Complex Networks

Enrico Denti

`enrico.denti@unibo.it`

Dipartimento di Informatica – Scienza e Ingegneria (DISI)

ALMA MATER STUDIORUM – Università di Bologna

14 dicembre 2012

Outline

- **Complex Networks**
  - Definition
  - Special cases: scale-free networks, small-world networks
  - Properties and indicators
  - Example: centrality indicators in a simple network
- **Case Studies**
  - author/co-authorship network
  - citation/collaboration networks
  - critical road network areas
  - social structure of organ transplant
  - emerging news networks
  - overlapping communities
  - …

Complex Networks

- **Complex networks**
  - *graphs (networks) with non-trivial topological features*
  - that map well onto many social, biological, technological networks.. …and more

- **"Non trivial features"…?**
  - connection between elements are neither regular, nor random
  - degree distribution, clustering coefficient, community/hierarchical structure
    - ▸ contrast: most mathematical models studied in the past (e.g. lattices) do *not* show these features

- **Special cases**
  - *scale-free networks* and *small-world networks* are special cases
    - ▸ *scale-free networks:* power-law degree distribution
    - ▸ *small-world networks:* short path lengths, high clustering

■ **Scale-free network**

- A special case of Complex Networks, where the ***degree distribution***, i.e. the probability that a node selected uniformly at random has a certain number of links (*degree*), follows a ***power law***:

$$p(k) = c\, k^{-\lambda}$$

- For most real networks, experimentally $2 < \lambda < 3$

  ▸ NB: most of reportedly "power laws" fail a rigorous statistical testing: still, their features are very different from network whose edges are independent and random (i.e. follow a Poisson distribution)

- Some vertices (***hubs***) have a much higher degree than the average

  ▸ but notice, there is no inherent threshold for a node to be viewed as a hub, otherwise the network would not be scale-free

  ▸ in contrast, network with a single well-defined scale are somewhat similar to a lattice in that every node has (roughly) the same degree

  ▸ Examples: World Wide Web, the network of Autonomous systems (ASs), some network of Internet routers, protein interaction networks, email ..

Complex Networks → Small world

- ## Small world network

  - A special case of Complex Networks, named after analogy with the *small-world phenomenon*

    - ▸ Origin: the six degrees of separation (Karinthy, 1929; Milgranm, 1967): two arbitrary people are connected by only six degrees of separation, i.e. the diameter of the graph of social connections is not larger than 6

  - In small-world netwoks (1998), a single parameter smoothly interpolates between a random graph to a lattice

    - ▸ the addition of a small number of long-range links is enough to turn a regular graph (in which the diameter is proportional to the size of the network) into a "small world" (in which the average number of edges between any two vertices is very small) while the clustering coefficient stays large

Complex Networks and Social Networks

- **Why Complex Networks are well suited for Social Networks?**

  - because they provide a *visual* and *mathematical analysis* of ***human-influenced relationships***

  - because the *social environment* can be expressed as ***patterns /regularities*** in relationships among the interesting units

- **How are they built?**

  - each network is represented by a graph

    - whose <u>*nodes*</u> represent individuals, organizations, information

    - whose <u>*edges*</u> represent some kind of relations (any kind: financial exchange, friends, web links, author/co-author, etc)

    - if the nodes represent people, the existence of an edge means that these people know each other in some way

- **How are they characterised?**

  - complex networks can be characterised both by ***global*** (network-level) ***properties*** and ***local*** (node-level) ***properties***

Network properties and Indicators

- **Global metrics** describe the network global characteristics
  - graph diameter
  - mean node distance
  - number of components
  - clusters
  - possible existence of small-world phenomena

  …etc

- **Local metrics** describe individual properties of network nodes
  - node degree
  - node position in a cluster
  - several *Centrality measures*
    - mainly *degree* centrality, *closeness* centrality, *betweenness* centrality
  - recent research proposed new "combined" centrality measures
    - *degree-degree, degree-closeness, degree-betweenness centrality*

Centrality measures

- **Degree centrality:** the number of edges adjacent to the node (normalised by dividing by N-1, where N is the number of nodes)
    - measures "how many connections" each individual has towards others
    - indicates an actor's communication activity and popularity

- **Closeness centrality:** the inverse of the sum of distances of a node to others (also called "farness") (normalised by dividing by N-1 as above)
    - measures "how close" an individual is to any other, giving higher importance to nodes in "nearest" positions
    - the idea behind is that it is easier to get information from closer nodes..

- **Betweenness centrality:** the number of shortest paths that pass thru the given node (normalised to the total number of shortest paths between any pair of nodes, regardless of passing thru the given node)
    - measures "how often" that node is in the shortest path to another
    - indicates a node's potential control of communication, its brokerage behaviour

- **Eigenvector centrality:** assigns relative scores to nodes based on the principle that connections to high-scoring nodes contribute more to the score of the node than equal connections to low-scoring nodes
    - indicates the "importance" of a node in terms of its "contribution" to the network

Other indicators

- As mentioned before, *degree distribution* expresses the probability that a given node has k connections

  - in most real networks, that distribution is roughly a power law

 $$p(k) = c\,k^{-\lambda} \qquad \text{with } 2 < \lambda < 3$$

- Another relevant property is the *clustering coefficient*

  $$C_i = \frac{2m_i}{k_i(k_i-1)}$$

  where m_i is the number of links between the k_i neighbours of node i

  - measures the ratio in which nodes tend to "group together"

    ▸ for instance, in author/co-author network, it indicates how much a node's collaborator has written a paper with one of its other collaborators

  - <u>The clustering coefficient of the whole network</u> is just the average of all C_i over the number of nodes in the network

    ▸ usable to identify small-world networks, which are expected to have high clustering and short average path lengths
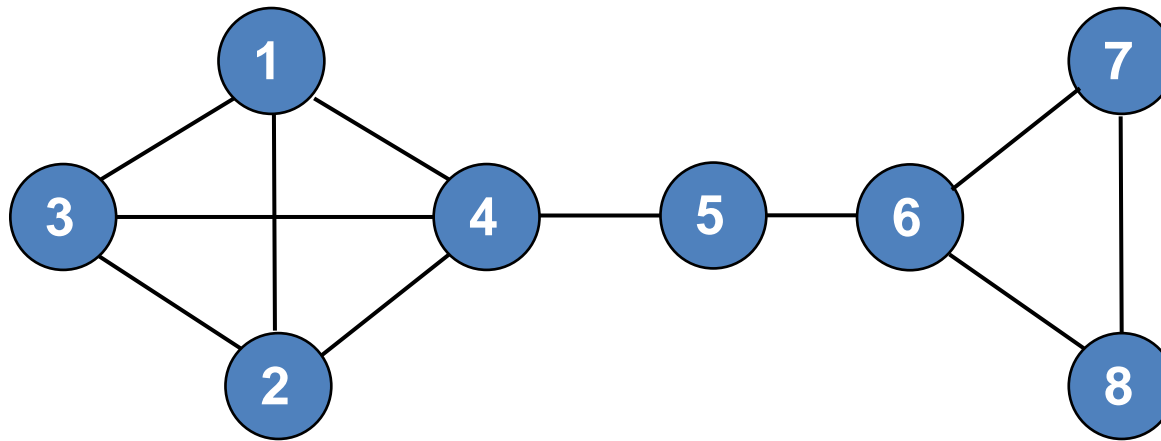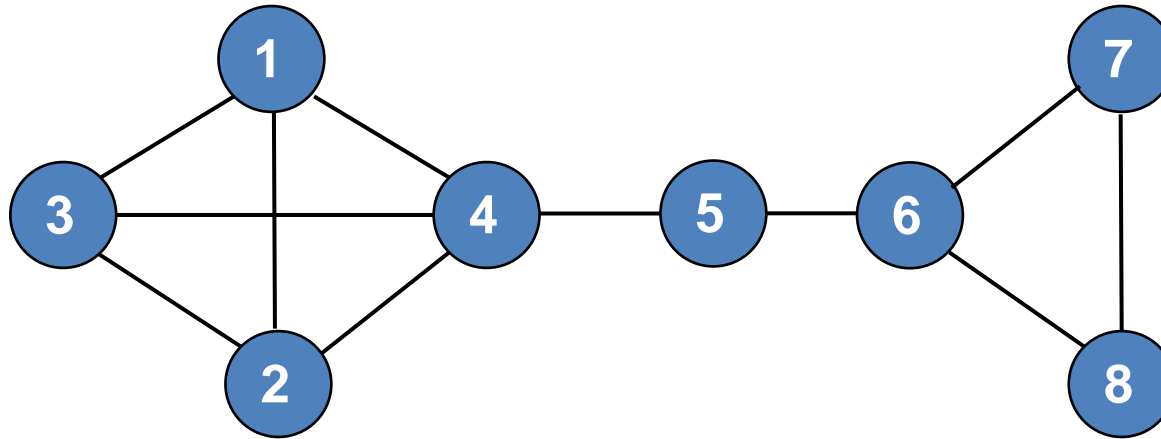
Node	Degree centrality	Closeness centrality	Betweenness centrality
1	3/7		
2	3/7		
3	3/7		
4	4/7		
5	2/7		
6	3/7		
7	2/7		
8	2/7		

