Clique and Link Percolation

Andrew Celi
Overview

- Classifications of Communities
- Overlapping Communities
- Clique Percolation
- Link Clustering
- Applications
Definitions

- Connectedness Hypothesis: A community is a connected subgraph
- Density Hypothesis: Nodes in a community are more likely to connect to each other than to outside nodes
Classifications of Communities

- Clique: A completely connected subgraph
- Strong Community: A connected subgraph whose nodes have more links to nodes within the same community than nodes outside the community
- Weak Community: A connected subgraph whose total internal degree exceeds the total external degree
Overlapping Communities
Overlapping Communities

- A node is rarely confined to a single community
- How can we detect overlapping communities?
Clique Percolation (CFinder)

1) A community is the union of overlapping, or adjacent, cliques
<table>
<thead>
<tr>
<th>1) A community is the union of overlapping, or adjacent, cliques</th>
<th>2) Two k-cliques are considered adjacent if they share k-1 nodes.</th>
</tr>
</thead>
</table>

![Diagram](image)
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1) A community is the union of overlapping, or adjacent, cliques</td>
<td>2) Two $k$-cliques are considered adjacent if they share $k-1$ nodes.</td>
</tr>
<tr>
<td>3) A $k$-clique community is the largest connected subgraph obtained by the union of all adjacent $k$-cliques</td>
<td></td>
</tr>
<tr>
<td>Clique Percolation (CFinder)</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>1) A community is the union of overlapping, or adjacent, cliques</td>
<td></td>
</tr>
<tr>
<td>2) Two k-cliques are considered adjacent if they share k-1 nodes.</td>
<td></td>
</tr>
<tr>
<td>3) A k-clique community is the largest connected subgraph obtained by the union of all adjacent k-cliques</td>
<td></td>
</tr>
<tr>
<td>4) K-cliques that cannot be reached from a particular k-clique belong to another k-clique community</td>
<td></td>
</tr>
</tbody>
</table>
Can these emerge by chance?

- Random networks can produce very large cliques if they have a very high density

\[ p_c(k) = \frac{1}{[(k-1)N]^{1/(k-1)}} \]

k = clique size, N = number of nodes

- Subcritical Communities: \( p < p_c(k) \)
- Supercritical Communities: \( p > p_c(k) \)
\[ p = 0.13, \quad p_c(3) = 0.16 \]

\[ p = 0.22, \quad p_c(3) = 0.16 \]
Link Clustering

- Nodes can belong to many communities but links typically are community specific.
- Algorithm finds link communities by exploring the similarity between their neighbors and themselves.
Link Clustering Algorithm

- Link Similarity can be defined as $S$:

$$S((i,k), (j,k)) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|}$$

- Identify link clusters using a matrix
Clique Percolation in Biological Networks

- CFinder: Locating cliques and overlapping modules in biological networks by Balázs Adamcsek, Gergely Palla, Illés J. Farkas, Imre Derényi, and Tamás Vicsek
- Identify which groups of proteins interact with each other
network of yeast PPI modules
node: module of proteins, link: overlap of modules

(a) network diagram

(b) retromer complex
Vps17, Vps29, Vps26, Vps35, Vps5

(c) nucleotide-excision repair factor 1 complex + Msh2
Rad10, Rad14, Msh2

(d) HOPS complex + Vps8
Vps11, Vps39, Vps18, Vps16, Vps33, Vps8, Vps41
enlarged portions of the network of modules

Zds1,2: chromatin silencing, cell polarity
prot. phosph. type 2A complex + putative member (Rts3)
Yck1: casein kinase, phosphorylation
cAMP-dep. protein kinase complex and its regulator
Hst1
Hos2
Hos4
Cdc55
Pph22
Cph1
Set3
Snt1
Sif2
Kap60
Kap95
Yip1
Ypt53
Eeb1
Ypt1
Yip4
Tpk3
Tpk2
Bcy1
Sir3
Tpd3
Rts1
Srl3
Srl4
Cdc12
Cdc42
Rgc1
Gic1
Gic2
Bem4
Cla4
Rga1
Cdc3
Cdc10
Gin4
Shs1
Kcc4
Bni5
Cdc11
Rtp4
Boi2
Boi1
Cdc24
Sec4
Yif1
Snx41
Sec15
Snr1

vesicle-mediated transp. except: Snx41 (transport at Golgi), Eeb1 (function: unknown)
establ. of cell polarity (6 of 103 total in the genome) and Far1 (cell-cycle arrest)
common biological process of Rsr1 and Sec15: bipolar bud site selection
establishment of cell polarity (10 of 103 total in the genome) and Csm1 (DNA repl.)
histone deacetyl. complex (part) function: negative reg. of meiosis (5 of 6 total in the genome included here)
protein carriers in the nucleus
septin ring (part) and its assembly (part)
References

Adamcsek, Balazs, et al. *CFinder: Locating Cliques and Overlapping Modules in Biological Networks.*