NetSci 2024

Overlapping Communities Detection by Commonality

Igor Kanovsky
Peres Academic Center

Dmitri Deinega
Ariel University

Vadim Levit
Ariel University

https://igorkan.github.io/

What a community is?

- There is no universally accepted definition for community in graphs.
- Graph partitioning problem. Dozens of algorithms and techniques.
- Sets of nodes that the algorithm finds are then called “clusters,” “communities,” “groups,” “classes,” or “modules”.
- This is Ok! The applications are context-depended.
- Do we need one more algorithm? If it has a use case!
- Some intuitive understanding of **overlapping communities** can be derived from social networks…

Natural Community!!

- Each node “knows” his community’s members: local property.
- Each node may belong to more than one community
- **Topologically**:
  two nodes **“surely”** belong to the same community if they have a significant number of common neighbors.

Nodes Commonality

- $N(i)$ is the neighborhood of a node i
- $commonality(N(i), N(j))$ is a function of two nodes to quantify the status of their common neighbors.
- Exists a threshold c_0 , if $commonality(i,j) > c_0$ nodes i,j “for suer” belongs to the same community.
- For different type of networks, *commonality* may be different functions
- Commonality may be a pretrained neural network.

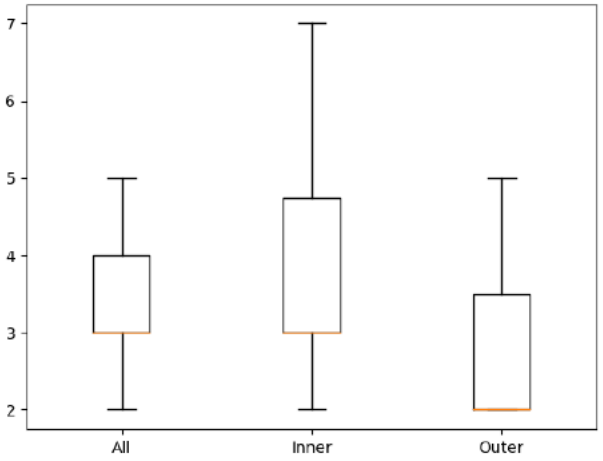
Commonality - Jaccard coefficient

$$c(i, j) = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|}$$

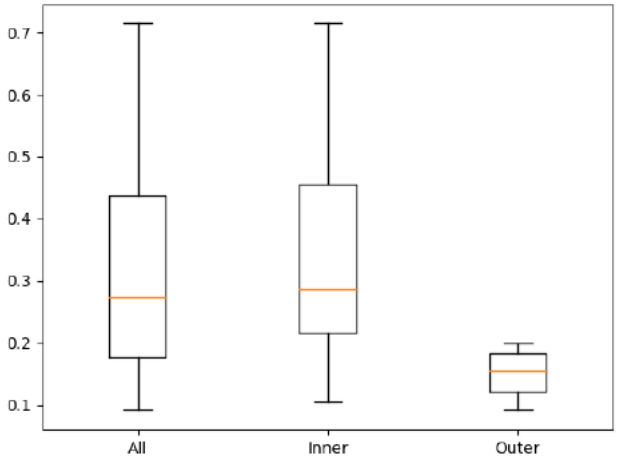
- *commonality* $c(i,j)$ of two nodes *i* and *j* is a fraction of common neighbors
- The simplest, but may be not the best
- $c($ $)=2/5$, $c($ $)=3/4$

