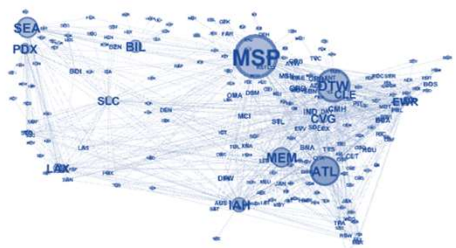
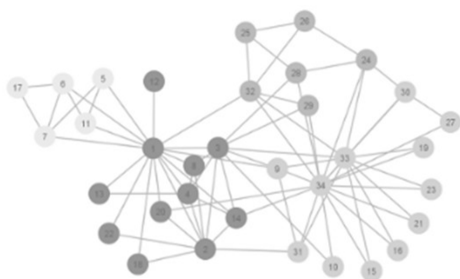
Prof. Ralucca Gera, rgera@nps.edu
Applied Mathematics Department,
Naval Postgraduate School



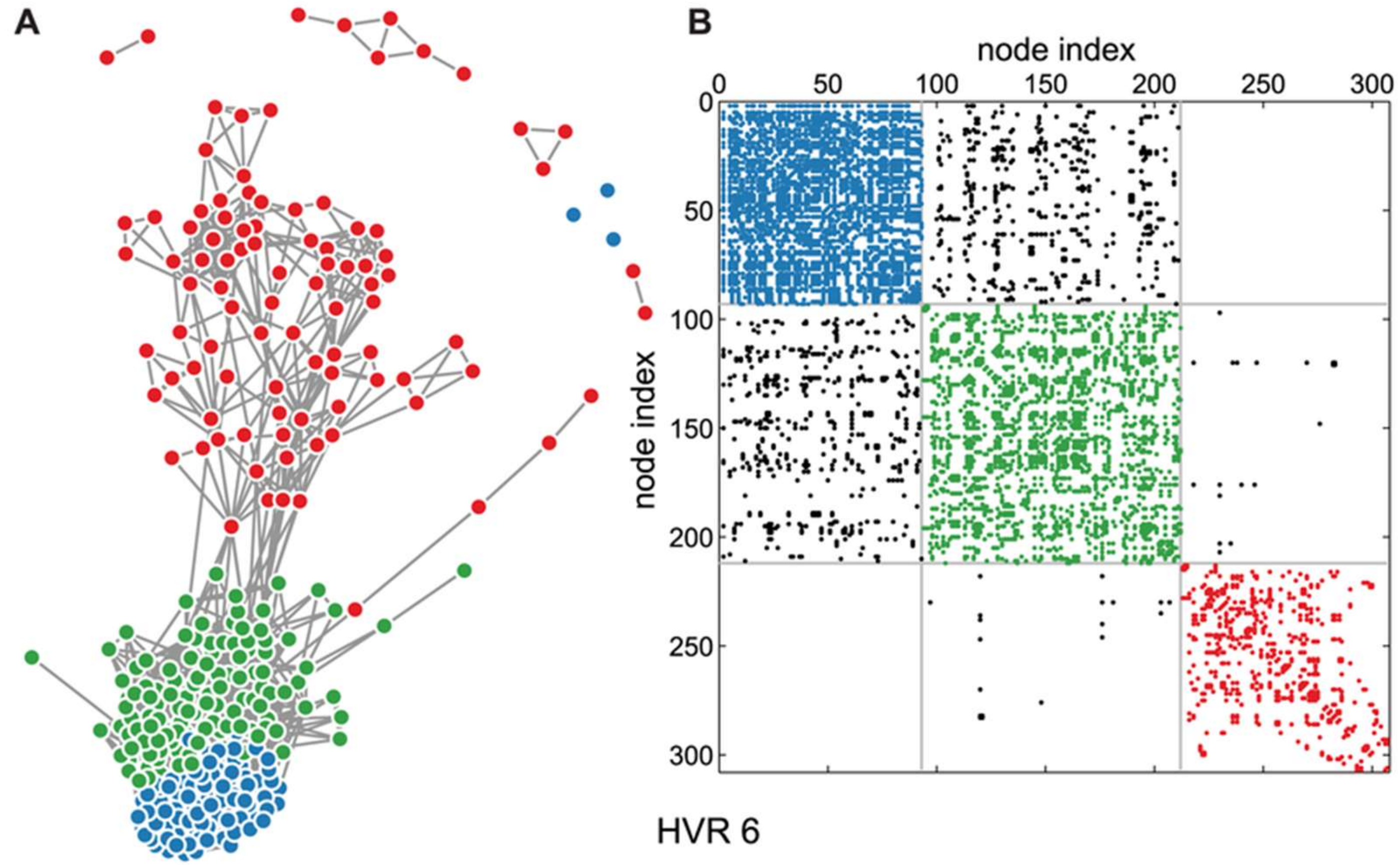


MA4404 Complex Networks

Groups of vertices and Core-periphery structure

Learning Outcomes

- Understand and contrast the different k-clique relaxation definitions:
  1. k-dense
  2. k-core
  3. k-plex
- Contrast macro-scale to meso-scale to micro-scale structure analysis.
- Determine which nodes are part of a densely connected core and which are part of a sparsely connected periphery:
  - A node belongs to a core if and only if it is well connected both to other core nodes and to peripheral nodes



Why?

- Most observed real networks have:
  - Heavy tail (powerlaw, exponential)
  - High clustering (high number of triangles especially in social networks, lower count otherwise)
  - Small average path (usually small diameter)
  - Communities/periphery/hierarchy
  - Homophily and assortative mixing (similar nodes tend to be adjacent)
- Where does the structure come from? How do we model it?

Macro and Meso Scale properties

Macro Scale properties (using all the interactions):

Small world (small average path, high clustering)	Powerlaw degree distr. (generally pref. attachment)

Meso Scale properties applying to groups (using k-clique, k-core, k-dense):

Community structure	Core-periphery structure

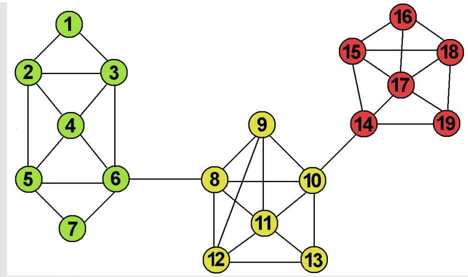
Micro Scale properties applying to small units:

Edge properties (such as who it connects, being a bridge)	Node properties (such as degree, cut-vertex)

Some local and global metrics pertaining to structure of networks

Structure they capture	Local Statistics	Global statistics
Direct influence General feel for the distribution of the edges	Vertex degree, in and out degree	Degree distribution
Closeness, distance between nodes	Geodesic (shortest path between two nodes) Distance (numerical value – length of a geodesic)	Diameter, radius, average path length
Connectedness of the network How critical are vertices to the connectedness of the graph? How much damage can a network take before disconnecting?	Existence of a bridge (cut-edge) Existence of a cut vertex	Vertex cut Edge cut
Tight node/edge neighborhoods, important nodes as a group	Clique, plex, core, community, k-dense (for edges)	Community detection Core-periphery structure

Groups and subgroups identifications



Some common approaches to subgroup identification and analysis:

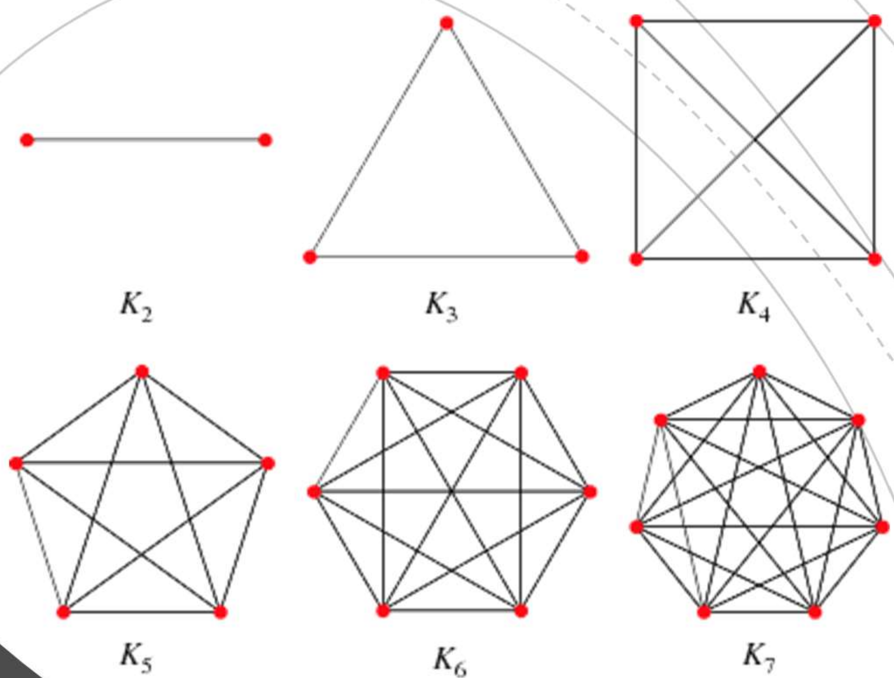
- K-cliques
- K-cores (k-shell)
- K-denseness
- Components
- Community detection

Communities are used to explore how large networks can be built up out of small and tight groups.

Core structure in a network is thus not merely densely connected but also tends to be "central" to the network(e.g.,in terms of short paths through the network)

0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0
0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0

Source: Guido Caldarelli, *Communities and Clustering in Some social Networks,* NetSci 2007 New York, May 20th 2007



k-clique

k-clique

A clique of size k: a complete subgraph on k nodes (i.e. s subset S of k nodes such that $\deg_{G[S]} v = k - 1$).

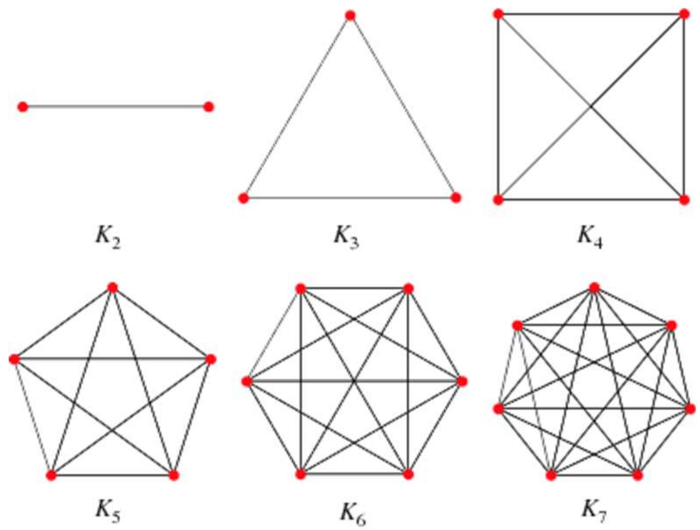
We usually search for the maximum cliques, or the node count in a maximum cliques (the clique number).

