A Primer on Social Network Analysis

SERGEI OBIEDKOV & CAMILLE ROTH

HSE Moscow — CAMS (CNRS/EHESS)

ICFCA 10 Agadir March 14-18
Structure

1 History and ontology
   - Historical landmarks
   - Social networks as graphs

2 SNA Concepts and Approaches
   - Position analysis
   - Community detection / topological analysis
   - Two-mode networks

3 SNA and FCA
   - Landmarks
   - Cliques
   - An empirical case: socio-semantic groups
The beginnings

- **Rapoport, 1940**
  → proportion of triangles

- **Forsyth & Katz, 1946**
  → introducing socio-matrices

- **Cartwright & Harary, 1956**
  → “balance theory”
The beginnings

- **Rapport, 1940**
  - proportion of triangles

- **Forsyth & Katz, 1946**
  - introducing socio-matrices

- **Cartwright & Harary, 1956**
  - “balance theory”

Example of matrix for this underlying graph:

```
+----------+----------+----------+----------+----------+
| A  B  C  D  E |
+----------+----------+----------+----------+----------+
| -  1  0.5  0  -1 |
| 1  -1 -1  2  0  |
| 0.5 -1 -1  0.5  0 |
| 0  0  0.5 -2  |
| -1  0  0  0  - |
+----------+----------+----------+----------+----------+
```
The beginnings

- **Rapoport, 1940**
  → proportion of triangles

- **Forsyth & Katz, 1946**
  → introducing socio-matrices

- **Cartwright & Harary, 1956**
  → “balance theory”
Interaction networks and one-mode data

Graphs: simple, directed, weighted...

\[ G = (V, E \subseteq V \times V) \]
Interaction networks and one-mode data

Graphs: simple, directed, weighted...

\[ G = (V, E \subseteq V \times V) \]
Affiliation networks and two mode-data

...using bipartite graphs

\[ G = (V_1, V_2, E \subseteq V_1 \times V_2) \]
Questionnaire

A. FCA theory
B. Conceptual Knowledge Processing
C. Concept Graphs
D. Lattice Drawing
E. Association Rules
F. Algorithmics
G. FCA and Software Engineering
H. Lattice Theory
I. Data Analysis
J. FCA and Data Mining
K. FCA and Logic
L. Philosophical Foundations
Representation

“Show me a graph”

Figure 1: The Dangers of Visual Representation
Distance

- definition
  → “distance” = number of steps

- empirical shape
  → unimodal law, low mean ("small-world")

- counter-intuitions
  1. the myth of the small distance...
     → “a knows b who knows c who knows d”
  2. the whole world within reach...
     → with 100 neighbors per person, what happens at distance 6?...

Transitivity and clustering

“degree centrality”

“betweenness centrality”
**Distance**

- **definition**
  → “distance” = number of steps

- **empirical shape**
  → unimodal law, low mean ("small-world")

- **counter-intuitions**
  1. the myth of the small distance...
     → “a knows b who knows c who knows d”
  2. the whole world within reach...
     → with 100 neighbors per person, what happens at distance 6?...

---

**Transitivity and clustering**

**“degree centrality”**

**“betweenness centrality”**
Transitivity and clustering

- “clustering” → generally high...

...denotes more finely the existence of transitivity

\[(Newman, 2002)\]

<table>
<thead>
<tr>
<th></th>
<th>MEDLINE</th>
<th>Complete</th>
<th>astro-ph</th>
<th>cond-mat</th>
<th>hep-th</th>
<th>SPIRES</th>
<th>NCSTRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total papers</td>
<td>2,163,923</td>
<td>98,502</td>
<td>22,029</td>
<td>22,016</td>
<td>19,985</td>
<td>66,652</td>
<td>13,169</td>
</tr>
<tr>
<td>Total authors</td>
<td>1,520,251</td>
<td>52,909</td>
<td>16,726</td>
<td>16,726</td>
<td>8,361</td>
<td>56,627</td>
<td>11,994</td>
</tr>
<tr>
<td>Clustering coefficient C</td>
<td>0.666 (7)</td>
<td>0.43 (1)</td>
<td>0.414 (6)</td>
<td>0.348 (6)</td>
<td>0.327 (2)</td>
<td>0.726 (8)</td>
<td>0.496 (6)</td>
</tr>
</tbody>
</table>