Degree centrality: the number of edges adjacent to the node (normalised by dividing by N-1)

Node	Degree centrality	Closeness centrality	Betweenness centrality
1	3/7	7/(1+1+1+2+3+4+4) = 7/16	
2	3/7	7/(1+1+1+2+3+4+4) = 7/16	
3	3/7	7/(1+1+1+2+3+4+4) = 7/16	
4	4/7	7/(1+1+1+1+2+3+3) = 7/12	
5	2/7	7/(2+2+2+1+1+2+2) = 7/12	
6	3/7	7/(3+3+3+2+1+1+1) = 7/14	
7	2/7	7/(4+4+4+3+2+1+1) = 7/19	
8	2/7	7/(4+4+4+3+2+1+1) = 7/19	

Closeness centrality: the inverse of the sum of distances of a node to others (normalised by dividing by N-1)

Node	Degree centrality	Closeness centrality	Betweenness centrality
1	3/7	7/16	0 / 42
2			0 / 42
3			0 / 42
4			24 / 42
5			24 / 42
6			20 / 42
7			0 / 42
8	2/7	7/19	0 / 42

Betweenness centrality: the number of shortest paths passing thru the given node (normalised to the total number of shortest paths between any pair of nodes)
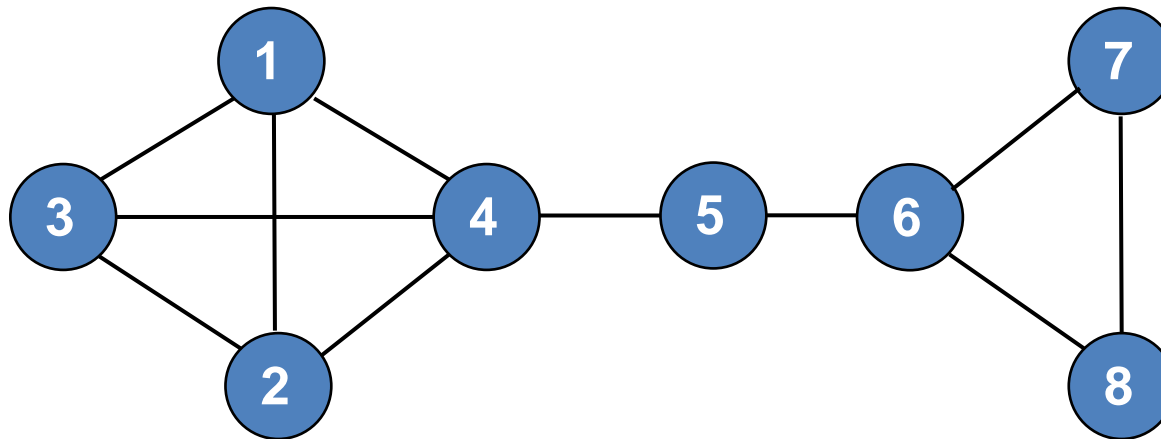
NOTE: the total number T of shortest paths between any pair of nodes is
$$T = (N-1)(N-2)$$
that is, equals to the number of pair of nodes not containing the given node.
In this case T=42

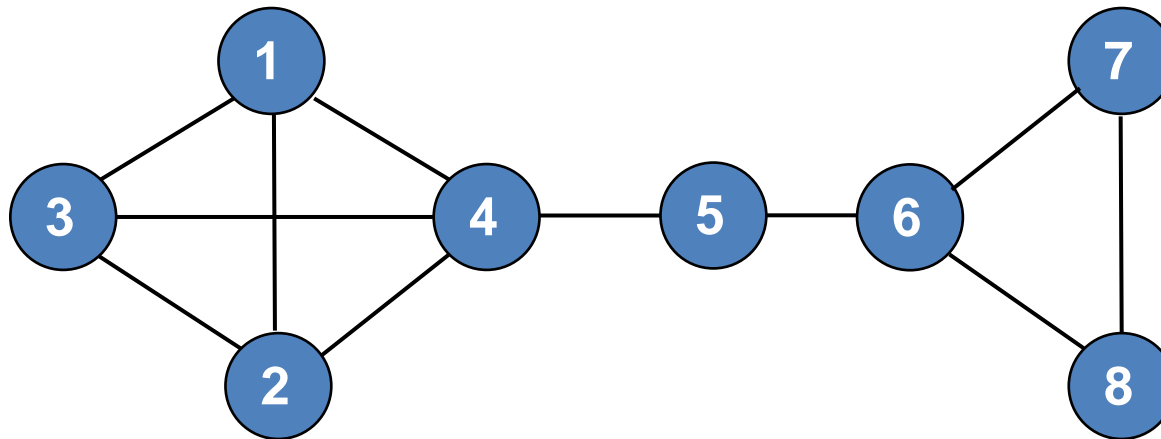| Node	Degree centrality	Closeness centrality	Betweenness centrality
1	0.429	0.438	0
	0.429	0.438	0
	0.429	0.438	0
	0.571	0.583	0.571
	0.286	0.583	0.571
	0.429	0.500	0.476
	0.286	0.368	0
8	0.286	0.368	0

Degree centrality measures how many con-nections an individual has towards others.
Indicates its communication activity and popularity

Betweenness centrality measures how often a node is in the shortest path to another, that is, its attitude in being a hub in communication, control

Closeness centrality measures how close an individual is to any other, giving higher importance to nodes in nearest positions

Final node placing *according to different centrality measures*

Node RANK	Degree centrality	Closeness centrality	Betweenness centrality
1	4	4, 5	4, 5
2	1, 2, 3, 6	6	6
3	5, 7, 8	1, 2, 3	1, 2, 3, 7, 8
4		7, 8	

The winner is..

TOP: Highest number of connections (critical in communication, roads…)

TOP: Highest number of neighbours
BOTTOM: the periphery of the empire

TOP: Most often passed through
BOTTOM: least relevant if cut away

Case Studies

Complex Networks are used in many social fields

- Identifying author/co-authorship network

- Identifying citation/collaboration networks

- Identifying critical road network areas

- Studying the social structure of organ transplant

- Mapping emerging news networks

- Detecting overlapping communities

- Modelling the Supreme Court influence (role of citations)

- Identifying subject matter experts in software repositories

- …

Co-authorship networks (Abbasi & Hossain 2012)

- Studies the usefulness of centrality measures in scholar performance

- Dataset: list of publications of five majoyr US universities (2001-2005)
  - Researchers' own publication lists, DBLP
  - Google Scholar, ACM portal (also used for citations)
  - 2139 publications, 1806 authors, 5310 co-authorship

- A co-authorship network was thus built and analyzed
  - centrality indicators were computed
  - correlation with h-index was studied (Spearman rank test)

- Result: correlation confirmed w.r.t. degree and betweenness centrality

Centrality indicator	Correlation with Sum of citations	Correlation with h-index
C_D (Degree)	0.332	0.311
C_C (Closeness)	-0.012	0.052
C_B (Betweenness)	0.338	0.501

Correlation is significant:
- at 0.01 for Degree and Betweenness
- at 0.05 for Closeness

- This work also led to the proposal of new, *hybrid* centrality measures
  - *degree-degree, degree-closeness, degree-betweenness centrality*
    - degree-* because they all consider a node's neighbours (degree)
    - but sum the values provided by one of the basic indicators

$$DD(k) = \sum_{i=1}^{n} C_D(i) \qquad DC(k) = \sum_{i=1}^{n} C_C(i) \qquad DB(k) = \sum_{i=1}^{n} C_B(i)$$

  - weighted versions also exist *(useful to represent repeated collaborations)*

Centrality	Correl. Sum Cit.	Correl. h-index
C_D *(Degree)*	**0.332**	**0.311**
C_C *(Closeness)*	-0.012	**0.052**
C_B *(Betweenness)*	**0.338**	**0.501**
DD *(Degree-Degree)*	**0.296**	**0.261**
DC *(Degree-Closeness)*	**0.303**	**0.295**
DB *(Degree-Betweenness)*	**0.203**	**0.255**

Correlation significant at 0.01 for the new hybrid measures

Interpretation

- All the indicators reveal the importance of an actor in a social network

  - importance = performance, power, social influence

- but Degree and Betweenness centralities appear to show higher correlation to other widely used indicators in this field

  - Betweenness centrality is known to better predict the preferential attachment by new entrants

- the proposed hybrid measures also show correlation

  - perhaps better predictors of attachment of new added nodes to existing ones?

