An Efficient Algorithm for Edge-Connectivity Degeneracy
Application to Influencial Node Identification

Stratis Limnios, Dimitrios Thilikos, Michalis Vazirgiannis

Computer Science Laboratory (LIX), École Polytechnique, France

Introduction

Goal: find highly connected subgraphs in terms of edge-connectivity:

- Provide a Degeneracy-like decomposition and ensure a more efficient computation
- Find dense subgraphs and good spreaders among them

Related work:

- Degenaracy frameworks for vairous similar tasks already exist as the k-core and k-truss decomposition (triangle-based degeneracy for the later)
- Several functions exist in the literature that aim to solve this problem but for most formulations those functions are also **hard to approximate.** [H. Gabow, Journal of Computer and System Sciences '1995]
- It was shown that most efficient spreaders are located within the k-core and the k-truss of the network [Kitsak et al., Nature Physics '10], [M. Rossi, WWW'15]

Contributions:

- An algorithm for an edge-connectivity based extension of the k-core, in order to compute it for specific k values
- Locate nodes that perform faster and wider epidemic spreading

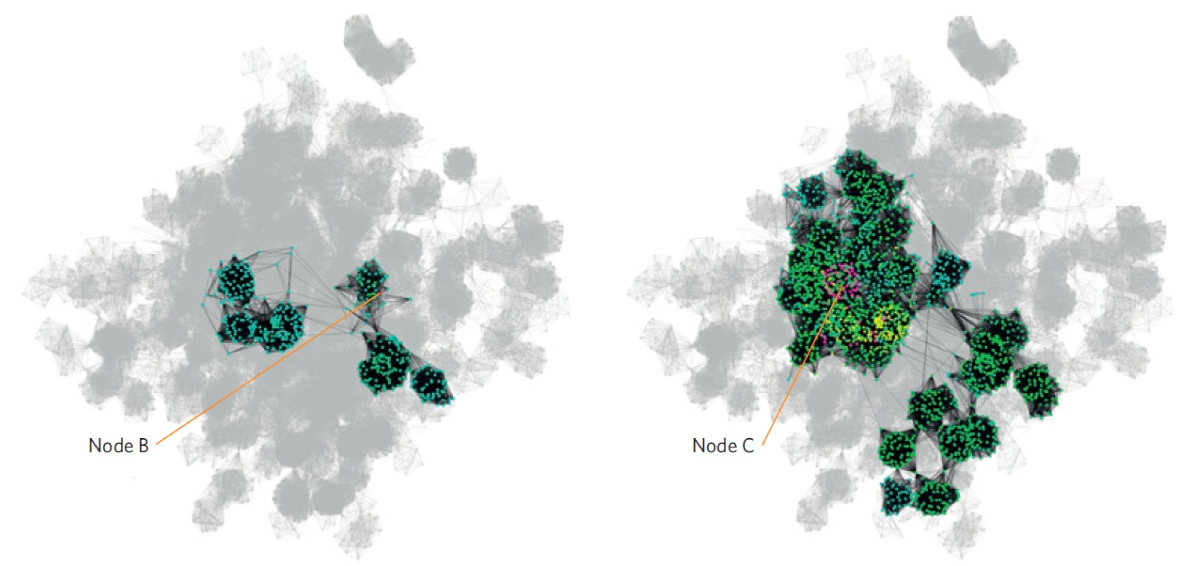


Figure: Influence of *Node B* having high degree and low k-core number versus influence of *Node C* having both high degree and k-core number [Kitsak et al., Nature Physics '10]

Preliminary Concepts and Definitions

DEFINITIONS: **k-core subgraph C_k, Core number c_v**:

- C_k is k-core subgraph of $G = (V, E)$ if it is a maximal connected subgraph in which all nodes have degree at least k
- Each node $v \in V$ has core number $c_v = k$, if it belongs to a k-core but not to a $(k+1)$-core

DEFINITION: **Edge-Connectivity** : A graph is d-edge connected if it has at least two vertices and for every two vertices there are d **edge disjoint paths** between them.

