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.
• 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/periiphery/hierarchy
  • Homophily and assortative mixing (similar nodes tend to be adjacent)

Why?

• 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

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

Some common approaches to subgroup identification and analysis:
• K-cliques
• K-cores (k-shell)
• K-denseness
• Components
• Community detection

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

\( k\)-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.
A clique of size $k$: a complete subgraph on $k$ nodes (i.e. a 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
k-core
A $k$-core of size $n$: maximal subset of $\alpha$ 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 $\alpha$ 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
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 $\alpha$ 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 $\alpha$ nodes part of the $\alpha$ group)
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 $\subset$ $k$-dense $\subset$ $k$-core
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 at least \( k-2 \) common neighbors.

• Identify a:
  • 2-dense
  • 3-dense
  • 4-dense
  • 5-dense
## Other extensions

### Definition of (locally) dense network structures

<table>
<thead>
<tr>
<th>Name of dense network structure</th>
<th>Definition</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clique</strong></td>
<td>A complete subgraph of size $k$, where complete means that any two of the $k$ elements are connected with each other</td>
<td>[36,37]</td>
</tr>
<tr>
<td><strong>$k$-clan</strong></td>
<td>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</td>
<td>[37,38,39]</td>
</tr>
<tr>
<td><strong>$k$-club</strong></td>
<td>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</td>
<td>[37,38,39]</td>
</tr>
<tr>
<td><strong>$k$-clique</strong></td>
<td>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</td>
<td>[37,38,39,40]</td>
</tr>
<tr>
<td><strong>$k$-clique community</strong></td>
<td>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)</td>
<td>[41,42]</td>
</tr>
<tr>
<td><strong>$k$-component</strong></td>
<td>A maximal connected subgraph, where all possible partitions of the subgraph must cut at least $k$ edges</td>
<td>[43]</td>
</tr>
<tr>
<td><strong>$k$-plex</strong></td>
<td>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</td>
<td>[37,44]</td>
</tr>
</tbody>
</table>

**References**

- [37,44]
- [37,45]
- [37,46]
- [37,47]
- [37,45]
- [48,49,50,51]
- [37,45,52]

---

Using them globally
A clique of size $k$: a complete subgraph on $k$ nodes (i.e. a 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 clique).
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.
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.
Core-periphery adjacency matrix

https://www.researchgate.net/figure/Stochastic-block-modeling-identifies-network-communities-HVR-6-is-shown-in-two-forms_fig2_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

Deciding on core-periphery

Not standardized, but generally the density of the \( k \)-core must be high, checked by the correlation, \( \rho \), defined as 

\[
\rho = \sum_{i,j} a_{ij} \delta_{ij},
\]

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

\[
\delta_{ij} = \begin{cases} 
1, & \text{if either node } i \text{ or } j \text{ is in the core} \\
0, & \text{otherwise}
\end{cases}
\]
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.
Possible structures

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

Finding the Core in Gephi

Under “Statistics” run “average degree” and then use “Filters”
1-core
4-core
Bring back the whole network
The core of the network

For this network the core is the 22-core, since the 23-core vanishes.
5. S. B. Seidman, Network structure and minimum degree, Social networks, vol. 5, no. 3, pp. 269287, 1983