commonality concept test

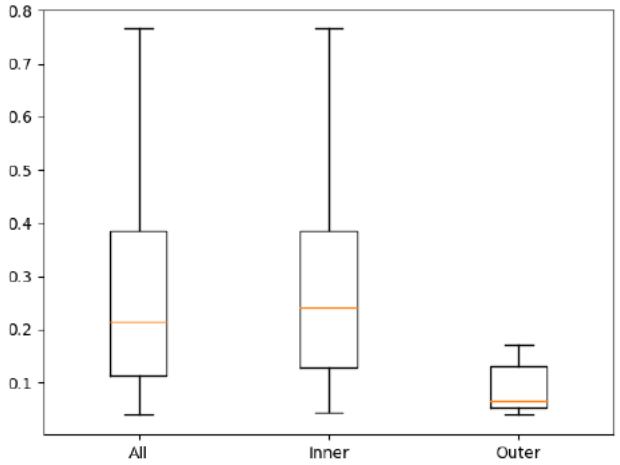
- calculate the commonality for different real-world networks with “ground truth”
- f1= $|N(i) \cap N(j)|$, f2= $\frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|}$, f3= $\frac{|N(i) \cap N(j)|^2}{|N(i) \cup N(j)|}$
- For Zachary’s Karate Club commonality distribution:



f1



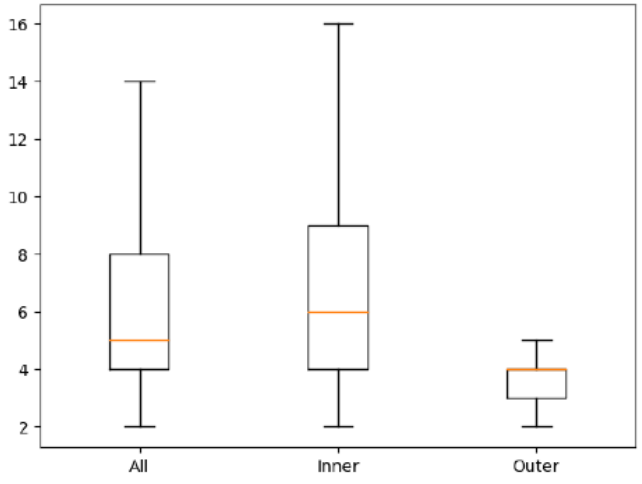
f2



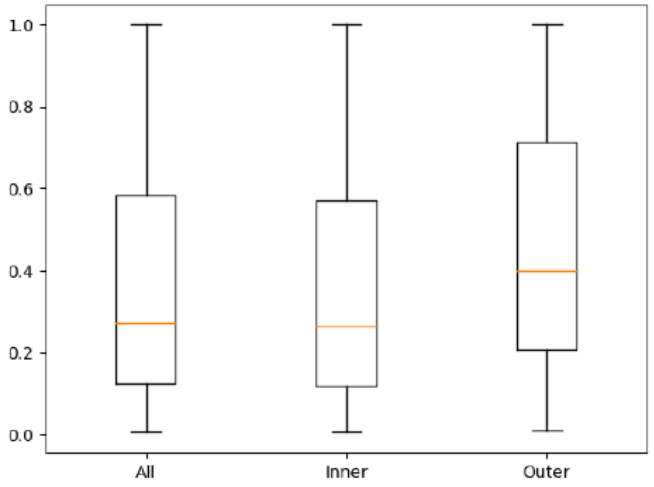
f3

commonality distribution

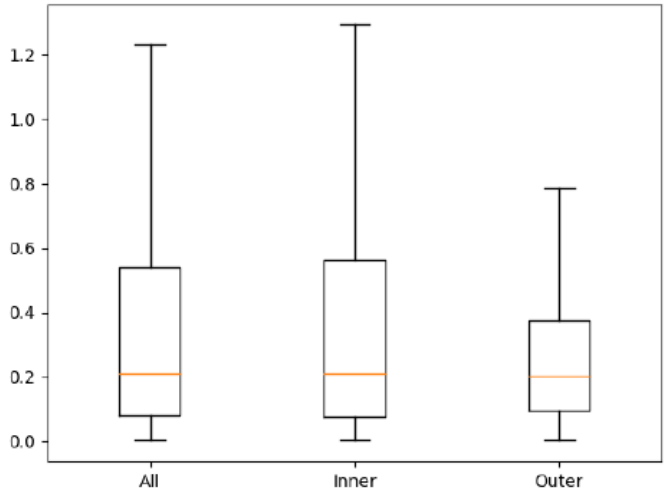
- For DBLP a co-authorship network Jaewon Yang and Jure Leskovec. “Defining and evaluating network communities based on ground-truth”. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. 2012, pp. 1–8.
- f1= $|N(i) \cap N(j)|$, f2= $\frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|}$, f3= $\frac{|N(i) \cap N(j)|^2}{|N(i) \cup N(j)|}$