Is it realistic and useful in large graphs?

Why is it hard to use this concept on real networks?

- Because one might not infer/know all the edges of the true network, so clique may exist but it may not be captured in the data to be analyzed
- Hard to find the largest clique in the network (decision problem for the clique number is NP-Complete)

A relaxed version of a clique might be just as useful in large networks.



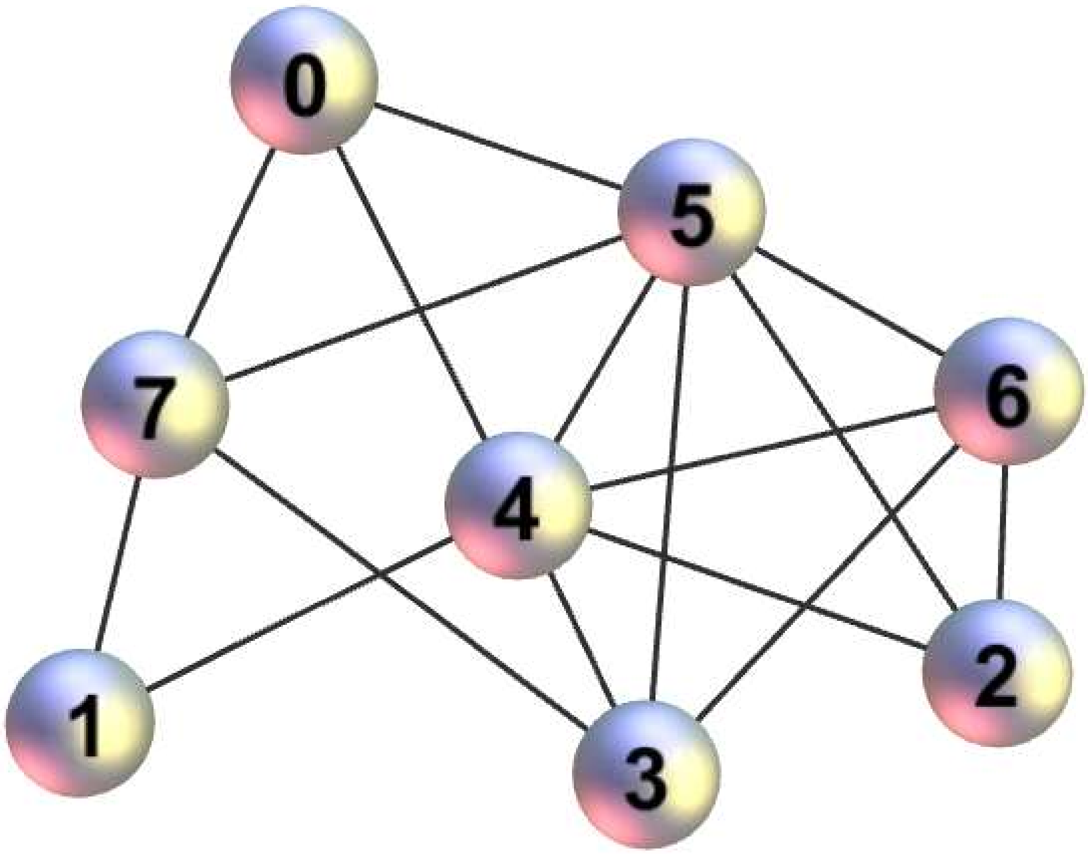
In class exercise

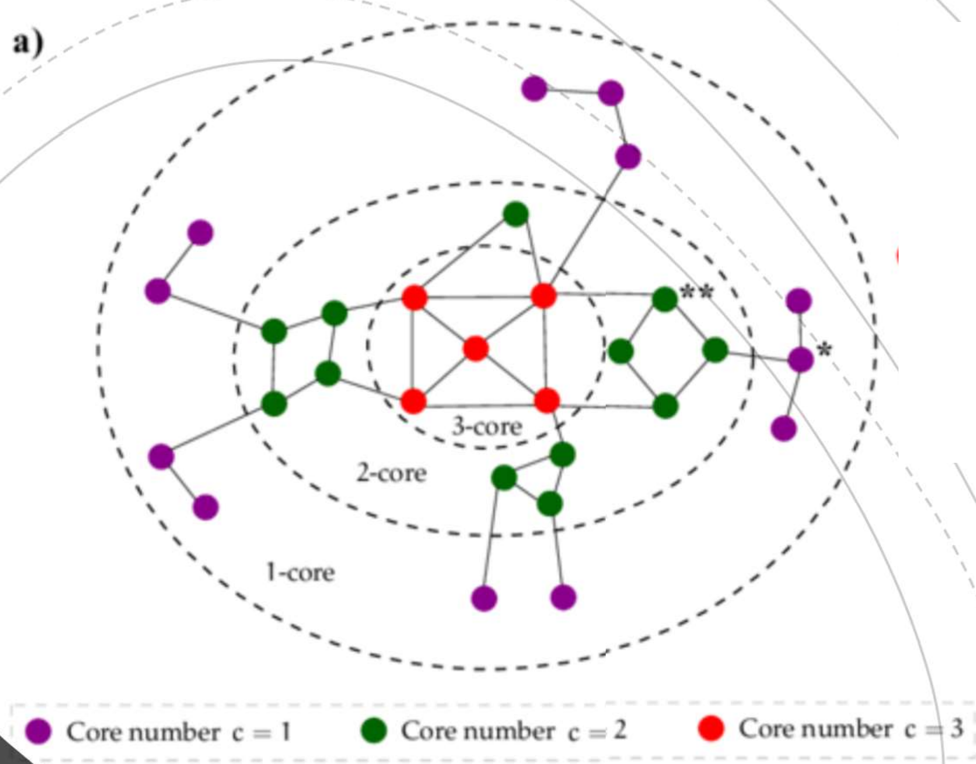
A clique of size k: a complete subgraph on k nodes (i.e. s subset S of k nodes such that $\deg_{G[S]} v = k - 1$).

Identify a:

- 1-clique
- 2-clique
- 3-clique
- 4-clique

Relaxed versions of a k -clique are k -dense and k -core




a)
**
*
3-core
2-core
1-core
Core number c = 1
Core number c = 2
Core number c = 3

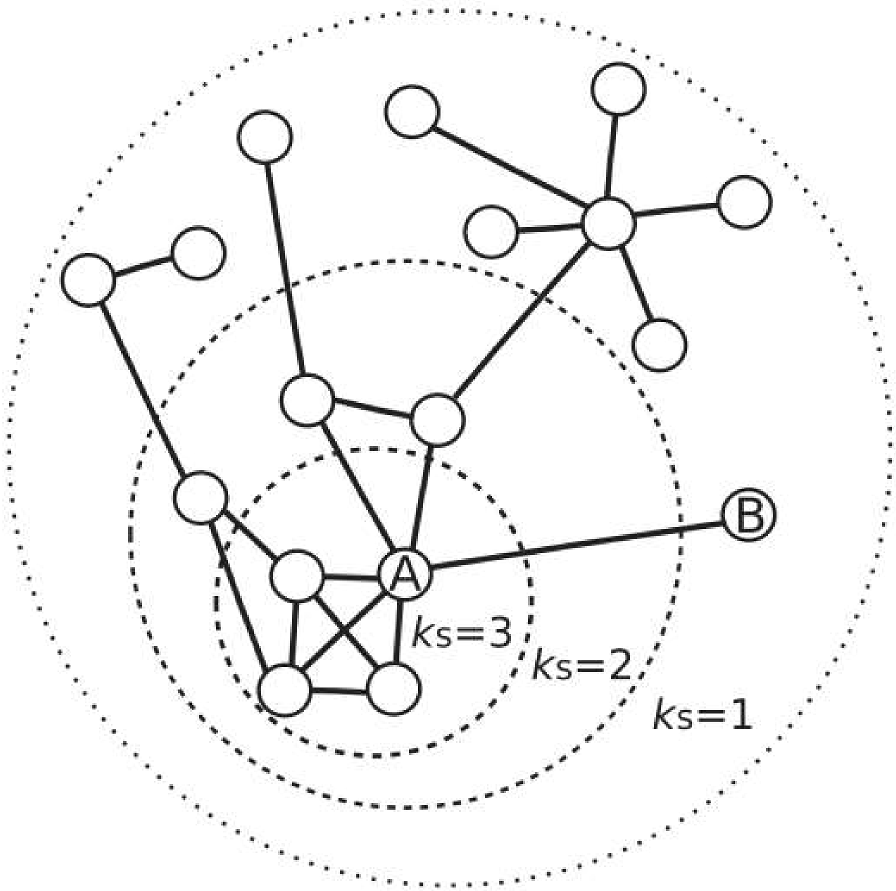
k-core

k-core

- A k-core of size n: maximal subset of α nodes ($\alpha \geq k+1$), each with $\deg_{G[S]} v \geq k$, where $G[S]$ is the subgraph induced by S
- Idea for a k-core: enough edges are present between the group of α nodes to make a group strong even if it is not a clique.

