How can we identify and analyze subgroups in affiliation networks?

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus
Overview

Today’s Agenda:

Analysis of Affiliation Networks

• Properties of Affiliation Networks
  – Properties of One-Mode Networks derived from Affiliation Networks
• Galois Lattices for Affiliation Networks
How can we identify groups and subgroups in a social graph?
Subgroups in Co-Affiliation Networks
Borgatti 1997

- We talked about cliques, clans and clubs in 1-mode
- The obvious next step would be to try to identify these subgroups in co-affiliation networks.
  - For example, we can search for cliques, n-cliques, n-clans, n-clubs.
- Unfortunately, these methods are not well suited for analysing a bipartite graph.
  - In fact, bipartite graphs contain no cliques
  - In contrast, bipartite graphs contain too many 2-cliques and 2-clans.
  - One of the problems is that, in the bipartite graph, all nodes of the same type are necessarily two links distant.

⇒ we need to consider special types of subgraphs which are more appropriate for two-mode data.
Subgroups in Co-Affiliation Networks

So while

- cliques
- n-cliques
- n-clans
- n-clubs
- bi-cliques

Are useful instruments to analyze one-mode networks, in two mode networks they are problematic.
Definition of a Clique
• A clique in a graph is a maximal complete subgraph of three or more nodes.

Remark:
• Restriction to at least three nodes ensures that dyads are not considered to be cliques
• Definition allows cliques to overlap

Informally:
• A collection of actors in which each actor is adjacent to the other members of the clique

Fig. 7.1. A graph and its cliques
n cliques
[Wasserman Faust 1994]

N-cliques require that the **geodesic distances** among members of a subgroup are small by defining a **cutoff value** \( n \) as the maximum length of geodesics connecting pairs of actors within the cohesive subgroup.

An n-clique is a maximal complete subgraph in which the largest geodesic distance between any two nodes is no greater than \( n \).

Which 2-cliques can you identify in the following graph?

**NOTE**: Geodesic distance between 4 and 5 "goes through" 6, a node which is not part of the 2-clique.
Reminder: Social Networks Examples

Why and How to Flash Your BIOS

This article is going to focus on the basics and explain ways to flash the BIOS, precautions and how to recover in case of a bad flash.

Why and How to Flash Your BIOS (Page 1 of 4) Flashing the BIOS is one of the most feared topics related to computers. Yes, people should be very cautious because it can be dangerous. This article is going to focus on the basics and explain ways to flash.
Transforming Two Mode Networks into One Mode Networks
[Wasserman Faust 1994]

- Two one mode (or co-affiliation) networks
  (folded from the children/party affiliation network)

\[ M_P = M_{PC} \times M_{PC}' \]

C...Children
P...Party

Fig. 8.5. Actor co-membership matrix for the six children

<table>
<thead>
<tr>
<th></th>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( n_3 )</th>
<th>( n_4 )</th>
<th>( n_5 )</th>
<th>( n_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 )</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>( n_2 )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( n_3 )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>( n_4 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( n_5 )</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>( n_6 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 8.6. Event overlap matrix for the three parties

<table>
<thead>
<tr>
<th></th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1 )</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>( m_2 )</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>( m_3 )</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

[Images taken from Wasserman Faust 1994]
Properties of Affiliation Networks
[Wasserman Faust 1994]

- Properties of Actors and Events
  - Rates of Participation
  - Size of Events

- Properties of One-Mode Networks that are derived from Affiliation Networks
  - Density
  - Reachability
  - Connectedness
  - Diameter
  - Cohesive Subsets of Actors or Events
  - (Reachability for Pairs of Actors)

But how are they influenced by their projections in two-mode networks?
Density
[Wasserman Faust 1994]

Reminder: Density in regular networks is the ratio of edges to vertices

- The **density** of a one-mode network derived from an affiliation network is a **function of the pairwise ties between actors or between events**.
- The number of overlap ties between events is, in part, a function of the number of events to which actors belong.
- An actor only creates a tie between a pair of events if it belongs to both events.
- An actor who belongs to
  - only one event creates no overlap ties between events
  - exactly two events creates a single tie
  - three events creates three ties
  - ...
- In General
  - An actor who belongs to n events creates n*(n-1)/2 ties
- Thus,
  - the rates of membership for actors influence the number of ties between events, and
  - the sizes of the events influence the number of ties between actors.
Reachability, Connectedness, Diameter

[Wasserman Faust 1994]

Reminder: Two nodes in a graph are adjacent if there is a line between them

• In an affiliation network,
  – no pair of actors is adjacent
  – No pair of events is adjacent

• no paths of length 1 between actors, but potentially paths of some longer length

• **Reachability** corresponds to path lengths between nodes

• An affiliation network is **connected** when all pairs of nodes (both actors and events) are reachable

• The **diameter** of an affiliation network is the length of the longest shortest path between any pair of nodes.
Reachability for Pairs of Actors

[Wasserman Faust 1994]

• In a valued graph, we can **define connectedness at level c** as the subsets of actors all of whom are connected at some minimum level c.
• Two nodes are **c-connected** (or reachable at level c) if there is a path between them in which all lines have a value of no less than c.

