I know who you clicked last summer

Svenja Schröder

December 30, 2007

Table of contents

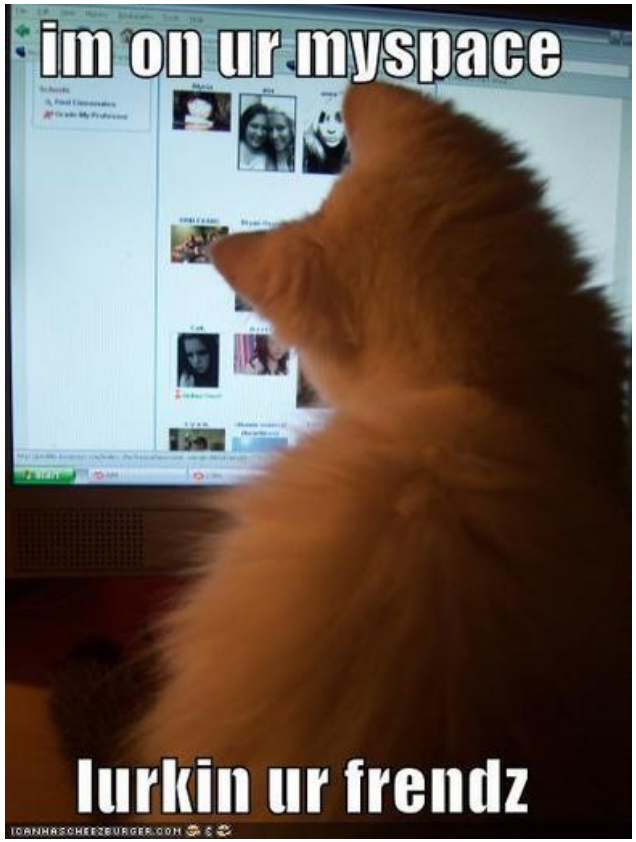
Introduction

Introduction

What is social network analysis?

- interdisciplinary approach: sociology, formal mathematics, statistics
- models actors and their relations (e. g. interactions) into an interpretable model
- precise methods for defining social concepts

Application fields



Application fields

- used in sociology, geography, social psychology, communication studies, information science, economics, biology, game theory ...
- analysis of all kinds of social structures
  - communities of actors
  - groups and subgroups
  - associations, societies, states
- popular example: social software
  - web2.0-communites

Small world experiment

- several experiments in the 70s by Stanley Milgram
- he examined the average path length for social networks of people in the US
  - packets were randomly sent to US citizens
  - they should forward it to another randomly chosen US citizen
  - if they didn't know each other personally, the sender should send the packet to a person of whom he/she thought he/she could know the receiver
- 'Six degrees of separation' myth

Basics

Basics

What does a social network look like?

- Social network = network = graph
  - many concepts of SNA are based on network or graph theory
- models social structures as a graph
  - consists of a set of **nodes** (vertices) and **edges** (ties)
  - edges are the most important part

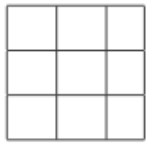
Nodes and edges

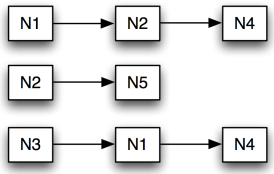
- **nodes** represent the actors or elements which are examined
  - sometimes with node weights (uncommon)
- **edges** are the relationships between the actors
  - differ in content, weighting and form
  - **content:** represent relationships, transactions or communication forms (e. g.)
  - **weighting:** can be weighted to express different intensity levels
  - **form:** *directed* or *undirected*

Types of networks

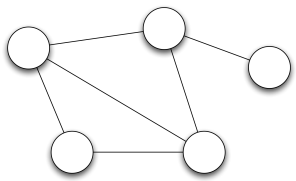
- **directed or undirected:** have directed or undirected edges (important for network flow)
- **one-mode:** one type of nodes (typical)
- **two-mode:** two types of nodes
  - edges only from one type of node to another type (directed)
- **bipartite / affiliation networks:** two-mode network with a set of actors and a set of affiliations
- **ego-centered:** from the view of one actor

Common network representations

- **Matrix** 
  - the classical, mathematical way
- **Adjacency list**



  - stores the neighbor or follower nodes for every node in a linked list
  - better to compute, less memory intense
- **Graphical visualisation**