THEOREM: **Menger** [K. Menger, Fundamenta Mathematicae '1927]:

- Let x and y be two distinct verticies of G then the size of the minimum edge cut for x and y is equal to the number of **pairwise edge-independent paths** from x to y.
- **Extension to subgraphs :** A maximal subgraph disconnected by no less than a k-edge cut is identical to a maximal subgraph with a minimum number k of edge-independent paths between any x, y pairs of nodes in the subgraph.

DEFINITIONS: **k-Edge-Connectivity Core Number**:

- As an application of **Menger's theorem** we define an **edge-connectivity function** in means of the minimal cut in a graph G :
$$\lambda(S) = min\{|\{e \in E(G) | \{e\} \cap S \neq \varnothing \wedge \{e\} \cap (V(G) \setminus S) \neq \varnothing\}|\}$$
A graph G is **d-edge connected** if and only if $\lambda(G) \geq d$.
- Let G be a graph. We define the **edge-connectivity degeneracy** of G as a follows:
$$\lambda^*(G) = \max\{\lambda(H) \mid H \subseteq G\}$$

FONDAMENTAL RESULTS : **Density and Degeneracy Inequalities** : With ϵ being $\epsilon(G) = |E(G)|\,|V(G)|$, and δ the minimal degree, $\delta(G) = min\{deg(v) | v \in V(G)\}$:
$$2\epsilon^*(G) \geq \delta^*(G) \geq \lambda^*(G) \geq \epsilon^*(G)$$

k-Edge-Connectivity Core Algorithm Illustration (k=3)