Algorithm for finding the core:

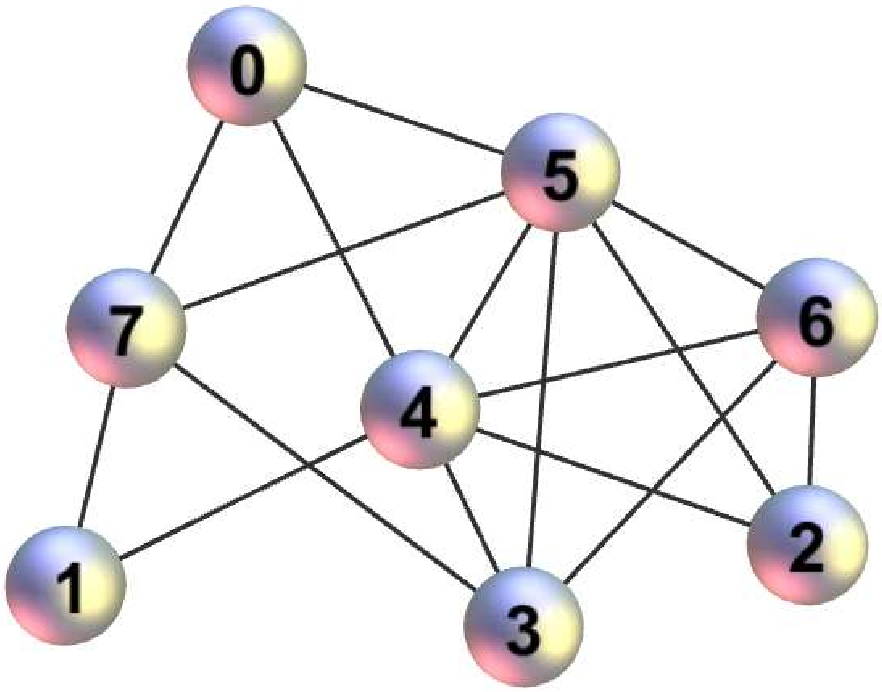
- eliminate lower order k-cores
- the k-core is subgraph of nodes associated with the highest k value



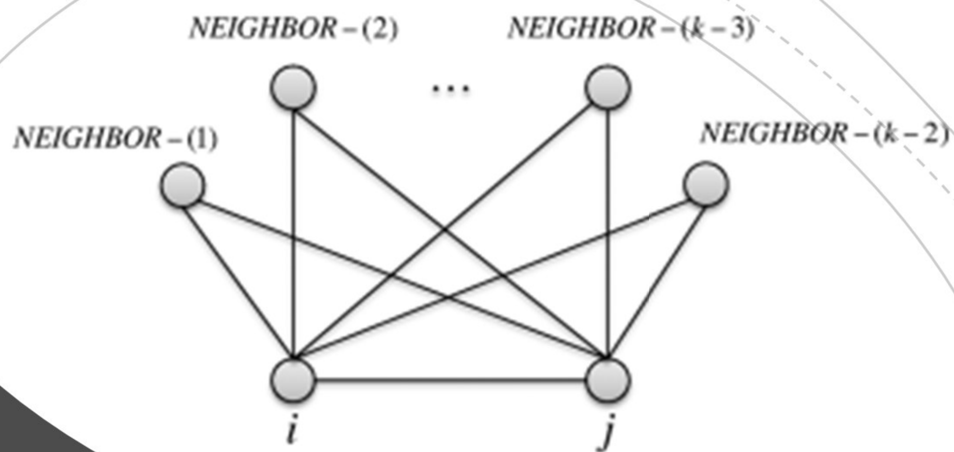
http://iopscience.iop.org/article/10.1088/1367-2630/14/8/083030

In-class exercise

- A k-core of size n: maximal subset of $\alpha \geq k+1$ nodes, each with $\deg_{G[S]} v \ \geq k$, where $G[S]$ is the subgraph induced by S
- Identify the:
  - 1-core
  - 2-core
  - 3-core
  - 4-core
  - the core.

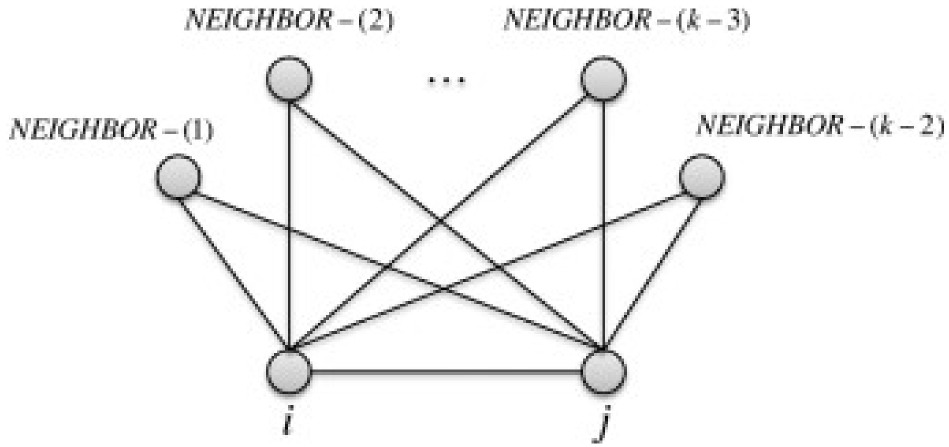


k-dense



k-dense

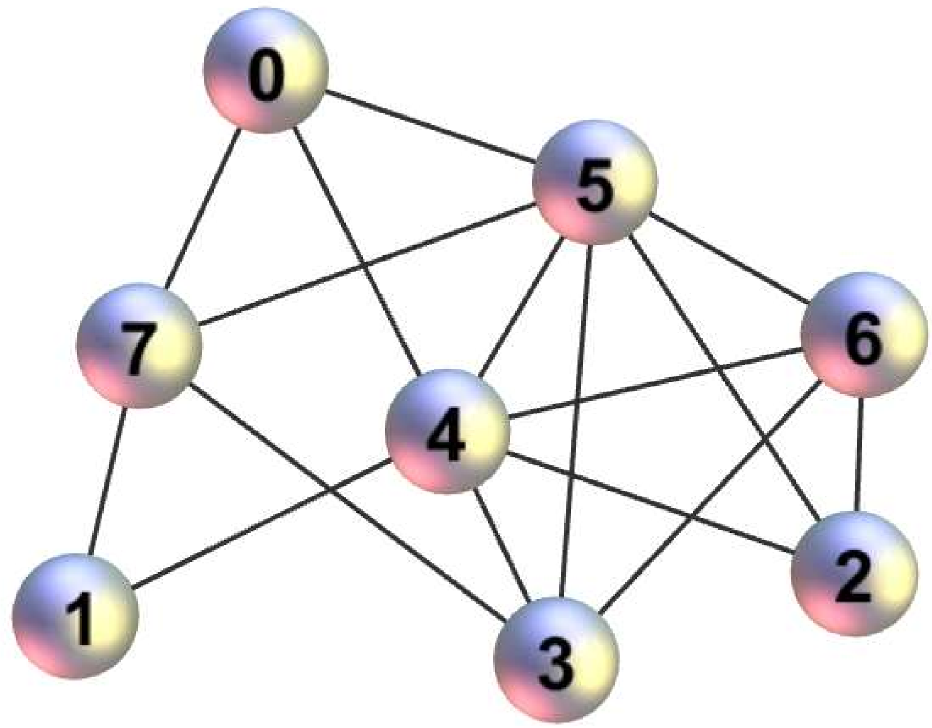
- A k- dense sub-graph is a group of some α vertices ($\alpha \geq k$), in which each pair of vertices {*i, j*} has at least k-2 common neighbors.



Idea: a relaxed k clique (k –dense looks at neighbors of edges/friendships rather than vertices, in making the α nodes part of the α group)

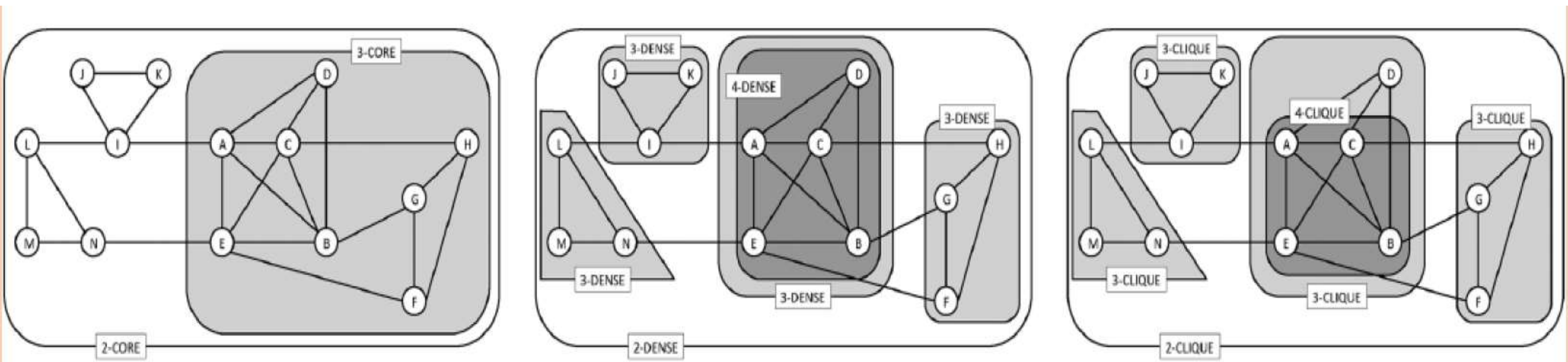
In class exercise

- A k- dense sub-graph is a group of some $\alpha \geq k$ vertices, in which each pair of vertices {*i, j*} has a least k-2 common neighbors.
- Identify a:
  - 2-dense
  - 3-dense
  - 4-dense
  - 5-dense

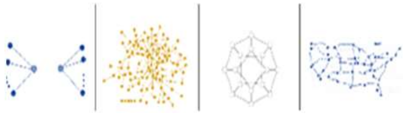


k-dense

- A k- dense sub-graph is a group of some $\alpha \geq k$ vertices, in which each pair of vertices {*i, j*} has at least k-2 common neighbors.