  - easy to grasp, intuitionally comprehensible

Paths in a network

- not only direct, but also indirect connections (path length > 1)
- **path:** 'walk' between two nodes without visiting a node twice
  - edge direction important in directed networks!
- **geodesic / distance:** shortest path between two nodes
- paths describe the flow of information and resources
- long path length can lead to distortion of information

Measures

Measures

- Network measures
- Actor measures

Density

- measure for connectivity in the network
- number of edges divided by the number of possible edges:
- $\Delta = \frac{E}{E_{max}}$
- subject of the network has influence on density (friendship networks vs. co-worker networks)

Degree

- very basic measure for influence of an actor
- undirected network:
  - **degree:** number of edges
- directed network:
  - **indegree:** number of incoming edges
  - **outdegree:** number of outgoing edges

Centrality and prestige

- describe importance, visibility and prominence of actors in the network
- prominent actors have big influence

Centrality

- measures involvement of actor in connections of other actors
  - works on undirected networks
- 3 types of centrality:
  - degree centrality
  - closeness centrality
  - betweenness centrality
- centrality measures are normed to network size $[0, 1]$

Betweenness centrality

- used most frequently
- measures for the number of shortest paths going through the actor
- $C_B(n_i) = \sum_{j<k,\ i\neq j,\ i\neq k} \frac{g_{jk}(n_i)}{g_{jk}}$

Prestige

- indicates how many direct or indirect votes an actor receives
  - = number of incoming relations (path $\geqq 1$)
  - needs a directed network
- Different types of prestige:
  - indegree prestige
  - proximity prestige
  - ...

What else is there?

- Clustering
- Clique analysis
- ...

Algorithms

Algorithms

Algorithms in SNA

- mostly borrowed from graph theory and interpreted accordingly
  - time complexity must be considered!
- this section: some example algorithms

Shortest paths: dijkstra algorithm

- classical algorithm in graph theory (Dijkstra, 1959)
- calculates shortest path between two nodes for positive edge weights
- used for calculation of centrality and prestige

Importance: HITS algorithm

- 'hypertext induced topic selection', hubs and authorities concept (Kleinberg, 1999)
- calculates importance based on links in graph structure
- every node gets a **hub** and an **authority** value
  - **hub:** node is linked often by nodes with high authority values

$$h(p) = \sum_{(q,p)\in E} a(q)$$

  - **authority:** node links many nodes with high hub values

$$a(p) = \sum_{(q,p)\in E} h(q)$$

  - mutually reinforcing relationship
- not classical, but yet interesting

Enhancing everything with an ontology

Enhancing everything with an ontology

Ontology

Gruber, 1993

'An ontology is an explicit specification of a conceptualization.'

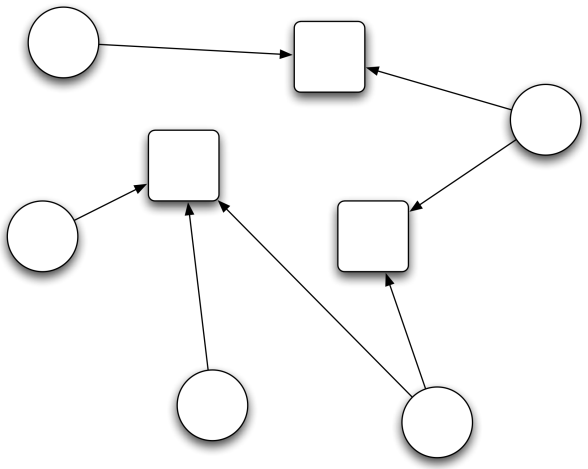
Gruber, 2000

'An ontology defines (specifies) the concepts, relationships, and other distinctions that are relevant for modeling a domain.'