f1

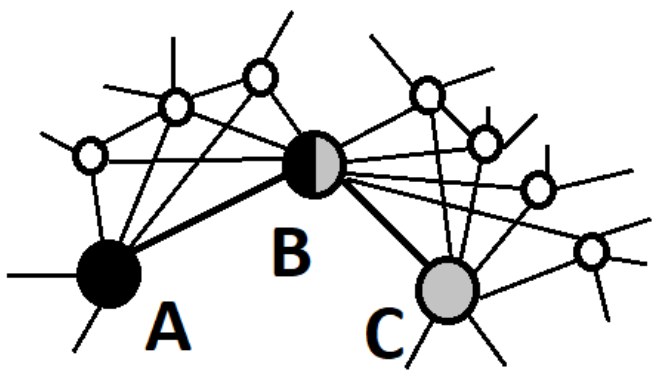


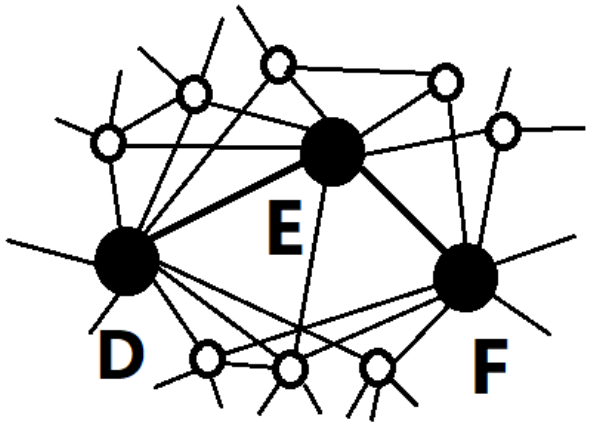
f2



f3

inner community path

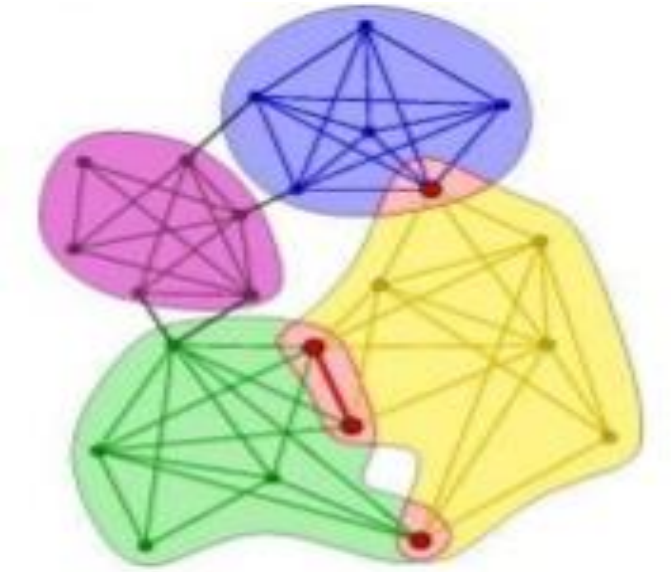




- A link (*i*,*j*) is inside a community “for sure” (inner link) if $c(i,j) > c_0$
- (A,B), (B,C), (D,E), (E,F) are "for sure" inner links. c(A,B)=3/7, c(B,C)=4/10, c(A,C)=1/13, c(D,E)=4/11, c(E,F)=3/11, c(D,F)=4/13.