“degree centrality”
“degree centrality”

- **degree centrality** = # of connections, of neighbors
  - generally corresponds to a social capital, or an attention
  - "measuring structural capital in that it facilitates some forms of social capital" (Coleman, 1988)

- empirically:
  1. heterogeneous ("power law", "Zipf-law", "scale-free", ...)
  2. dynamically stable

“betweenness centrality”
Distance

Transitivity and clustering

“degree centrality”

“betweenness centrality”

- **intuitively:** describes "intermediary" positions
- **precisely:**
  
  \[ BC_z = \sum_{(i,j) \in V \times V} \frac{C_z(i,j)}{C(i,j)} \]
Behavior estimation: processus of link creation

Those may influence interaction:

- structural properties (degree, etc.)...
- ...as well as non-structural properties (homophily, etc.)

Histories

“Econometric” models (regressions)

Scoring methods

  → common neighbors, degree, etc.
- “Implicit Structure and the Dynamics of Blogspace”, Adar
Behavior estimation: processus of link creation

Histograms

Propension to create/receive a link with respect to parameters related, notably, to the network structure... \( P(L|d) \)

- trivial estimator: \( \Pi(d) = \frac{\nu(d)}{N(d)} = \frac{\text{proportion of new dyads of type } d}{\text{proportion of couples of type } d} \)
- empirical interaction propensions

“Econometric” models (regressions)

Scoring methods
Behavior estimation: processus of link creation

Histograms

Propension to create/receive a link with respect to parameters related, notably, to the network structure...\(P(L|d)\)

- trivial estimator: \(\Pi(d) = \frac{\nu(d)}{N(d)} = \frac{\text{proportion of new dyads of type } d}{\text{proportion of couples of type } d}\)

- empirical interaction propensions

“Econometric” models (regressions)

Scoring methods
Behavior estimation: processus of link creation

Histograms

Propensity to create/receive a link with respect to parameters related, notably, to the network structure... \( P(L|d) \)

- Trivial estimator: \( \Pi(d) = \frac{\nu(d)}{N(d)} = \frac{\text{proportion of new dyads of type } d}{\text{proportion of couples of type } d} \)
- Empirical interaction propensities

“Econometric” models (regressions)

- \textit{logit} model for the existence of a link with respect to parameters...

\[ \text{prob}(y_{ij,t} = 1 | \sum_{J_{st}} y_{ij,t} = 1, X_{j,t}, W_{ij,t}) = \frac{\exp(\beta X_{j,t} + \gamma W_{ij,t})}{\sum_{j,t} \exp(\beta X_{j,t} + \gamma W_{ij,t})} \]
Behavior estimation: processus of link creation

Histograms

“Econometric” models (regressions)

- logit model for the existence of a link with respect to parameters...

\[
\text{prob}\left(y_{ij,t} = 1 \mid \sum_{j,t} y_{ij,t} = 1, X_{j,t}, W_{ij,t}\right) = \frac{\exp(\beta X_{j,t} + \gamma W_{ij,t})}{\sum_{j,t} \exp(\beta X_{j,t} + \gamma W_{ij,t})}
\]

- objective function

\[
f_i(X, \beta) = \sum_{k=1}^{L} \beta_k s_{i,k}(X)
\]

that agents optimize their interaction choices (Snijders, 2001)
Behavior estimation: processus of link creation

Histograms

“Econometric” models (regressions)

Scoring methods

  → common neighbors, degree, etc.

