node has degree *k*

N_k = # nodes with degree k

$P(k) = N_k / N$ ➔ plot







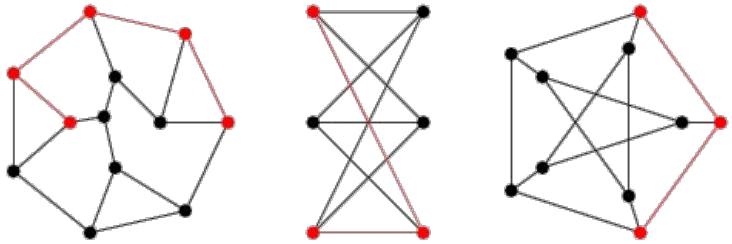
Paths

A *path is* a sequence of nodes in which each node is adjacent to the next one

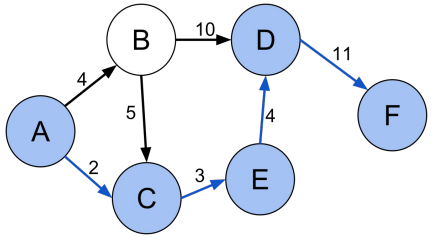
$P_{i0,in}$ of length *n* between nodes i_0 and i_n is an ordered collection of *n+1* nodes and *n* links

$$P_n = \{i_0, i_1, i_2, ..., i_n\}$$

$$P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), ..., (i_{n-1}, i_n)\}$$



Examples of phats in an undirected graph.

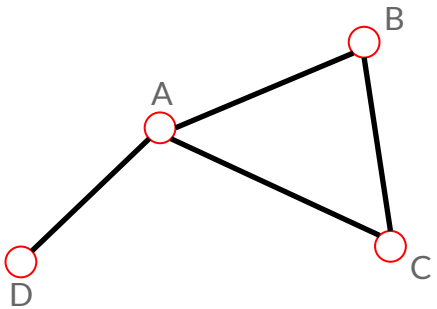


In a directed graph, the path can follow only the direction of an arrow.

Distance in a Graph

Undirected graphs

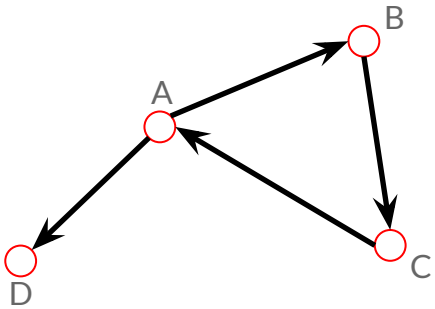
The *distance* (*shortest path*, *geodesic path*) between two nodes is defined as the number of edges along the shortest path connecting them.



*If the two nodes are disconnected, the distance is infinity.

Directed graphs (DiGraphs)

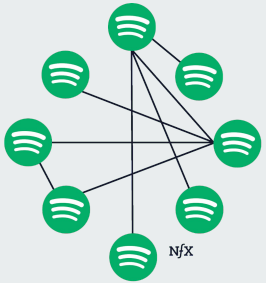
Each path needs to follow the direction of the arrows.



Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

Network Density

Ratio of existing edges over possible ones.

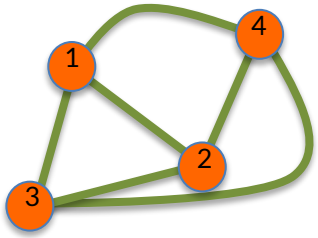


Low Density

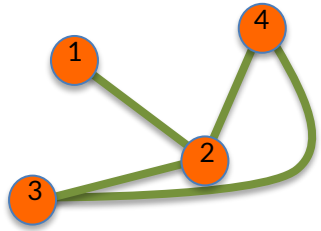
High Density

$$d(G) = \frac{L}{L_{max}}$$

Examples



Complete graph
d(G) = 1



Sparse graph
d(G) = 2/3

Chapter 2

Centrality & Tie Strength

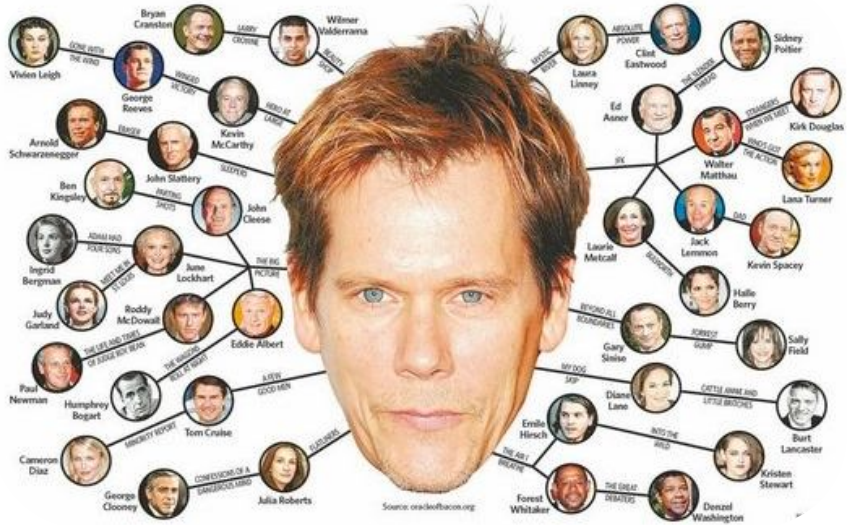
How important is a node in a network?

