Social Network Analysis and CI (Collective Intelligence)

A. Geyer-Schulz, B. Hoser
Institute of Information Systems and Management
Universität Karlsruhe (TH)
Germany
Content

- Social Network Analysis

- Social Network Analysis and CI
  - The Social Self
  - Social Perception
  - Social Influence
  - Group Processes

- Two Examples of SNA Applied
  - Topic Trend Detection
  - Fraud Detection in Markets
Which source to trust

directed friendship network
Social Network Analysis

- Individuals (actors) are not isolates regarding their actions. They always act within the possibilities and constraints given by their social environment.

- Examples
  - Smoking in groups of high school kids
  - Fashion
  - Trading at the stock market

- Interactions are modelled as networks.

- Methods from such fields as graph theory, mathematics, physics, sociology, social psychology are used to analyze these networks.
Short Overview of history

• Since the 1930ies sociometry is used (Moreno)
• Visualization of interactions between members of groups or between groups as graphs
• Analysis and methodological tool among others: Social Network Analysis.
• Research topic started at around the 1970ies (Freeman, Wellman, Wasserman, Faust, Bonacich, etc.)
• Has boomed in recent years due to paradigm shift from individual actors to networks and the internet with its data availability and recently with the hype on social network sites.
A few points on SNA

- **Description of networks:**
  - Denseness, connectedness, structuredness, randomness
- **Description of actors:**
  - Central (diverse centrality measures depending on the goal of the analysis), bridges, isolates, etc.
- **Models on network formation**
  - Homophily (’birds of a feather flock together’), balance theory, strength of weak ties, etc.
- **We focus on centrality measures, especially eigenvector centrality:**
Eigenvector centrality

- An actor is called central if he is connected to central actors → recursive

- History:
  - 1953: Katz
    - Transfer of eigensystem approach to social networks with symmetric relationships
  - 1972 /2001: Bonacich,
    - Approach to analyze asymmetric networks by introducing an exogenous factor inherent to the actor apart from his network connections

- Our approach (2004):
  - Use of complex-valued adjacency matrices for asymmetric communication networks.
Complex hermitian adjacency matrix

\[ H = (A + i A^T) e^{-i\pi/4} \]

- \( A \) : real valued adjacency matrix of graph \( G \), \( a_{ii} = 0 \).
- \( G(E,V,w) \): \( e_{ij} \in E \) edges with weights \( a_{ij} \) if \( v_i \rightarrow v_j \) \( \forall v_i, v_j \in V \)
- \( A^T \) : transpose of \( A \)
- \(-1 = i^2\) imaginary unit
- \( e^{-i\pi/4} \) : rotation (or scaling) factor
Characteristics of Hermitian Eigensystem

(1) \( H = H^* \) (Hermitian)
   \[ \rightarrow \lambda_i \in \mathbb{R}, \forall \ i \]
   \[ \rightarrow \text{Since trace}(H) = 0 \rightarrow \lambda_i \in \mathbb{R}^\pm \]

(2) \( HH^* = H^*H \) (normal)
   \[ \rightarrow \langle x_i, x_j \rangle = c \delta_{ij}, \text{with } \delta_{ij} = \begin{cases} 1; & i=j \\ 0; & i \neq j \end{cases} \]
   \[ \rightarrow x_{ij} \in \mathbb{C} \]
   \[ \rightarrow \text{For all rotations} \]

(3) Spectral decomposition (complete)
   \[ \rightarrow \sum \lambda_i P_i = H, \ P_i = x_i x_i^* , \forall i \]
Characteristics of Hilbert Space

- Complete normed inner product space
- Norm: $\langle x, x \rangle = ||x||^2 := 1$ (normalized)
- Distance:
  $$d(x,y) = ||x-y||^2 = \langle x-y, x-y \rangle$$
  $$= \langle x, x \rangle + \langle y, y \rangle - \langle x, y \rangle - \langle y, x \rangle$$
  $$= 2 - 2\text{Re}(\langle x, y \rangle)$$
- if $\text{Re}(\langle x, y \rangle) \approx 1 \rightarrow d \approx 0$
Interpretation

