Emerging communities in networks – a flow of ties

Małgorzata J. Krawczyk, Przemysław Gawroński, Krzysztof Kułakowski

Complex Systems Group
Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
Communities (...) occur in many networked systems from biology, computer science, engineering, economics, politics, etc. In protein-protein interaction networks, communities are likely to group proteins having the same specific function within the cell (...), in metabolic networks they may be related to functional modules such as cycles and pathways in food webs they may identify compartments and so on.

[S. Fortunato, Community detection in graphs, Phys. Reports 486, 75-174 (2010)]

"Community detection"

[Web of Science, 16.06.2014]
Are there any communities?

...there must be more edges "inside" the community than edges linking vertices of the community with the rest of the graph.

\[ Q = \frac{1}{w} \sum_{ij} \left( A(i, j) - \frac{k_i k_j}{w} \right) \delta(c_i, c_j) \]

where

\[ k_i = \sum_j A(i, j) \]

\[ w = \sum_{ij} A(i, j) \]

FIG. 4: Splitting of the Zachary club network. Squares and circles indicate the two communities observed by Zachary, colors denote the further subdivision found by our algorithm.

[S. Fortunato, Community detection in graphs, Physics Reports 486 (2010) 75]
Outline

1. The goal:
   a spontaneous flow towards solution

2. Examples:
   - Heider balance
   - communities in the Sierpinski triangle with a bit of noise

[alejka.pl/siedmiomilowe_buty.html]
[www.giantsteps-project.eu/]
Cognitive dissonance is a state of conflict in the mind, whereby you have two opposing views at the same time.

An example:

Sarah just bought a new car. Admittedly, she spent much more money than she should have and feels regretful and actually a little embarrassed as well (buyer’s remorse). 😞

Rather than continue feeling these undesirable emotions, she decides that the car is less likely to break down than her older one, and will actually save her loads of cash in the long run. 😊

[L. Festinger, A Theory of Cognitive Dissonance, Stanford UP 1957]
[psychohawks.wordpress.com/]
[lasombradecharvaka.blogspot.com/]
Heider balance

A friend of my friend is my friend
A friend of my enemy is my enemy
An enemy of my friend is my enemy
An enemy of my enemy is my friend

[E. Aronson, V. Cope, My enemy’s enemy is my friend, J. of Personality & Soc. Psych. 8 (1968) 8]
[compendium.open.ac.uk/]
How the balance can be restored

Cartwright - Harary theory

...on a complete graph, balanced societies are remarkably simple:

either all individuals are mutual friends (“utopia”),
or the network segregates into two mutually antagonistic but internally friendly cliques - a “bipolar” state.

An example:

How the major relationship changed from 1872 to 1907 (Antal et al, 2006)

[D. Cartwright, F. Harary, Psychol. Rev. 63 (1956) 277]
Discrete algorithms for complete graphs:

1. **Local Triad Dynamics (LTD)**: operates at randomly selected triads

   - Select a link randomly and check how the change of its sign does influence the number of imbalanced triads.
   - If the number decreases, the change is accepted.
   - If the number increases, the change is withdrawn.
   - If the number remains the same, change the link with probability $\frac{1}{2}$.

2. **Constrained Triad Dynamics (CTD)**: the whole number of imbalanced triads is controlled
   - Select a link randomly and check how the change of its sign does influence the number of imbalanced triads.
   - If the number decreases, OK.
   - If the number increases, withdraw the change.
   - If the number is the same, change the link with probability $\frac{1}{2}$.

LTD: - always leads to a balanced state, but which one?
- psychologically not plausible: the dissonance can increase

CTD: - jammed states

Examples of jammed states for 9 nodes
(only positive links are shown)

\[ U = -\binom{N}{3}^{-1} \sum_{ijk} s_{ij} s_{jk} s_{ki} \]

Jammed states for 26 nodes

Proposition - a flow

\[
\frac{dx_{ij}}{dt} = G\left(\frac{x_{ij}}{R}\right) \sum_k x_{ik} x_{kj}
\]

\[
G(x) = 1 - x^2
\]

keeps the relation \(x_{ij}\) within the range \((-R,R)\)

positive if \(x_{ik}\) and \(x_{kj}\) both friendly or both hostile, elsewhere negative

For \(G(x)=1\), jammed states do not appear (Marvel et al).