We can measure nodes importance using so-called centrality.

Bad term:
nothing to do with being central in general

Usage:

- Some centralities have straightforward interpretation
- Centralities can be used as node features for machine learning on graph



https://oracleofbacon.org/

Where are you?

It is always possible, once *fixed a context*, to measure our distance from a “focal” node.

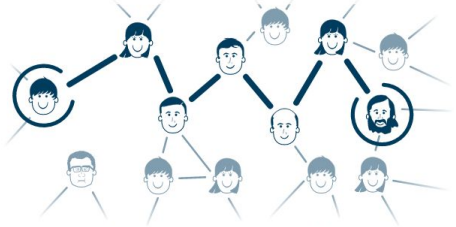
For instance:

Movie Stars:
- Bacon number

Researchers:
- Erdos number

Are such “focal” nodes really different from the others?

Co-authorship distance computation



Find the path between two authors:

Paul Erdős | Giulio Rossetti

Paul Erdős
co-authored 2 papers with Shlomo Moran
co-authored 7 papers with Ronny Lempel
co-authored 4 papers with Fabrizio Silvestri
co-authored 2 papers with **Giulio Rossetti**
distance = 4

Name	Erdős number	Bacon number	Erdős–Bacon number
Daniel Kleitman	1	2	3[4]
Bruce Reznick	1	2	3[69]
Albert M. Chan	3[21][22][23]	1[24]	4
Nicholas Metropolis	2[10]	2[68]	4
Steven Strogatz	3[79][80][81]	1[a][c][82]	4[a][c]
Robert J. Marks II	3[44][45][46]	2[61][62][63]	5
Tom Porter	3 (in two ways)[6][7]	2[a][c][8][9]	5[a][c]
Richard Thaler	3[86][87][88]	2[89][a][90]	5
Doron Zeilberger	2[93][35]	3[a][94][95][96]	5[a][c]
Misha Collins	4[25][26][27][28]	2[29][30]	6
William A. Dembski	4[44][45][46][47]	2[a][48]	6[a]
Richard Feynman	3	3	6[10]
Ken Goldberg	3[49][50][51]	3[52][53][54]	6
Stephen Hawking	4	2[a]	6[20]

00	Degree	How many friends do you have?
01	Eigenvector	Are you connected to important nodes?
02	PageRank	How many important interactions do you have?
03	Closeness	What's your average distance w.r.t. the rest of the network?
04	Harmonic	What's your harmonic average distance w.r.t. the rest of the network?
05	Betwenness	How much do you help the network to stay connected?

Connectivity-based centralities

Geometric centralities

Each centrality measures is a proxy of an underlying network process.

If such a process is irrelevant for the network than the centrality measure makes no sense

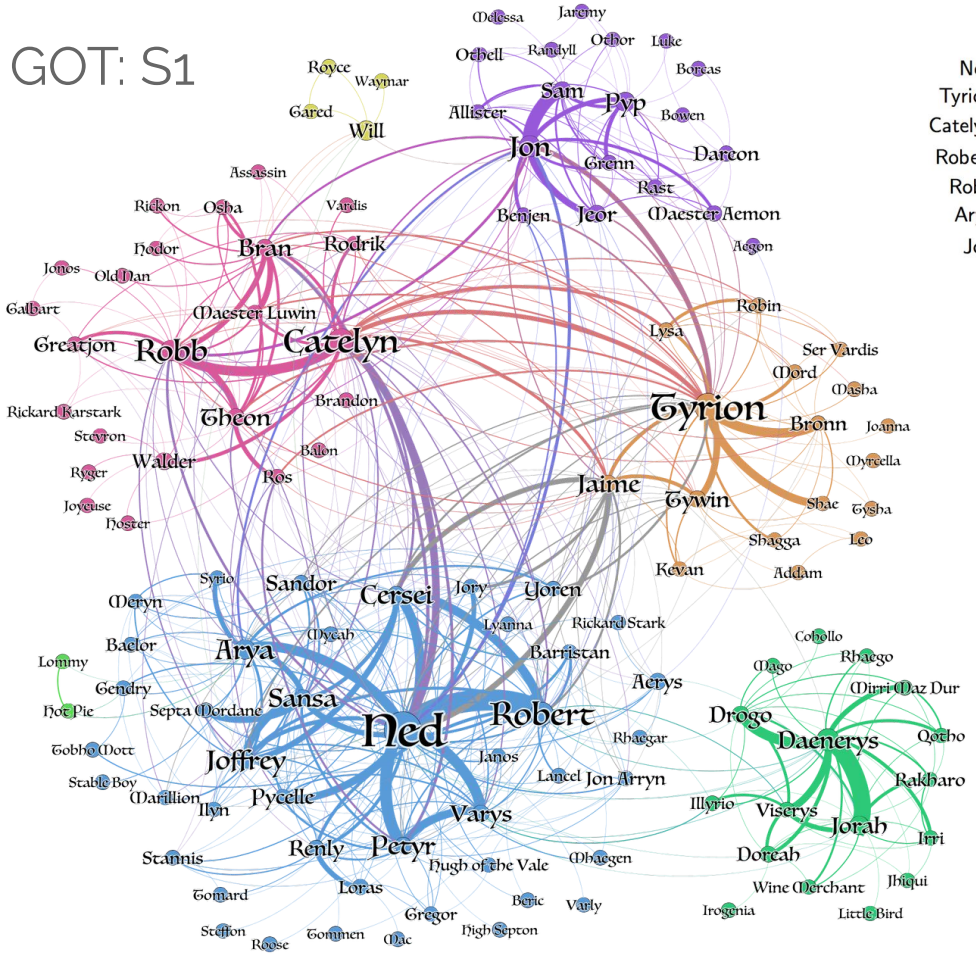
- E.g. If information does not spread through shortest paths, betweenness centrality is irrelevant