- Eigensystem $Hx = \lambda x$ still describes the recursive definition of centrality.
- Eigenvalues can be interpreted as weights of the orthogonal projectors $P$. Thus the higher the absolute eigenvalue, the more relevant $P$.
- The orthogonal projectors $P$ define independent communication behavior patterns within the network.
- The value of each component of each eigenvector is complex. The absolute value gives the relative relevance of actor $i$ on communication pattern $k$. The phase gives the direction of behavior with respect to all other actors.
Summary: Social Network Analysis

- SNA is a methodological approach to analyze networks of actors and the assumption that no one acts outside his or her social environment.
- SNA can provide models to simulate or explain behavior in networks based on the analysis.
- For collective intelligence SNA provides part of the social context.
SNA and Collective Intelligence

• The Social Self
• Social Perception
• Social Influence
• Group Processes
Social Self
Self-Awareness and Behavior

Self-focusing persons
- Private Self-Consciousness (I think about myself)
- Public Self-Consciousness (I am concerned about what others think about me)

Self-focusing situations

Self-awareness

Accessibility of self-discrepancies

Expectation for discrepancy reduction

Match behavior to standards
- Personal standards
- Social standards or norms

Withdraw from self-awareness

HIGH

LOW
Social Perception

Observation → Attribution → Snap judgments
- Need for self-esteem triggers motivational biases (ingroup vs. outgroup)
- Personal attribution (automatic)
- Situational attribution (cognitive effort)

Dispositions → Integration
- Emotions
- Mood Priming

Impressions

Confirmation
- Self-fulfilling prophecies

Perceiver

co-funded by the European Union
Social Perception: Social Identity

- Personal identity
  - Need for self-esteem
    - Social identity
    - Group achievements
      - Ingroup favoritism
      - Outgroup derogation
    - Personal achievements
      - Self-esteem
Social Influence

Sources of influence:
- Strength (status, ability, relationship)
- Immediacy
- Number (up to a point)

Targets of influence:
- Strength (status, ability, relationship)
- Immediacy
- Number (up to a point)
Social Processes: Groupthink

Preconditions:
- High cohesiveness
- Group structure: homogenous members, isolation, directive leadership, unsystematic procedures
- Stressful situation

Groupthink

Symptoms
- Overestimation of group
- Close-mindedness
- Increased pressures toward uniformity: Mindguards and pressure on dissenters, self-censorship, illusion of unanimity

Consequences
- Defective decision-making: Incomplete survey of alternatives, incomplete survey of objectives, failure to examine risks of preferred choice, failure to reappraise initially rejected alternatives, poor information search, selective bias in processing information at hand, failure to work out contingency plans
- High probability of a bad decision
Social Processes: Help in Emergency

Emergency

Step 1: Notice that something is happening
- Distraction
  Stop fooling around kids, we are here to eat.
- Self-concerns
  I'm late for a very important date.

Step 2: Interpret event as an emergency
- Ambiguity
  Is she really sick or just drunk?
- Relationship between attacker and victim
  They have to resolve their own family problems.
- Pluralistic ignorance
  No one else seems concerned.

Step 3: Take responsibility for helping
- Diffusion of responsibility
  Someone else must have called the police.
- Lack of competence
  I'm not trained to handle this, and whom can I call?