- but also imply harder calculations…

Author / Citation networks
(Divakarmuthy & Menezes 2012)

- Studies collaboration network extracting data from ACM Digital Library

  - Different from co-authorship: there, if two people have co-authored a paper, then they certainly know each other

  - Instead, just citing someone else's work has no such implication

- The problem with citation counting

  - citations count often assumed as a measure of the impact of a work

    - in favor: high correlation between citations and Nobel prize winners (1992)

    - in contrast: there are works that are never cited

- A similar problem in another field: music

  - composers have always been credited when their song is recorded

  - songs written, but never recorded, are not listed in music DB $\cong$ *inexistent*

- Goal: investigate the effect of uncited works on collaboration

  - comparing two collaboration (co-author) networks

  - a standard one vs a "citation-based" one *which excluded uncited papers*

- **The complex network**
  - globally speaking, a bipartite graph linking authors and paper
  - but focus is just on the author projection

- **Interpretation of indicators**
  - the clustering coefficient is very high → *small world*
    - ▸ groups of highly collaborative people, with few connections outside group
  - the network follows a *power law* with exponent (2.28) in the typical range
  - degree centrality (C_D) measures an author's connectivity
    - ▸ tend to form triads (cliques of degree 3): if A knows B and B knows C, there is a very high chance that A comes to know C, too
  - closeness centrality (C_C) may reveal authors well connected to their immediate neighbours (C_D high) but part of an isolated clique (C_C low)
    - ▸ It is also useful to reveal "*potential collaborativeness*": a person may have little collaborations, but the structure of the network makes him/her a good candidate to acquire new ones → to improve his connectivity (C_D)

- **Basic analysis**
  - nodes (authors) with highest scores were considered
  - in most cases, top performers in the "plain" and "citation" networks were the same… but not always
    - ▸ some people are in the "top ten" of degree, but their connections come from never-cited papers
  - so, the citation-based network is a better representation…(?)

- **Further analysis**
  - some people have high degree because they have many collaborators, but these collaborators *do not* have collaborators themselves
  - to correct this aspect, a good idea is to focus on the *core of the network*
    - ▸ the network is filtered by removing away edges below a given threshold, and deleting the nodes remaining isolated (0-degree nodes)
    - ▸ threshold increased until only 10000 nodes (out of 50000) remained
    - ▸ *giant components* (largest connected subgraphs) were then extracted

- Results

rank	ACM Network		ACM-C Network	
	degree	closeness	degree	closeness
1	**Jack Dongarra**	**Scott Shenker**	**Jack Dongarra**	**Scott Shenker**
2	**Hector Garcia-Molina**	**Hari Balakrishnan**	**Hector Garcia-Molina**	**Christos Papadimitriou**
3	**Luca Benini**	**David Culler**	**Alberto Vincentelli**	**Prabhakar Raghavan**
4	Mateo Valero	**Joseph Hellerstein**	**Luca Benini**	Jeffrey Ullman
5	**Alberto Vincentelli**	**Christos Papadimitriou**	**David Culler**	**Hari Balakrishnan**
6	Andrew Byun Kahng	Ion Stoica	**Michael Stonebraker**	Mihalis Yannakakis
7	Milind Tambe	**Prabhakar Raghavan**	Robert Brayton	Rajeev Motwani
8	**Micha Sharir**	Thomas Anderson	Gerhard Weikum	**Randy Katz**
9	**David Culler**	**Randy Katz**	**Micha Sharir**	**Michael Stonebraker**
10	Thomas Henzinger	Li Feng Zhang	**Scott Shenker**	**Joseph Hellerstein**

Table 3: Ranking of authors according to ("collaborativeness") and closeness centrality ("potential collaborativeness"). Note that the list of authors for degree is still quite stable when we consider only citations. However we notice that closeness now also remains stable. Authors that appear in two or more rankings are shown in **bold**.

- Issues worth noting

  - some "top ten" people disappeared after pruning the network to its core → they were very connected, but with very weak connections

  - the pruned network is not influenced by "sporadic" collaborations

- Context: complex networks aim to understand functionality by analyzing the network structure

  - the topological structure of road network affects critical traffic jam areas

  - the topology of social networks affects the spread of information/diseases

  - the topology of electrical grids affects robustness of energy distribution

- Goal: identifying the *driving nodes* of a network, i.e. the nodes that have to be controlled in order to control the entire network

  - claim: these nodes depend on the network topology, not on its dynamics

  - network centralities are key parameters to measure node's importance

- New question: if we remove/add a node in the network, how do other nodes modify their functionality because of this removal?

  - some networks naturally modify over time (financial, social…)

  - other networks change for specific reasons (electrical failures, road closures..)

  - **goal: to quantify the influence of single nodes** in different parts of the network → new notions: *centrality interference* and *robustness*

Critical road network areas
(Scardoni & Laudanna 2012)

- Perspective: node-by-node modifications
  - a single node modification can be irrelevant to the overall organization of the network (e.g. its scale-free structure)
  - but can profoundly modify the properties of one or more nodes in different regions of the network (e.g. changing its modular structure)

- Focus on centralities
  - adding/removing a node influences the centrality values of other nodes
  - so, the effects of single node alterations can be calculated by *analyzing modifications of centralities values*
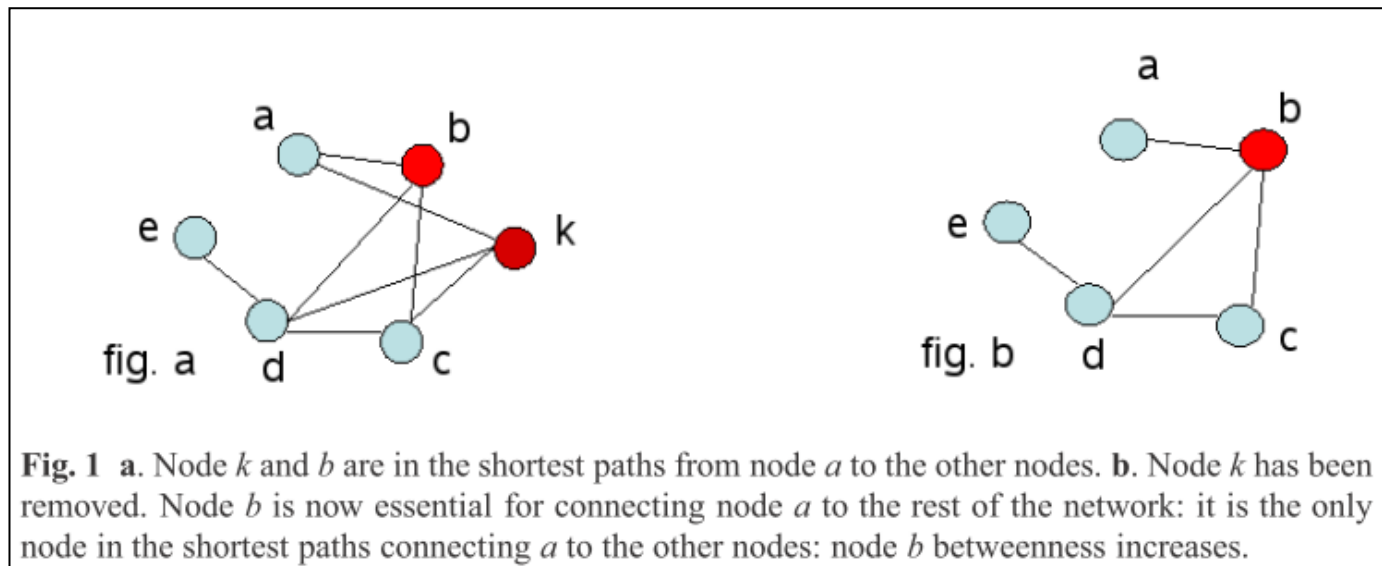
- Idea: introducing *centralities interference* and *robustness*
  - basic step: definition of *interference* for the <u>betweenness centrality</u>
  - possible extension to other centrality measures
  - Assumption: all definitions consider connected networks (i.e. networks where each node is reachable from all the others) and that remain connected also after nodes deletion.

- Why focusing on betweenness centrality?

  - because it naturally captures the idea of a node "being essential" in going from a place to another

  - if a node is removed, the betweenness value of those that can "replace" it (i.e. that are on the alternative shortest paths) increases.

  - this is what is called *betweenness interference*: adding/removing a node "interferes" with the betweenness value of others

Fig. 1 **a**. Node k and b are in the shortest paths from node a to the other nodes. **b**. Node k has been removed. Node b is now essential for connecting node a to the rest of the network: it is the only node in the shortest paths connecting a to the other nodes: node b betweenness increases.

Critical road network areas
(Scardoni & Laudanna 2012)

- The **betweenness interference** *of node i with respect to node n* in the network G is

$$IntBtw(i,n,G) = relBtw(G,n) - relBtw(G', n)$$

where G' is the network obtained from G removing node i and its edges

- The interference value can be positive or negative:

  - if it is negative, the role of node *n* in the network is higher when the node *i* is *not* present (Interpretation: the presence of node *i* is "negative" for node *n* to play a "central role" in the network)

  - if it is positive, the role of node *n* in the network is higher when node *i* is also present (Interpretation; the presence of node *i* is "positive" for node *n* to play a "central role" in the network)

- Similar definitions could be given w.r.t. other centrality types

- By extension, *modulus interference* is also defined:

$$ModIntBtw(i,n,G) = | Btw(G,n) - Btw(G', n) |$$

(Note that absolute interference values, rather then relative values, are used)

Critical road network areas
## (Scardoni & Laudanna 2012)

- **Next step: to quantify the interference of a node w.r.t. the entire network**

  - How important is node i for the functionality of the entire network?