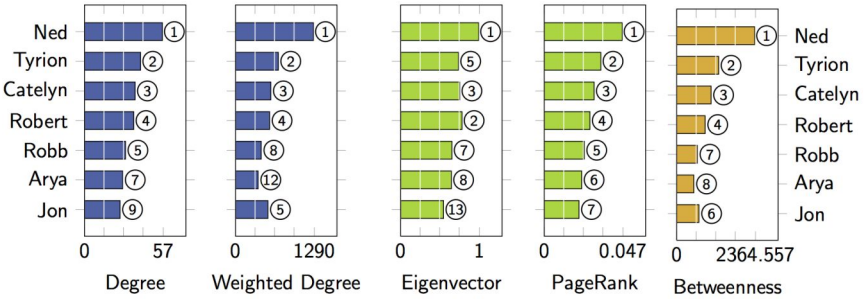
Centrality measures should be used with caution (a) for exploratory purposes and (b) for characterisation

Understanding Centralities

Data and Viz @mathbeveridge
www.networkofthrones.wordpress.com

GOT: S1





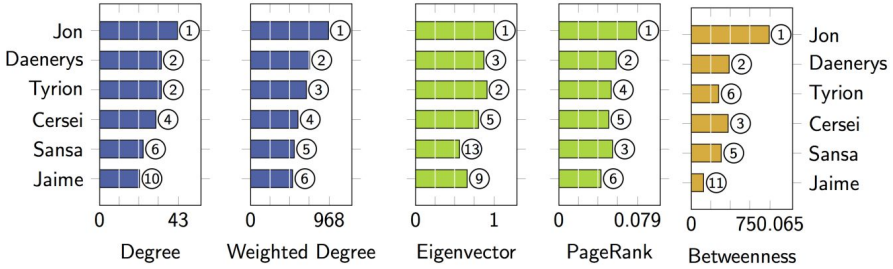
Node Label: PageRank
Node Size: Betweenness Centrality

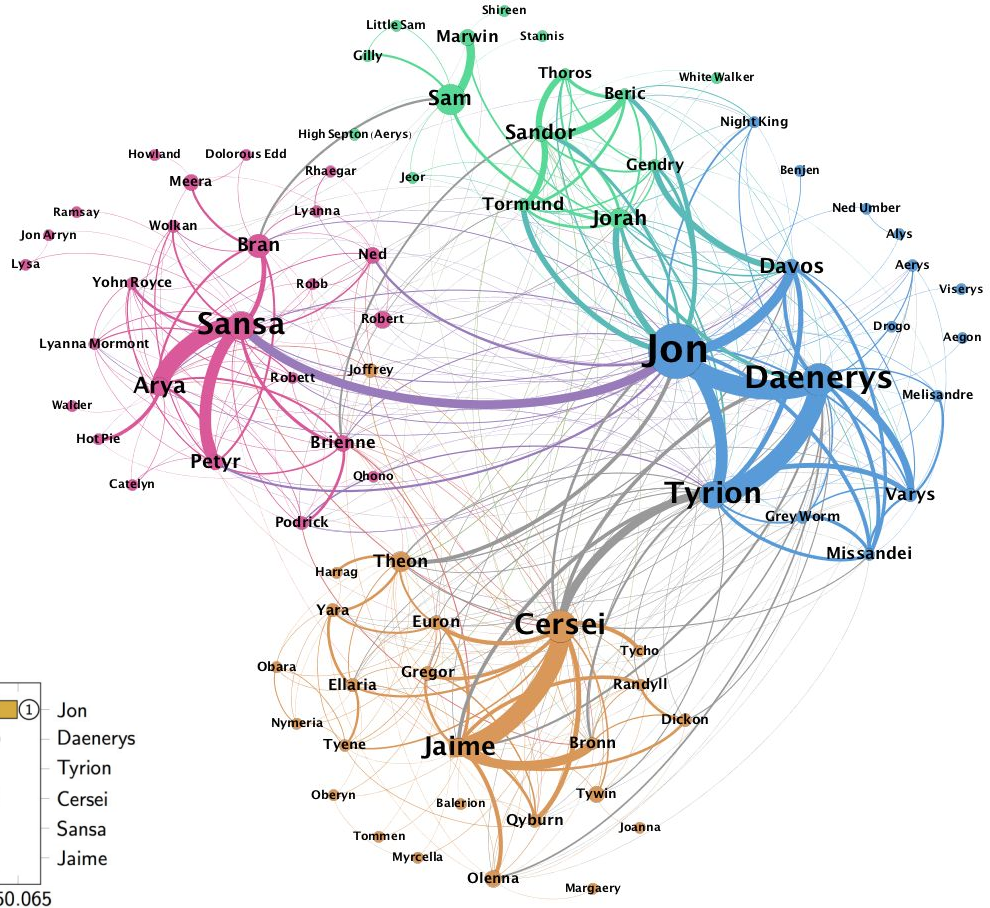
Edge Size: #interactions
Colors: Community (with Louvain)

GOT: S7

Node Label: PageRank
Node Size: Betweenness Centrality

Edge Size: #interactions
Colors: Community (with Louvain)





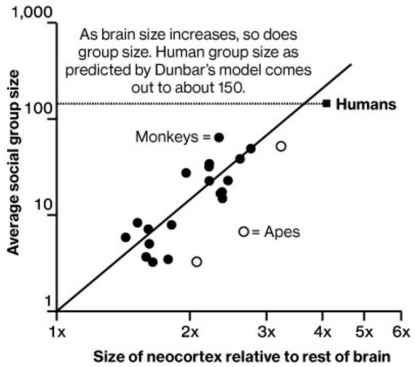
The Strength of Weak Ties

How many friends does one person needs?