Step 4: Decide how to help

Step 5: Provide help
- Audience inhibition
  I'll look like a fool.
- Costs exceed rewards
  What if I do something wrong? He'll sue me.
Topic trend detection:

- Research with industry partners (Siemens and Münchner Rückversicherung)
- Goal: to find hot topics being discussed in newsgroups about mobile phones, respectively about health relevant issues discussed in blogs.
- Approach: find (eigenvector)-central actors in newsgroup/blog network; find the relevant words/phrases they used and combine these two inputs to define 'hot topics by relevant people'.
**Topic Trends**

- **Research question**
  - Finding topic trends generated and sustained over time by relevant people within the networks

- **Approach**
  - Classical content analysis
  - Enriched by social network analysis information
  - Model network of words used by actor
Authors use of words (reduced)

news@domain: vibra, stummschaltung, etc.
raldo: bedienung, rufumleitung

henklbr: 'english', browser, video, stream

var

ID
Results: Authors use of words (full)
**Discussion - Static**

- Eigensystem analysis finds structure and ranking in a given data set.
- Communication networks: structure of the communication between vertices/agents is analyzed.
- We can identify the relevant vertices/agents based on the complete group, and on the substructures in which they mainly participate.
- By using the directional information we can now find the clusters of agents and identify them.
- In the case of author to author networks: Each author can be assigned to a certain subgroup based on his behavior.
- In the case of company to company networks: Each company can be assigned to a certain subgroup based on the behavior of the authors.
- In the case of authors’ use-of-words: Subgroups consist of authors and words. The clusters here are built from the common use of the words by all authors within the subgroup.
Discussion – Time dependent

• ‘Trace’ vertices/agents or groups over time.
• Shifts can be made visible. These shifts reflect the changing relevance in communication.
• Shifts can be used when looking for topic shifts. Words which are used by ‘rising’ subgroups may be more important than words used by ‘declining’ subgroups.
Improvements

- Multiple identities/synonyms
  - Authors: same person - different email addresses
  - Companies: Telekom, T-kom, t-kom
  - Words: misspelling of words, language/translation

- Words (not complete):
  - Intelligent stemming: siemen_
  - Elimination of stop words
  - Filtering (for example frequency based) of most frequent and of very rare words to eliminate 'auch', 'schreib' etc. or a positive list

- Time:
  - Time stamp correction
Fraud Detection

- In forecasting markets with prizes for the best traders as incentive, two types of fraud (behavior not consistent with market regulations) can be expected:
  - Money transfer (ring of traders, multiple accounts)
  - Price manipulations (in or outgoing stars, potentially with losses)

(Examples by courtesy of Jan Schröder, FSM (Forecasting Strategy Markets))
**Money Transfer**

The 2006 State Parliament Elections in Baden-Württemberg in Germany

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Market opening</td>
<td>2006-01-12 16:23:44</td>
</tr>
<tr>
<td>Market close</td>
<td>2006-03-26 18:00:00</td>
</tr>
<tr>
<td>Number of traders (at least one sell or buy transaction)</td>
<td>306 traders</td>
</tr>
<tr>
<td>Number of traders (at least one sell transaction)</td>
<td>190 traders</td>
</tr>
<tr>
<td>Number of traders (at least one buy transaction)</td>
<td>291 traders</td>
</tr>
<tr>
<td>Number of transactions</td>
<td>10786 transactions</td>
</tr>
<tr>
<td>Number of shares</td>
<td>7 shares</td>
</tr>
<tr>
<td>Average volume per trade</td>
<td>214.6 shares</td>
</tr>
<tr>
<td>Average moneyflow per trade</td>
<td>2,462.1 MU</td>
</tr>
<tr>
<td>Moneyflow in total</td>
<td>26,556,378 MU</td>
</tr>
<tr>
<td>Shareflow in total</td>
<td>2,314,197 shares</td>
</tr>
</tbody>
</table>
Share Prices / Election Forecast

[Graph showing price trends over time for different political parties]
And the Winners are ...