k - clique ⊂ k - dense ⊂ k-1 – core



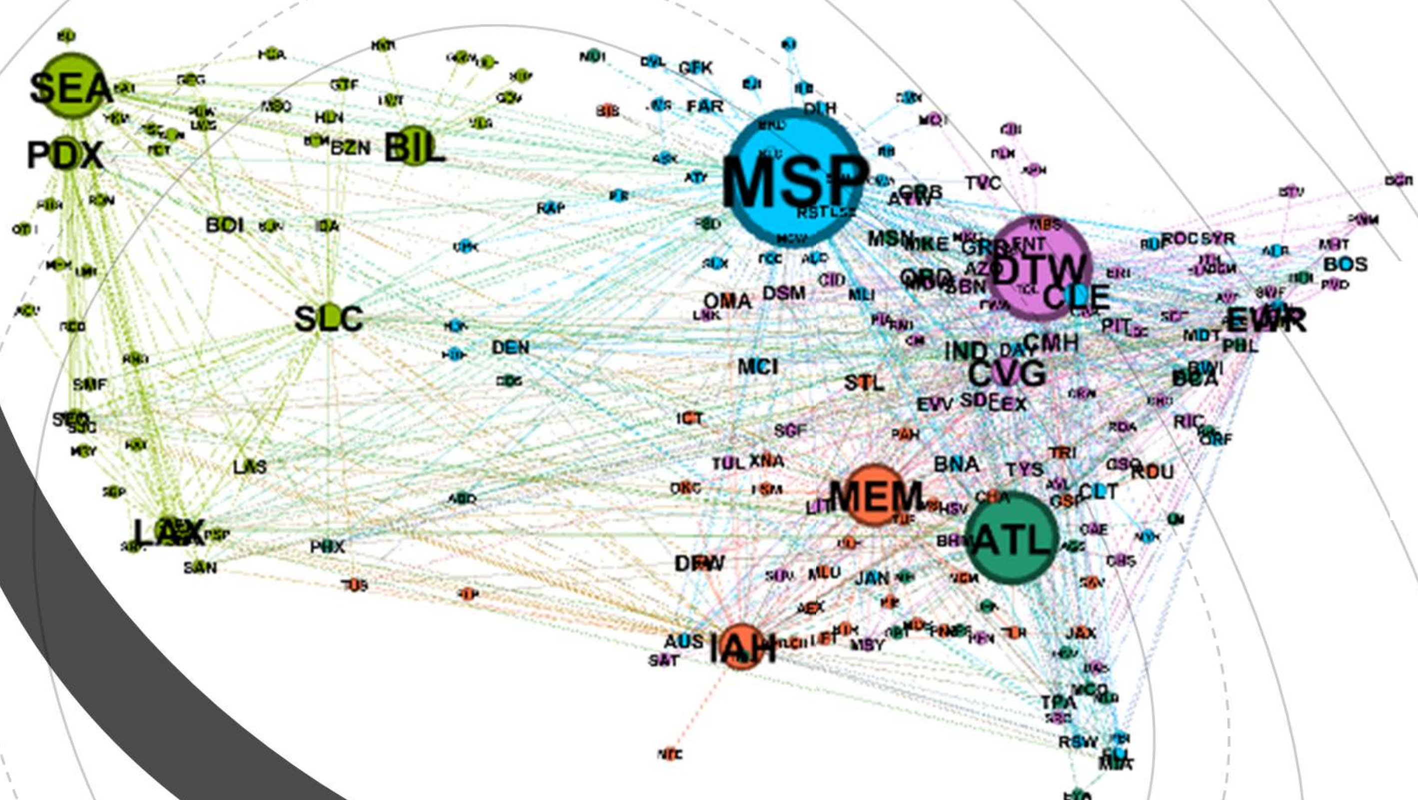
Other extensions

Table 1.

Definition of (locally) dense network structures

Name of dense network structure	Definition	References
Clique	A complete subgraph of size *k*, where complete means that any two of the *k* elements are connected with each other	[36,37]
k-**clan**	A maximal connected subgraph having a subgraph-diameter $\leq k$, where the subgraph-diameter is the maximal number of links amongst the shortest paths *inside* the subgraph connecting any two elements of the subgraph	[37,38,39]
k-**club**	A connected subgraph, where the distance between elements of the subgraph $\leq k$, and where no further elements can be added that have a distance $\leq k$ from all the existing elements of the subgraph	[37,38,39]
k-**clique**	A maximal connected subgraph having a diameter $\leq k$, where the diameter is the maximal number of links amongst the shortest paths (including those *outside* the subgraph), which connect any two elements of the subgraph	[37,38,39,40]
k-**clique community**	A union of all cliques with *k* elements that can be reached from each other through a series of adjacent cliques with *k* elements, where two adjacent cliques with *k* elements share $k-1$ elements (note that in this definition the term *k*-clique is also often used, which means a clique with *k* elements, and not the *k*-clique as defined in this set of definitions; the definition may be extended to include variable overlap between cliques)	[41,42]
k-**component**	A maximal connected subgraph, where all possible partitions of the subgraph must cut at least *k* edges	[43]
k-**plex**	A maximal connected subgraph, where each of the *n* elements of the subgraph is linked to at least $n-k$ other elements in the same subgraph	[37,44]
k-**plex**	A maximal connected subgraph, where each of the *n* elements of the subgraph is linked to at least $n-k$ other elements in the same subgraph	[37,44]
Strong LS-set	A maximal connected subgraph, where each subset of elements of the subgraph (including the individual elements themselves) have more connections with other elements of the subgraph than with elements outside the subgraph	[37,45]
LS-set	a maximal connected subgraph, where each element of the subgraph has more connections with other elements of the subgraph than with elements outside of the subgraph	[37,45,46]
lambda-set	a maximal connected subgraph, where each element of the subgraph has a larger element-connectivity with other elements of the subgraph than with elements outside of the subgraph (where element-connectivity means the minimum number of elements that must be removed from the network in order to leave no path between the two elements)	[37,47]
weak (modified) LS-set	a maximal connected subgraph, where the sum of the inter-modular links of the subgraph is smaller than the sum of the intra-modular edges	[37,45]
k-**truss** **or** *k*-**dense subgraph**	the largest subgraph, where every edge is contained in at least $(k-2)$ triangles within the subgraph	[48,49,50,51]
k-**core**	a maximal connected subgraph, where the elements of the subgraph are connected to at least *k* other elements of the same subgraph; alternatively: the union of all *k*-shells with indices greater or equal *k*, where the *k*-shell is defined as the set of consecutively removed nodes and belonging links having a degree $\leq k$	[37,45,52]

https://academic.oup.com/comnet/article/doi/10.1093/comnet/cnt016/2392115/Structure-and-dynamics-of-core-periphery-networks



Using them globally

Communities vs. core/dense/clique

A clique of size k: a complete subgraph on k nodes (i.e. s subset S of k nodes such that $\deg_{G[S]} v = k - 1$).

A k-core of size n: maximal subset of $\alpha \geq k + 1$ nodes, each with $\deg_{G[S]} v \; \geq k$, where $G[S]$ is the subgraph induced by S

A k- dense sub-graph is a group of some $\alpha \geq k$ vertices, in which each pair of vertices {*i, j*} has at least k-2 common neighbors.

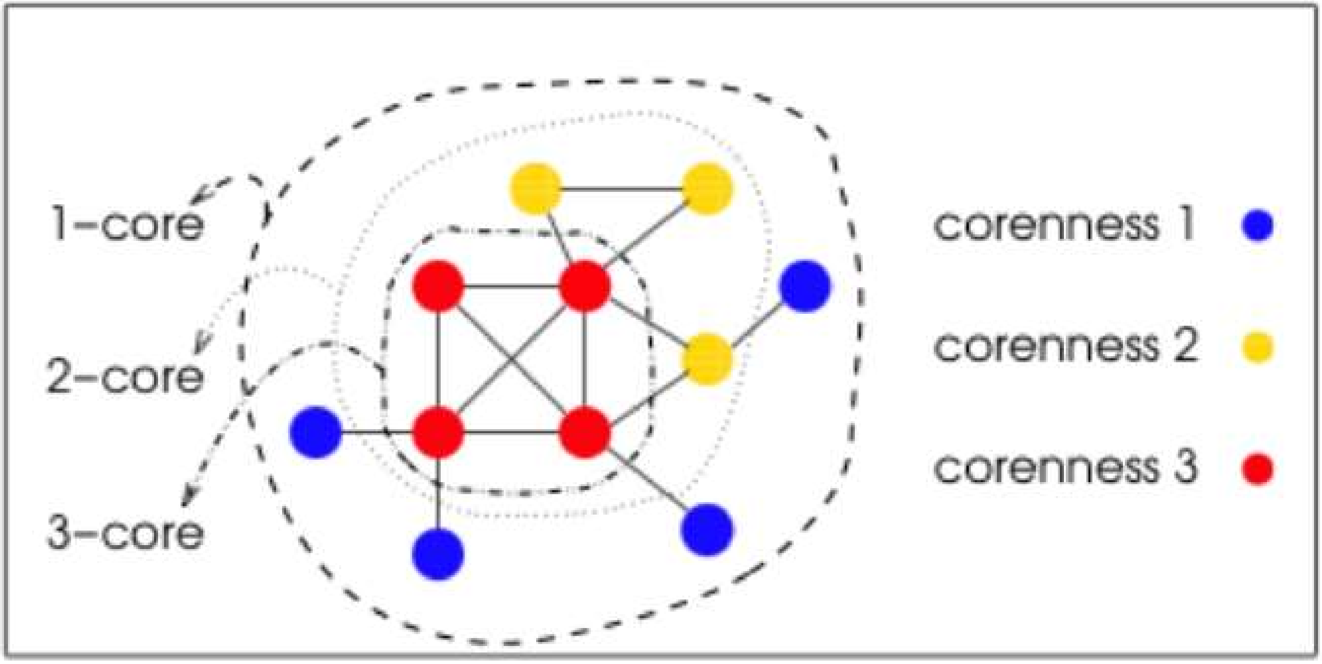
K-core/dense/clique: look at the connections inside the group of nodes

Communities look both at internal and external ties (high internal and low external ties)

Core-periphery decomposition is also looking at internal and external to the core (doesn't have to be a k-core)

	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	0	0	0	0	0
2	1		1	1	0	1	1	1	0	0
3	1	1		1	0	0	0	1	1	0
4	1	1	1		1	0	0	0	0	1
5	1	0	0	1		0	0	0	0	0
6	0	1	0	0	0		0	0	0	0
7	0	1	0	0	0	0		0	0	0
8	0	1	1	0	0	0	0		0	0
9	0	0	1	0	0	0	0	0		0
10	0	0	0	1	0	0	0	0	0	

K-core (k-shell) decomposition



The decomposition identifies the shells for different k-values.