• Basis for the k-neighbourhood graph $G_k$ / KNC Plot

![Diagram](image)
Cohesive Subsets of Actors or Events

[Wasserman Faust 1994]

Reminder: a clique is a maximal complete subgraph of three or more nodes

• In a valued graph, we can define a **clique at level c** as a maximal complete subgraph of three or more nodes, all of which are adjacent at level c

• That is all pairs of nodes have lines between them with values that are greater than or equal to c. *By increasing c, we can locate more cohesive subgroups.*

• A **clique at level c** is a subgraph in which all pairs of actors share memberships in no fewer than c events

• Basis for the k-neighbourhood graph G_k / KNC Plot

![Diagram of a clique at level 1]
Affiliation Networks
[Wasserman Faust 1994]

"Two mode" networks
- Links only between the two modes
- Folding
- K-neighbourhood graph
- KNC Plot

But:
By folding, information is lost
Redundancy
[latapy 2006, p 22]

redundancy coefficient of $v$:

$$rc(v) = \frac{|\{\{u, w\} \subseteq N(v), \exists v' \neq v, (v', u) \in E \text{ and } (v', w) \in E\}|}{|N(v)|(|N(v)| - 1)/2}.$$  

In other words, the redundancy coefficient of $v$ is 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, see Figure 9; this is why we call this the redundancy. If it is equal to 1 then the projection would be exactly the same without $v$; if it is 0 it means that none of its neighbours would be linked together in the projection.  

Figure 9: Example of redundancy computation. From left to right: a bipartite graph, its $\perp$-projection, and the $\perp$-projection obtained if the node $A$ is first removed. Only two links disappear, leading to $rc(A) = \frac{2}{6} = 0.666 \ldots$.
Affiliation Networks
[Wasserman Faust 1994]

So:

How can we simultaneously analyze **Actors AND Events** in Affiliation Networks?

Can we show both,
- the relationships among the entities within each mode, and also
- how the two modes are associated with each other?
Galois Lattices
[Freeman White 1993]

A satisfactory representation should facilitate the visualization of three kinds of patterning:
1. the actor-event structure,
2. the actor-actor structure, and
3. the event-event structure

at the same time.
Évariste Galois
1811-1832

- A republican (fighting the French king)
- Not allowed to enter Ecole Polytechnique twice
- No recognition of his work as a mathematician (not considered for the Academy of Sciences Grand Prize in Mathematics for his ideas on solutions for quintic equations)
- Then he focused on politics / sentenced to prison for marching against the king
- Romance with a mysterious woman who was engaged
- Died in a pistol duel with her fiance at the age of 20
- A letter to his friends
  "I beg my patriots, my friends, not to reproach me for dying otherwise than for my country. I died the victim of an infamous coquette and her two dupes. It is in a miserable piece of slander that I end my life. Oh! Why die for something so little, so contemptible? I call on heaven to witness that only under compulsion and force have I yielded to a provocation which I have tried to avert by every means". [Fermat’s Last Theorem]
- Spent the night before the duel writing down his mathematical achievements ("I have no time!“ see image to the right)
Galois Lattices (or Galois Connections)  
[Wasserman Faust 1994, Freeman White 1993]

- A long history in Mathematics  
- Introduced by Birkhoff in 1940 (cf. Birkhoff „Lattice Theory“ 1967)  
- Affiliation networks focus on subsets and the duality of the relationship between actors and events  
- The idea of subsets refers both to subsets of actors contained in events and subsets of events that actors attend.  
- The idea of duality refers to the complementary perspectives of relations  
  - between actors as participants in events, and  
  - between events as collections of actors.  
- Galois lattices incorporate both ideas.
A Lattice
[Wasserman Faust 1994]

• Galois lattices are special kind of lattices

Consider a set of elements $N = \{n_1, n_2, \ldots n_g\}$ and a binary relation „less than or equal“ ($\leq$) that is reflexive, antisymmetric and transitive.

