Directed communities

Weighted communities

$k$-clique percolation and clustering in directed and weighted networks

Gergely Palla$^1$  Dániel Ábel$^2$  Imre Derényi$^2$
Illés Farkas$^1$  Péter Pollner$^1$  Tamás Vicsek$^{1,2}$

$^1$Statistical and Biological Physics Research Group, Hungarian Academy of Sciences,
$^2$Department of Biological Physics, Eötvös University, Hungary

March 2008, Louvain-la-Neuve
Outline

- Introduction
  - The Clique Percolation Method (CPM)
  - Phase transition in the Erdős-Rényi graph

- Directed communities
  - Relative in- and out degree
  - Directed CPM
  - Results

- Weighted communities
  - Weights in the original CPM
  - Weighted CPM
  - Results
The Clique Percolation Method (CPM)

Definitions

- **$k$-clique**: a complete (fully connected) subgraph of $k$ vertices.
- **$k$-clique adjacency**: two $k$-cliques are adjacent if they share $k - 1$ vertices, i.e., if they differ only in a single node.
The Clique Percolation Method (CPM)

Definitions

- **k-clique**: a complete (fully connected) subgraph of $k$ vertices.
- **k-clique adjacency**: two $k$-cliques are adjacent if they share $k - 1$ vertices, i.e., if they differ only in a single node.
The Clique Percolation Method (CPM)

Definitions

- **k-clique**: a complete (fully connected) subgraph of $k$ vertices.
- **k-clique adjacency**: two $k$-cliques are adjacent if they share $k - 1$ vertices, i.e., if they differ only in a single node.

Vicsek group

Directed and weighted communities
**Definition**

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.
Definition

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:
**Definition**

- *k-clique community*: the union of *k*-cliques that can be reached from one to the other through a sequence of adjacent *k*-cliques.

**Illustration:**

![Diagram of a network illustrating a k-clique community]
Definition

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:

[Diagram showing a network with $k$-cliques highlighted in green, illustrating the concept of $k$-clique communities.]
Definition

- *k-clique community*: the union of *k*-cliques that can be reached from one to the other through a sequence of adjacent *k*-cliques.

Illustration:


**Definition**

- **k-clique community**: the union of \( k \)-cliques that can be reached from one to the other through a sequence of adjacent \( k \)-cliques.

**Illustration:**

![Vicsek group](image-url)
Definition

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.
Definition

- *k-clique community*: the union of *k*-cliques that can be reached from one to the other through a sequence of adjacent *k*-cliques.

Illustration:
Definition

- **k-clique community**: the union of k-cliques that can be reached from one to the other through a sequence of adjacent k-cliques.

Illustration:

[Diagram showing a network of nodes and connections, illustrating a k-clique community.]
Definition

- **k-clique community**: the union of k-cliques that can be reached from one to the other through a sequence of adjacent k-cliques.

Illustration:
**Definition**

- *k-clique community*: the union of \( k \)-cliques that can be reached from one to the other through a sequence of adjacent \( k \)-cliques.

**Illustration:**

[Diagram showing a network of nodes and edges, representing a k-clique community]
**Definition**

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

**Illustration:**

[Diagram of a network with nodes and edges, illustrating a $k$-clique community.]
Definition

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:

[Vicsek group]

Directed and weighted communities
Definition

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:
Definition

- *k-clique community*: the union of *k*-cliques that can be reached from one to the other through a sequence of adjacent *k*-cliques.

Illustration:

Vicsek group
Definition

- **k-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:
Definition

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:
Definition

- **k-clique community**: the union of \( k \)-cliques that can be reached from one to the other through a sequence of adjacent \( k \)-cliques.
Definition

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:
**Definition**

- **$k$-clique community**: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

**Illustration:**

[Diagram of a Vicsek group showing directed and weighted communities]
Definition

$k$-clique community: the union of $k$-cliques that can be reached from one to the other through a sequence of adjacent $k$-cliques.

Illustration:

same at $k = 4$: 

Vicsek group
Advantages of the CPM

The main advantages of the CPM:

- Allows overlaps between the communities.
- The definition is based on the density of the links.
- It is local. (No resolution limit).
The Erdős-Rényi graph:

- $N$ nodes,
- every pair is independently linked with probability $p$.