Generally (but not well defined): the core of the network (the k-core for the largest k) and the outer periphery (last layer: 1-core taking away the 2-core). There are modifications where several top values of k make the core.

http://3.bp.blogspot.com/-TIjz3nstWD0/ToGwUGivEjI/AAAAAAAAsWw/etkwklnPNw4/s1600/k-cores.png

The shells in the k-core and degree

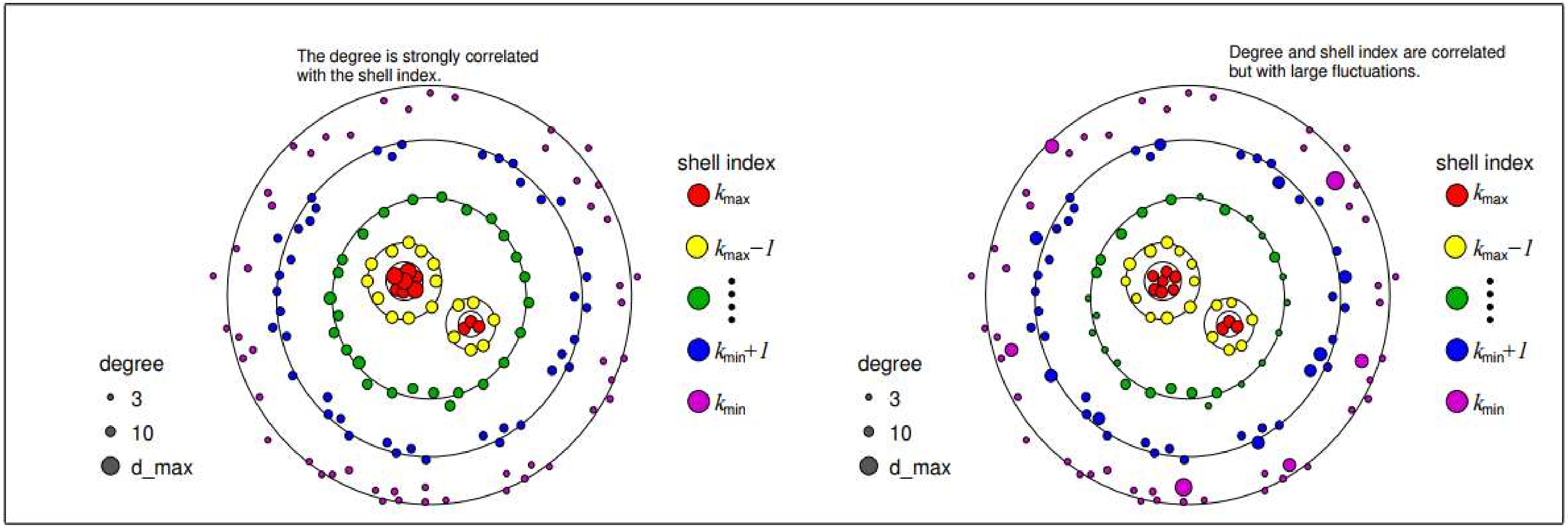
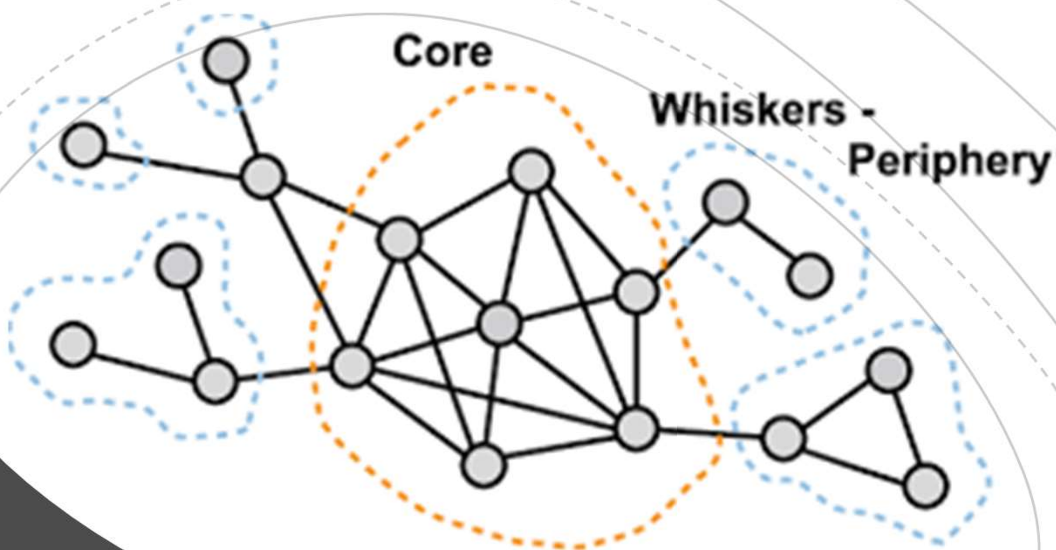


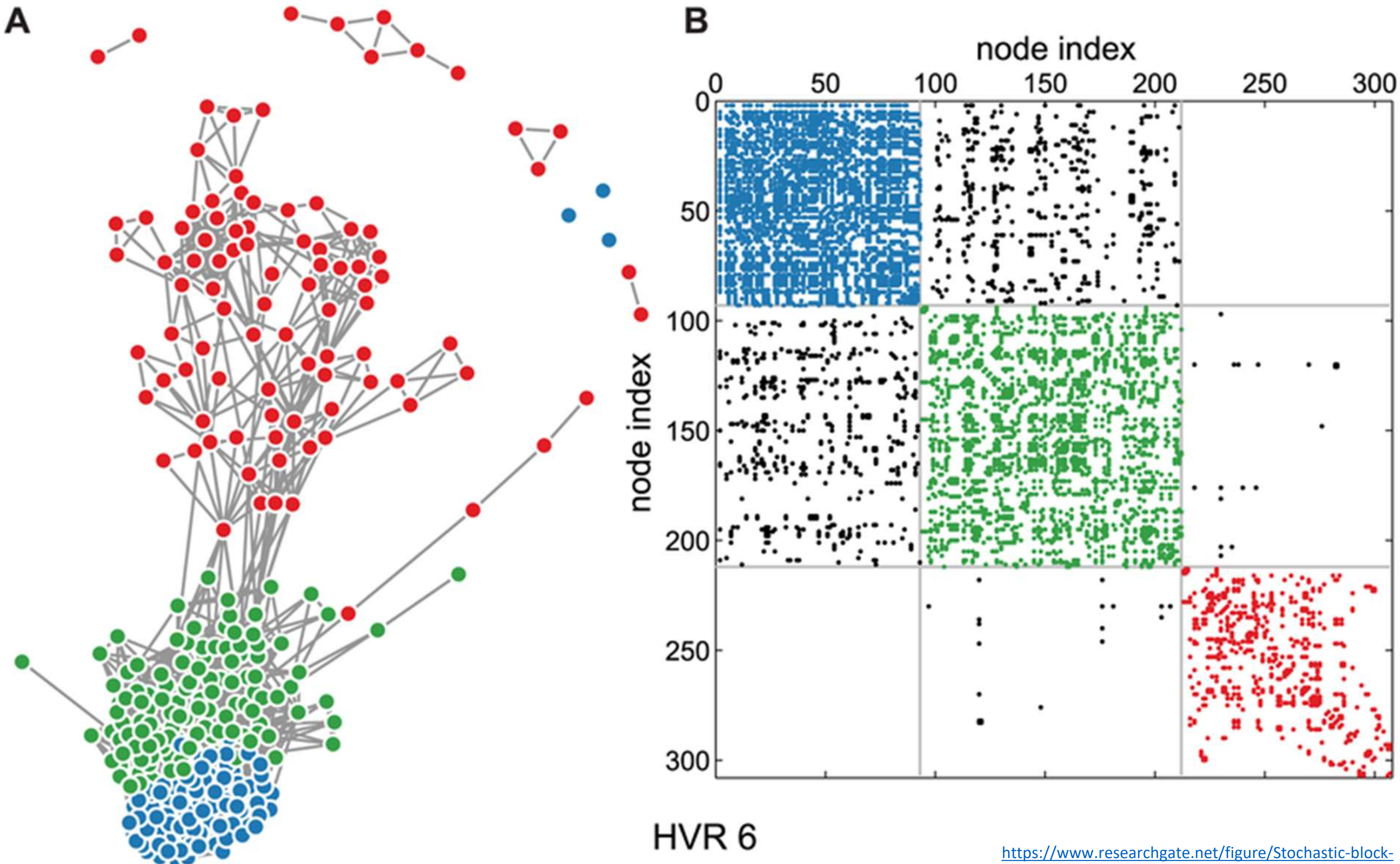
Figure 3: Correlations between shell index and degree. On the left, we report a graph with strong correlation: the size of the nodes grows from the periphery to the center, in correspondence with the shell index. In the right-hand case, the degree-index correlations are blurred by large fluctuations, as stressed by the presence of hubs in the external shells.

http://papers.nips.cc/paper/2789-large-scale-networks-fingerprinting-and-visualization-using-the-k-core-decomposition.pdf



Core-periphery

Core-periphery adjacency matrix

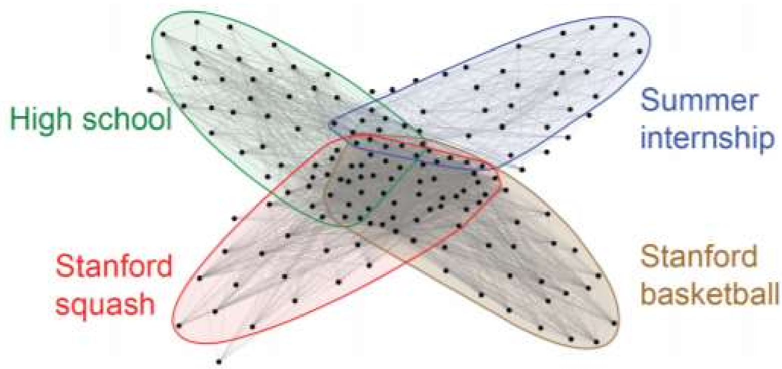


https://www.researchgate.net/figure/Stochastic-block-modeling-identifies-network-communities-HVR-6-is-shown-in-two-forms_fig7_257839768

Core-periphery decomposition

- The core-periphery decomposition captures the notion that many networks decompose into:
  - a densely connected core, and
  - a sparsely connected periphery (see Ref [6] & [12]).
- The core-periphery structure is a pervasive and crucial characteristic of large networks [13], [14], [15].