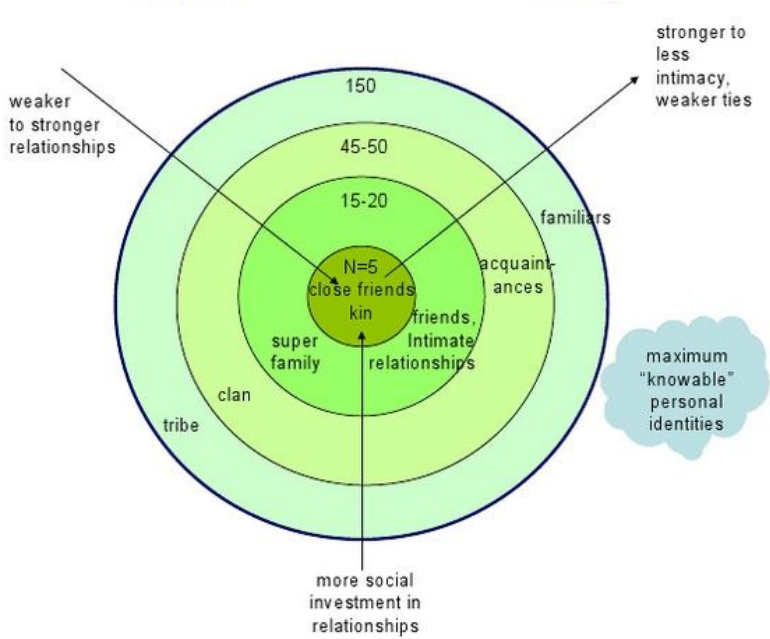
Not all ties in a social graph are the same

Dunbar's Number (Sociological Theory)

a suggested cognitive limit to the number of people with whom one can maintain stable social relationships



Considering the average human brain size and extrapolating from the results of primates, humans can comfortably maintain 150 stable relationships



In Dunbar's own words:

"the number of people you would not feel embarrassed about joining uninvited for a drink if you happened to bump into them in a bar"

Dunbar, Robin IM. Neocortex size as a constraint on group size in primates. Journal of human evolution 22.6 (1992): 469-493.

Dunbar, Robin. How many friends does one person need?: Dunbar's number and other evolutionary quirks. Faber & Faber, 2010.

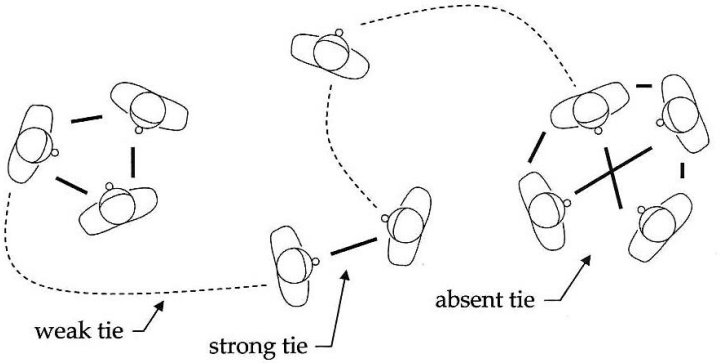
The strength of weak ties

Mark S. **Granovetter**, 1973

- (PhD Thesis)
  "How people get to know about new jobs?"
- Answer: Through personal contacts

Unexpected result:

Often acquaintances, **not** close friends… but *why*?



How to measure tie strength?

Granovetter's **dimensions of tie strength**:

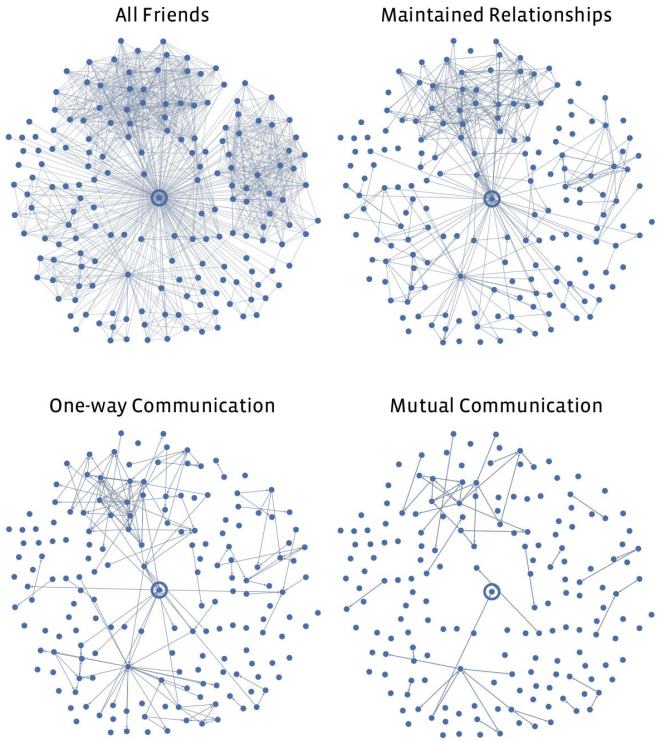
- the amount of time spent interacting with someone,
- the level of intimacy,
- the level of emotional intensity,
- and the level of reciprocity.