A giant $k$-clique percolation cluster can be found if $p \geq p_c(k)$.

The order parameter of the phase transition is the size of the giant cluster:

- The number of nodes, $N^*$ \quad \rightarrow \quad \Phi \equiv \frac{N^*}{N}$,
- The number of $k$-cliques, $\mathcal{N}^*$ \quad \rightarrow \quad \Psi \equiv \frac{\mathcal{N}^*}{\mathcal{N}}$. 
The Erdős-Rényi graph:
- $N$ nodes,
- every pair is independently linked with probability $p$.

A giant $k$-clique percolation cluster can be found if $p \geq p_c(k)$.

The order parameter of the phase transition is the size of the giant cluster:
- The number of nodes, $N^*$  $\rightarrow$  $\Phi \equiv N^*/N$,
- The number of $k$-cliques, $N^*$  $\rightarrow$  $\Psi \equiv N^*/\mathcal{N}$.
The Erdős-Rényi graph:
- \( N \) nodes,
- every pair is independently linked with probability \( p \).

A giant \( k \)-clique percolation cluster can be found if \( p \geq p_c(k) \).

The order parameter of the phase transition is the size of the giant cluster:
- The number of nodes, \( N^* \) \( \rightarrow \) \( \Phi \equiv N^*/N \),
- The number of \( k \)-cliques, \( N^* \) \( \rightarrow \) \( \Psi \equiv N^*/N \).
Numerical results:

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

Direction of the links:
- Direction of some kind of flow (e.g. information, energy).
- Asymmetrical relation (e.g. superior-inferior).

Out-hubs in communities represent “sources”, whereas in-hubs correspond to “drains”:
Directed links

Direction of the links:
- Direction of some kind of flow (e.g. information, energy).
- Asymmetrical relation (e.g. superior-inferior).

Out-hubs in communities represent "sources", whereas in-hubs correspond to "drains":

[Diagram showing direction of links with source/top and drain/bottom]

Vicsek group
Directed and weighted communities
Directed links

Direction of the links:

- Direction of some kind of flow (e.g. information, energy).
- Asymmetrical relation (e.g. superior-inferior).

Out-hubs in communities represent “sources”, whereas in-hubs correspond to “drains”:
We define the relative in-degree and relative out-degree of node \( i \) in community \( \alpha \) as

\[
D_{i,\text{in}}^{\alpha} \equiv \frac{d_{i,\text{in}}^{\alpha}}{d_{i,\text{in}}^{\alpha} + d_{i,\text{out}}^{\alpha}},
\]

\[
D_{i,\text{out}}^{\alpha} \equiv \frac{d_{i,\text{out}}^{\alpha}}{d_{i,\text{in}}^{\alpha} + d_{i,\text{out}}^{\alpha}},
\]

For weighted networks these can be replaced by the relative in-strength and relative out-strength:

\[
W_{i,\text{in}}^{\alpha} \equiv \frac{w_{i,\text{in}}^{\alpha}}{w_{i,\text{in}}^{\alpha} + w_{i,\text{out}}^{\alpha}},
\]

\[
W_{i,\text{out}}^{\alpha} \equiv \frac{w_{i,\text{out}}^{\alpha}}{w_{i,\text{in}}^{\alpha} + w_{i,\text{out}}^{\alpha}},
\]
Relative in- and out-degree

We define the relative in-degree and relative out-degree of node $i$ in community $\alpha$ as

$$D_{i,\text{in}}^{\alpha} \equiv \frac{d_{i,\text{in}}^{\alpha}}{d_{i,\text{in}}^{\alpha} + d_{i,\text{out}}^{\alpha}},$$

$$D_{i,\text{out}}^{\alpha} \equiv \frac{d_{i,\text{out}}^{\alpha}}{d_{i,\text{in}}^{\alpha} + d_{i,\text{out}}^{\alpha}},$$

For weighted networks these can be replaced by the relative in-strength and relative out-strength:

$$W_{i,\text{in}}^{\alpha} \equiv \frac{w_{i,\text{in}}^{\alpha}}{w_{i,\text{in}}^{\alpha} + w_{i,\text{out}}^{\alpha}},$$

$$W_{i,\text{out}}^{\alpha} \equiv \frac{w_{i,\text{out}}^{\alpha}}{w_{i,\text{in}}^{\alpha} + w_{i,\text{out}}^{\alpha}},$$
Directed $k$-cliques?

