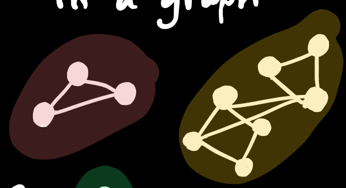
GraphT 4

Previously

1. Connected components in a graph



3

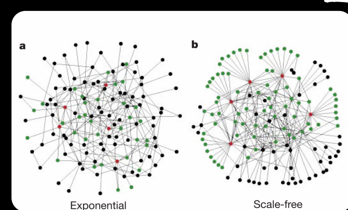
2. Evaluating Graph resilience to attacks via fragmentation

(percolation threshold)

3. Case study:

Exponential graphs (homogeneous) | Scale-free graphs (inhomogeneous)

[Albert et al. 2000]

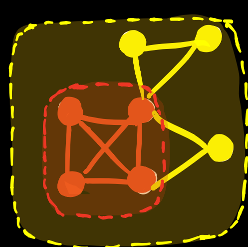


Today

1. Core and periphery organization in a graph

{ k-cores and s-cores

2. graph core decomposition



3. Rich club coefficient

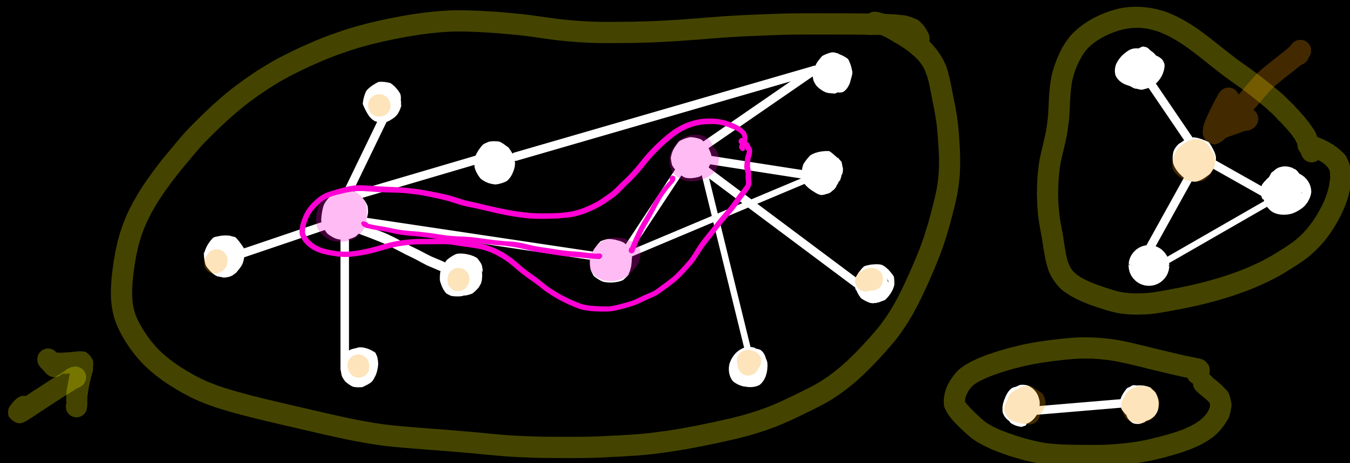
weighted | unweighted

1) Core & periphery organization

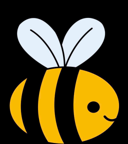
- Connected components offer a fairly course description of the core of a graph.

☹ Since real-world graphs comprise one large component spanning most nodes → does not allow us to identify subsets of graph elements that act as a critical backbone or information-processing core.

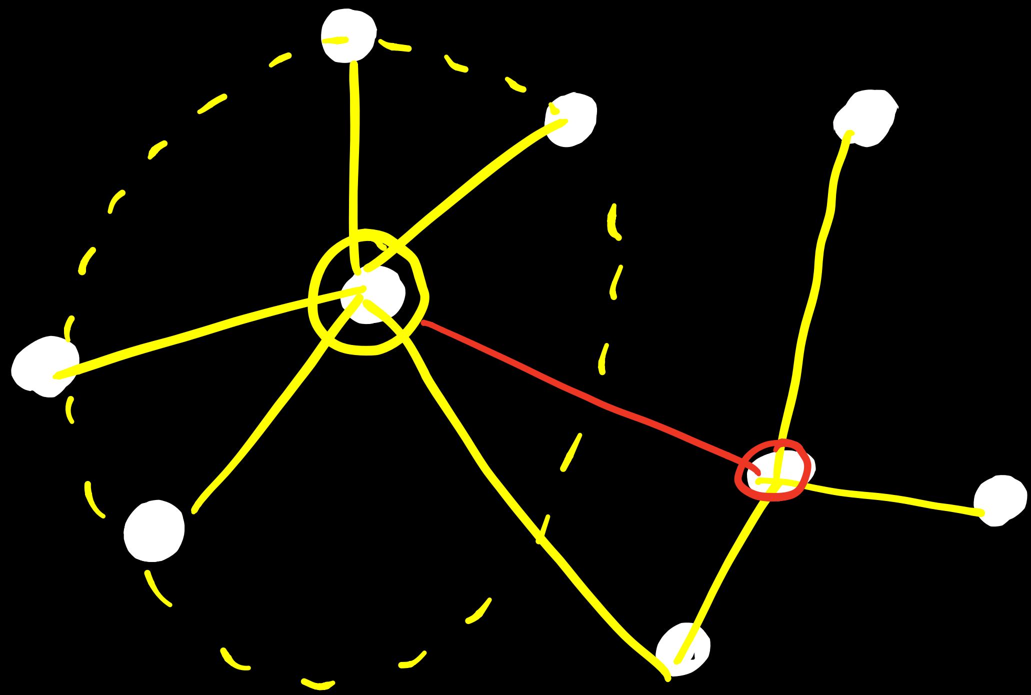
☹ → Using CC, we cannot detect densely connected and topologically central subgraph that exists within the node-connected component.