Granovetter, Mark S. "The strength of weak ties." Social networks. Academic Press, 1977. 347-367.

Ties Strength on Facebook

Different types of connections

- Mutual communication:
  both user sent messages eachother
- One-way communication:
  user messages where not reciprocated
- Maintained relationship:
  user clicked on content produced by his friend (no communication)

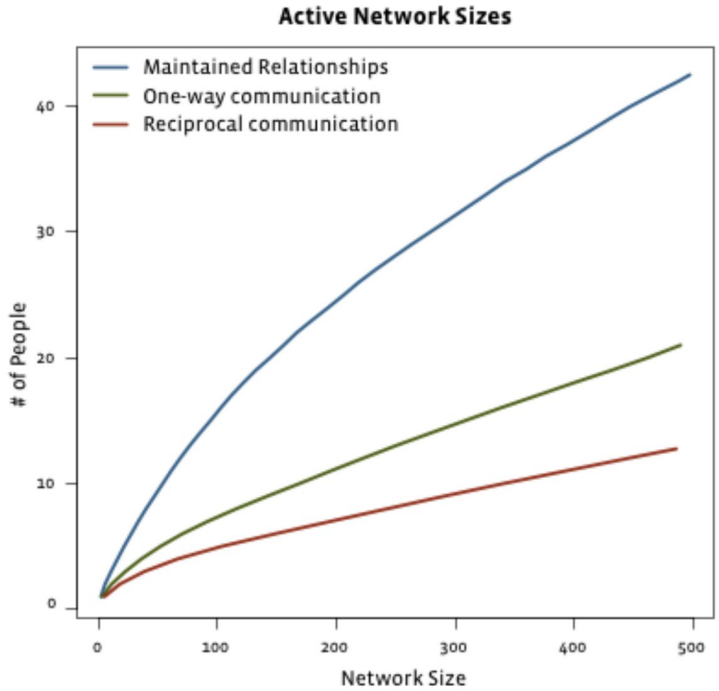


Cameron Marlow, Lee Byron, Tom Lento, and Itamar Rosenn. Maintained relationships on Facebook, 2009. http://overstated. net/2009/03/09/maintainedrelationships-on-facebook

Does tie strength affect network size?

Tie strength allows to:

- discriminate different type of contacts,
- categorize them by the involvement required to nurture them



Chapter 3

Community Discovery & Homophily

But…what is it exactly a community?

Unfortunately does not exist a universally shared definition of what a community is…

A *general* idea is that a community should represent:

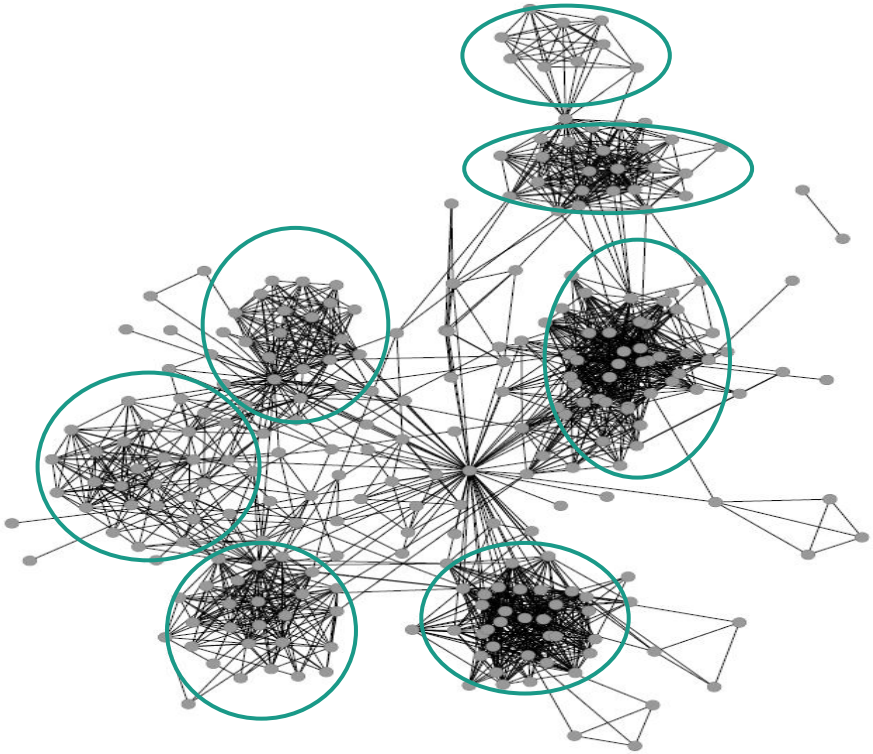
> *"A set of entities where each entity is closer, in the network sense, to the other entities within the community than to the entities outside it."*

or, equivalently

> *"A set of nodes more tightly connected within each other than with nodes belonging to other sets."*