Formally
• Reflexive: $n_i \leq n_i$
• Antisymmetric: $n_i \leq n_j$ and $n_j \leq n_i$ iff $n_j = n_i$
• Transitive: $n_i \leq n_j$ and $n_j \leq n_k$ implies $n_i \leq n_k$

Such a system defines a **partial order** on the set $N$.
(Partially Ordered Sets, POS, poset)
A Lattice
[Wasserman Faust 1994]

For any pair of elements, \( n_i, n_j \), we define their

- **lower bound** as that element \( n_k \) such that \( n_k \leq n_i \) and \( n_k \leq n_j \)
- **upper bound** as that element \( n_k \) such that \( n_i \leq n_k \) and \( n_j \leq n_k \)

With that,

- A lower bound is called a **greatest lower bound** \( n_k \) (or meet/infimum) of \( n_i, n_j \)
  if \( n_i \leq n_k \) for all lower bounds \( n_l \) of \( n_i, n_j \)
- An upper bound is called a **least upper bound** \( n_k \) (or join/supremum) of \( n_i, n_j \)
  if \( n_k \leq n_l \) for all upper bounds \( n_l \) of \( n_i, n_j \)

A **lattice** consists of a set of elements \( N \), a binary relation \( \leq \) that is reflexive, antisymmetric and transitive and each pair of elements \( n_i, n_j \), has both a least upper bound and a greatest lower bound.

A **lattice** is thus a partially ordered set in which each pair of elements has both a meet and a join.
“⊂” as our relation

What is the greatest lower bound (meet) of Allison and Eliot? And what does it mean?

Each point represents a subset of parties

Both attended all parties that Keith attended (P3).

Drew: {Party 2}
Keith: {Party 3}
Sarah: {Party 1, Party 2}
Eliot: {Party 2, Party 3}
Allison: {Party 1, Party 3}
Ross: {Party 1, Party 2, Party 3}

Fig. 8.10. Relationships among children as subsets of birthday parties.
What is the least upper bound (join) of Allison and Eliot? And what does it mean?

Ross attended all parties that Allison and Eliot attended.

Drew: \{Party 2\}
Keith: \{Party 3\}
Sarah: \{Party 1, Party 2\}
Eliot: \{Party 2, Party 3\}
Allison: \{Party 1, Party 3\}
Ross: \{Party 1, Party 2, Party 3\}

Fig. 8.10. Relationships among children as subsets of birthday parties.
Ross attended all parties that Eliot attended as well.

Elliot attended all parties that Keith and Drew attended as well.

Drew: \{Party 2\}
Keith: \{Party 3\}
Sarah: \{Party 1, Party 2\}
Elliot: \{Party 2, Party 3\}
Allison: \{Party 1, Party 3\}
Ross: \{Party 1, Party 2, Party 3\}

Fig. 8.10. Relationships among children as subsets of birthday parties.
Lattice
[Wasserman Faust 1994]

Who attended the most parties?
Who attended the least parties?

Fig. 8.10. Relationships among children as subsets of birthday parties.
Did Sarah attend all parties that Eliot and Drew attended?
Lattice
[Wasserman Faust 1994]

Each point represents a subset of children

Party 1: \{Allison, Ross, Sarah\}
Party 2: \{Drew, Eliot, Ross, Sarah\}
Party 3: \{Allison, Eliot, Keith, Ross\}

Fig. 8.9. Relationships among birthday parties as subsets of children
A Galois Lattice
[Wasserman Faust 1994]

A Galois lattice (also called a Galois connection) focuses on the relation between two sets.

- A relation $\lambda$ is defined on pairs from the Cartesian product $N \times M$.
- $\lambda$ is thus defined on pairs, a relation $n_i \in N \lambda m_j \in M$

We let the sets $N$ and $M$ be the set of actors and the set of events, and let $\lambda$ be the relation of affiliation.

- Thus, $n_i \lambda m_j$ if actor $i$ is affiliated with event $j$.

We also have $\lambda^{-1}$ where $m_j \lambda^{-1} n_i$ if event $j$ contains actor $i$. 
A Galois Lattice
[Wasserman Faust 1994]

Just as we have considered an individual actor and the subset of events with which it is affiliated, we can also consider a subset of actors and the subset of events with which all of these actors are affiliated.

We can define two mappings:
• \( \uparrow : N_s \rightarrow M_s \) from a subset of actors \( N_s \subseteq N \) to a subset of events \( M_s \subseteq M \) such that \( n_i \lor m_j \) for all \( n_i \in N_i \) and all \( m_j \in M_j \).
  - In terms of an affiliation network, the \( \uparrow \) mapping goes from a subset of actors to that subset of events with which all of the actors in the subset are affiliated.
  - For example, if there is no event with which all actors in subset \( N_s \) are affiliated, then \( \uparrow(N_s) = 0 \)
• \( \downarrow \) mapping can be defined analogously.
Galois Lattice
[Wasserman Faust 1994]

Fig. 8.11. Galois lattice of children and birthday parties

Which parties did Eliot attend?
Galois Lattice

[Wasserman Faust 1994]

Which parties did both Eliot and Allison attend?