  - A node can interfere with high value with respect to few nodes and can have low interference value with respect to many others…

  - … but the opposite can also occur, with a node interfering with significant values with respect to *the most of* the nodes in the network

- New concepts: *max interference* and *global interference*

  - maxIntBtw(i,G) = max { IntBtw(i,n,G) } (n ≠ i)

    ▸ If it is high, at least one node is consitently affected by node I

  - IntBtw(i,G) = $\sum$ IntBtw(i,n,G) (n ≠ i)

    ▸ If it is high, the node interferes with high values with respect to the most of nodes in the network.

  - To be able to compare different networks, these two values can be normalized as usual, dividing them by N−1

Critical road network areas
(Scardoni & Laudanna 2012)

- **Reversing the question: which nodes can affect node n?**
  - Important if node n is <u>known</u> to play a central role in the network
- **New notions: *robustness*, *competition* and *dependence* of a node**
  - **RobBtw(n,G) = 1/ (max { | IntBtw(i,n,G) | })   (n ≠ i)**
    - ▸ note that the *absolute value* of betweenness interference is used here
    - ▸ <u>If robustness is low</u>, the node can be easily "attacked" by removing or adding particular nodes; <u>if it is high</u>, no node removal/addition that can affect its betweenness value – and consequently its functionality.
  - Positive and negative robustness could be defined, but would be counter-intuitive: better to go with their reciprocal – ***dependence*** and ***competition***
    - ▸ **DepBtw(n,G)   = max { IntBtw(i,n,G) }   where IntBtw(i,n,G) ≥ 0 (n ≠ i)**
    - ▸ **CompBtw(n,G) = max { | IntBtw(i,n,G) | }   where IntBtw(i,n,G) ≤ 0 (n ≠ i)**
  - High dependence means that the node is "central" only because of the presence of at least another node in the network
  - High competition means that the central role of the node can be improved by removing a particular node from the network (*competitor nodes*)
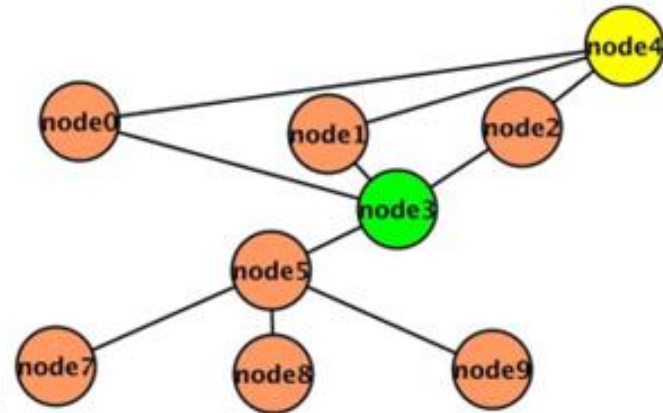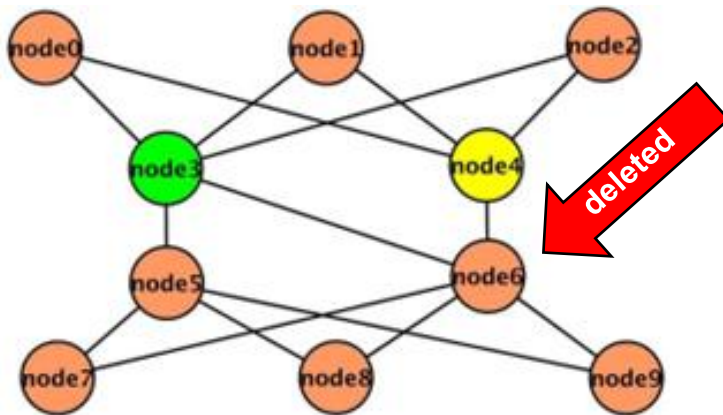
Left:

- node3 (green) and node6 have the highest values of betweenness (25.64); node4 (yellow) and node5 have a lower value (12, the third highest value)

- Robustness: node3 has a higher value (0.046) than node4 (0.036), so it is more likely to remain "central" in the network if some other node is removed

Right (node6 removed):

- node4 (yellow) loses this "central" role and becomes peripheric

  - in fact, its dependence on node6 is quite high (0.1143, with a relative dependence 0.76: that is, it would lose 76% of its starting betweenness value if node6 is removed)

- node3 (green) maintains its "central" role

  - in fact, it has the highest dependence (0.1) on node5 (relative: 31%), not node6
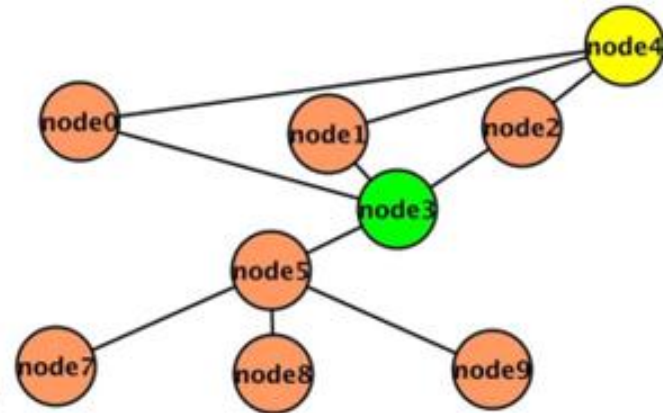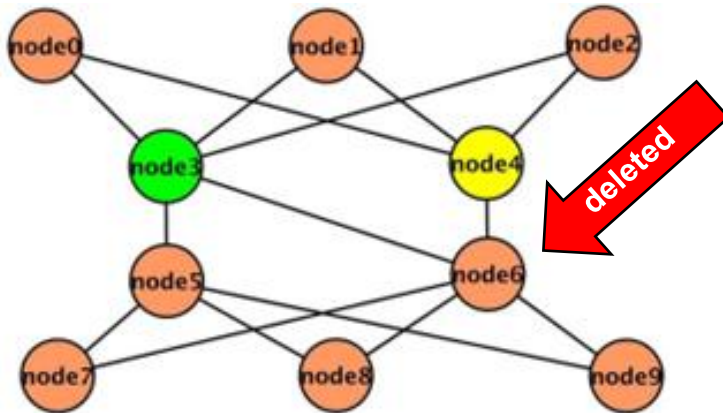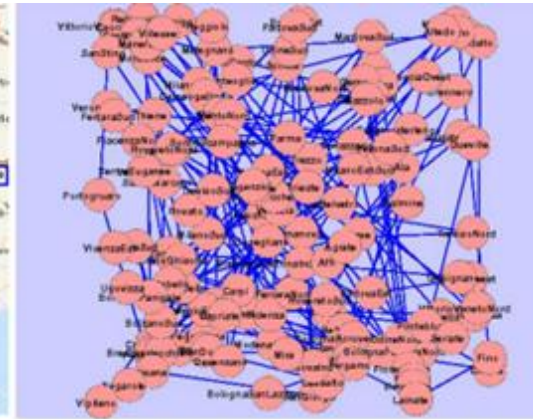
Competition:

- node3 and node4 are "competitors" in the network
  - intuitively, missing one of the two nodes, its role can be replaced by the other
  - in fact, if node3 is removed, node4 becomes the only connection between the "top" and "bottom" of the network; the same for node3 if node4 is removed

- competition values: node4 = 0.2786, node3= 0.2162
  - the highest value of node3 depends on deletion of node4
  - the highest value of node4 depends on deletion of node3
  - the competition value of node4 (0.2786) is higher then node3's (0.2162) because the node4's starting betweenness value is lower (12) than node3's (25.64), so the increase of betweenness is higher (+185%)

Critical road network areas
(Scardoni & Laudanna 2012)

- **Application: Italy's north-east highway network**
  - 136 nodes, 144 edges; distance <u>in minutes</u> between highway exits
  - <u>Betweenness interference</u> evaluated at three points: Melegnano, Como, Mestre

Melegnano

Node name	Betweenness Interference
MilanoSud	0.33
Lodi	0.25
Sesto	0.24
Casalpusterlengo	0.23
PiacenzaNord	0.2
Fidenza	0.18
Fiorenzuola	0.18
Parma	0.18
PiacenzaSud	0.18
ReggioEmilia	0.18

Como

Node name	Betweenness Interference
Fino	0.08
Lainate	0.07
MilanoNord	
Agrate	0.04
Cavenago	0.04
MilanoEst	0.04
Monza	0.04
Trezzo	0.04
Bergamo	0.03
Capriate	0.03

Mestre

Node name	Betweenness Interference
Mira	1.77
PadovaEst	0.06
Grisignano	0.04
Montebello	0.04
Montecchio	0.04
PadovaOv...	0.04
Soave	0.04
VeronaEst	0.04
VicenzaEst	0.04

> Melegnano is a critical node towards Bologna: in fact, the first ten <u>positive interference</u> values are all the towns between Milano Sud and Parma.