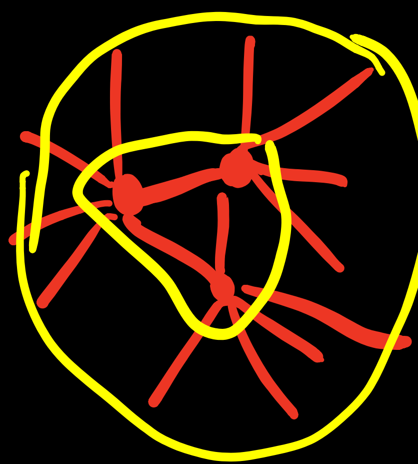


- A graph G with a clear distinction between core and topologically peripheral nodes will show the following properties:
  1. Core nodes should occupy a topologically central position in the graph.
  2. Core nodes should be highly interconnected with each other.
  3. Peripheral nodes should be (at least) moderately connected to core nodes, but sparsely interconnected with each other.



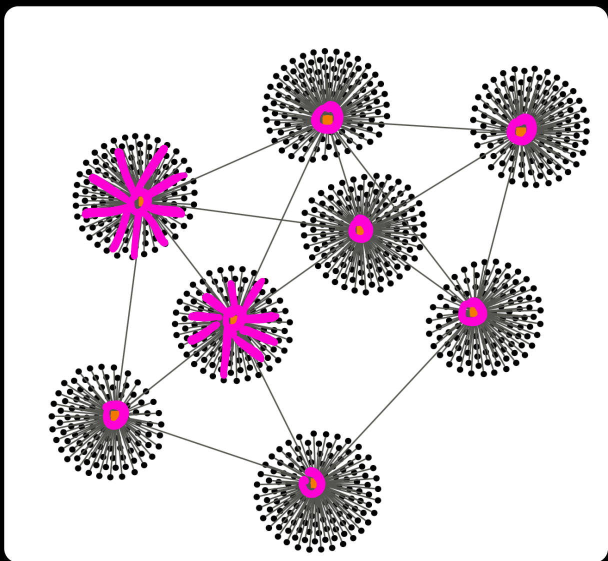
" Can you give an example of a real-world graph with a core-periphery organization?
(define node & edge)





- E.g., such structure appear when resources are scarce.
  → there is some cost involved in forming connections between nodes.

☺ → a core is a cost-effective solution for the integration of distributed graph nodes in the periphery.



[Peixoto et al. 2012]

S.P. Borgatti, M.G. Everett / Social Networks 21 (1999) 375–395

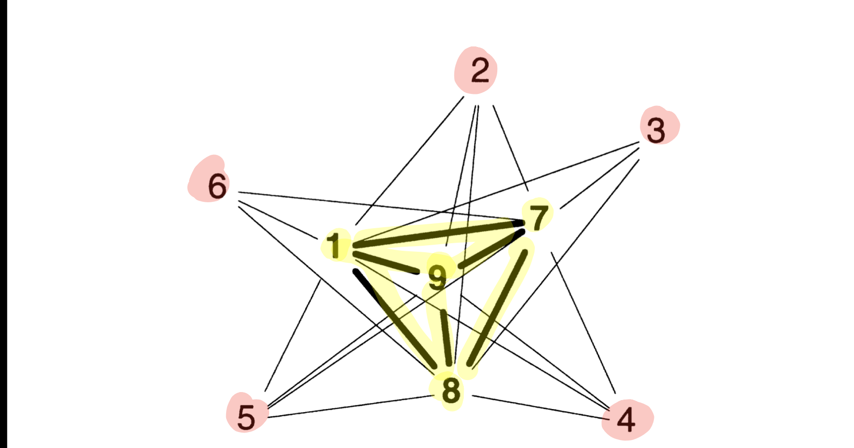


Fig. 3. Core/periphery structure.

[Borgatti et al. 1999]

	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	

[Borgatti et al. 1999]

Blocked adjacency matrix to emphasize the core-periphery pattern.

- core nodes are adjacent to other core nodes
- core nodes are adjacent to some periphery nodes
- periphery nodes do not connect to other periphery nodes

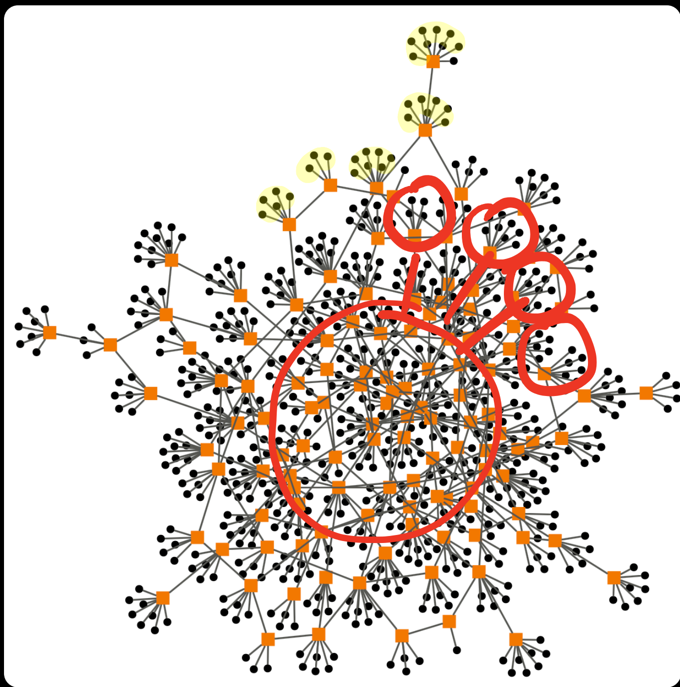
	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	1	1	1	1	1
2	1		1	1	1	1	1	1	1	1
3	1	1		1	1	1	1	1	1	1
4	1	1	1		1	1	1	1	1	1
5	1	1	1	1		0	0	0	0	0
6	1	1	1	1	0		0	0	0	0
7	1	1	1	1	0	0		0	0	0
8	1	1	1	1	0	0	0		0	0
9	1	1	1	1	0	0	0	0		0
10	1	1	1	1	0	0	0	0	0	

[Borgatti et al. 1999]

Idealized core-periphery structure

→ maximally centralized graph

{ 1-blocks with high density
{ 0-blocks with a few ties



[Peixoto et al. 2012]

→ A strong core-periphery organization optimizes robustness to random node failures (Peixoto and Bornholdt, 2012)

↳ the core becomes a focal point for data integration.

