Social Network Analysis using Formal Concept Analysis

Rokia Missaoui
LARIM
Université du Québec en Outaouais (UQO)
Canada
http://w3.uqo.ca/missaoui

January 30, 2013
Collaboration

IBM CBAP - University of Ottawa
LATECE - Université du Québec à Montréal
Bern University of Applied Sciences
Université Cheikh Anta Diop, Dakar
ERIC - Université Lumière Lyon 2
LIMOS - Université Blaise Pascal
Outline

- Introduction
- Social network analysis (SNA)
- Formal concept analysis (FCA)
- Using FCA for SNA
  - Community and core/peripheral node detection
  - Concept and association rule mining
- Complex structure management
  - Heterogeneous information networks
  - Multimodal data networks
- Conclusion
Introduction

Social network

Concept (Galois) lattice
Introduction

- **Big data and complex structures**
  - Performance and scalability issues
  - Visualization issues
- **Each user has his/her own needs for data analysis**
- **Data evolution and partitioning**
  - Need for incremental algorithms and operations on structures (lattices and graphs)
- **Solutions**
  - Efficient algorithms and implementations
  - Data selection and decomposition, nested structures
  - Pattern management, browsing, ....
Social Network

- **Definition**
  - A social structure of nodes (actors) that are related to each other by various ties such as friendship, affinity, collaboration, ...

- **Different types of graphs**
  - Simple, directed, weighted, or labeled graphs
  - One-mode or many-mode (multidimensional) data
  - Heterogeneous information networks with more than one type of nodes and/or links
One-mode vs. Two-mode Networks

(Sun & Han, 2012)
Social Network Analysis

- Many topics
  - Position analysis. E.g., leader or mediator, core/peripheral actor
  - Influence computation and maximization
  - Network reorganization
  - Link prediction and recommendation
  - Community detection and evolution, etc.
One-mode Data

- Interaction networks
  - A graph $G = (V, E)$, where $V$ is a set of vertices/nodes and $E$ a set of edges/links
  - E.g., friendship, co-authorship

- Example

http://mathworld.wolfram.com/AdjacencyMatrix.html
One-mode Data

**Example. Adjacency matrix**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Position Analysis

Centrality measure

- Degree
  How many people can this person reach directly?

- Betweenness
  How likely is this person to be the most direct route between two people in the network?

- Closeness
  How fast can this person reach everyone in the network?

- Eigenvector
  How well is this person connected to other well-connected people?

Interpretation in social networks

CNM Social Media Module – Giorgos Cheliotis (gcheliotis@nus.edu.sg)
The eccentricity of a node $i$ is the greatest geodesic distance between $i$ and any other node in the network.
Community Detection

- Find clusters in networks
  - E.g., research communities, Web groups, ...

- Methods
  - Hierarchical clustering
  - Girvan–Newman algorithm
  - Modularity maximization
  - Clique based methods (e.g., clique percolation method, Freeman’s approach)
  - Biclustering (e.g., block-modeling)
  - Spectral graph partitioning, etc.
Community Detection

- Algebraic approaches
  - Clique and $n$-cliques
  - Structural and regular equivalence
  - $k$-cores and $k$-components

- Algorithmic approaches
  - Larger definition of community: dense connections within a group but sparser ones between groups
  - Partition construction
  - Many algorithms (e.g. modularity maximization, ...)

Community Detection

- Cliques and n-cliques in undirected graphs
  - Clique: subgraph of at least three nodes which are all directly connected to one another
  - A maximal clique: a clique which does not exist within a larger one
  - \(n\)-clique: set of nodes such that the shortest distance between each pair of them is no longer than \(n\).
Community Detection

- **Algorithmic approaches**
  - **Agglomerative** using for instance similarity measures to produce dendrograms
  - **Divisive** using e.g. edge betweenness centrality
  - Graph exploration methods such as **clique percolation** which produces **overlapping cliques**
Community Detection

- An example
Community Detection

Dendrogram (produced by UCI NET)
Link Prediction

- Objective
  - Predict the link to be created between two nodes, based on the network topology and possibly other features
  - Hard when the network is sparse

- Examples
  - Predict a future link between two Web pages, two researchers, ...

- Methods
  - Learning algorithms (e.g., classification) and probabilistic models (e.g., Bayesian networks)
  - Collective prediction, e.g., Markov random field model
  - A proximity-based approach by Liben-Nowell & Kleinberg, etc.
Formal Concept Analysis

FCA (Ganter & Wille, 1999)

- Based on lattice and order theory; a conceptual clustering approach; a data mining framework for concepts and association rule computation

Achievements

- Efficient algorithms for lattice construction
- Association rule mining: minimal implication basis, succinct representation of association rules, etc...
- Extensions to FCA: logical, fuzzy, rough, and relational concept analysis
- Generalization to $n$ dimensions: triadic and polyadic CA
- Many applications in different domains (SNA, CS, ....)
Formal Concept Analysis

- **Formal context** := \((G, M, I)\) with \(I \subseteq G \times M\).
  