Communities in Complex Networks

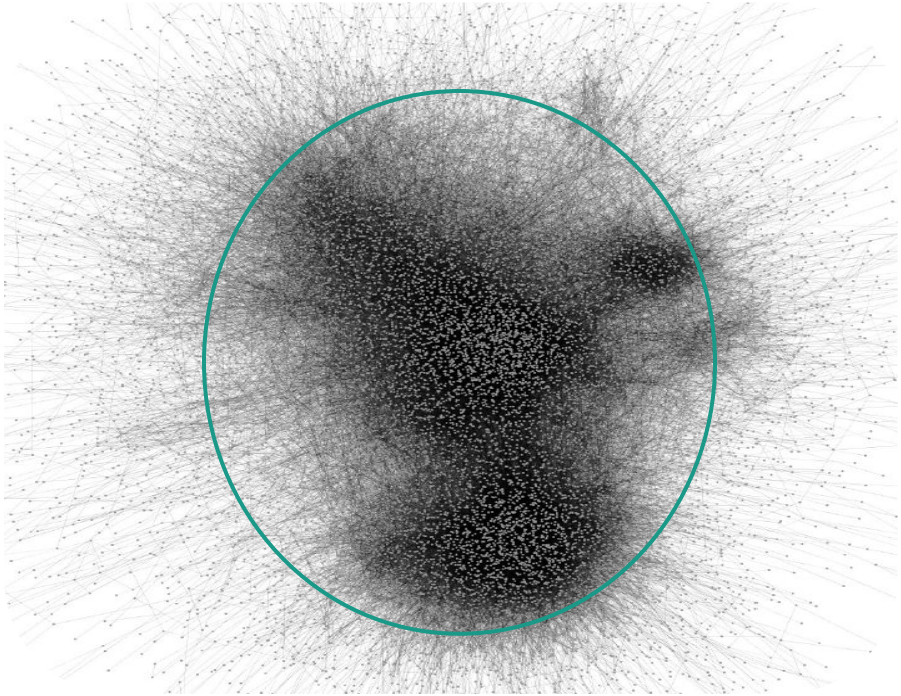
In simple, small, networks it is easy identify them by looking at the structure…



Real world networks? Too complex for visual analysis

We can't easily identify (e.g., visually) different communities

We need automated procedures!



Do Birds of a Feather Flock Together?

Homophilic behaviors in complex networks

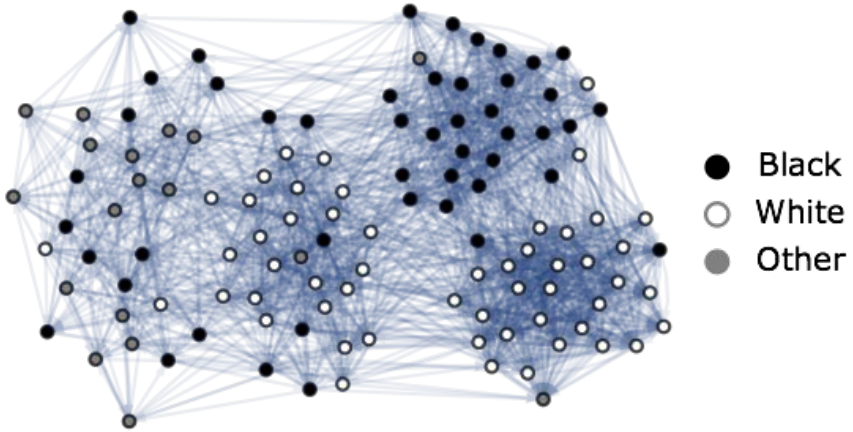
Homophily

Property of (social) networks that nodes of the same attitude tends to be connected with a higher probability than expected

- It appears as correlation between vertex properties of x(i) and x(j) if (i,j) ∈ E

Disassortative mixing:

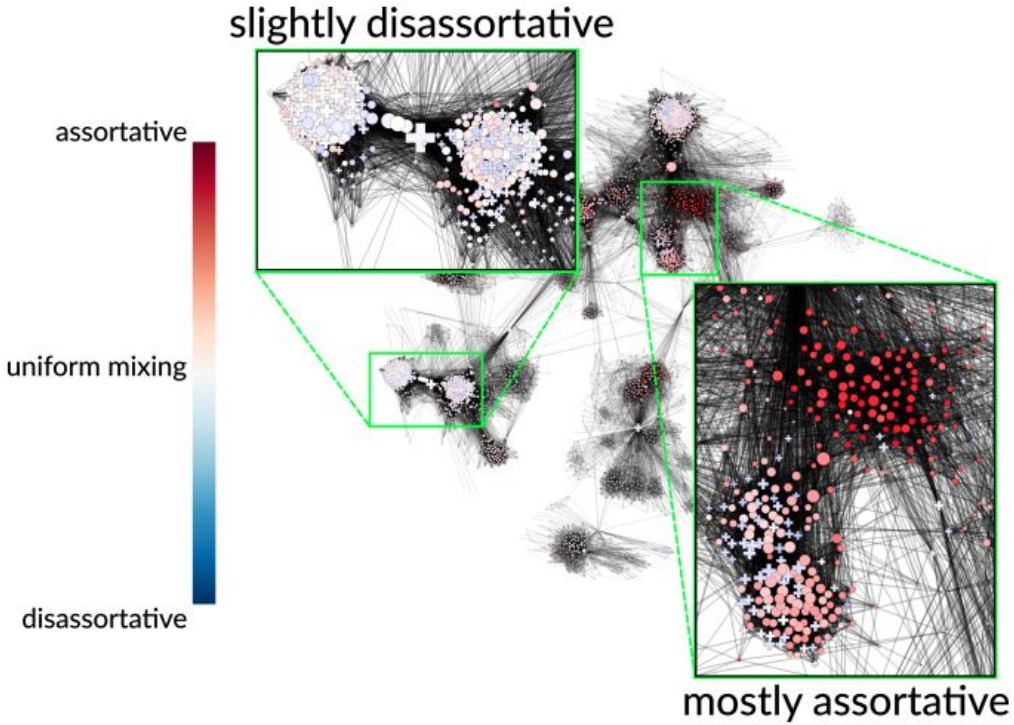
Contrary of homophily: dissimilar nodes tend to be connected
(e.g., sexual networks, predator-prey)