- “Implicit Structure and the Dynamics of Blogspace”, Adar, Adamic et al., 2007
  → classifiers based on common neighbors, content, etc.

\[
\sum_{\ell=1}^{\infty} \beta^\ell \cdot |\text{paths}_{x,y}^{(\ell)}|\]

Figure 5: A visualization of all blogs linking to the Giant Microbes site with all explicit and implicit links.
**Algebraic approaches**

1. **Cliques** (Luce, 1949), n-cliques

2. **Structure theorem** (Cartwright & Harary, 1956, on Heider, 1946)

3. **Structural equivalence** (Lorrain & White, 1971)

4. **k-cores and k-components** (White & Harary, 2001)

→ drawbacks: rigidity and, often, computation

**Algorithmic approaches**
Algebraic approaches

1. Cliques (Luce, 1949), n-cliques

2. Structure theorem (Cartwright & Harary, 1956, on Heider, 1946)

3. Structural equivalence (Lorrain & White, 1971)

4. k-cores and k-components (White & Harary, 2001)

→ drawbacks: rigidity and, often, computation

Algorithmic approaches
Algebraic approaches

1. Cliques (Luce, 1949), n-cliques

2. Structure theorem (Cartwright & Harary, 1956, on Heider, 1946)

3. Structural equivalence (Lorrain & White, 1971)

4. k-cores and k-components (White & Harary, 2001)

→ drawbacks: rigidity and, often, computation

Algorithmic approaches

Sergei OBIEDKOV & Camille ROTH
### Algebraic approaches

1. **Cliques** *(Luce, 1949), n-cliques*

2. **Structure theorem** *(Cartwright & Harary, 1956, on Heider, 1946)*

3. **Structural equivalence** *(Lorrain & White, 1971)*

4. **k-cores and k-components** *(White & Harary, 2001)*

→ drawbacks: rigidity and, often, computation

### Algorithmic approaches
Algebraic approaches

1. Cliques (Luce, 1949), n-cliques

2. Structure theorem (Cartwright & Harary, 1956, on Heider, 1946)

3. Structural equivalence (Lorrain & White, 1971)

4. k-cores and k-components (White & Harary, 2001)

→ **drawbacks**: rigidity and, often, computation

Algorithmic approaches
Algebraic approaches

Algorithmic approaches

→ impulse from hard sciences + increase in computational capabilities

1 **broad definition:**

- “its members should have many relations with each other and few with non-members” (Alba, 1973)
- “groups of vertices within which connections are dense, but between which connections are sparser” (Newman, 2004)

2 **algorithms for building partitions:**

- aggregative, e.g. using similarity measures
- desaggregative, e.g. using betweenness centrality

→ drawback: no explicit definition, distinct from the algorithm
Algebraic approaches

Algorithmic approaches

→ *impulse from hard sciences + increase in computational capabilities*

1 **broad definition:**

- “its members should have many relations with each other and few with non-members” (Alba, 1973)
- “groups of vertices within which connections are dense, but between which connections are sparser” (Newman, 2004)

2 **algorithms** for building partitions:

- aggregative, e.g. using similarity measures
Algorithmic approaches

→ *impulse from hard sciences + increase in computational capabilities*

1 **broad definition:**

- “its members should have many relations with each other and few with non-members” (Alba, 1973)
- “groups of vertices within which connections are dense, but between which connections are sparser” (Newman, 2004)

2 **algorithms** for building partitions:

- aggregative, e.g. using similarity measures
- desaggregative, e.g. using betweenness centrality

3 **algorithms** exploring the graph (e.g. clique percolation)

→ **drawback:** no explicit definition, distinct from the algorithm
Algebraic approaches

Algorithmic approaches

→ impulse from hard sciences + increase in computational capabilities

1 broad definition:

- “its members should have many relations with each other and few with non-members” (Alba, 1973)
- “groups of vertices within which connections are dense, but between which connections are sparser” (Newman, 2004)

2 algorithms for building partitions:

- aggregative, e.g. using similarity measures
- desaggregative, e.g. using betweenness centrality

3 algorithms exploring the graph (e.g. clique percolation)

→ drawback: no explicit definition, distinct from the algorithm
Algorithmic approaches

→ impulse from hard sciences + increase in computational capabilities

1 broad definition:

- “its members should have many relations with each other and few with non-members” (Alba, 1973)
- “groups of vertices within which connections are dense, but between which connections are sparser” (Newman, 2004)

2 algorithms for building partitions:

- aggregative, e.g. using similarity measures
- desaggregative, e.g. using betweenness centrality

3 algorithms exploring the graph (e.g. clique percolation)

→ drawback: no explicit definition, distinct from the algorithm
Network classes

Amaral et al., 2000 – “Classes of small-world networks”

Milo et al., 2004 – “Superfamilies of evolved and designed networks”
Network classes

Amaral et al., 2000 – “Classes of small-world networks”

Milo et al., 2004 – “Superfamilies of evolved and designed networks”
Discussion thread networks

Dorat, Latapy, Conein & Auray, 2005: “Multi-level analysis of a discussion list network”

Fig. 5. The three levels at which we will consider our data. From left to right: the threads (trees), the labelled threads, and the interaction network. Notice that we removed the loops (here, \((f, f)\)) and that we do not consider multiple links (for instance here \((a, b)\)).