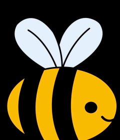
Evolution of robust network topologies: Emergence of central backbones

Tiago P. Peixoto* and Stefan Bornholdt†

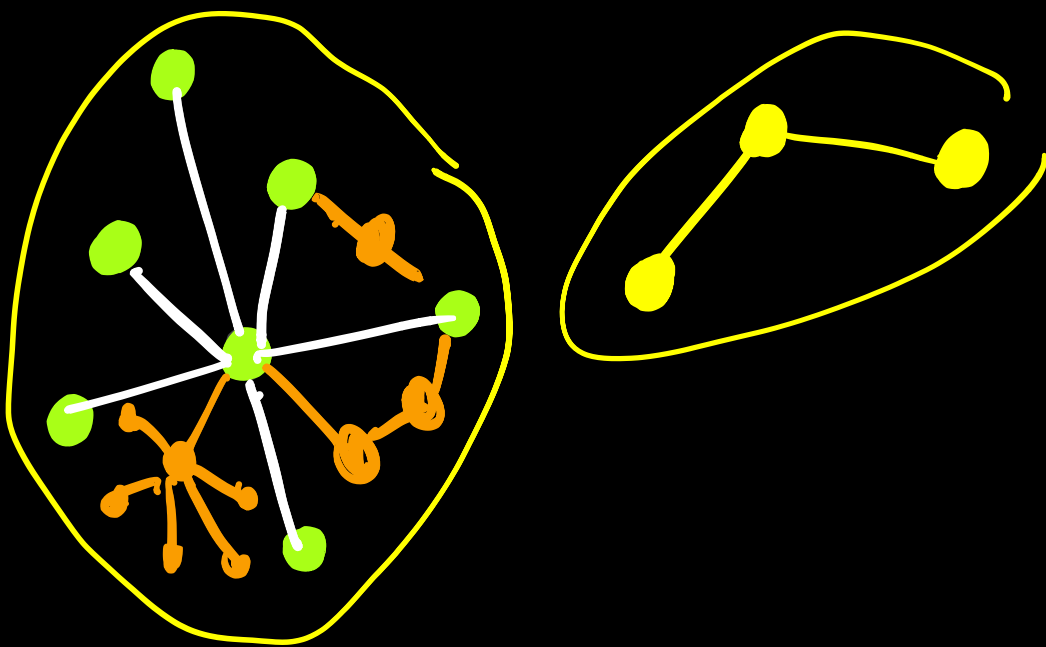
Institut für Theoretische Physik, Universität Bremen, Hochschulring 18, D-28359 Bremen, Germany

We model the robustness against random failure or intentional attack of networks with arbitrary large-scale structure. We construct a block-based model which incorporates — in a general fashion — both connectivity and interdependence links, as well as arbitrary degree distributions and block correlations. By optimizing the percolation properties of this general class of networks, we identify a simple core-periphery structure as the topology most robust against random failure. In such networks, a distinct and small "core" of nodes with higher degree is responsible for most of the connectivity, functioning as a central "backbone" of the system. This centralized topology remains the optimal structure when other constraints are imposed, such as a given fraction of interdependence links and fixed degree distributions. This distinguishes simple centralized topologies as the most likely to emerge, when robustness against failure is the dominant evolutionary force.

[Peixoto et al. 2012]



" So far we have explored 3 topological scales to examine graph structure. What are they? "



node centrality measures → microscale topology

core & periphery → mesoscale topology

components → macroscale topology

K-cores and s-cores

- To characterize the graph structure at multiple scales, one can decompose the graph into k-core modules (Seidman, 1983).
- Idea = nodes with degree less than k are removed until all remaining nodes in the network have at least k connections (k degree).

  → a k-core = subgraph where all nodes have a degree of at least k

  highly interconnected nodes

k-core graph decomposition

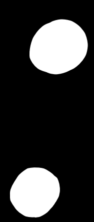
step 1 : remove all nodes with degree 1

recalculate degree of remaining nodes and only nodes v with $d(v) > 1$ are retained to define 2-core of the graph.

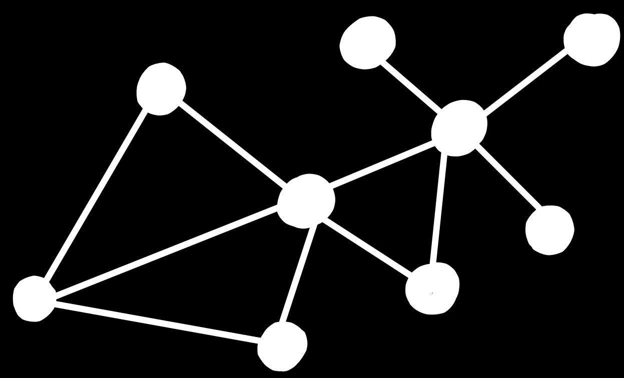
Step 2 : nodes v with $d(v) \leqslant 2$ (and their edges) are removed to define the 3-core.

Step 3 : nodes v with $d(v) \leqslant 3$ (and their edges) are removed to define the 4-core and so on.

At each step, if removal of a node and its edges causes the degree of one of its neighbors to drop below k, then that neighbor is also removed from the graph.