Examples of Vertex properties

age, gender, nationality,
political beliefs, socioeconomic status,
obesity, ...

Homophily can be a link creation mechanism or consequence of social influence (and it is difficult to distinguish)

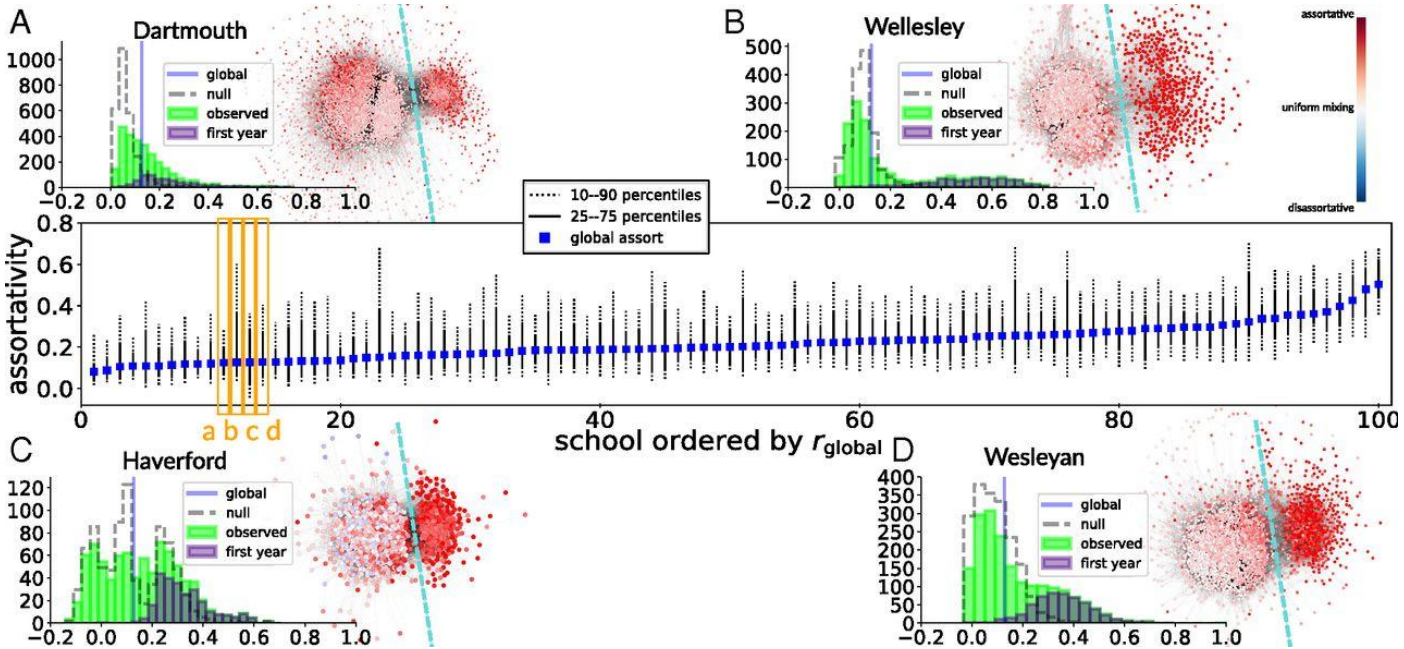


Local assortativity of gender in a sample of Facebook friendships (McAuley and Leskovec 2012).

Different regions of the graph exhibit strikingly different patterns, suggesting that a single variable, e.g. global assortativity (Newman's), would provide a poor description of the system.

Limits of a global assortativity score

Evaluation on real data



Facebook100

Distribution of local assortativity for the "dorm" node feature

Chapter 4

What can we do with networks?

Migration **S**tudies

Brain Drain, Turkey Crisis…

(H2020 project, HummingBird)

Where (and why) scholars move?