Welser et al., 2005: “Visualizing the Signatures of Social Roles in Online Discussion Groups”

Figure 3a: Exemplary local network for an Answer Person

Figure 3b: Exemplary local network for a Discussion Person
Tableau 1. Quatre modèles de production du public des blogs

<table>
<thead>
<tr>
<th>Enonciation/Enoncé</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personne/Contenu</td>
<td>+++</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Enoncé</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Identité</td>
<td>Pseudonyme</td>
<td>Samon</td>
<td>Signature</td>
<td>Identité civile</td>
</tr>
<tr>
<td>Mode de communication</td>
<td>Partage des intimités</td>
<td>Communication continue</td>
<td>Affinités électorales</td>
<td>Échange d’opinions</td>
</tr>
<tr>
<td>Forme relationnelle</td>
<td>Halo</td>
<td>Clan/Nébuleuse</td>
<td>Communauté</td>
<td>Espace public</td>
</tr>
<tr>
<td>Public du blog</td>
<td>Halo</td>
<td>Peu</td>
<td>Communauté</td>
<td>Espace public</td>
</tr>
<tr>
<td>Liens externes vers web</td>
<td>Peu</td>
<td>Très peu</td>
<td>Nombreux</td>
<td>Très nombreux</td>
</tr>
<tr>
<td>Type de blogs</td>
<td>Journaux intimes</td>
<td>Blogs d’adolescents, blogs de famille, de voyages</td>
<td>Blogs de collectionneurs, de fans, de critiques, etc.</td>
<td>Blogs journalistiques, politiques, citoyens</td>
</tr>
<tr>
<td>Forme du réseau</td>
<td>Etoile</td>
<td>Clan</td>
<td>Communauté</td>
<td>Polarisation</td>
</tr>
<tr>
<td>Taille</td>
<td>Très petit</td>
<td>Petit</td>
<td>Important</td>
<td>Très important</td>
</tr>
<tr>
<td>Densité</td>
<td>Très faible</td>
<td>Très forte</td>
<td>Assez forte</td>
<td>Forte avec</td>
</tr>
</tbody>
</table>

Figure S. Blogroll et citation links for (a) DTW, (b) UAE, (c) Kuwait. Nodes are colored red if they are in the community and are sized by indegree.
Blog networks

**Cardon, Delaunay-Teterel, 2006**

**Ali-Hasan, Adamic, 2007**

**McGlohon, Leskovec, Faloutsos, Hurst, Glance, 2007**

Fig. 2: Common cascade shapes, ordered by the frequency in the dataset

(a) First vs. second PC

(b) Second vs. third PC

In other words, the conversation mass for a blogger equals:

the total number of posts in all conversation trees below the point in which the blogger contributed, summed over all conversation trees in which the blogger appears.

3.2 Principal component analysis

Given many vectors in $D$-dimensional space, how can visualize them, when the dimensionality $D$ is high? This is exactly where Principal Component Analysis (PCA) helps. PCA will find the optimal 2-dimensional plane to project the data points, maintaining the pair-wise distances as best as possible. PCA is even more powerful than that: it can give us a sorted list of directions ("principal components") on which we can project. See [12] or [14] for more details.

3.3 Clustering blogs by CascadeType

Our first experiments involved performing PCA on a large, sparse matrix where rows represented blogs and columns represented different types of cascades. Each entry was a count, and in order to reduce the variance, we took the log of each count. Our dataset consisted of 44,791 blogs with 8,965 cascade types. It was of interest to impose social networks upon the blogs, based on what topics the blogs tended to focus on. We hand-classified a sample of the blogs in the icwsm data by topic, and found that we could often separate communities based on this analysis. For the purposes of visualization we chose to focus on two of the larger communities, politically conservative blogs and "humorous" blogs (such as blogs for different web-comics and humorists). Figure 3(a) shows these blogs plotted on the first two principal components, and Figure 3(b) shows them plotted on the second and third principal components. Ovals are drawn around the main clusters. We notice a distinct separation between the conservative community and the humor community; this means that the two communities engage in very different conversation patterns.