First 8 highest ranks in the highscore for the market for the 2006 state parliament elections in Baden-Württemberg in Germany

<table>
<thead>
<tr>
<th>Rank</th>
<th>Username</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>elfriede</td>
<td>226688.71</td>
</tr>
<tr>
<td>2</td>
<td>herrritie</td>
<td>176265.35</td>
</tr>
<tr>
<td>3</td>
<td>gruener</td>
<td>172738.39</td>
</tr>
<tr>
<td>4</td>
<td>MarcEichler</td>
<td>166769.19</td>
</tr>
<tr>
<td>5</td>
<td>potato joe</td>
<td>164285.52</td>
</tr>
<tr>
<td>6</td>
<td>Maio_Shan</td>
<td>156898.10</td>
</tr>
<tr>
<td>7</td>
<td>Ritvars</td>
<td>155599.08</td>
</tr>
<tr>
<td>8</td>
<td>henning</td>
<td>147850.64</td>
</tr>
</tbody>
</table>
Are they honest?
No, elfriede (#1) used 4 accounts
And henning (#8) used two!
Price Manipulation

Statistical Data of the 2007 national parliament elections in Switzerland

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Market opening</td>
<td>2007-09-11 10:00:00</td>
</tr>
<tr>
<td>Market close</td>
<td>2007-10-21 12:00:00</td>
</tr>
<tr>
<td>Number of traders (at least one sell or buy transaction)</td>
<td>511 traders</td>
</tr>
<tr>
<td>Number of traders (at least one sell transaction)</td>
<td>310 traders</td>
</tr>
<tr>
<td>Number of traders (at least one buy transaction)</td>
<td>487 traders</td>
</tr>
<tr>
<td>Number of transactions</td>
<td>16,421 transactions</td>
</tr>
<tr>
<td>Number of shares</td>
<td>8 shares</td>
</tr>
<tr>
<td>Average volume per trade</td>
<td>350.7 shares</td>
</tr>
<tr>
<td>Average moneyflow per trade</td>
<td>3,781.4 MU</td>
</tr>
<tr>
<td>Moneyflow in total</td>
<td>62,093,811 MU</td>
</tr>
<tr>
<td>Shareflow in total</td>
<td>5,759,656 shares</td>
</tr>
</tbody>
</table>
A new party (GLP), the forecasts are far off: Trader 3224 is a manipulator.

Lots of inbound trades:

Outbound trades
Literature

- Bettina Hoser, Jan Schröder, Andreas Geyer-Schulz, Maximilian Viermetz, Michal Subacz. Topic trend detection in newsgroups. KI (Künstliche Intelligenz) 3, p.37-40. 2007
Institute of Information Systems and Management

Universität Karlsruhe (TH)
D-76128 Karlsruhe

Phone: +49 . 721 . 608 – 8402 or 8407
Fax: +49 . 721 . 608 - 8403
Social Intelligence

- Three interconnected layers comprise social intelligence
  - Content of social interaction (e.g. email content)
  - Meta communication layer (e.g. Choice of communication channel, wording, empathy, ...)
  - Our focus: Structural Information derived from Social Interaction e.g.:
    - Communication patterns (email, chat)
    - Link structures between personal profiles
    - Resource sharing (Collaboration)
    - Ranking functions (friend vs. buddy, trust, etc.)
    - Choice behavior (indirect interaction by choice of products, friends, etc.)
Social Intelligence

• Help to base individual actions on results of social interaction e.g.:
  ▪ Choose travel destination not only based on recommendations derived from information retrieval, but also from personal relationship (e.g. friendship) with information source
  ▪ Validate information about possible emergency not only based on information retrieval from pictures but also from trustworthiness of source ranked by social network (generated on past experience)
Social “Media Intelligence”

- Social context (community) determines e.g. tags for media objects
- Assignment of semantics is a social process
- Support media intelligence by social intelligence, e.g.:
  - present tags that were used by close contacts within social network relevant to topic. Example: Picture of sunset with clouds
    - Travel community: sunset in the pacific
    - Meteorology community: cumulus clouds over pacific ocean
    - Pilot community: flight conditions over the pacific
  - Present media based on tags used by given social network