Melegnano

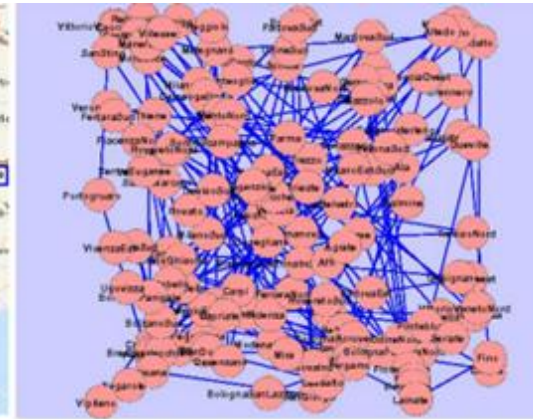| Node name	Betweenness Interference
BresciaCentro	-0.23
BresciaOvest	-0.2
Ospitaletto	-0.2
Grumello	-0.19
Manerbio	-0.19
Palazzolo	-0.19
Ponteoglio	-0.19
Pontevico	-0.19
Rovato	-0.19
Bergamo	-0.18

Como

Node name	Betweenness Interference
PadovaEst	-0.03
Grisignano	-0.02
Mestre	-0.02
Mira	-0.02
Mirano	-0.02
Montebello	-0.02
Montecchio	-0.02
PadovaOvest	-0.02
VeronaSud	-0.02
VicenzaEst	-0.02

Mestre

Node name	Betweenness Interference
Spinea	-1.72
Preganziol	-1.66
Ala	0.0
Belluno	0.0
BolognaArcoveggio	0.0
BolognaFiera	0.0
BolognaPanigale	0.0
BolognaSanLazzaro	0.0
BolzanoNord	0.0

> If Melegnano is removed (road block-ed), the alternative paths are revealed by the <u>negative interference</u> of Mele-gnano: the area around Brescia Centro

Critical road network areas
(Scardoni & Laudanna 2012)

NOTICE: the interference analysis predicts that the Brescia area nodes are undirectly influenced by Melegnano *even if these nodes are far from Melegnano in the network.*
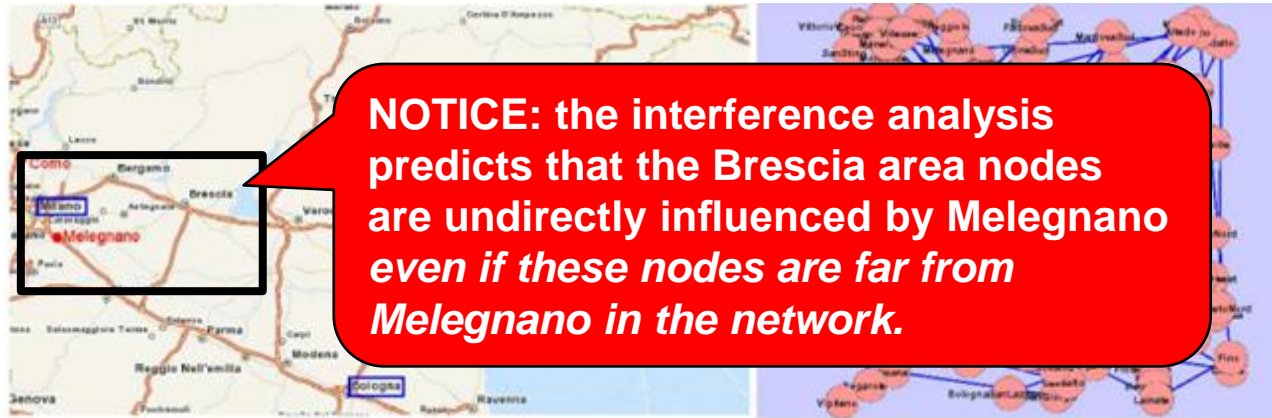
The shortest road from Milan to Bologna passes through Melegnano, but if this node is blocked the shortest road is the one passing through Brescia (blue road)

■ **Application: Italy's north-east highway network**

- 136 nodes, 144 edges; distance <u>in minutes</u> between highway exits
- <u>Betweenness interference</u> evaluated at three points: Melegnano, Como, Mestre

Melegnano		Como		Mestre	
Node name	Betweenness Interference	Node name	Betweenness Interference	Node name	Betweenness Interference
MilanoSud	0.33	Fino	0.08	Mira	1.77
Lodi	0.25	Lainate	0.07	PadovaEst	
Sesto	0.24	MilanoNord	0.05	Mirano	0.05
Casalpusterlengo	0.23	Agrate	0.04	Grisignano	0.04
PiacenzaNord	0.2	Cavenago	0.04	Montebello	0.04
Fidenza	0.18	MilanoEst	0.04	Montecchio	0.04
Fiorenzuola	0.18	Monza	0.04	PadovaOv...	0.04
Parma	0.18	Trezzo	0.04	Soave	0.04
PiacenzaSud	0.18	Bergamo	0.03	VeronaEst	0.04
ReggioEmilia	0.18	Capriate	0.03	VicenzaEst	0.04

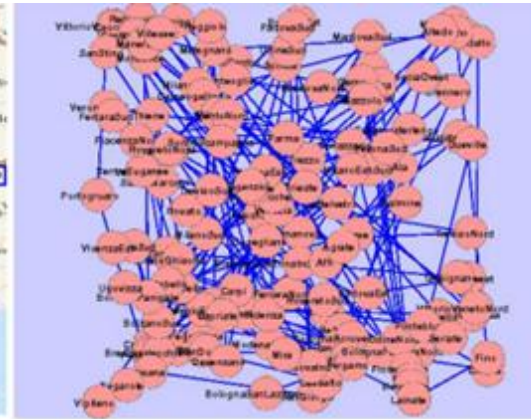| Melegnano		Como		Mestre	
Node name	Betweenness Interference	Node name	Betweenness Interference	Node name	Betweenness Interference
BresciaCentro	-0.23	PadovaEst	-0.03	Spinea	-1.72
BresciaOvest	-0.2	Grisignano	-0.02	Preganziol	-1.66
Ospitaletto	-0.2	Mestre	-0.02	Venezia	-1.06
Grumello	-0.19	Mira	-0.02	Ala	0.0
Manerbio	-0.19	Mirano	-0.02	Belluno	0.0
Palazzolo	-0.19	Montebello	-0.02	BolognaArcoveggio	0.0
Ponteoglio	-0.19	Montecchio	-0.02	BolognaFiera	0.0
Pontevico	-0.19	PadovaOvest	-0.02	BolognaPanigale	0.0
Rovato	-0.19	VeronaSud	-0.02	BolognaSanLazzaro	0.0
Bergamo	-0.18	VicenzaEst	-0.02	BolzanoNord	0.0

> Como is a peripheral node: its interference is high only for its neighbour, and susbstantially smaller than Melegnano's (Como interference value is just 0.08, vs Melegnano's value of 0.32)

> If Como is blocked, the road to Switzerland is blocked, so the only alternative path is on the east corridor

- ■ **Application: Italy's north-east highway network**
  - ● 136 nodes, 144 edges; distance <u>in minutes</u> between highway exits
  - ● <u>Betweenness interference</u> evaluated at three points: Melegnano, Como, Mestre

Melegnano		Como		Mestre	
Node name	Betweenness Interference	Node name	Betweenness Interference	Node name	Betweenness Interference
MilanoSud	0.33	Fino	0.08	Mira	1.77
Lodi	0.25	Lainate	0.07	PadovaEst	
Sesto	0.24	MilanoNord	0.05	Mirano	0.05
Casalpusterlengo	0.23	Agrate	0.04	Grisignano	0.04
PiacenzaNord	0.2	Cavenago	0.04	Montebello	0.04
Fidenza	0.18	MilanoEst	0.04	Montecchio	0.04
Fiorenzuola	0.18	Monza	0.04	PadovaOv...	0.04
Parma	0.18	Trezzo	0.04	Soave	0.04
PiacenzaSud	0.18	Bergamo	0.03	VeronaEst	0.04
ReggioEmilia	0.18	Capriate	0.03	VicenzaEst	0.04

Melegnano		Como		Mestre	
Node name	Betweenness Interference	Node name	Betweenness Interference	Node name	Betweenness Interference
BresciaCentro	-0.23	PadovaEst	-0.03	Spinea	-1.72
BresciaOvest	-0.2	Grisignano	-0.02	Preganziol	-1.66
Ospitaletto	-0.2	Mestre	-0.02	Venezia	-1.06
Grumello	-0.19	Mira	-0.02	Ala	0.0
Manerbio	-0.19	Mirano	-0.02	Belluno	0.0
Palazzolo	-0.19	Montebello	-0.02	BolognaArcoveggio	0.0
Ponteoglio	-0.19	Montecchio	-0.02	BolognaFiera	0.0
Pontevico	-0.19	PadovaOvest	-0.02	BolognaPanigale	0.0
Rovato	-0.19	VeronaSud	-0.02	BolognaSanLazzaro	0.0
Bergamo	-0.18	VicenzaEst	-0.02	BolzanoNord	0.0

Mestre is a well known key connection between Trieste and Milano, Bologna. The negative impact of its blocking only affects its neighbours, thanks to the existence of the "passante" alternative.