3.4 Observations

Based on our CascadeType analysis, we make the following observations:

Observation 1. Communities often cluster around the same types of cascades, with distinct conversation patterns. It seems that conservative blogs and the "humorous" blogs form separate clusters. We believe this is the case because conservative blogs tend to form deep, chainlike graphs whereas the humorous blogs form stars. Some similar observations may be made for other communities; we used these two because they were the most distinct. This result shows that blogs...
Interlock positions

(Robins & Alexander, 2004)

Bipartite clustering

Table 1. Graph statistics for the United States network infrastructure. (Median geodesics and counts of configurations in the observed network; means and standard deviations for simulated graph distribution; standardized z-score for observed value; median of simulated median geodesics.)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Observed</th>
<th>Simulated distribution</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{A2}$ (bipartite company two-stars)</td>
<td>3892</td>
<td>3387 (58)</td>
<td>+8.71</td>
</tr>
<tr>
<td>$S_{A3}$ (bipartite company three-stars)</td>
<td>8696</td>
<td>6110 (326)</td>
<td>+7.93</td>
</tr>
<tr>
<td>$S_{P2}$ (bipartite director two-stars)</td>
<td>1144</td>
<td>1132 (11)</td>
<td>+1.01</td>
</tr>
<tr>
<td>$S_{P3}$ (bipartite director three-stars)</td>
<td>535</td>
<td>502 (31)</td>
<td>+1.06</td>
</tr>
<tr>
<td>$L_3$ (bipartite three-paths)</td>
<td>14786</td>
<td>12272 (248)</td>
<td>+10.14</td>
</tr>
<tr>
<td>$C_4$ (multiple links)</td>
<td>856</td>
<td>98 (19)</td>
<td>+39.89</td>
</tr>
<tr>
<td>$C$ (bipartite clustering)</td>
<td>0.232</td>
<td>0.032 (0.006)</td>
<td>+32.75</td>
</tr>
</tbody>
</table>

Overlap/redundancy

(Latapy, Magnien, del Vecchio, 2007)

“the fraction of pairs of neighbours of v linked to another node than v. In the projection, these nodes would be linked together even if v were not there.”
**Interlock positions**

(Robins & Alexander, 2004)

Table 1. Graph statistics for the United States network infrastructure. (Median geodesics and counts of configurations in the observed network; means and standard deviations for simulated graph distribution; standardized z-score for observed value; median of simulated median geodesics.)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Observed</th>
<th>Simulated distribution</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{A2}$ (bipartite company two-stars)</td>
<td>3892</td>
<td>3387 (58)</td>
<td>+8.71</td>
</tr>
<tr>
<td>$S_{A3}$ (bipartite company three-stars)</td>
<td>8696</td>
<td>6110 (326)</td>
<td>+7.93</td>
</tr>
<tr>
<td>$S_{P2}$ (bipartite director two-stars)</td>
<td>1144</td>
<td>1132 (11)</td>
<td>+1.01</td>
</tr>
<tr>
<td>$S_{P3}$ (bipartite director three-stars)</td>
<td>535</td>
<td>502 (31)</td>
<td>+1.06</td>
</tr>
<tr>
<td>$L_3$ (bipartite three-paths)</td>
<td>14786</td>
<td>12272 (248)</td>
<td>+10.14</td>
</tr>
<tr>
<td>$C_4$ (multiple links)</td>
<td>856</td>
<td>98 (19)</td>
<td>+39.89</td>
</tr>
<tr>
<td>$C$ (bipartite clustering)</td>
<td>0.232</td>
<td>0.032 (0.006)</td>
<td>+32.75</td>
</tr>
</tbody>
</table>

**Overlap/redundancy**

(Latapy, Magnien, del Vecchhio, 2007)