A test: Natchez women data


<table>
<thead>
<tr>
<th>Names of Participants of Group I</th>
<th>Code Numbers and Dates of Social Events Reported in Old City Herald</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) 6/27 (2) 3/2 (3) 4/12 (4) 9/26 (5) 2/25 (6) 5/19 (7) 3/15 (8) 9/16 (9) 4/6 (10) 6/10 (11) 2/23 (12) 4/7 (13) 11/21 (14) 8/3</td>
</tr>
<tr>
<td>1. Mrs. Evelyn Jefferson.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>2. Miss Laura Mandeville.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>3. Miss Theresa Anderson.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>4. Miss Brenda Rogers.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>5. Miss Charlotte McDowd.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>6. Miss Frances Anderson.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>7. Miss Eleanor Nye.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>8. Miss Pearl Oglethorpe.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>9. Miss Ruth DeSand.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>10. Miss Verne Sanderson.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>11. Miss Myra Liddell.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>12. Miss Katherine Rogers.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>13. Mrs. Sylvia Avondale.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>14. Mrs. Nora Fayette.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>15. Mrs. Helen Lloyd.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>16. Mrs. Dorothy Murchison.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>17. Mrs. Olivia Carleton.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
<tr>
<td>18. Mrs. Flora Price.</td>
<td>X X X X X X X X X X X X X X</td>
</tr>
</tbody>
</table>
### Meta-analysis – comparison of 21 methods

<table>
<thead>
<tr>
<th>Code</th>
<th>Analysis</th>
<th>Closeness to the Matching Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DGG41 Davis, Gardner and Gardner, Ethnography</td>
<td>0.920</td>
</tr>
<tr>
<td>2</td>
<td>HOM50 Homans, Intuition</td>
<td>0.854</td>
</tr>
<tr>
<td>3</td>
<td>P&amp;C72 Phillips and Conviser, Information Theory</td>
<td>0.968</td>
</tr>
<tr>
<td>4</td>
<td>BGR74 Breiger, Algebra</td>
<td>0.933</td>
</tr>
<tr>
<td>5</td>
<td>BBA75 Breiger, Boorman and Arabie, CONCOR</td>
<td>0.927</td>
</tr>
<tr>
<td>6</td>
<td>BCH78 Bonacich, Boolean Algebra</td>
<td>0.841</td>
</tr>
<tr>
<td>7</td>
<td>DOR79 Doreian, Algebraic Topology</td>
<td>0.923</td>
</tr>
<tr>
<td>8</td>
<td>BCH91 Bonacich, Correspondence Analysis</td>
<td>0.968</td>
</tr>
<tr>
<td>9</td>
<td>FRE92 Freeman, G-Transitivity</td>
<td>0.926</td>
</tr>
<tr>
<td>10</td>
<td>E&amp;B93 Everett and Borgatti, Regular Coloring</td>
<td>0.916</td>
</tr>
<tr>
<td>11</td>
<td>FR193 Freeman, Genetic Algorithm 1</td>
<td>0.968</td>
</tr>
<tr>
<td>12</td>
<td>FR293 Freeman, Genetic Algorithm 2</td>
<td>0.926</td>
</tr>
<tr>
<td>13</td>
<td>FW193 Freeman and White, Galois Lattice</td>
<td>0.917</td>
</tr>
<tr>
<td>14</td>
<td>FW293 Freeman and White, Galois Sub-Lattice</td>
<td>0.954</td>
</tr>
<tr>
<td>15</td>
<td>BE197 Borgatti and Everett, Bi-Clique</td>
<td>0.916</td>
</tr>
<tr>
<td>16</td>
<td>BE297 Borgatti and Everett, Taboo Search</td>
<td>0.968</td>
</tr>
<tr>
<td>17</td>
<td>BE397 Borgatti and Everett, Genetic Algorithm</td>
<td>0.968</td>
</tr>
<tr>
<td>18</td>
<td>S&amp;F99 Skvoretz and Faust, p* Model</td>
<td>0.957</td>
</tr>
<tr>
<td>19</td>
<td>ROB00 Roberts, SVD with Normalization</td>
<td>0.968</td>
</tr>
<tr>
<td>20</td>
<td>OSB00 Osbourn, VERI Algorithm</td>
<td>0.543</td>
</tr>
<tr>
<td>21</td>
<td>NEW01 Newman, Weighted Co-Attendance</td>
<td>0.932</td>
</tr>
</tbody>
</table>