- If overlapping communities are considered: the network core forms as a result of many overlapping communities



http://ilpubs.stanford.edu:8090/1103/2/paper-IEEE-full.pdf

Measuring core-periphery

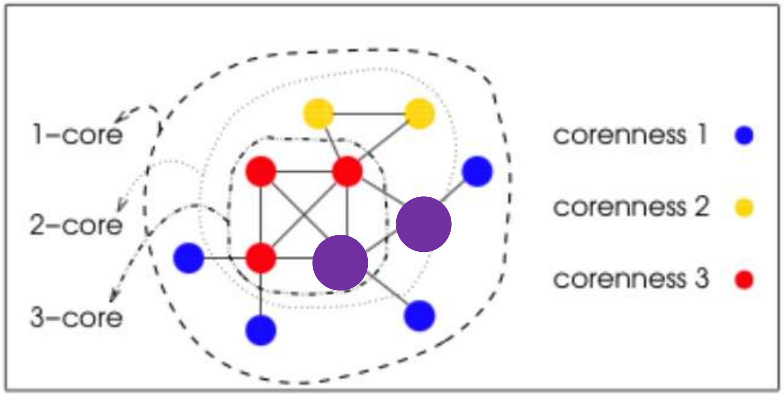
Not standardized, but generally the density of the k-core must be high, checked against the ideal matrix for core-periphery. This is computer by the correlation, ρ, defined as $\rho = \sum_{i,j} a_{ij}\delta_{ij}$,

where a_{ij} is the (i,j) adjacency matrix entry of the network, and

$$\delta_{ij} = \begin{cases} 1, if\ either\ node\ i\ or\ j\ is\ in\ the\ core \\ 0, \qquad\qquad\qquad\qquad otherwise \end{cases}$$

The ideal core-periphery matrix

	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	1	1	1	1	1
2	1		1	1	1	1	1	1	1	1
3	1	1		1	1	1	1	1	1	1
4	1	1	1		1	1	1	1	1	1
5	1	1	1	1		0	0	0	0	0
6	1	1	1	1	0		0	0	0	0
7	1	1	1	1	0	0		0	0	0
8	1	1	1	1	0	0	0		0	0
9	1	1	1	1	0	0	0	0		0
10	1	1	1	1	0	0	0	0	0	



http://www.sciencedirect.com/science/article/pii/S0378873399000192

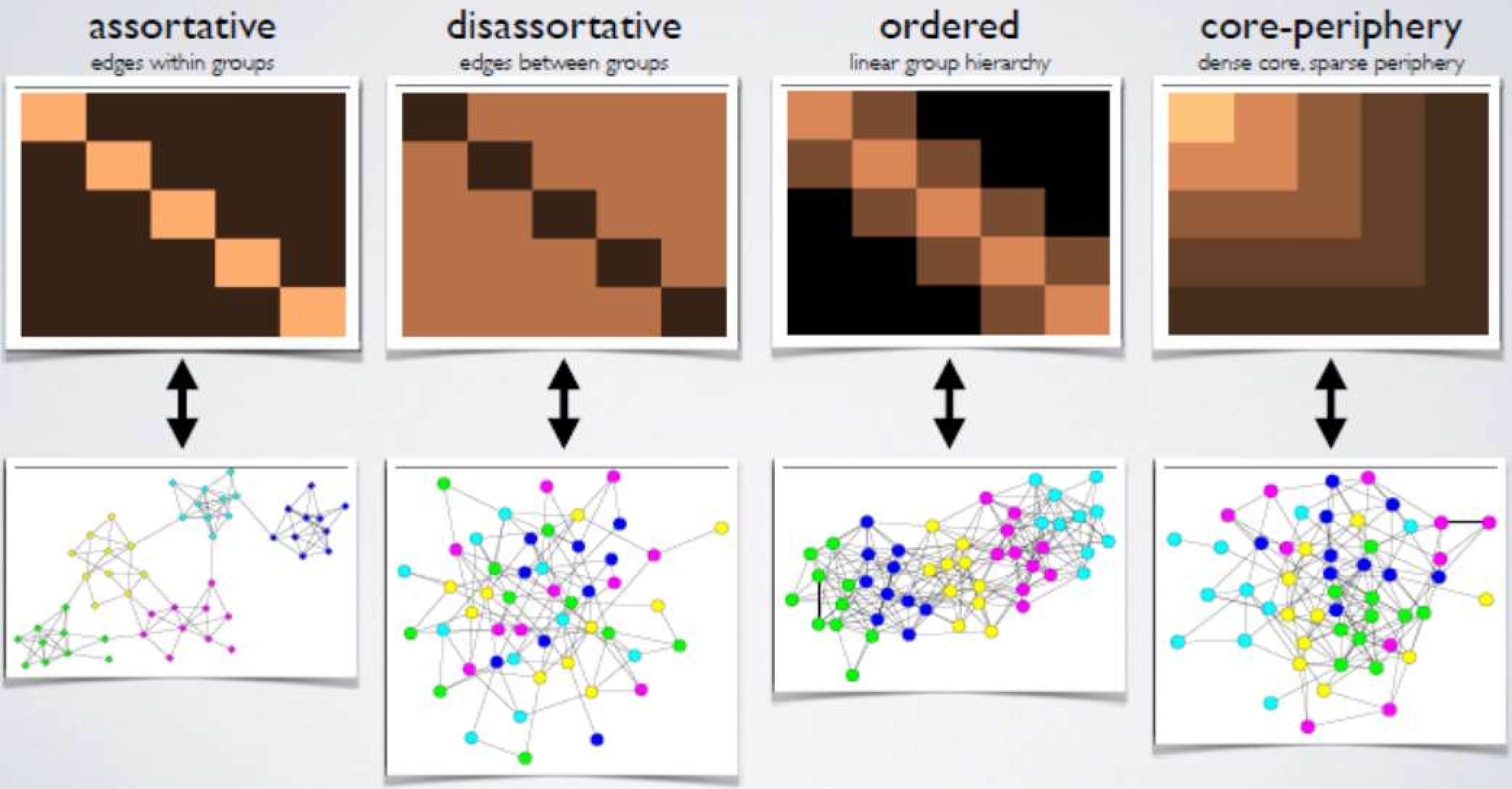
Extensions of core-periphery?!

Limitation:

- There are just two classes of nodes: core and periphery.
- Is a three-class partition consisting of core, semi-periphery, and periphery more realistic?
- Or even partitioning with more classes?
- The problem becomes more difficult as the number of classes is increased, and good justification is needed.

http://www.sciencedirect.com/science/article/pii/S0378873399000192

Possible structures

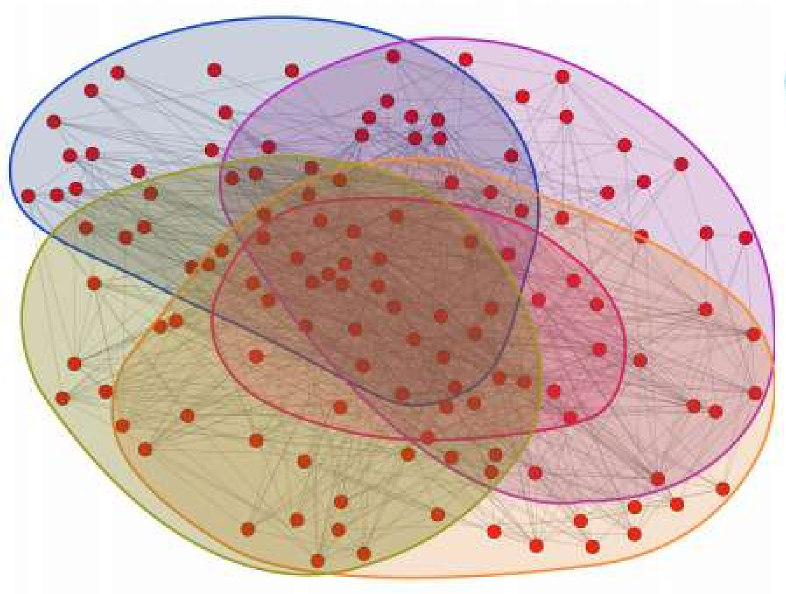


dark shade = 0 (nonadjacent)
light shade = 1 (adjacent)

From Aaron Clauset and Mason Porter

Core and communities

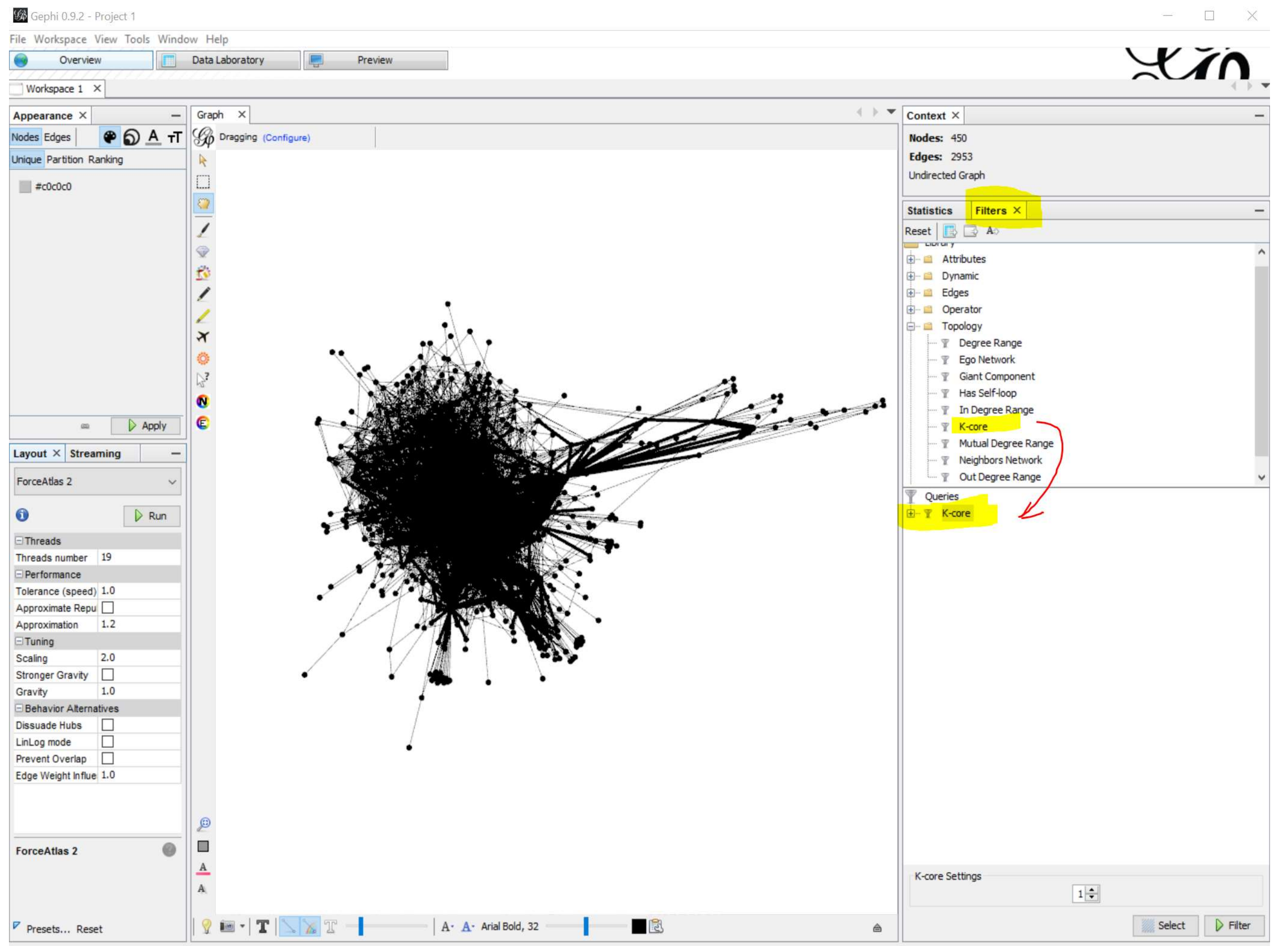
- The network core was traditionally viewed as a single giant community (lacking internal communities, see references [7], [8], [9], [10]).
- Yang and Leskovec (2014, reference [11]) showed that dense cores form as a result of many overlapping communities.
- General observations:
  - foodweb, social, and web networks exhibit a single dominant core, while
  - protein-protein interaction and product co-purchasing networks contain many local cores formed around the central core