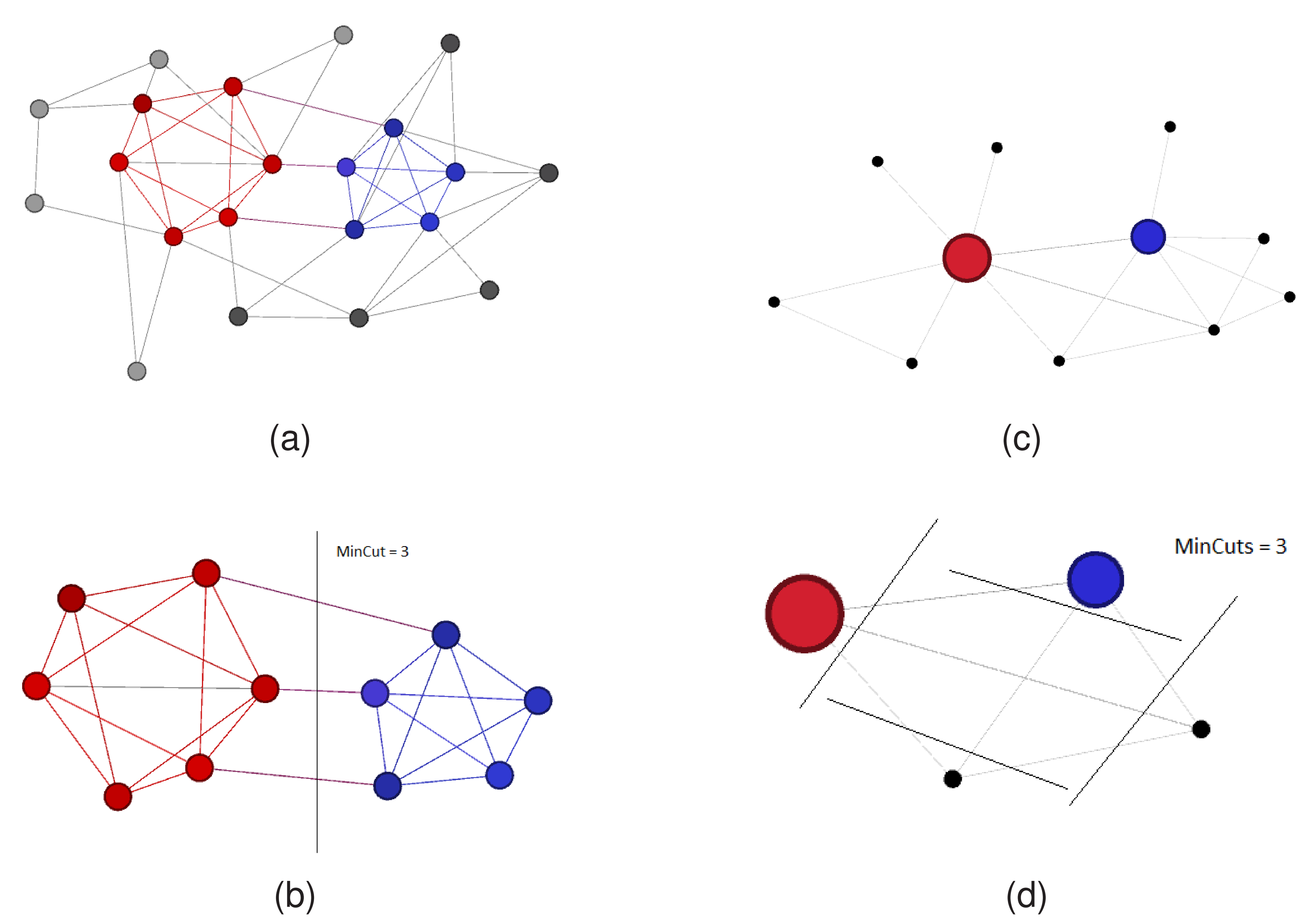


Figure: Illustration of the algorithm on an artificial graph for k=3 (a) Compute the deepest k-core in the given graph, (b) Find all the minimal cuts $\leq$3 (c) Contract the unbroken parts into single nodes and find the k-core of the obtained graph, (d) Find again all the minimal cuts $\leq$3

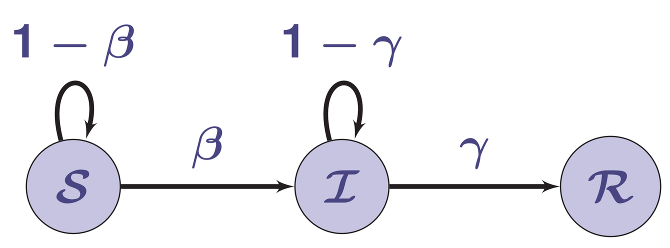
- Compute the deepest core, then find all the subgraphs that cannot be split by a minimal cut inferior to the threshold k,
- Contract the obtained subgraphs into the original graph and start over,
- The algorithm terminates after finding only singletons.

Proposed Methodology

- Aim to identify those **single spreaders** in a network that will achieve an efficient spreading of information

How to simulate the spreading process?

- We apply the **S**usceptible-**I**nfected-**R**ecovered (SIR) model [Easley & Kleinberg, Cambridge University Press '10]
  - Set one node to be infected (single spreader), as chosen by different methods
  - Infected nodes can infect their susceptible neighbors with probability β
  - A node that has been previously infected can recover from the disease with a probability γ