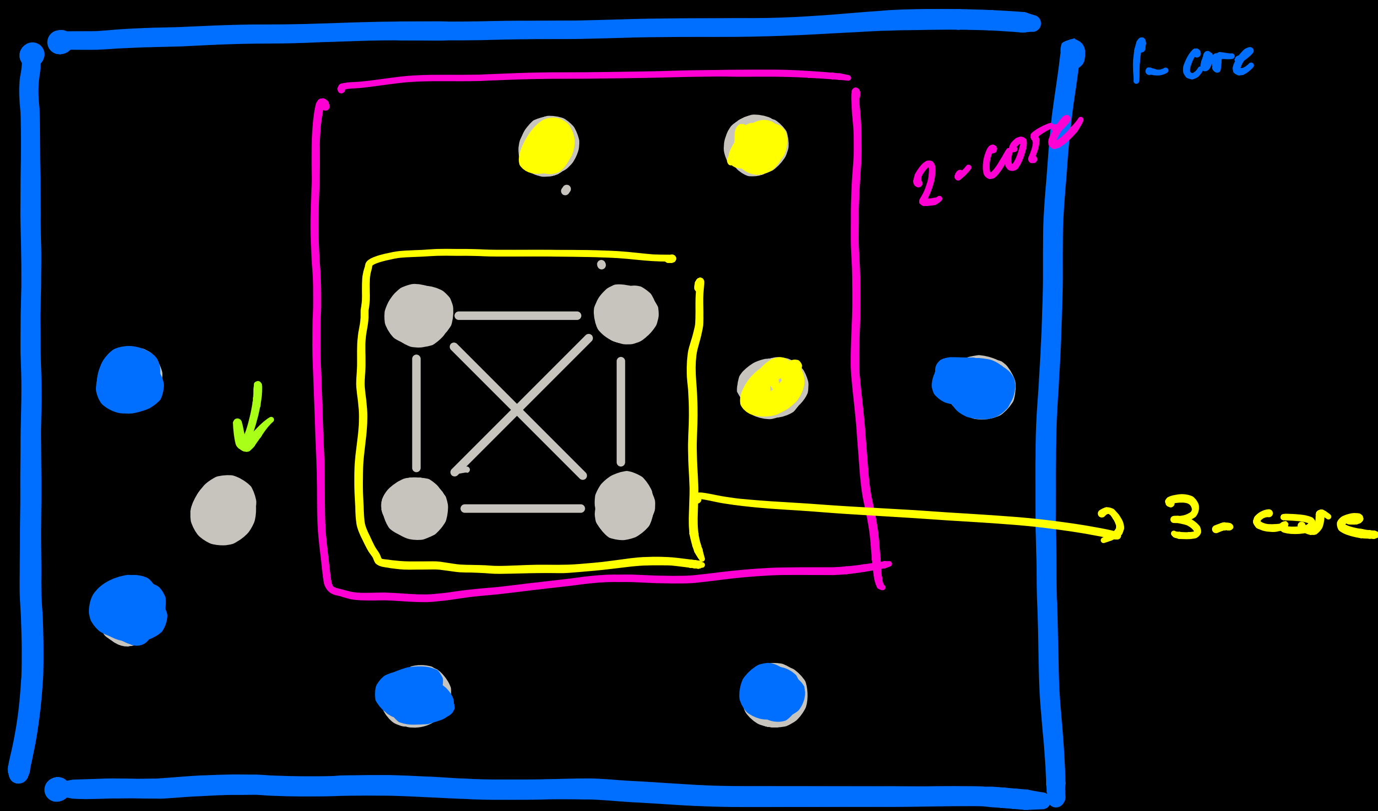
G

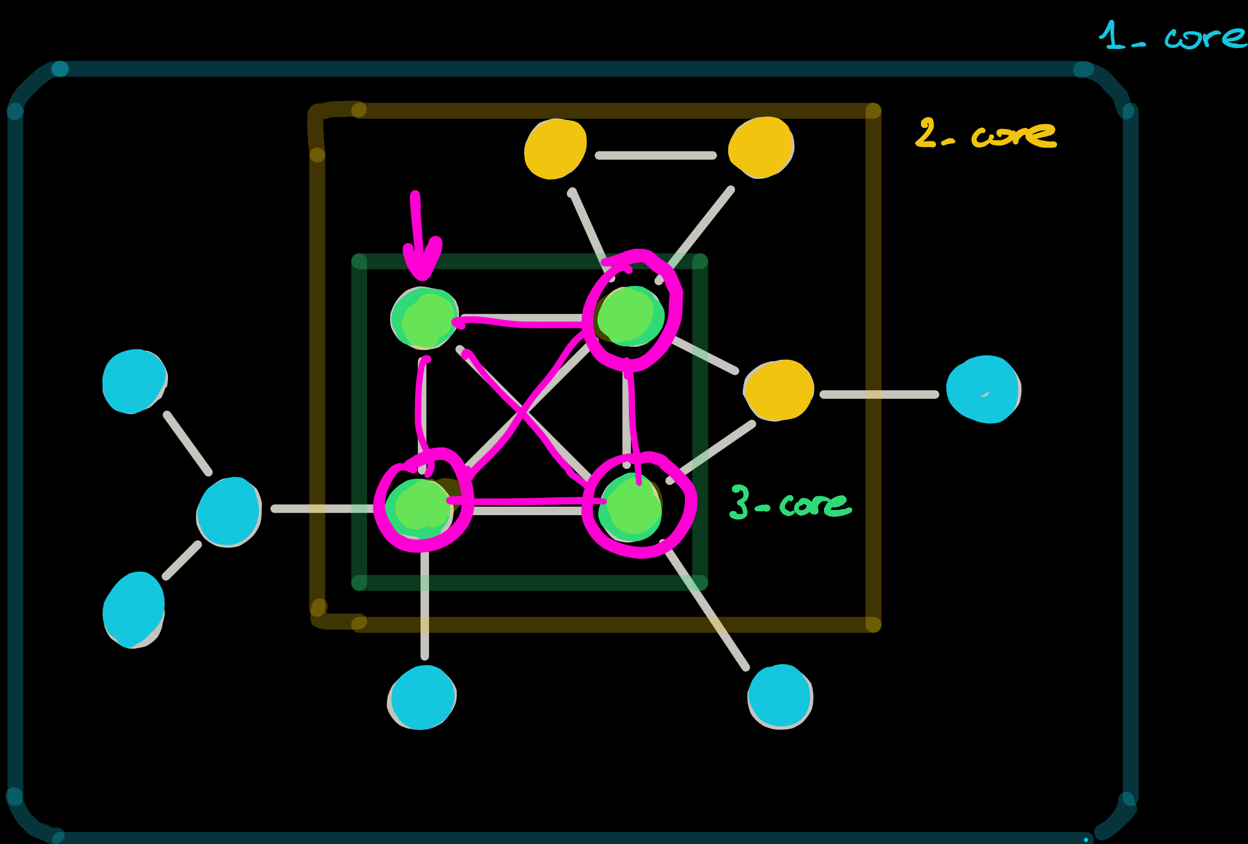


G

2-core

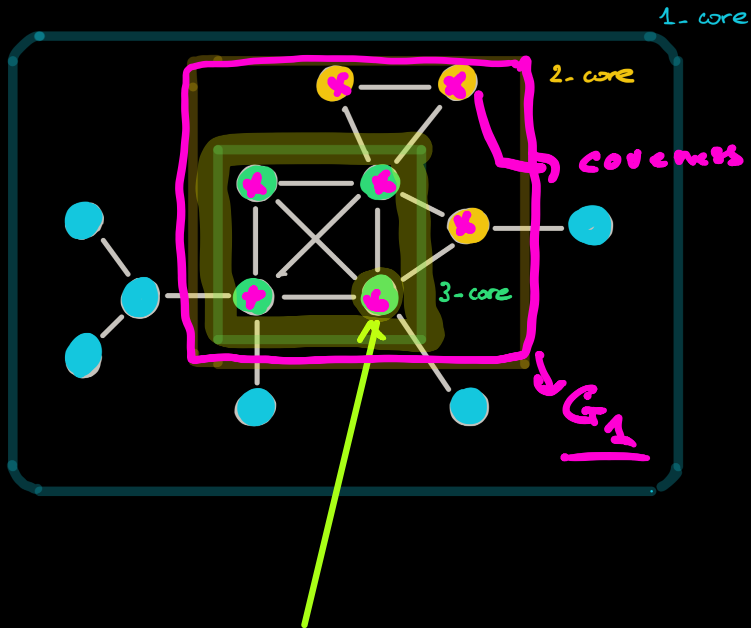
Find the k-cores of the graph below for k=1,2,3.





"What is the relationship between k+1-core and k-core?"

k+1-core is a subgraph of k-core.



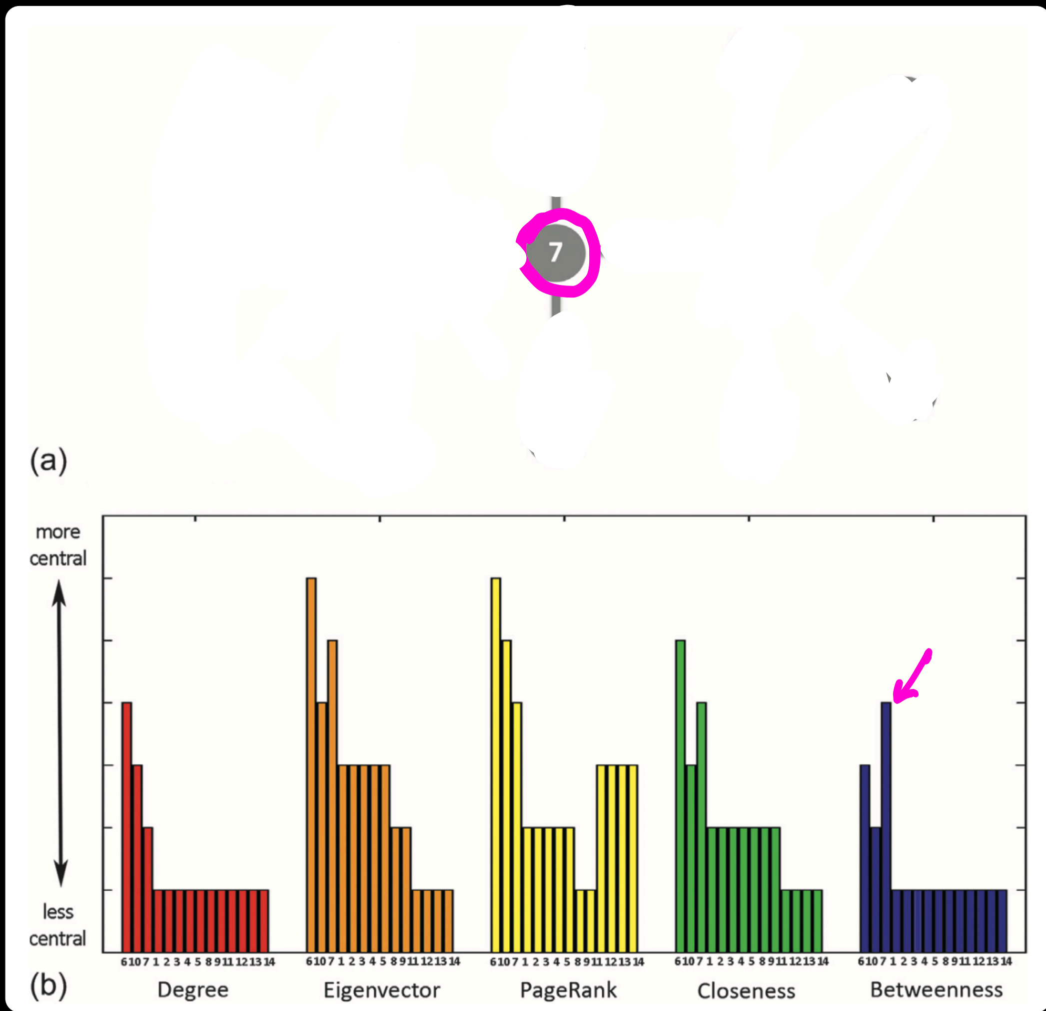
* node coreness: a measure that quantifies the participation of a node to the network core.

How can we measure "node coreness" in a k-core? Any ideas?

we can use centrality measures to index coreness.

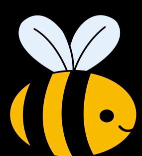
but not all measures are suitable.

e.g., a node with high betweeness centrality and a low-degree is unlikely to belong to an inner k-core of a graph.