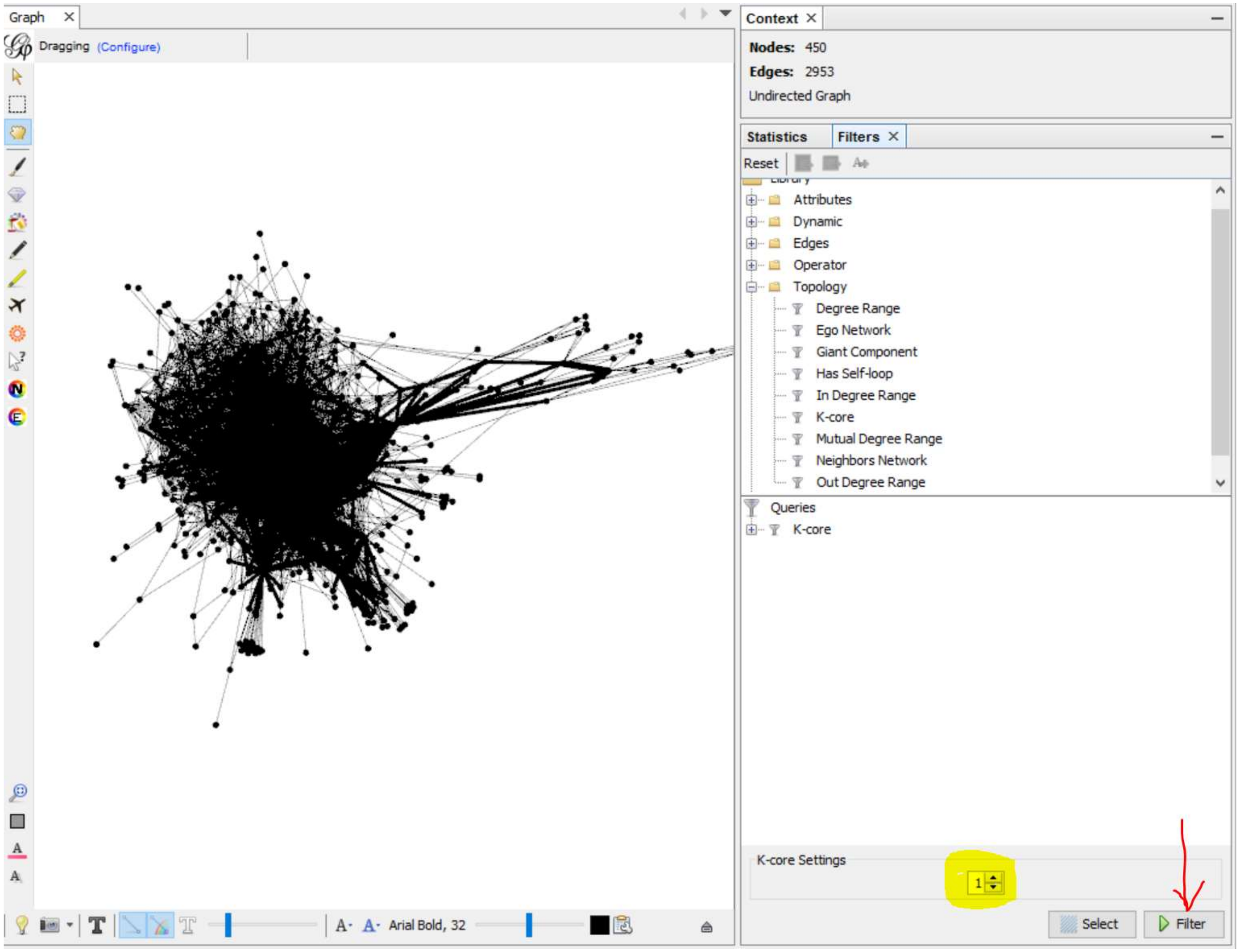
http://ilpubs.stanford.edu:8090/1103/2/paper-IEEE-full.pdf

Finding the Core in Gephi

Under “Statistics” run “average degree” and then use “Filters”