If Mestre is blocked, the *passante* road via Spinea is the clear alternative.
<u>But if Spinea is blocked</u>, its high negative interference value (-1.7) with Mestre predicts that this will be even more congested than before the *passante*

Summing up…

- The interference analysis identified critical areas in roads network

- This doesn't result in a real dynamic prediction of traffic jam: it analyses the network structure to identify the parts of the network that can be more easily affected by particular modification of single nodes

  - traffic jam, closure of an exit, work in progress, …

- Interference and robustness analysis can be applied to other domains

  - biological networks, social networks, electtrical grids, transportation…

  - for any case study, the methodology and the interpretation of the analysis strictly depends on the kind of network and the kind of centrality used

- Interference and robustness allows understanding how a network *locally rearranges itself* when nodes are removed or added

- In perspective, interference and robustness usable as a base for new clusterization (by grouping nodes based on their interference value)?

 - perhaps less purely mathematical, but more contextual-oriented..?

The social structure of organ transplant
(Venugopal, Stoner, Cadeiras & Menezes 2012)

- Purpose: modelling the current structure of organ transplantation to analyse the allocation system

    - **question: are organs kept locally whenever possible?**

    - currently, the USA are divided in 11 regions of similar population

    - major criteria for allocation: severity level (first inside region, then outside)

    - six organs considered

        ‣ intestine, lung, pancreas, heart, liver, kidney

- Donor → Recipient (geographical) *directed* network

    - States as nodes, edges as transplantation *directed* relationships

        ‣ State intended as "State of residence" of the donor / recipient

        ‣ weighted edges: +1 for each transplant done

- Goal: find highly-connected (sub)groups of States

    - to this end, the *undirected* version of the network will be used
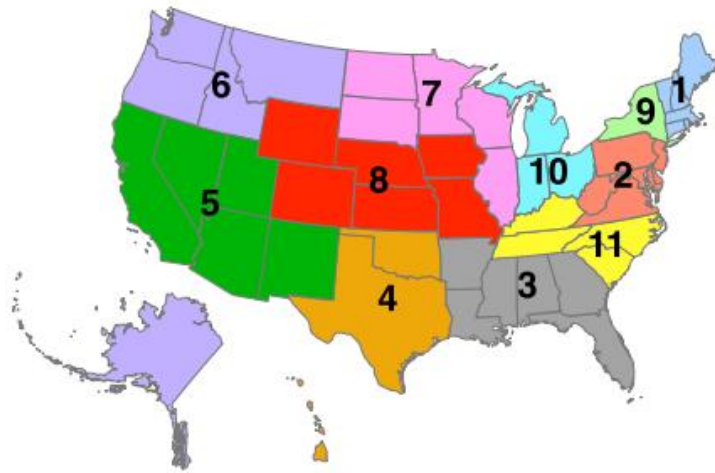
11 regions
50 US States + 6 territories

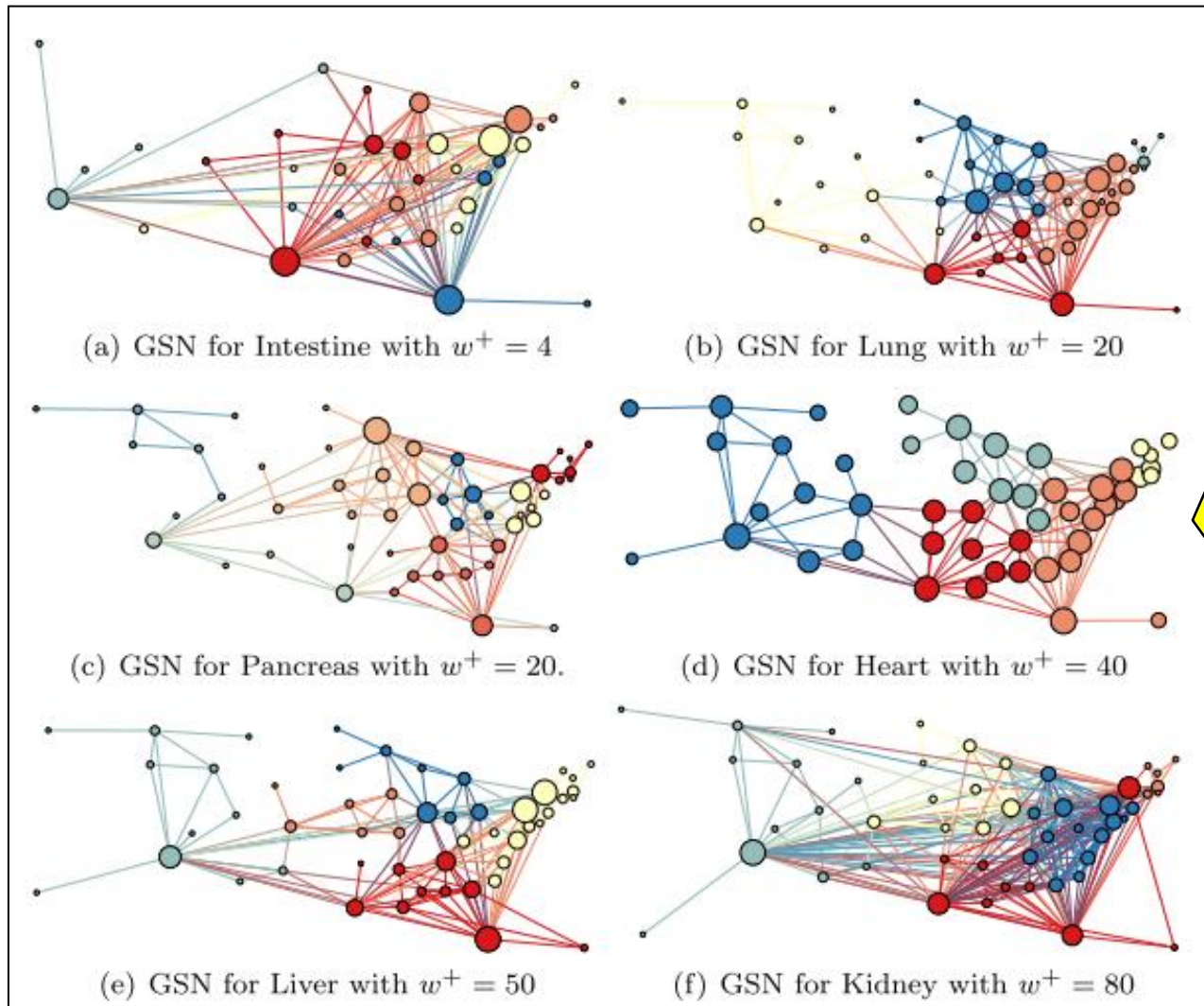The resulting network has 56 nodes (50 States + 6 territories)
It is almost fully connected, as it can expected by construction: this does not reveal any interesting feature.

Table 1 All networks below have 56 nodes which include the 50 states in the Unites states plus its territories. Graph density is a measure of the number of edges in the network in relation to the max number of edges it could have (in this case 1,540).

GSN	Edges	Graph Density	Communities
Intestine	519	0.337	11
Lung	882	0.573	4
Pancreas	940	0.610	7
Heart	980	0.636	3
Liver	1,169	0.759	4
Kidney	1,224	0.866	4

The social structure of organ transplant
(Venugopal, Stoner, Cadeiras & Menezes 2012)

(a) GSN for Intestine with $w^+ = 4$

(b) GSN for Lung with $w^+ = 20$

(c) GSN for Pancreas with $w^+ = 20$.

(d) GSN for Heart with $w^+ = 40$

(e) GSN for Liver with $w^+ = 50$

(f) GSN for Kidney with $w^+ = 80$

Trick: to apply a threshold and remove the edges below the given threshold.