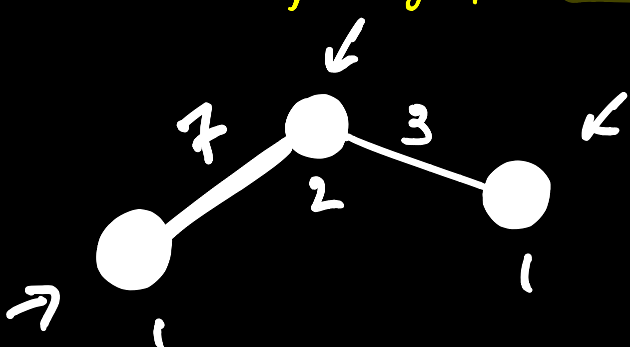


Generalization to s-core

- In weighted networks, an s-core decomposition defines the s-core of a graph comprising all nodes with a strength greater than s.



Do you think that s-cores and k-cores of a graph converge (i.e., are same)?



PHYSICAL REVIEW E **88**, 062819 (2013)

s-core network decomposition: A generalization of *k*-core analysis to weighted networks

Marius Eidsaa and Eivind Almaas*

Department of Biotechnology, NTNU - Norwegian University of Science and Technology, N-7491 Trondheim, Norway

(Received 14 March 2013; revised manuscript received 15 November 2013; published 30 December 2013)

A broad range of systems spanning biology, technology, and social phenomena may be represented and analyzed as complex networks. Recent studies of such networks using k-core decomposition have uncovered groups of nodes that play important roles. Here, we present s-core analysis, a generalization of k-core (or k-shell) analysis to complex networks where the links have different strengths or weights. We demonstrate the s-core decomposition approach on two random networks (ER and configuration model with scale-free degree distribution) where the link weights are (i) random, (ii) correlated, and (iii) anticorrelated with the node degrees. Finally, we apply the s-core decomposition approach to the protein-interaction network of the yeast *Saccharomyces cerevisiae* in the context of two gene-expression experiments: oxidative stress in response to cumene hydroperoxide (CHP), and fermentation stress response (FSR). We find that the innermost s-cores are (i) different from innermost k-cores, (ii) different for the two stress conditions CHP and FSR, and (iii) enriched with proteins whose biological functions give insight into how yeast manages these specific stresses.

DOI: 10.1103/PhysRevE.88.062819 PACS number(s): 89.75.Hc, 05.10.−a, 05.40.Fb, 87.18.−h

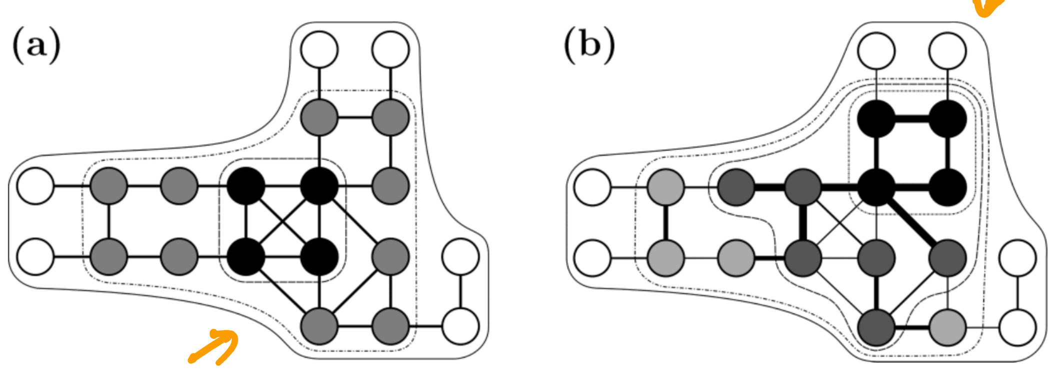
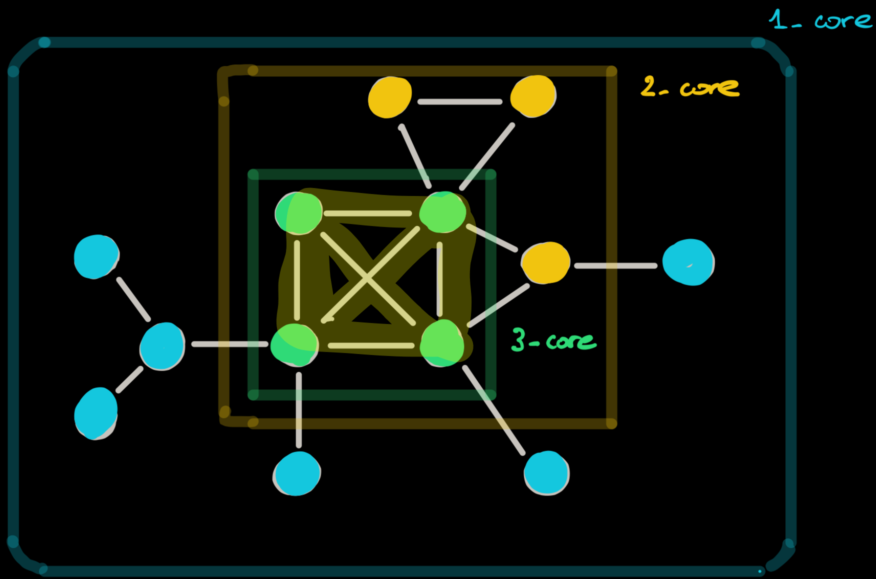


FIG. 1. Illustration of network decomposition: (a) k-core decomposition using node degree. (b) s-core decomposition using node strength. All nodes inside the continuous lines belong to the same core. The link thickness indicates connection strength. Nodes in the same k- or s-shell have identical shading. From outermost towards the core: white, gray, dark-gray (only in the s-core), and black.

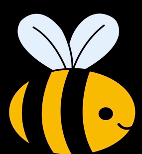
[Eidsaa et al. 2013]

Rich Clubs

- Many of core decomposition strategy rely on the node degree as the defining feature of a core.
  ↳ help identify a highly interconnected set of hub nodes.