Comparing undirected and directed connections:

- Undirected: $A$ <-> $B$
- Directed: $A$ -> $B$ and $B$ -> $A$
- Weighted: $A$ <-> $B$ with a directed edge $A$ -> $B$

In case of $k$-cliques:
- $k(k-1)/2$ links $\rightarrow 3^{k(k-1)/2}$ possible configurations.
- However, we would like the $k$-clique to have some kind of directionality as a whole as well.
Directed $k$-cliques?

Comparing undirected and directed connections:

In case of $k$-cliques:
- $k(k-1)/2$ links $\rightarrow 3^{k(k-1)/2}$ possible configurations.
- However, we would like the $k$-clique to have some kind of directionality as a whole as well.
A directed $k$-clique has to fulfil the following conditions:

**In the absence of double links:**
- Any directed link in the $k$-clique points from a node with a higher order (larger restricted out-degree) to a node with a lower order.
- The $k$-clique contains no directed loops.
- The restricted out-degree of each node in the $k$-clique is different.

**If double links are present:**
It is possible to eliminate the double links in such a way that the single links fulfil the above conditions.
A directed $k$-clique has to fulfil the following conditions:

**In the absence of double links:**
- Any directed link in the $k$-clique points from a node with a higher order (larger restricted out-degree) to a node with a lower order.
- The $k$-clique contains no directed loops.
- The restricted out-degree of each node in the $k$-clique is different.

**If double links are present:**
It is possible to eliminate the double links in such a way that the single links fulfil the above conditions.
Definition

A directed $k$-clique has to fulfil the following conditions:

In the absence of double links:

- Any directed link in the $k$-clique points from a node with a higher order (larger restricted out-degree) to a node with a lower order.
- The $k$-clique contains no directed loops.
- The restricted out-degree of each node in the $k$-clique is different.

If double links are present:

It is possible to eliminate the double links in such a way that the single links fulfil the above conditions.
A directed $k$-clique has to fulfil the following conditions:

**In the absence of double links:**
- Any directed link in the $k$-clique points from a node with a higher order (larger restricted out-degree) to a node with a lower order.
- The $k$-clique contains no directed loops.
- The restricted out-degree of each node in the $k$-clique is different.

**If double links are present:**
It is possible to eliminate the double links in such a way that the single links fulfil the above conditions.
Definition

A directed $k$-clique has to fulfil the following conditions:

In the absence of double links:
- Any directed link in the $k$-clique points from a node with a higher order (larger restricted out-degree) to a node with a lower order.
- The $k$-clique contains no directed loops.
- The restricted out-degree of each node in the $k$-clique is different.

If double links are present:
It is possible to eliminate the double links in such a way that the single links fulfil the above conditions.
contains double links?

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>c)</td>
<td>1 (or 2)</td>
<td>2 (or 1)</td>
</tr>
<tr>
<td>d)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Directed CPM

directed k-clique?

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>b)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>c)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>d)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Vicsek group
Directed and weighted communities
Phase transition in the directed E-R graph

The directed E-R graph:

- $N$ nodes,
- The $N(N - 1)$ possible “places” for the directed links are filled independently with probability $p$.

Theoretical prediction of the critical point for the appearance of a giant directed $k$-clique percolation cluster:

$$\rho_{c_{\text{theor}}} = \frac{1}{[Nk(k - 1)]^{\frac{1}{k - 1}}}.$$  

Order parameters: $\Phi$, $\Psi$ (same as in the undirected case).
Phase transition in the directed E-R graph

The directed E-R graph:

- $N$ nodes,
- The $N(N - 1)$ possible “places” for the directed links are filled independently with probability $p$.