Bloomberg

Economics

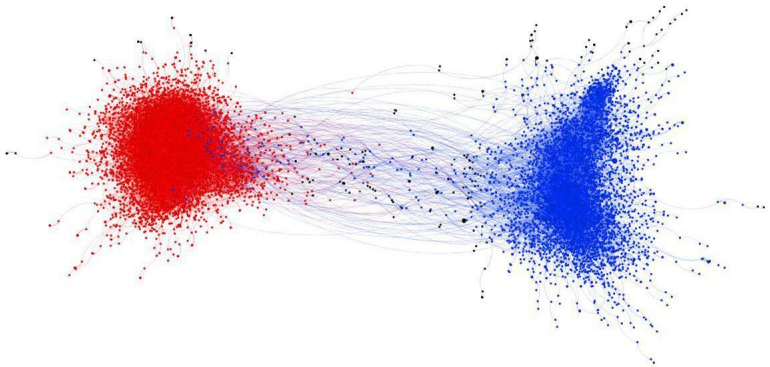
Europe Risks New Migration Meltdown as Erdogan Opens Floodgates

By Cagan Koc, Selcan Hacaoglu, and Nikos Chrysoloras

29 febbraio 2020, 12:37 CET *Updated on 1 marzo 2020, 09:21 CET*

Polluted **I**nformation **E**nvs

Echo Chambers, Social Pressure & Fake News

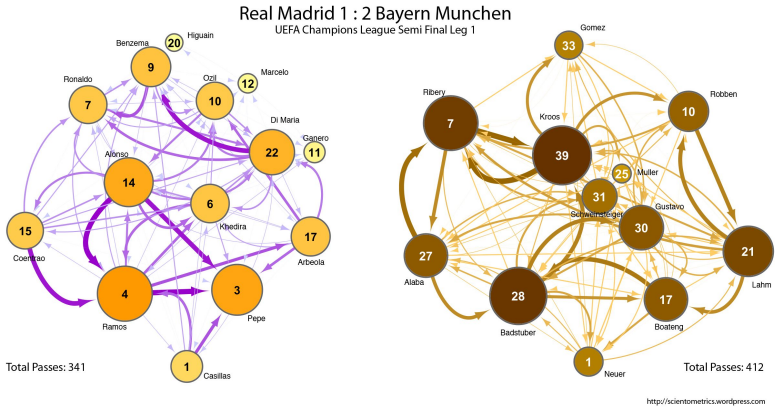


Identifying, modeling & dismantle echo chambers
Modeling Opinion Dynamics of Fake news perception
...

Can we model how fake news spread?
Is peer pressure playing a role?

Science of **S**uccess

Sport, Hit-Savyy, Start-ups,...



Modeling Sport with networks

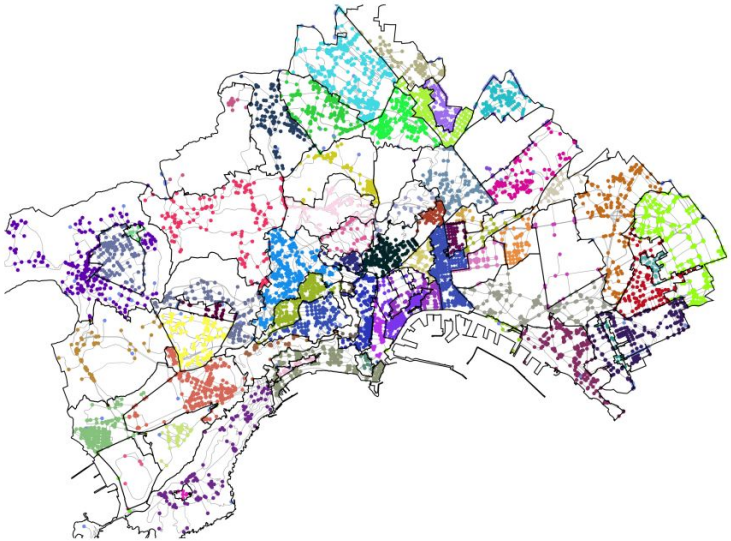
Revising Rogers Innovators and their role...

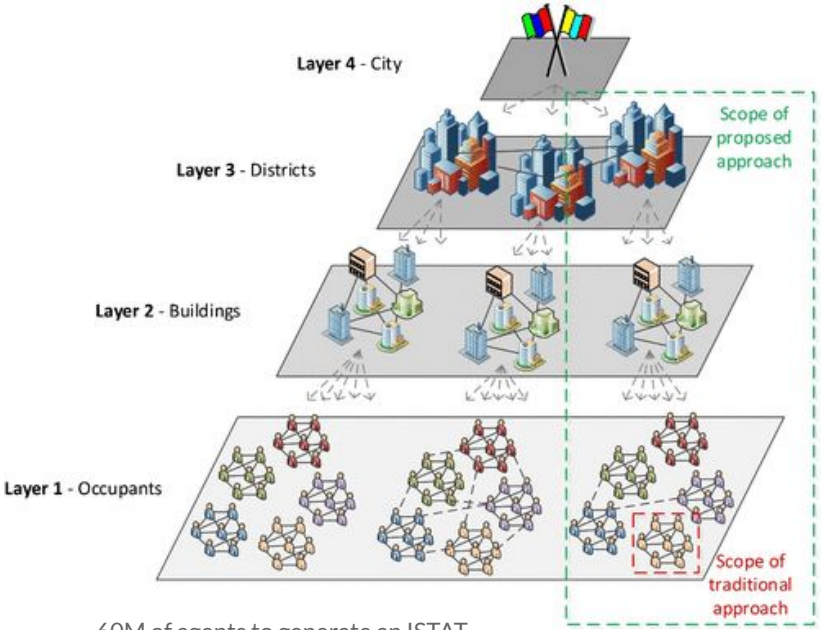
Can we predict start-ups success?

(one active project EPO, one H2020 Proposal, OrAIcle)

Others

Mobility, Urban planning,...





60M of agents to generate an ISTAT data-driven social proxy for Italy

Chapter 5

Let's play!

Viz: https://gephi.org/gephi-lite/

Data: https://networkrepository.com/index.php

Python: http://sobigdata.eu/