- A path is inside a community “for sure” (inner path) if it consists of inner links and for each couple of nodes (*p*,*l*) on distance two $c(p,l)>c_0$
- (D,E,F) is inner path.

Natural Community - definition



- Natural community is a set of nodes having an inner path between each couple of them.
- Natural communities are overlapping.
- Clusters and communities may be different
- Some of links inside a community may have a small *c* value. They are not part of inner path.

Algorithm for a native community detection

Input: network $G=(V,E)$, threshold value c_0;
Output: A community $C \subset G$.

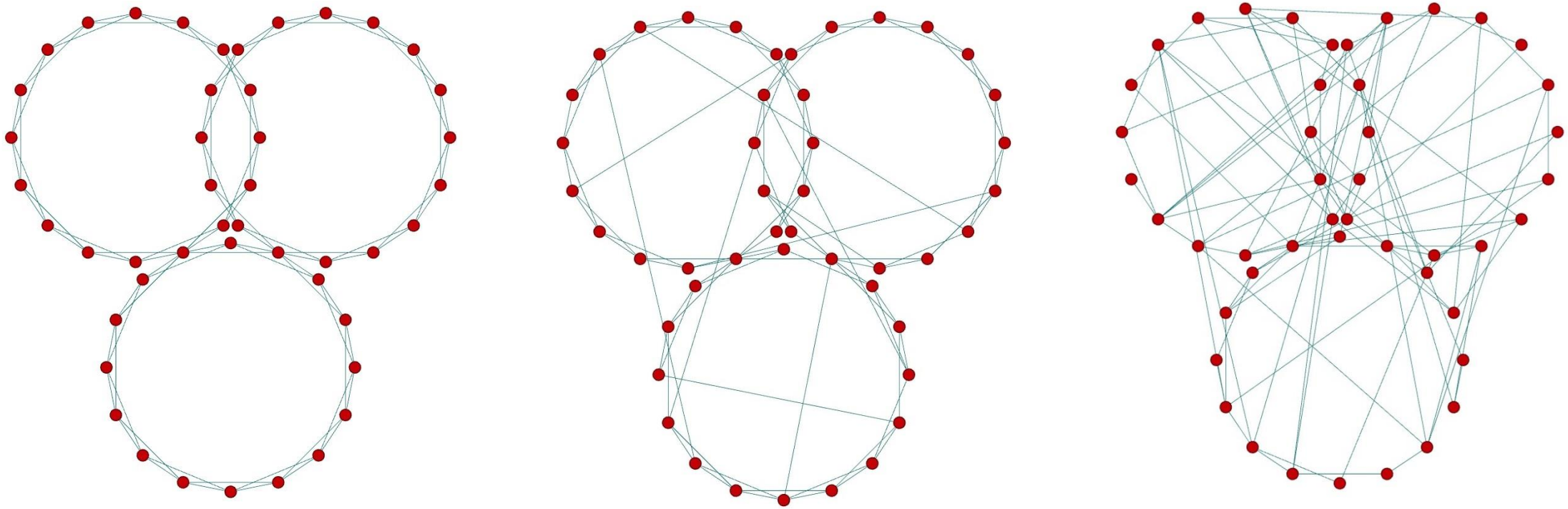
1. Start with arbitrary node v , C={v};
2. Loop for each $w \in N(v)$
   if $c(v,w) > c_0$ put (v,w) into queue Q; break;
3. loop while Q is not empty
   a. pop (v,w) from Q; add w to C;
   b. loop for each $u \in N(w)$
      if $(c(w,u) > c_0$ and $c(w,v) > c_0$ and $u \notin C)$
      put (w, u) into Q;