“the fraction of pairs of neighbours of v linked to another node than v. In the projection, these nodes would be linked together even if v were not there.”
Social Networks and Concept Lattices

- Special issue of *Social Networks* in 1996
- Applications in sociology
  - Duquenne and Mohr
- Web communities
  - Rome and Haralick 2005
- Citation analysis
  - Tilley and Eklund (2007)
- Folksonomies and triadic contexts
  - Gerd Stumme’s talk on Thursday
- Data mining in social networks
  - Various authors
- ...
Overlapping Cliques
Freeman 1996

- Two actors are in relation $C$ if they belong to the same clique of size $k$
- **Groups** are the equivalence classes of $C^+$ (the transitive closure of $C$)
  - $xC^+y \iff xa_1Ca_2, \ldots, a_{n-1}Ca_n, a_nCy$
  - if $k = 2$, groups are the connected components of the network graph
Two actors are in relation $C$ if they belong to the same clique of size $k$.

Groups are the equivalence classes of $C^+$ (the transitive closure of $C$).

Clique lattice:
- Build a formal context with actors as objects and maximal cliques as attributes.
- Actors at the bottom are at the core of group, while peripheral members are closer to the top.
In highly connected networks, $C^+$ has only one equivalence class embracing all actors.

Increasing $k$, we miss small groups.

It is unnatural to forbid intersection of groups.

Idea: Find *bridging* cliques and remove them.
In highly connected networks, $C^+$ has only one equivalence class embracing all actors.

- Increasing $k$, we miss small groups.
- It is unnatural to forbid intersection of groups.

Idea: Find bridging cliques and remove them.
In highly connected networks, $C^+$ has only one equivalence class embracing all actors

Increasing $k$, we miss small groups

It is unnatural to forbid intersection of groups

Idea: Find *bridging* cliques and remove them
Define a **bridging clique** as one whose attribute concept belongs to maximal chains of different lengths.

Two actors are in relation $\hat{C}$ if they belong to the same non-bridging clique.

**Groups** are the equivalence classes of $\hat{C}^+$.
Rationale: Bridging cliques make positions of individuals in groups ambiguous
Problem:
- Lots of bridging cliques in real data
- Many actors belong to bridging cliques only

Define level- \( i \) concepts to be concepts that are at distance \( k \) from the top concept in the covering graph of the clique lattice.

Two actors are in the relation \( C_i \) if they belong to the extent of the same level- \( i \) concept.

Level- \( i \) groups are equivalence classes of \( C_i^+ \).
Problem:
- Lots of bridging cliques in real data
- Many actors belong to bridging cliques only

Define level-\(i\) concepts to be concepts that are at distance \(k\) from the top concept in the covering graph of the clique lattice.

Two actors are in the relation \(C_i\) if they belong to the extent of the same level-\(i\) concept.

Level-\(i\) groups are equivalence classes of \(C_i^+\).
Epistemic communities

Embryologists working on the zebrafish

- Expert-based description: (i) Biochemical signaling mechanisms, involving pathways and receptors, (ii) Comparative studies, (iii) Brain, nervous system.
Epistemic communities

**Embryologists working on the zebrafish**

- **Expert-based description:** (i) Biochemical signaling mechanisms, involving pathways and receptors, (ii) Comparative studies, (iii) Brain, nervous system.
Epistemic communities

Embryologists working on the zebrafish

- **Expert-based description:** (i) Biochemical signaling mechanisms, involving pathways and receptors, (ii) Comparative studies, (iii) Brain, nervous system.
Epistemic communities

Embryologists working on the zebrafish

- **Expert-based description:** (i) Biochemical signaling mechanisms, involving pathways and receptors, (ii) Comparative studies, (iii) Brain, nervous system.
Stability index

Intensional stability [Kuznetsov 1990, 2003]

How strongly the concept intent depends on particular objects of the concept extent:

\[ \sigma(A, B) = \frac{|\{C \subseteq A \mid C' = B\}|}{2|A|} \]
Stability index

Intensional stability [Kuznetsov 1990, 2003]
How strongly the concept intent depends on particular objects of the concept extent:

$$\sigma(A, B) = \frac{|\{C \subseteq A \mid C' = B\}|}{2|A|}$$

- The probability of preserving the intent after removing an arbitrary number of objects from the context:

$$\sigma(A, B) = \frac{|\{K_H \mid H \subseteq G \text{ and } B = B^{l_Hl_H}\}|}{2|G|}$$
Stability index