- We can define a core based on the density of connectivity between the hubs of a given graph G.



"Do high-degree nodes, which represent potential graph hubs, show higher interconnectivity with each other than expected by chance?"

Q_1: Do the CEOs of Fortune 500 companies show a statistically significant tendency to socialize or do business with each other?

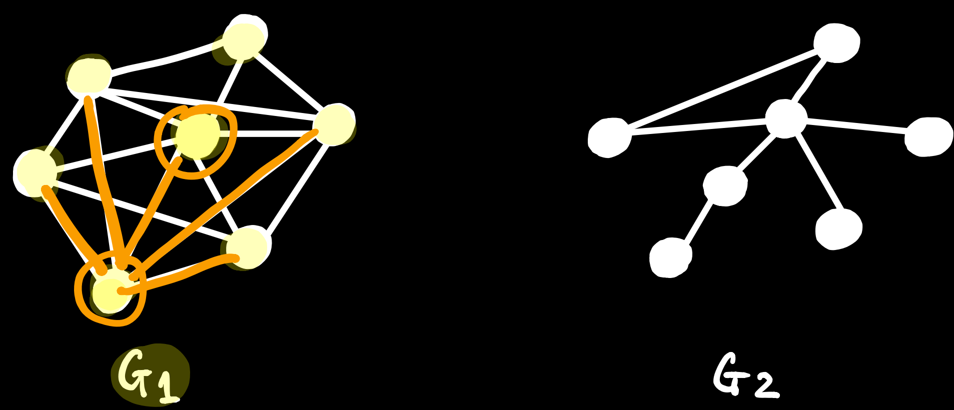
Q_2: Do the most highly cited scientists in a particular discipline are more likely to cite each other?

Q_3: Do major air transportation hubs are highly interconnected?

These are the questions addressed by an analysis of the rich-club properties of a graph.

Unweighted rich-club coefficient

- In a given graph G_1, the "richest" nodes are expected to be densely connected, which points to an interlinked core of high degree nodes.



Which graph is richer? G_1 or G_2?

Rich club coefficient definition

$\varphi(k)$ at each level of degree k is defined as:

G

$$\varphi(k) = \frac{E_{>k}}{N_{>k}(N_{>k}-1)/2}$$

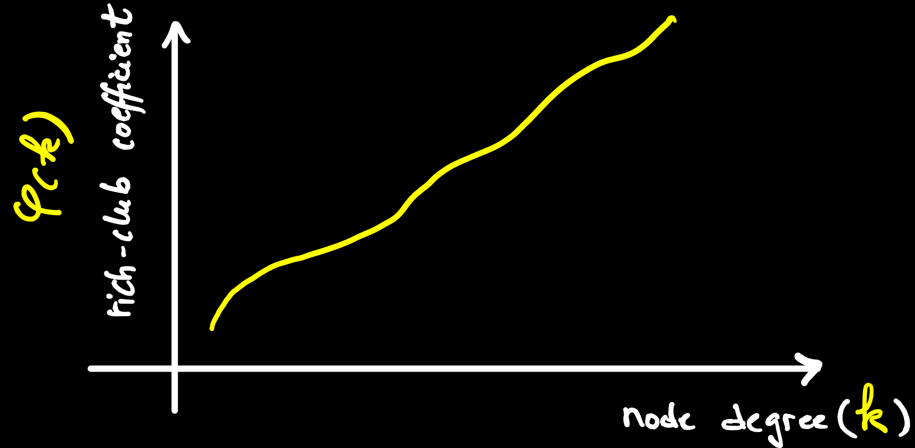
number of edges at nodes of degree $>k$

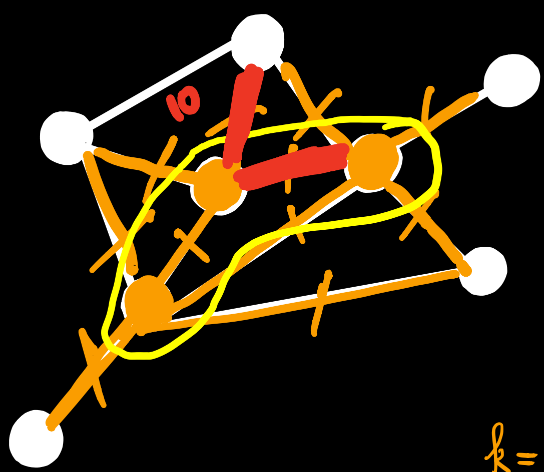
total number of edges in this undirected subgraph

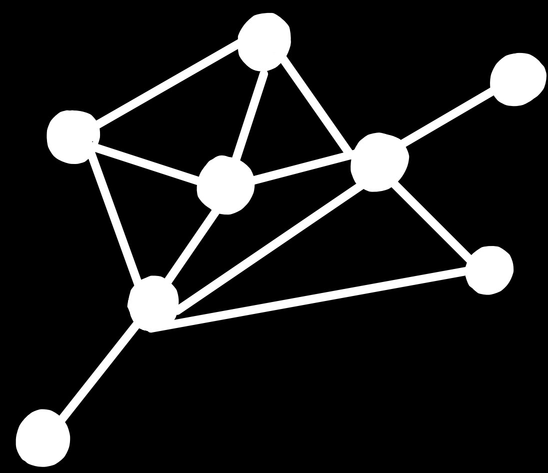
- If $\varphi(k) = 0 \Rightarrow$ no edges of nodes with degree $>k$.

  a cut-off value of "richness", where richness is defined in terms of degree.

- $\varphi(k)$ quantifies the connectivity between the richest nodes in a graph, for a given level of k.







$k = 3$

$$\varphi(k) = \frac{E_{>k}}{\boxed{N_{>k}(N_{>k}-1)/2}}$$

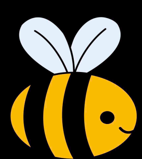
number of edges at nodes of degree > k

← "total" number of edges in this undirected subgraph

$\varphi(k=3) = ?$

$11/3$ $\quad 11 \Big/ \left(\frac{3 \times 2}{2}\right)$

Weighted rich-club coefficient



"How to generalize this to weighted graphs?"