Intuitive motivation: relations are important if they are strong (that is, if all edges have a weight at least *w*, then edges of weight *w* bring no actual information and should be discarded).
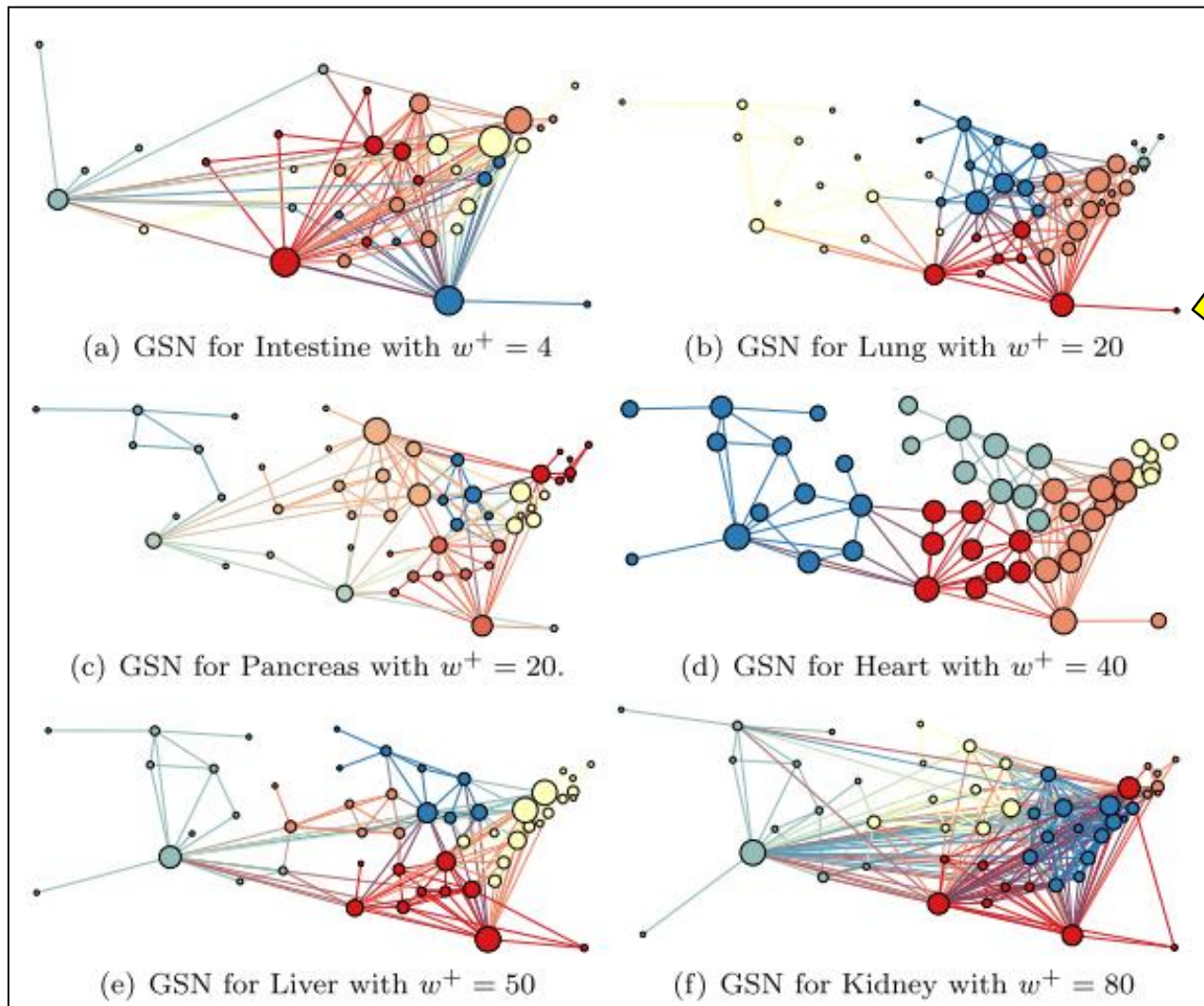
How to choose a threshold: increase *w* until the community structure becomes clear.

RESULT: the six networks shown here for each organ

Significant differences appear evident among the different networks.

(a) GSN for Intestine with $w^+ = 4$

(b) GSN for Lung with $w^+ = 20$

(c) GSN for Pancreas with $w^+ = 20$.

(d) GSN for Heart with $w^+ = 40$

(e) GSN for Liver with $w^+ = 50$

(f) GSN for Kidney with $w^+ = 80$

Observation 1 – Figure 2d

Heart transplant shows <u>five communities</u> with very well defined borders

These borders are related to the <u>11 regions</u>, but the number of communities is lower, as some regions are very small and close to each other.
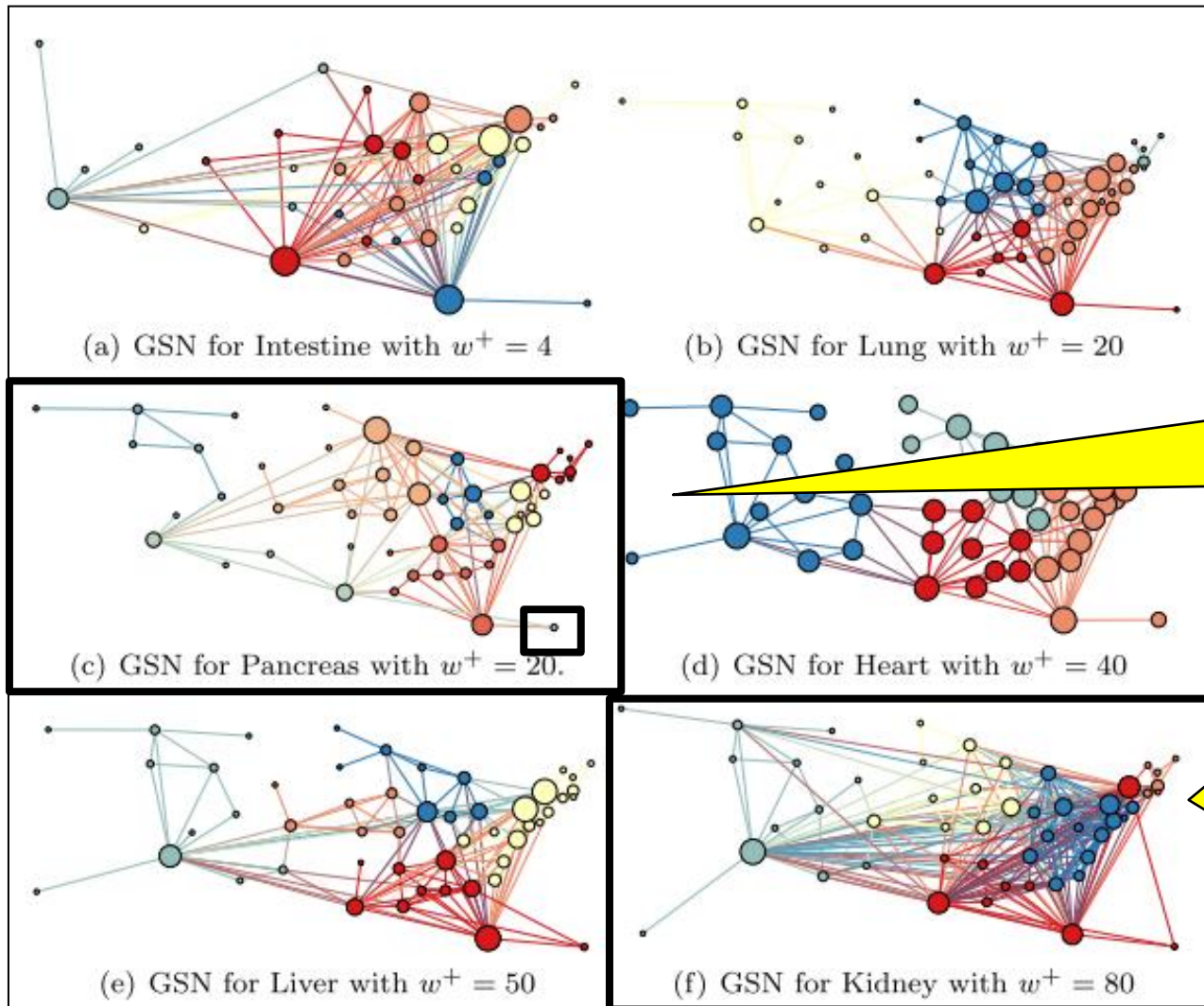
Observation 2 – Fig. 2a/2d

The geographical divisions between communities is better for the heart than for intestine → the "distance" aspect of organ policies is better respected for heart than for intestine.

Observation 3 – all

<u>Better communities</u>: heart, lung, pancreas, liver

<u>Need improvement</u>: intestine, kidney

(a) GSN for Intestine with $w^+ = 4$

(b) GSN for Lung with $w^+ = 20$

(c) GSN for Pancreas with $w^+ = 20$.

(d) GSN for Heart with $w^+ = 40$

(e) GSN for Liver with $w^+ = 50$

(f) GSN for Kidney with $w^+ = 80$

CURIOSITY: the pancreas network shows *Puerto Rico linked to a community in the southwest* of the country, rather than the more natural southeast community

The kidney network is actually okay *except for a node in the northeast* (State of New York) *which is part of the south-east community* (Florida, Texas, Puerto Rico, and others).