  \(G :=\) set of **objects** and \(M :=\) set of **attributes**.

- Derivation. \(A \subseteq G\) and \(B \subseteq M\).
  
  \(A' := \{m \in M \mid \forall g \in A \quad glm\}\)
  
  \(B' := \{g \in G \mid \forall m \in B \quad glm\}\).

- **Formal concept** := a pair \((A, B)\) with \(A' = B\) and \(B' = A\).
  
  \(A :=\) **extent** of \((A, B)\) and \(B :=\) **intent** of \((A, B)\).

- \(\mathcal{B}(G, M, I) :=\) set of all concepts of \((G, M, I)\).

- **Concept hierarchy**
  
  \((A, B) \leq (C, D) : \iff A \subseteq C \quad (\iff D \subseteq B)\).

- \(\mathcal{B}(G, M, I) := (\mathcal{B}(G, M, I), \leq)\)
Formal Concept Analysis

**Theorem**

\( \mathbb{B}(G, M, I) \) is a complete lattice in which infimum and supremum are given by:

\[
\bigwedge_{t \in T} (A_t, B_t) = \left( \bigcap_{t \in T} A_t, \left( \bigcup_{t \in T} B_t \right)^\prime \right)
\]

\[
\bigvee_{t \in T} (A_t, B_t) = \left( \left( \bigcup_{t \in T} A_t \right)^\prime, \bigcap_{t \in T} B_t \right).
\]

\( \mathbb{B}(G, M, I) \) is called the **concept lattice** of the context \((G, M, I)\).
Formal Concept Analysis

Formal context

EVENT

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Lattice with reduced labeling

A → ¬B

C → D

6 → 4

C → A [0.5]

Concept (Galois) lattice
Applying FCA for SNA
Applying FCA for SNA

Main contributions

- Analysis of affiliation networks (Freeman & White, 1993)
- Special issue of Social Networks in 1996
  - E.g., analysis of interaction networks using cliques and FCA (Freeman, 1996)
- Stability index of a concept (Kuznetsov 2007)
- Web communities (Rome & Haralick, 2005)
- Folksonomy analysis (Jäschke et al., 2006)
- Workshop on SNA using FCA (Obiedkov et al., 2007)
- Citation analysis (Tilley & Eklund, 2007)
- FCA in Sociology (Duquenne & Mohr, 2008), etc.
Freeman’s Approach to Group Detection

- Extract maximal cliques from a one-mode data network
- Form a formal context where objects are individuals and attributes are maximal cliques
- Construct the concept lattice
- Identify bridging cliques and edges,
- Eliminate bridging edges to produce communities
- Central actors are near the bottom of the lattice while peripheral ones are in the upper part
### Formal context

Objects are **actors** and attributes are **maximal cliques**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Concept Lattice

Infimum = (∅, {A, ..., H})

Supremum = ({1, ..., 15}, ∅)

Lattice with full labelling

Rokia Missaoui
EGC’2013 - Toulouse
Concept Lattice

Concept \([\{6, 7, 14\}, G]\)

Peripheral actor

Cliques A, B, C, ..., H at the 1st level

Peripheral actors

Central nodes

Lattice with reduced labelling

Rokia Missaoui  EGC’2013 - Toulouse
Bridging Cliques & Edges

- Bridging clique
- Bridging edge
Community Detection

Deletion of bridging edges

Central nodes
2, 3, 4, 5, 9, 6, 14

Community
Community Detection

- Limits of Freeman’s approach
  - The notion of clique is too restrictive: no cliques → no communities!
  - There may be many bridging edges
  - Some nodes (even core ones) are lost after the removal of bridging edges

- Improvement in (Falzon, 2000)
  - All the lattice layers are exploited rather than the clique (first) layer only
  - No node is lost
Two-mode Data

- $G=(V_1 \cup V_2, E \subseteq V_1 \times V_2)$, bipartite graph
- E.g., Southern **women** attending **events**
Two-mode Data

- Participation of women to events (Davis)

![Two-mode Data Table](image)

**FIGURE 5.** Davis, Gardner, and Gardner’s two mode data.
Concept Lattice

Actors: 1, .. 18
Events: A, .. N

Three event groups:
G1 = {A, B, .. E}
G2 = {F, G, H, I}
G3 = {J, K, L, M, N}

Central actors:
1, 2, 3, 4, 12, 13, 14, 15

Implications:

J → L: If an actor attends Event J, he does so for Event L
5 → 3, 4: Events attended by Actor 5 are also attended by 3 & 4

Concept ({3, 9}, ({E, G, H, I})
Conversion to one-mode Data

- Projection using matrix multiplication
  - $A \times A^T$ gives the number of events co-attended by both the row and the column women

Evelyn and Theresa co-attended 7 events
Conversion to one-mode Data

- Projection using matrix multiplication
  - $A^T \times A$ gives the number of women who attended both the row event and the column event

14 women attended event E8 and one woman attended both Events E8 and E11
Two-mode Data

- Detection of overlapping communities
  - Crampes et Plantié, 2012
  - Only the concepts of the first two layers of the concept lattice are generated
  - Measures such as cohesion, separation and autonomy are used to define communities from concepts
Biclustering

- Dual-projection approach (Everett & Borgatti, 2012)

Women in this group are structurally equivalent to core events.
Three-mode Data

- How FCA can be helpful?
  - Triadic concept analysis (Lehmann & Wille, 1995)
    - Triadic contexts, concepts and diagrams
    - Concept trilattices and their visualization
  - Triadic implications (Biedermann, 1998)
  - Polyadic concept analysis (Voutsadakis, 2002)
Three-mode Data

- More recent work

- Different types of triadic implications and research topics to explore (Ganter & Obiedkov, 2004)
- Algorithm TRIAS for triadic concept generation (Jäschke et al., 2006)
- Two algorithms for triadic concept generation: RSM and Cube Miner (Ji et al., 2006)
- Data Peeler for n-set computation (Cerf et al., 2008)
- Inter-dimensional rules (Nguyen et al., 2010)
- Triadic concept analysis with fuzzy attributes (Belohlávek et al., 2010)
A triadic context $\mathbb{K} := (K_1, K_2, K_3, Y)$ where $Y \subseteq K_1 \times K_2 \times K_3$. The elements of $K_1$, $K_2$ and $K_3$ are called (formal) objects, attributes and conditions, respectively. A triple $(a_1, a_2, a_3)$ in $Y$ means that object $a_1$ has attribute $a_2$ under condition $a_3$.

Triadic concept or (closed) tri-set

It is a triple $(A_1, A_2, A_3)$ with $A_1 \subseteq K_1$, $A_2 \subseteq K_2$, $A_3 \subseteq K_3$ and $A_1 \times A_2 \times A_3 \subseteq Y$ such that no $A_i$ (for $i=1, 3$) can be augmented without violating this condition. The subsets $A_1$, $A_2$ and $A_3$ are called the extent, the intent and the modus of the triadic concept $(A_1, A_2, A_3)$ respectively.
Three-mode Data

- Three mode (tridimensional) data: objects, attributes and conditions
- E.g. events (1 .. 5), researchers (P, N, R, K, S) and roles (a, b, c, d)
Three-mode Data

- Triadic concepts:
  - (12345, PRK, a), (12345, P, ad), (14, PN, bd), ...

- Rules
  - Any role (e.g., event organizer) played by S is also played by P
  - Whenever N attends events as a speaker (a) and PC member (d), then P does so
Heterogeneous information Networks

(Sun & Han, 2012)

Definition 1.1  (Information network) An information network is defined as a directed graph $G = (V, E)$ with an object type mapping function $\tau : V \rightarrow A$ and a link type mapping function $\phi : E \rightarrow R$, where each object $v \in V$ belongs to one particular object type $\tau(v) \in A$, each link $e \in E$ belongs to a particular relation $\phi(e) \in R$, and if two links belong to the same relation type, the two links share the same starting object type as well as the ending object type.

Heterogeneous IN when the number of object or link types is $>1$
Conclusion

- Observations
  - Many studies in FCA can be usefully exploited for mining social networks: negation, ontology-based analysis, visualization (e.g., nested line diagrams), context transformation and decomposition, ...

- Usefulness of FCA extensions
  - Triadic concept analysis (Lehmann & Wille 1995)
  - Logical CA (Ferré, 2000)
  - Relational CA (Rouane-Hacene et al., 2012),
  - Rough CA, fuzzy CA, etc.


Michel Crampes & Michel Plantié. Détection de communautés chevauchantes dans les graphes bipartis, MARAMI’2012.


Lucia Falzon. Determining Groups from the Clique Structure in Large Social Networks. Social Networks, 22, p. 159-172, 2000

References

References

- Yizhou Sun and Jiawei Han, Mining Heterogeneous Information Networks: Principles and Methodologies, Morgan & Claypool Publishers, 2012.