$N(v)$ is set of the node v neighbors

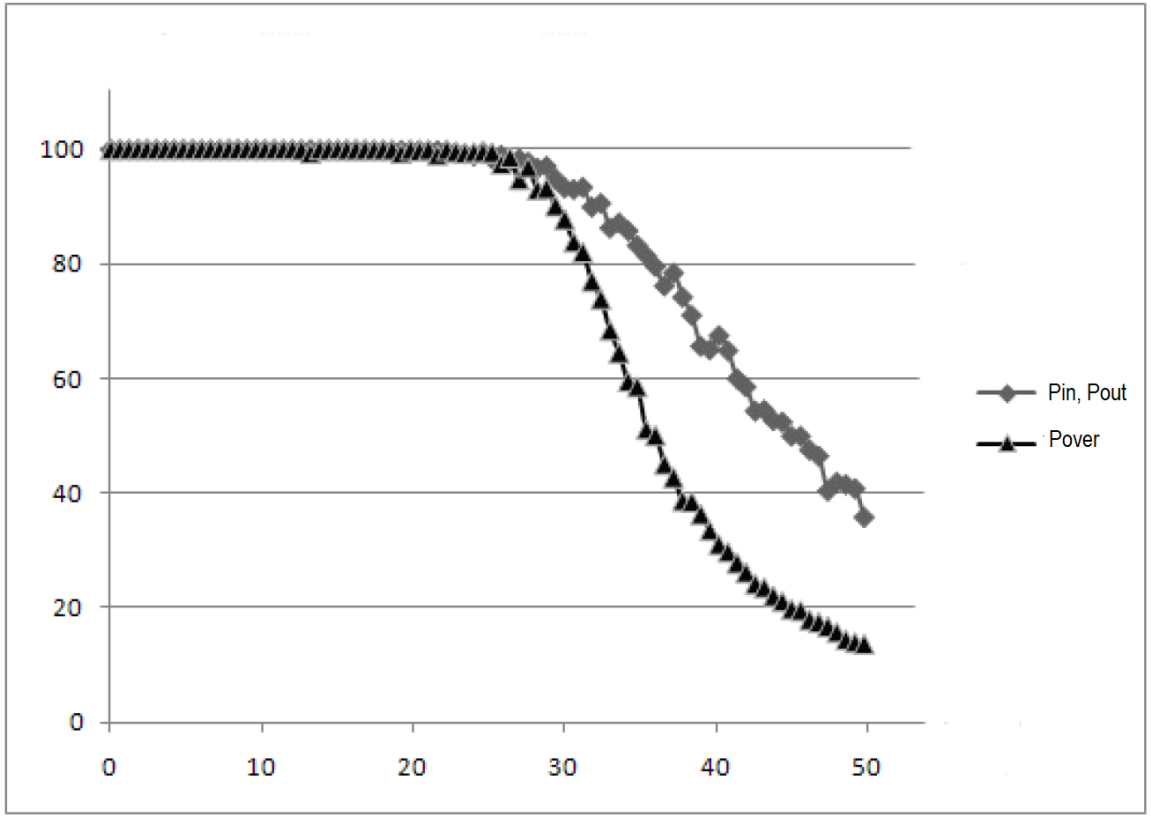
https://github.com/netcomdet/Network-Community-Detection

Synthetic data Test Case: Small World Graph Extension

- Collection of ring lattices with randomly reconnected P_{in} links inside the ring lattice and P_{out} reconnected links between the rings
- For overlapping case P_{over} randomly chosen common for rings nodes.



Simulation for the Test Case



- % of nodes recognized as correct communities' members as function of % randomized links for 16 rings - communities

Conclusion

- **Commonality** quantifies the potential of two nodes to belong to the same community, based on their shared neighbors.
- An **Inner Path** is defined as a path within a community where each set of three adjacent nodes exhibits high commonality.
- A **Natural Community** is a collection of nodes interconnected by inner paths.
- **The algorithm** developed from these definitions is straightforward, utilized local data, efficient and effective. It also exhibits stability in the face of random link perturbations.

Thank you.

igork@pac.ac.il https://igorkan.github.il/ @igorkan