Drew: {Party 2}
Keith: {Party 3}
Sarah: {Party 1, Party 2}
Eliot: {Party 2, Party 3}
Allison: {Party 1, Party 3}
Ross: {Party 1, Party 2, Party 3}

Fig. 8.11. Galois lattice of children and birthday parties
Galois Lattice
[Wasserman Faust 1994]

Who attended party 1?

Party 1: \{Allison, Ross, Sarah\}
Party 2: \{Drew, Eliot, Ross, Sarah\}
Party 3: \{Allison, Eliot, Keith, Ross\}

Fig. 8.11. Galois lattice of children and birthday parties
Galois Lattice
[Freeman White 1993]

• Reduced and Full Labeling

<table>
<thead>
<tr>
<th>EVENT</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTOR</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

FIGURE 2. Hypothetical two mode data.

FIGURE 3. Lattice of the data of Figure 2—full labeling.

FIGURE 4. Lattice of the data of Figure 2—reduced labeling.
Galois Lattice - Example

[Freeman White 1993]

**FIGURE 5.** Davis, Gardner, and Gardner's two mode data.

**FIGURE 6.** Lattice of the Davis, Gardner, and Gardner data.
What can we do with Galois Lattices?

1. We can see the pattern of participation of actors in events.
2. We can see the downward containment structures of events.
3. We can see the upward containment structures of actors.
4. We can distinguish classes of events.
5. We can see the segregation of actors by the event classes.
Galois Lattice
[Freeman White 1993]

What can we do with Galois Lattices?

1. We can see the pattern of participation of actors in events.
   - Each actor (or set of actors) participated in those events labeled at or above her labeled point in the line diagram and each event (or set of events) included all the actors labeled at or below its point.
   - Thus the relation $\lambda (I)$ is displayed, and the original data are completely recoverable from the diagram.

2. We can see the downward containment structures of events.
   - The uppermost set of seven labeled events (E, F, G, H, I, K, and L) are the events that involved the largest sets of actors.
   - Other events are contained in the lower intersections (meets) of these events. Event C is a second level event: It is contained in event E, and events A, B, and D are, in turn, third level events; they are contained in C (and therefore in E).
   - Similarly, event J is second level, contained in L, and M and N are third level, contained in J.

3. We can see the upward containment structures of actors.
   - The lowest labeled actors (1, 2, 3, 4, 13, 14, and 15) are primary. They are the actors who were active in the largest sets of events.

4. We can distinguish classes of events.
   - Two sets of events $E_1 = \{A, B, C, D, E\}$ and $E_2 = \{J, K, L, M, N\}$ share no common actor. This is shown by the fact that their lower bound falls at the bottommost point, the point that contains no common actors. Therefore, $E_1$ and $E_2$ are group-defining events. In contrast. The four events $E_3 = \{F, G, H, I\}$ each share at least one actor with events in $E_1$, and at least one actor with events in $E_2$; they might be called bridging events.

5. We can see the segregation of actors by the event classes.
   - The nonoverlapping event sets $E_1$ and $E_2$ segregate all but two of the actors into two sets $A_1 = \{1, 2, 3, 4, 5, 6, 7, 9\}$ and $A_2 = \{10, 11, 12, 13, 14, 15, 17, 18\}$. Actors from these different subsets never interact in the non-overlapping events.
Advantages of Galois Lattices
[Wasserman Faust 1994]

• Focus on subsets
  – Especially appropriate for representing affiliation networks

• Complementary relationships between actors and events displayed at the same time

• Patterns in the relationships between actors and events may be more apparent in the Galois lattice

⇒ Galois lattice serves much the same function as a graph or sociogram (which serves as a representation of a one-mode network)
Shortcomings of Galois Lattices

[Wasserman Faust 1994]

• Visual display of Galois Lattices can become quite complex
• No unique „best“ visual representation for a given Galois lattice
• Although the vertical dimension represents the degrees of subset inclusion relationships, the horizontal dimension is arbitrary.
• Properties and further analyses of Galois lattices (unlike networks) are not well developed

➤ Galois lattices are primarily an exploratory representation of an affiliation network, from which one might be able to see patterns in the data.
ConExp

Download: http://sourceforge.net/projects/conexp
Project Website: http://conexp.sourceforge.net/
Documentation: http://conexp.sourceforge.net/users/documentation/

• Import your network data using .csv format
  (see „Opening existing documents“ in the documentation)
• Experiment with „Drawing Options“ and layouting the lattice manually
• Show „multi-labels“ for objects and attributes
• Interpret the lattice
• Save it in .cex file
Any questions?

See you next week!