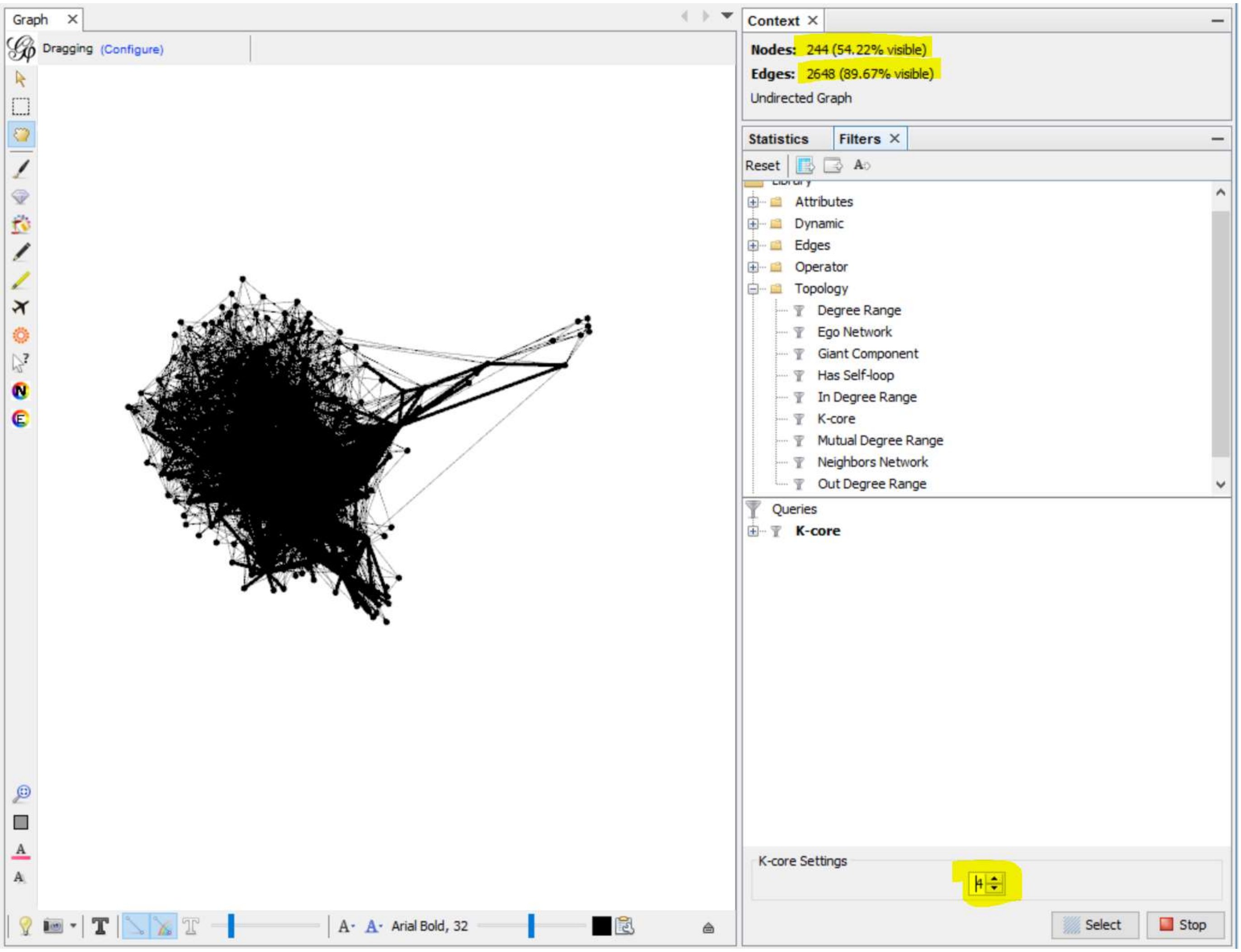


1-core

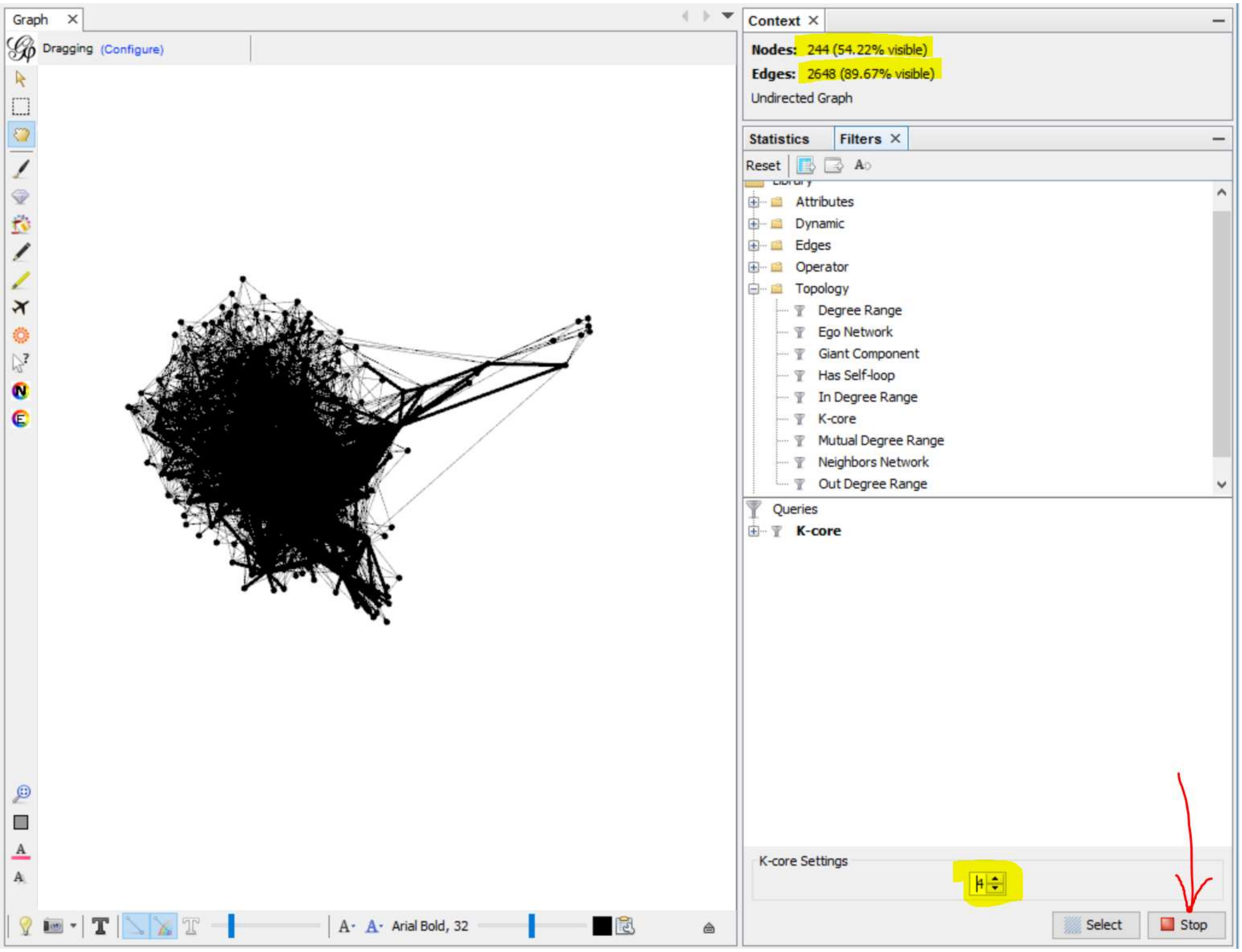


Graph
Dragging (Configure)
Arial Bold, 32
Context
Nodes: 450
Edges: 2953
Undirected Graph
Statistics
Filters
Reset
Attributes
Dynamic
Edges
Operator
Topology
Degree Range
Ego Network
Giant Component
Has Self-loop
In Degree Range
K-core
Mutual Degree Range
Neighbors Network
Out Degree Range
Queries
K-core
K-core Settings
1
Select
Filter

4-core

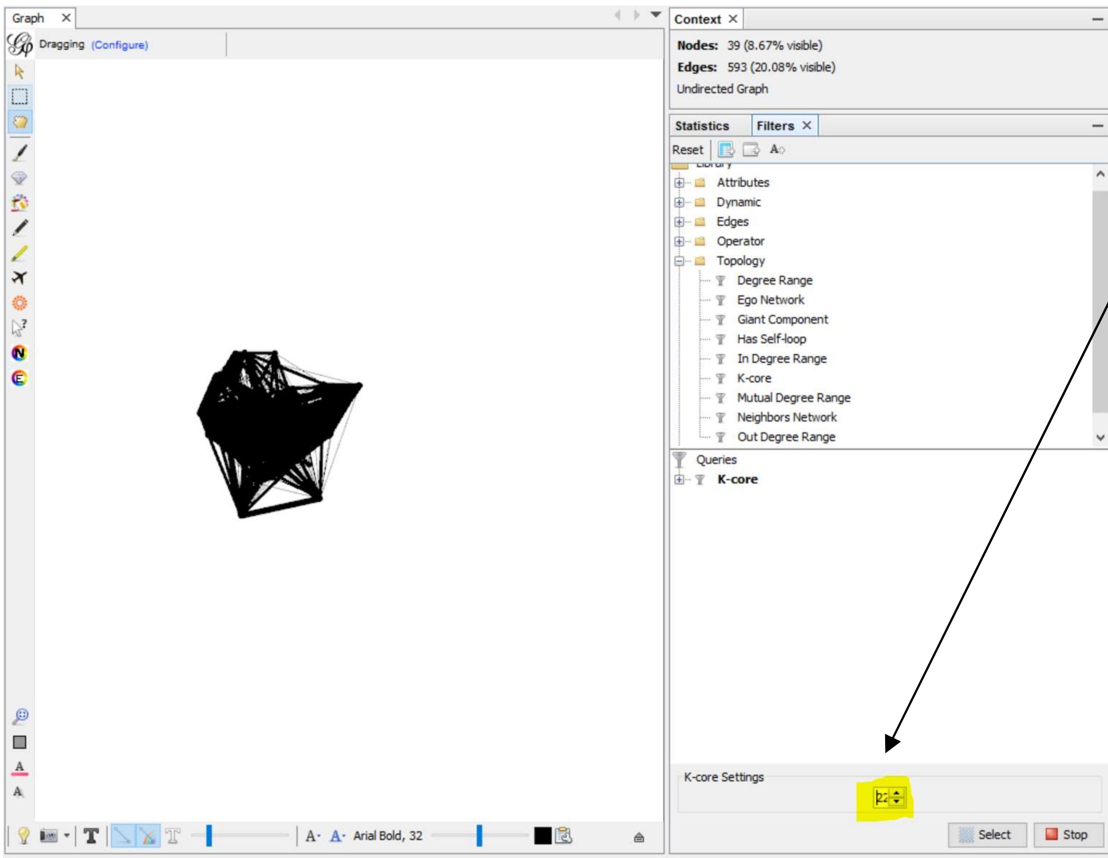

Graph
Dragging (Configure)
Arial Bold, 32
Context
Nodes: 244 (54.22% visible)
Edges: 2648 (89.67% visible)
Undirected Graph
Statistics
Filters
Reset
Attributes
Dynamic
Edges
Operator
Topology
Degree Range
Ego Network
Giant Component
Has Self-loop
In Degree Range
K-core
Mutual Degree Range
Neighbors Network
Out Degree Range
Queries
K-core
K-core Settings
4
Select
Stop

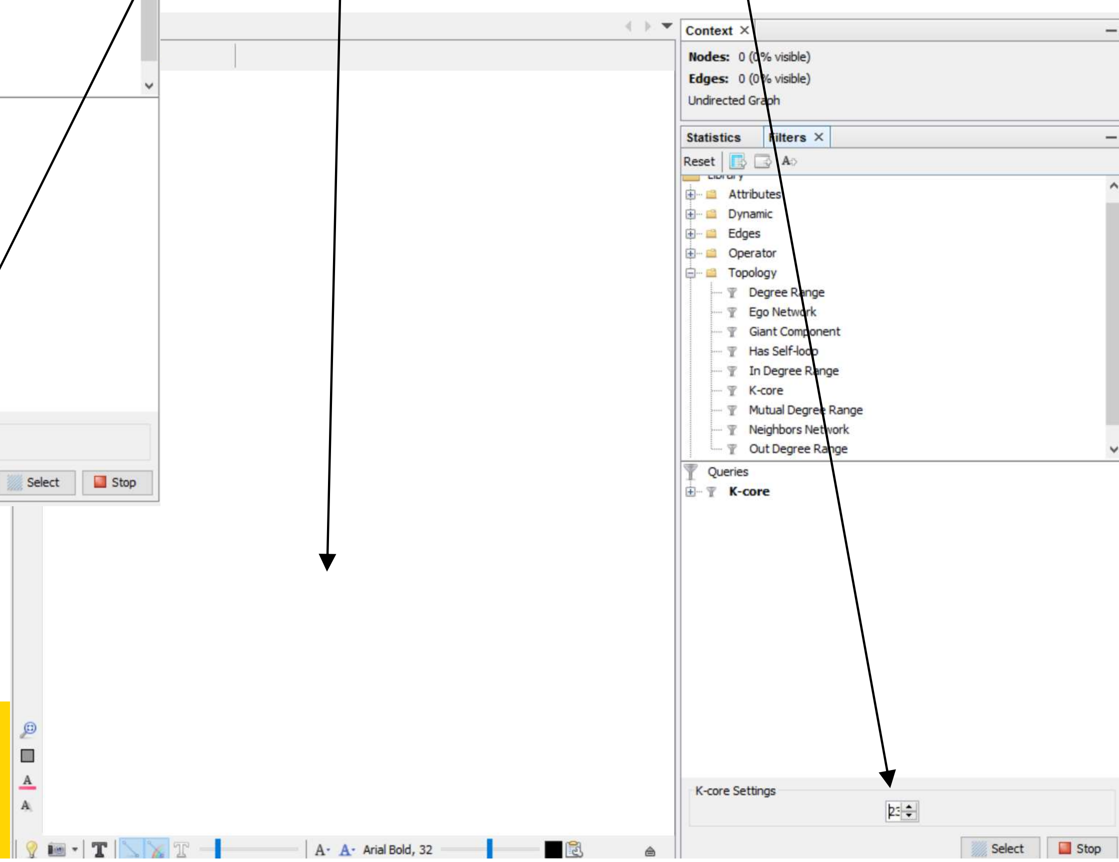
Bring back the whole network



The core of the network

For this network the core is the 22-core, since the 23-core vanishes





Let’s practice in Gephi!

Main References

1. M. E. Newman, Analysis of weighted networks Physical Review E, vol. 70, no. 5, 2004.
2. Borgatti, Stephen P., and Martin G. Everett. "Models of core/periphery structures“ Social networks 21.4 (2000): 375-395.
3. Csermely, Peter, et al. "Structure and dynamics of core/periphery networks.“ Journal of Complex Networks 1.2 (2013): 93-123.
4. Kitsak, Maksim, et al. "Identification of influential spreaders in complex networks." Nature Physics 6.11 (2010): 888-893
5. S. B. Seidman, Network structure and minimum degree, Social networks, vol. 5, no. 3, pp. 269287, 1983
6. Borgatti, Stephen P., and Martin G. Everett. "Models of core/periphery structures." *Social networks* 21.4 (2000): 375-395.
7. J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, “Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters,” Internet Mathematics, vol. 6, no. 1, pp. 29–123, 2009.
8. A. Clauset, M. Newman, and C. Moore, “Finding community structure in very large networks,” Physical Review E, vol. 70, p. 066111, 2004.
9. M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, “Demon: a local-first discovery method for overlapping communities,” in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2012, pp. 615–623.
10. J. Leskovec, K. Lang, and M. Mahoney, “Empirical comparison of algorithms for network community detection,” in Proceedings of the International Conference on World Wide Web (WWW), 2010
11. Jaewon Yang and Jure Leskovec. “Overlapping Communities Explain Core-Periphery Organization of Networks” Proceedings of the IEEE (2014)
12. P. Holme, “Core-periphery organization of complex networks,” Physical Review E, vol. 72, p. 046111, 2005.
13. F. D. Rossa, F. Dercole, and C. Piccardi, “Profiling coreperiphery network structure by random walkers,” Scientific Reports, vol. 3, 2013.
14. J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, “Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters,” Internet Mathematics, vol. 6, no. 1, pp. 29–123, 2009.
15. M. P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, “Core-periphery structure in networks,” SIAM Journal of Applied Mathematics, vol. 74, no. 1, pp. 167–190, 2014