---

A test: Zachary karate club

The obtained split almost exactly as real:

\[ 1, 8, 11, 14, 17, 18, 20, 22 \]

vs

\[ 9, 10, 15, 16, 19, 21, 23-34 \]

for \( R=5 \) and any \( \varepsilon \), if only \(-R < x_{ij} < R\)


[http://vlado.fmf.uni-lj.si/pub/networks/data/Ucinet/UciData.htm#zachary]


A test: Zachary karate club

\[ x_{ij} \rightarrow x_{ij} - \varepsilon \]

\[ \frac{dx_{ij}}{dt} = [1 - (x_{ij} / R)^2] \sum_k x_{ik} x_{kj} \]

The obtained split almost exactly as real:

1-8,11-14,17,18,20,22  
vs  
9,10,15,16,19,21,23-34

for \( R=5 \) and any \( \varepsilon \), if only \(-R < x_{ij} < R\)

[http://vlado.fmf.uni-lj.si/pub/networks/data/Ucinet/UciData.htm#zachary]  
Zachary karate club + calculated relations

\[ x_{ij} \]

\[ \text{time} \]
More communities

\[ \frac{dA_{ij}}{dt} = G(A_{ij}) \sum_{k \neq i, j} (A_{ik}A_{kj} - \beta) \]

where \( A \) - the connectivity matrix,

\[ G(x) = \Theta(x)\Theta(1 - x) \]

FIG. 4. Results for clusters of different sizes: (a) \( N=110 \), clusters of 50, 48, and 12 nodes, (b) \( N=130 \), clusters of 22, 34, 11, 10, and 53 nodes, (c) \( N=130 \), clusters of 19, 60, 45, and 6 nodes, (d) \( N=110 \), clusters of 18, 22, 30, 38, and 2 nodes [symbols denote: \( \square \), the Newman algorithm; \( \times \), Eq. (2) with \( \beta=0.25 \); \( \bullet \), Eq. (2) with \( \beta=0.4 \)].

[M. J. Krawczyk, Differential equations as a tool for community identification, PRE 77 (2008) 065701R]
Figure 3: The Sierpinski triangle network with indication of communities for $N = 42$, here we have 9 communities of 3 nodes each (black nodes) and 15 communities of 1 node each (white nodes).

[M.J.Krawczyk, Communities and classes in symmetric fractals, IJMPC (2014), in print (arXiv:1404.7416)]
How to detect overlapping nodes?

- add a small noise to the initial matrix of connectivity
- run the equations many times
- construct a histogram: how often a given node belongs to particular community?
- if more than one community appear with the largest peaks, the node is overlapping
Sierpiński triangle with noise, N=15
$N = 15$

$\beta = 0.15$

No noise:
- isolated communities

$Q = 0.12$

Noise $a = 0.01$:

$\langle Q \rangle = 0.19$

- communities

$Q_{\text{max}} = 0.25$

No noise:
- communities

$Q = 0.26$
Conclusions:

- the designed flow of links drives the system to a set of separate clusters

- the 4-th axiom of Heider precludes more than two communities

- an application of noise breaks symmetry, allows to reveal overlapping nodes and leads to better partitions
Thank you