Intensional stability [Kuznetsov 1990, 2003]

How strongly the concept intent depends on particular objects of the concept extent:

$$\sigma(A, B) = \frac{|\{C \subseteq A \mid C' = B\}|}{2|A|}$$

- Is the topic well-established or is it likely to disappear when loosing some of its supporters?
- A similar idea could be used for identifying core and peripheral community members
**Stability index**

**Extensional stability**

How strongly the concept extent depends on particular attributes of the concept intent:

\[
\sigma(A, B) = \frac{|\{D \subseteq B \mid D' = A\}|}{2|B|}
\]

- The stability of communities as groups of people
- How much do these people really have in common?
Separation index

- Are members of this community sufficiently different from other actors?
- Are attributes describing the community sufficiently different from other attributes?

Separation index of concept \((A, B)\)

\[
s(A, B) = \frac{|A||B|}{\sum_{g \in A} |g'| + \sum_{m \in B} |m'| - |A||B|}
\]
Multirelational networks

- E.g., bloggers can be in various relations with each other
  
  $$x \text{ cites/comments/is friends with } y$$

- Find all dependencies such as
  
  if $$x$$ is friends with $$y$$ and $$y$$ cites $$z$$ then $$x$$ comments $$z$$

**FCA:** Finding dependencies for $$k$$ variables [Bernhard Ganter]

- Objects are all $$k$$-tuples of actors
- Attributes are $$x_iR x_j$$, where $$i, j \leq k$$ and $$R$$ ranges over the relations in the network
- $$(a_1, \ldots, a_k)$$ is related to $$x_iR x_j$$ if $$a_iR a_j$$ holds in the network
- The implications of this context are exactly the dependencies over $$k$$ variables in the network
- Drawback: Computationally heavy and admits no exceptions
E.g., bloggers can be in various relations with each other

- \( x \) cites/comments/is friends with \( y \)

Find all dependencies such as

- if \( x \) is friends with \( y \) and \( y \) cites \( z \) then \( x \) comments \( z \)

**FCA:** Finding dependencies for \( k \) variables [Bernhard Ganter]

- Objects are all \( k \)-tuples of actors
- Attributes are \( x_iR_{x_j} \), where \( i, j \leq k \) and \( R \) ranges over the relations in the network
- \((a_1, \ldots, a_k)\) is related to \( x_iR_{x_j} \) if \( a_iR_{a_j} \) holds in the network
- The implications of this context are exactly the dependencies over \( k \) variables in the network
- Drawback: Computationally heavy and admits no exceptions
Multirelational networks

- E.g., bloggers can be in various relations with each other
  
  $x$ cites/comments/is friends with $y$

- Find all dependencies such as
  
  if $x$ is friends with $y$ and $y$ cites $z$ then $x$ comments $z$

**FCA:** Finding dependencies for $k$ variables [Bernhard Ganter]

- Objects are all $k$-tuples of actors
- Attributes are $x_i R x_j$, where $i, j \leq k$ and $R$ ranges over the relations in the network
- $(a_1, \ldots, a_k)$ is related to $x_i R x_j$ if $a_i R a_j$ holds in the network
- The implications of this context are exactly the dependencies over $k$ variables in the network
- Drawback: Computationally heavy and admits no exceptions
Multirelational networks

- E.g., bloggers can be in various relations with each other

  $x$ cites/comments/is friends with $y$

- Find all dependencies such as

  if $x$ is friends with $y$ and $y$ cites $z$ then $x$ comments $z$

**FCA:** Finding dependencies for $k$ variables [Bernhard Ganter]

- Objects are all $k$-tuples of actors
- Attributes are $x_i R x_j$, where $i, j \leq k$ and $R$ ranges over the relations in the network
- $(a_1, \ldots, a_k)$ is related to $x_i R x_j$ if $a_i R a_j$ holds in the network
- The implications of this context are exactly the dependencies over $k$ variables in the network
- Drawback: Computationally heavy and admits no exceptions