- The rich-club effect can be generalized to weighted graphs (Opsahl et al. 2008):
  1. Rank nodes according to some "richness" measure (e.g., degree)
  2. Apply a threshold to define a subgraph with only nodes that higher than a certain rank.

3. Compute the total sum of weights on the links between the nodes comprising this subgraph.

4. Compute the sum of weights for the same number of edges, but this time for the most highly weighted edges in the whole graph.

5. Take the ratio of the two quantities computed in steps 3 and 4.

$$\varphi^{w}(r) = \frac{W_{>r}}{\sum_{\ell=1}^{E_{>r}} w_\ell^{rank}}$$

Sum of weights on the edges of the subgraph of nodes with rank $> r$

Number of edges in this subgraph

a vector of all edge weights in the graph ranked from largest to smallest weight

Sum over the same number of edges in w^{rank} that are contained in the subgraph used to compute $W_{>r}$

- If $\varphi^{w}(r) = 1$ $\Rightarrow$ sum of weights of edges between the richest nodes is maximal $\Rightarrow$ the most highly weighted edges are the edges of the rich-club.

- If $\varphi^{w}(r) < 1$ $\Rightarrow$ quantifies the fraction of the strongest weights in the graph that exist on edges between the richest nodes.

- $\varphi^{w}(r)$ assesses whether the most highly weighted edges in G are found on links between the richest nodes.

RESEARCH ARTICLE

NETWORK SCIENCES

Systematic inequality and hierarchy in faculty hiring networks

10.1126/sciadv.1400005

Aaron Clauset,[1,2,3]* Samuel Arbesman,[4] Daniel B. Larremore[5,6]

The faculty job market plays a fundamental role in shaping research priorities, educational outcomes, and career trajectories among scientists and institutions. However, a quantitative understanding of faculty hiring as a system is lacking. Using a simple technique to extract the institutional prestige ranking that best explains an observed faculty hiring network—who hires whose graduates as faculty—we present and analyze comprehensive placement data on nearly 19,000 regular faculty in three disparate disciplines. Across disciplines, we find that faculty hiring follows a common and steeply hierarchical structure that reflects profound social inequality. Furthermore, doctoral prestige alone better predicts ultimate placement than a *U.S. News & World Report* rank, women generally place worse than men, and increased institutional prestige leads to increased faculty production, better faculty placement, and a more influential position within the discipline. These results advance our ability to quantify the influence of prestige in academia and shed new light on the academic system.

APPLIED PHYSICS LETTERS **91**, 084103 (2007)

Rich-club phenomenon across complex network hierarchies

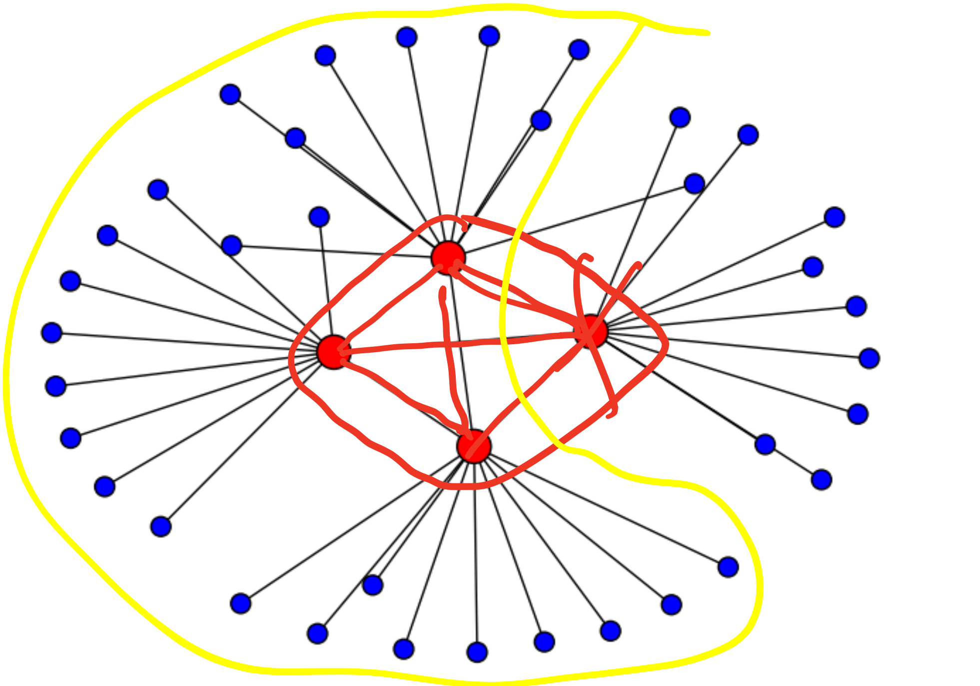
Julian J. McAuley[a)]
Statistical Machine Learning Program, NICTA, Canberra, Australian Capital Territory 0200, Australia and RSISE, Australian National University, Canberra, Australian Capital Territory 0200, Australia

Luciano da Fontoura Costa
Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos SP 13560-970, Brazil

Tibério S. Caetano
Statistical Machine Learning Program, NICTA, Canberra, Australian Capital Territory 0200, Australia and RSISE, Australian National University, Canberra, Australian Capital Territory 0200, Australia

(Received 17 February 2007; accepted 31 July 2007; published online 23 August 2007)

The “rich-club phenomenon” in complex networks is characterized when nodes of higher degree are more interconnected than nodes with lower degree. The presence of this phenomenon may indicate several interesting high-level network properties, such as tolerance to hub failures. Here, the authors investigate the existence of this phenomenon across the hierarchies of several real-world networks. Their simulations reveal that the presence or absence of this phenomenon in a network does not imply its presence or absence in the network’s successive hierarchies, and that this behavior is even nonmonotonic in some cases. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2773951]



why having a rich club in a graph makes it more robust ?