- philosophical term, later used in KI and knowledge modeling
- today: one of the basic ideas of semantic web proposed by Tim Berners-Lee
- can be expressed as a graph

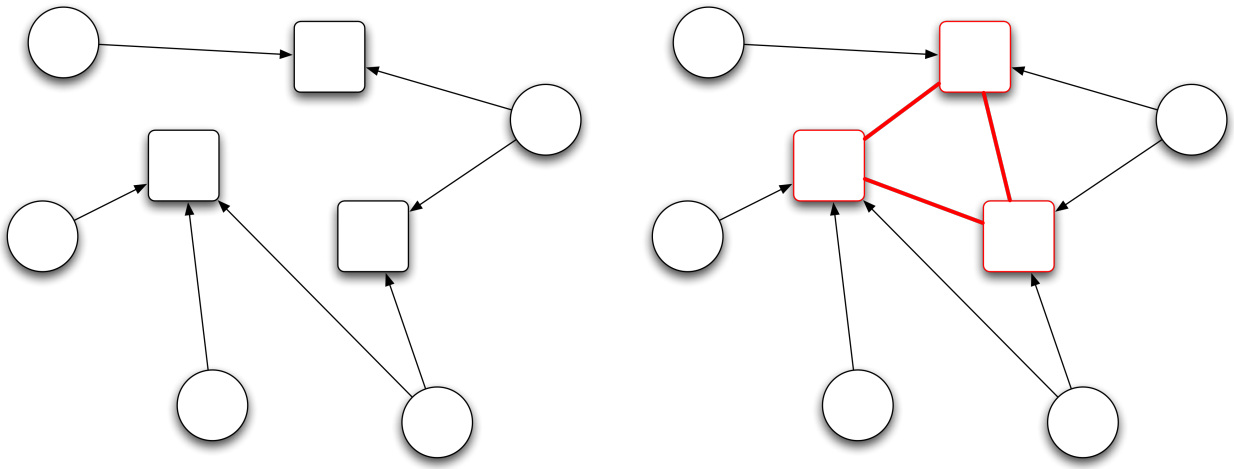
Why an ontology?

- e. g. widely used in recommender systems
- new ‘dimension’ in bipartite networks



Why an ontology?

- e. g. widely used in recommender systems
- new ‘dimension’ in bipartite networks



Ontology based recommendations

- categorization of data items
- structuring of meta data
- overcomes the *cold start problem* of a recommender system

Example

Example

The example...

- Sputnik data (www.openbeacon.org)

Generating social events

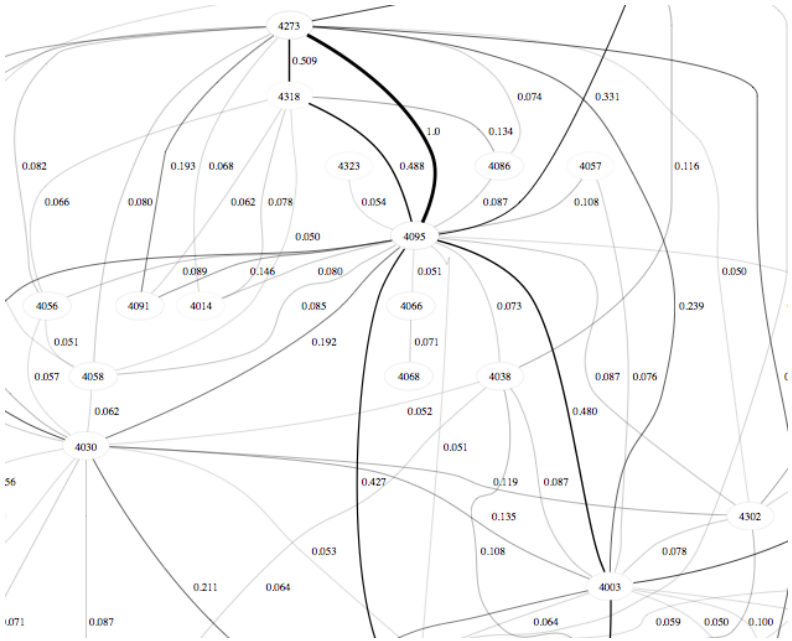
- 3 types of 'social events':
  - followed
  - stayed
  - talk attendance
- additionally generated to the other Sputnik data this year

Calculated measures

- **nodes:**
  - degree
  - betweenness centrality
- **edges:**
  - normalized followed value
  - normalized stayed value
  - simultaneously attended talks
  - ‘friendship factor’

Sorry...

- . . . not yet finished
- still needs some tuning



List of references

- Wasserman and Faust: 'Social Network Analysis'
- Jansen: 'Einführung in die Netzwerkanalyse. Grundlagen, Methoden, Forschungsbeispiele'
- Wikipedia

...

KTHXBYE!

For further questions: *svenja@23bit.net* or *http://sv.23bit.net*