Mapping emerging news networks in the Bay Area
(Ramos, Gunes, Mensing & Ryfe 2012)

- Purpose: mapping the changes in the news ecology of the SFO area
  - studying the relationships between 143 local news sites
    - connections between news organizations
    - connections between journalists and their sources
    - connections between users of the news sites

- Context
  - from 1990, newspaper circulation has declined by 31%, and people watching evening news on a major network has declined by 57%
  - professional staff in newspapers has declined by 25% in 10 years
  - but 60% of all Americans now access news online in a typical day
  - → the centralized, one-way distribution model of mass produced news is changing towards a more decentralized, interactive, integrated structure

- The reconfiguring of media relationships online is reshaping the structure of social communication, altering the role and function of message, audience, producer and production.

- **Goal:** *discovering the emerging shapes* of the network
  - looking for the formation of small worlds or other notable configurations
  - the San Francisco bay area is a good testbed because of its role in online innovation and the high broadband penetration

- **Connections between news organizations**
  - methodology: using search engines and links on news sites
    - ‣ a 143-item list was made of sites that <u>produce news</u> AND are <u>updated at least weekly</u> AND <u>are based</u> in SF area
    - ‣ the list included several different sources – from traditional media outlets, to small blogs and other non-traditional formats.
  - web crawling (via a modified WebSPHINX crawler) was then exploited to discover the connections between news organizations: 118 sites crawled
    - ‣ links to <u>outside destinations</u> were recorded (with the number of times linked)
    - ‣ a link was considered duplicate if identical up to the subdomain level (www.example.com/page1.html = www.example.com/page2.html www.example.com/page1.html ≠ other.example.com)

[Connections between news organizations (cont'd)]

- A directed, weighted graph was first generated for each news site

  - each site and its linked sites were nodes

  - a directed weighted edge was drawn from the seed site to the outside site

    - weight = how many times that domain was linked to from the seed site

- Then, all such graphs were merged into a larger graph

- Several parameters were now measured

  - *in-degree* and *out-degree* to reveal authorities (hubs) organizations

  - *betweenness* to reveal sites linking otherwise-unconnected sites

  - PageRank to reveal the "more important" sites (in terms of link number)
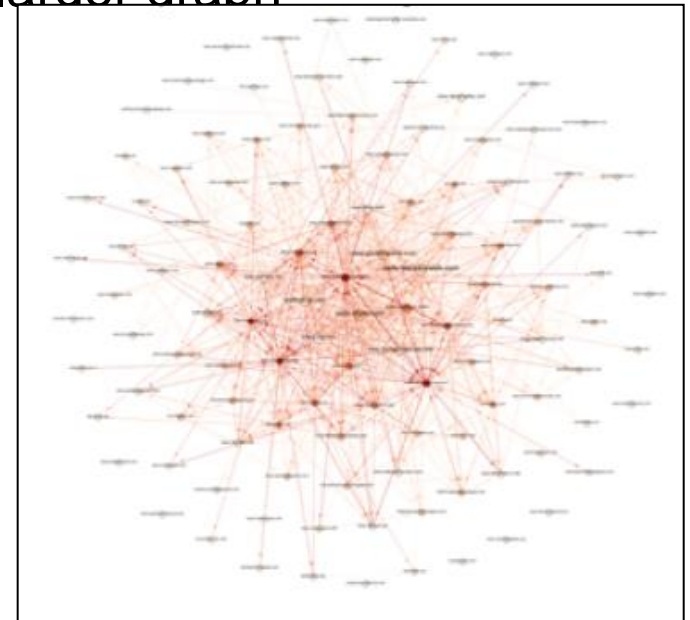
- **Connections between news organizations: Results**

    - Average in/out degree = 6
      (but 43 sites not linking to any other)

    - the sites with the most incoming links (i.e., authorities) were traditional news organizations (San Francisco Chronicle, San Jose Mercury..)

- the <u>top eight sites</u> for ***incoming*** **links** (<u>traditional news organizations</u>) have *a very low number of outgoing links*

    - maybe because traditional news reporters tend to follow the mass media model of news reporting, without adapting to a networked structure…?

- the sites with ***the most outgoing links*** are generally <u>independent or non-traditional organizations</u>

    - fits the expectation that non-traditional/newly-created news sites are more likely to utilize the full potential of the network, provide links to sources and supplemental information about the story being reported, etc

    - traditional news organizations have highest page ranks, but alternative news site (e.g. The Easy Bay Express) have the highest betweenness: *they work as bridges between different types of news organizations*

- **Connections between journalists and community**

  - again, based on web crawling, with site-specific rules to find authors' info – but only crawling the larger sites with clearly-identified author credits

  - the author/external reference pair was recorded, avoiding duplicates

- The obtained graph is bipartite – the journalists on the one side, the external sites referenced in their articles on the other

  - properties mesured: the degrees of journalists (which journalists used links most often) and the degrees of sites (which sites are linked to most often)

- Results

  - **Bay Citizen's**: 4-75 external links (the top five journalists: 30); tends to utilize newer journalistic practices providing references to outside sources, but clearly these practices are unevenly spread among writers

  - **Blogging Bayport Alameda:** a local blog written by one (very prolific) author, thousands of external links

  - **San Francisco Chronicle's:** larger size but fewer journalists, not many links to outside sources (average 3-14 links, the top five journalists 8-14); _no other sites are heavily linked_. Journalists who did utilize external links more frequently are freelancers and other non-staff.

Mapping emerging news networks in the Bay Area
(Ramos, Gunes, Mensing & Ryfe 2012)

- **Connections between users/commenters**
  - <u>purpose</u>: find the connections between users/commenters on news web sites and how they interact with news organizations and other users
  - <u>reason</u>: the ability for the reader to comment directly on a news story is a recent, appreciated feature found on most sites, which highlights the collaborative nature of the Internet.
  - <u>focus</u>: a small subset of sites with the story comment features and a large population of registered users

- Measured properties
  - *the degrees of users*, to determine the most prolific in commenting
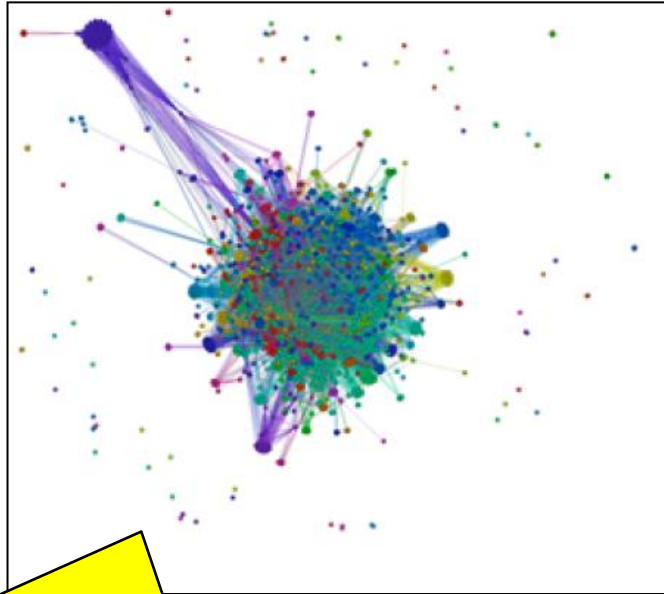 - *the degree of the stories*, to determine the most "interesting" stories

- Further graph obtained by deleting the story nodes, *to reveal clusters*
  - shows user-user connections, i.e. users that commented the same stories
    - ▸ edge weight = the number of stories they both commented on
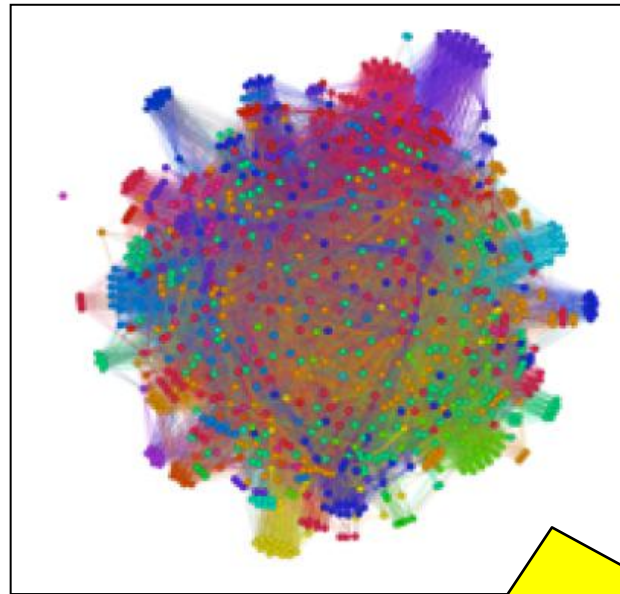
■ Connections between users/commenters

Berkeleyside: fairly active community, with the top 4 users commenting on over 100 stories.

In the derived user-user graph, a user was linked, on average, to **19** other users.

142 different communities were identified (including single-user communities), the top two containing about the 30% of users → a large portion of users frequently comment on the same stories.

Socket site: a niche site (real estate), with a smaller but more closely knit community, with the top 4 users commenting on 80-100 stories each.

In the derived user-user graph, a user was linked, on average, to **57** other users.

19 different communities were identified, with the top three containing about the 53% of users → tighter knit group probably due to the specific site focus