Visualisation

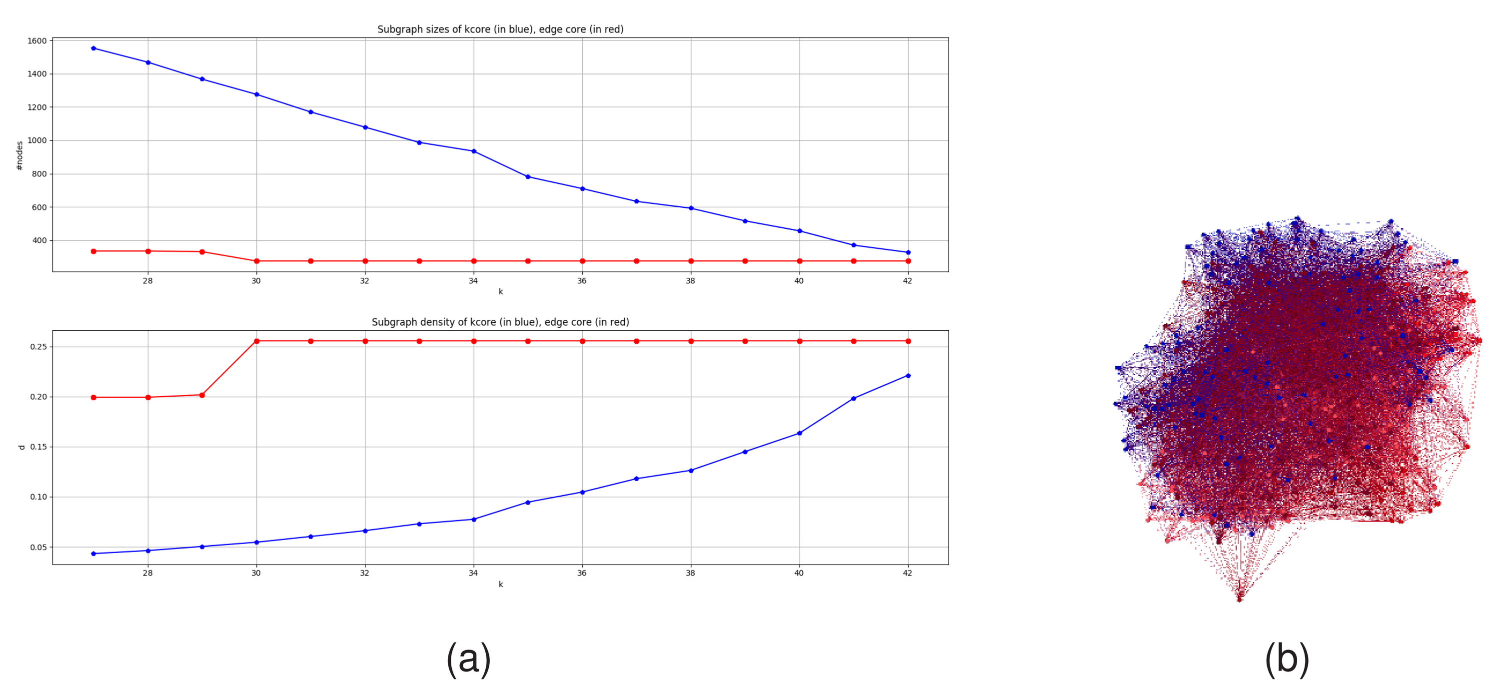


Figure: (a) Density and size comparison between equivalent edge-connectivity core subgraphs (red line) and k-core subgraphs (blue line), (b) representation of the 22^{nd} edge connectivity core (red) inside the 40^{th}-core (blue)

Datasets

Network name	# Nodes	# Edges	k-core	K-truss	$\lvert C\rvert - \lvert T\rvert$	$\lvert T\rvert$	Epidemic threshold
EMAIL-ENRON	33,696	180,811	43	22	230	45	0.0084
WIKI-VOTE	7,066	100,736	53	23	286	50	0.0072

Experimental Evaluation

		Time Step				
	Method	2	6	10	*Final step*	*Max step*
EMAIL-ENRON	edge core	**21.47**	**350.13**	250.93	**2,647.74**	**28**
	truss	8.44	204.08	355.84	2,596.52	33
	core	4.78	152.55	**364.13**	2,465.60	37
	top degree	6.89	155.48	357.08	2,471.67	36
WIKI-VOTE	edge core	**5.15**	**24.72**	**52.74**	**626.09**	**34**
	truss	2.92	15.27	42.46	560.66	52
	core	1.92	10.65	32.40	466.01	57
	top degree	2.43	12.05	35.55	502.88	62

Table: Average number of infected nodes for some steps of the SIR model, using β close to the epidemic threshold of each graph and $\gamma = 0.8$

- The Edge-core method achieves higher infection rate during the first steps
- The total number of infected nodes at the end of the process is larger, while the fade out occurs earlier

stratis.limnios@polytechnique.edu