Theoretical prediction of the critical point for the appearance of a giant directed $k$-clique percolation cluster:

$$p_c^{\text{theor}} = \frac{1}{[Nk(k - 1)]^{\frac{1}{k-1}}}.$$  

Order parameters: $\Phi$, $\Psi$ (same as in the undirected case).
Introduction
Directed communities
Weighted communities

Numerical results

(a) $\Phi(p)$ for $k = 4$ and different $N$ values:
- $N = 25$
- $N = 50$
- $N = 100$
- $N = 200$
- $N = 400$
- $N = 800$
- $N = 1600$

(b) $\Psi(p)$ for $k = 4$ and different $N$ values:
- $N = 50$
- $N = 100$
- $N = 200$
- $N = 400$
- $N = 800$
- $N = 1600$

(c) $\chi(p)$ for $k = 4$ and different $N$ values:
- $N = 25$
- $N = 50$
- $N = 100$
- $N = 200$
- $N = 400$
- $N = 800$
- $N = 1600$

(d) $P_{C_{\text{num}}} / P_{C_{\text{theor}}}$ for $k = 3, 4, 5, 6$:
- $k = 3$
- $k = 4$
- $k = 5$
- $k = 6$

Vicsek group
Directed and weighted communities
Introduction
Directed communities
Weighted communities

Relative in- and out-degree
Directed CPM
Results

Word association network

Local picture of the communities:
The number of hits in Google as a function of $W_{i,\text{out}}^\alpha$:
Local picture of the communities:
Comparing overlaps

Membership number in function of $D^{\alpha}_{i,\text{out}}$:

Community overlaps:
- word association net, Google’s web pages $\rightarrow$ in-hubs,
- e-mail net, transcription regulatory network $\rightarrow$ out-hubs.
In the original CPM we can take into account the weights by ignoring links weaker than a certain threshold $w^*$. Changing $w^*$ and $k$ is similar to changing the resolution in a microscope.

**Optimal $k$-clique size and $w^*$**

Where the community structure is as highly structured as possible: just below the critical point of the appearance of a giant $k$-clique community.
Link weights in the original CPM

In the original CPM we can take into account the weights by ignoring links weaker than a certain threshold $w^*$. Changing $w^*$ and $k$ is similar to changing the resolution in a microscope.

Optimal $k$-clique size and $w^*$

Where the community structure is as highly structured as possible: just below the critical point of the appearance of a giant $k$-clique community.
In the original CPM we can take into account the weights by ignoring links weaker than a certain threshold $w^*$. Changing $w^*$ and $k$ is similar to changing the resolution in a microscope.

**Optimal $k$-clique size and $w^*$**

Where the community structure is as highly structured as possible: just below the critical point of the appearance of a giant $k$-clique community.
The intensity $I$ of a sub-graph is defined as the geometrical mean of its link weights.

For a $k$-clique $C$: $I(C) = \left( \prod_{\substack{i<j \\ i,j \in C}} w_{ij} \right)^{2/k/(k-1)}$

**Weighted $k$-clique**

A $k$-clique with an intensity greater or equal to a given intensity threshold $I^*$. 
Percolation transition in the E-R graph

A weighted E-R graph:

- N nodes,
- every pair is linked independently with uniform probability $p$,
- each link is assigned a weight chosen randomly from a uniform distribution on the $(0, 1]$ interval.

The critical linking probability is a function of the intensity threshold. At $I = 0$ we recover $p_c(I = 0) = [N(k - 1)]^{-1/(k-1)}$. 
Percolation transition in the E-R graph

A weighted E-R graph:

- N nodes,
- every pair is linked independently with uniform probability $p$,
- each link is assigned a weight chosen randomly from a uniform distribution on the $(0, 1]$ interval.

The critical linking probability is a function of the intensity threshold. At $I = 0$ we recover $p_c(I = 0) = [N(k - 1)]^{-1/(k-1)}$. 
Results

Directed and weighted communities
New York Stock Exchange graph:

- We studied the pre-computed stock correlation matrix containing the averaged correlation between the daily logarithmic returns.
- The correlation coefficients can be used as link weights. We kept only the strongest 3%.
Summary

- Directed communities:
  - Relative in- and out-degree,
  - Directed $k$-cliques.

- Weighted communities:
  - $k$-clique intensity.

- Publications:
  - New Journal of Physics 9, 180 (2007),

- Downloadable community finding software:
  - http://cfinder.org