Social Network Analysis using Graph Metrics of Web-based Social Networks

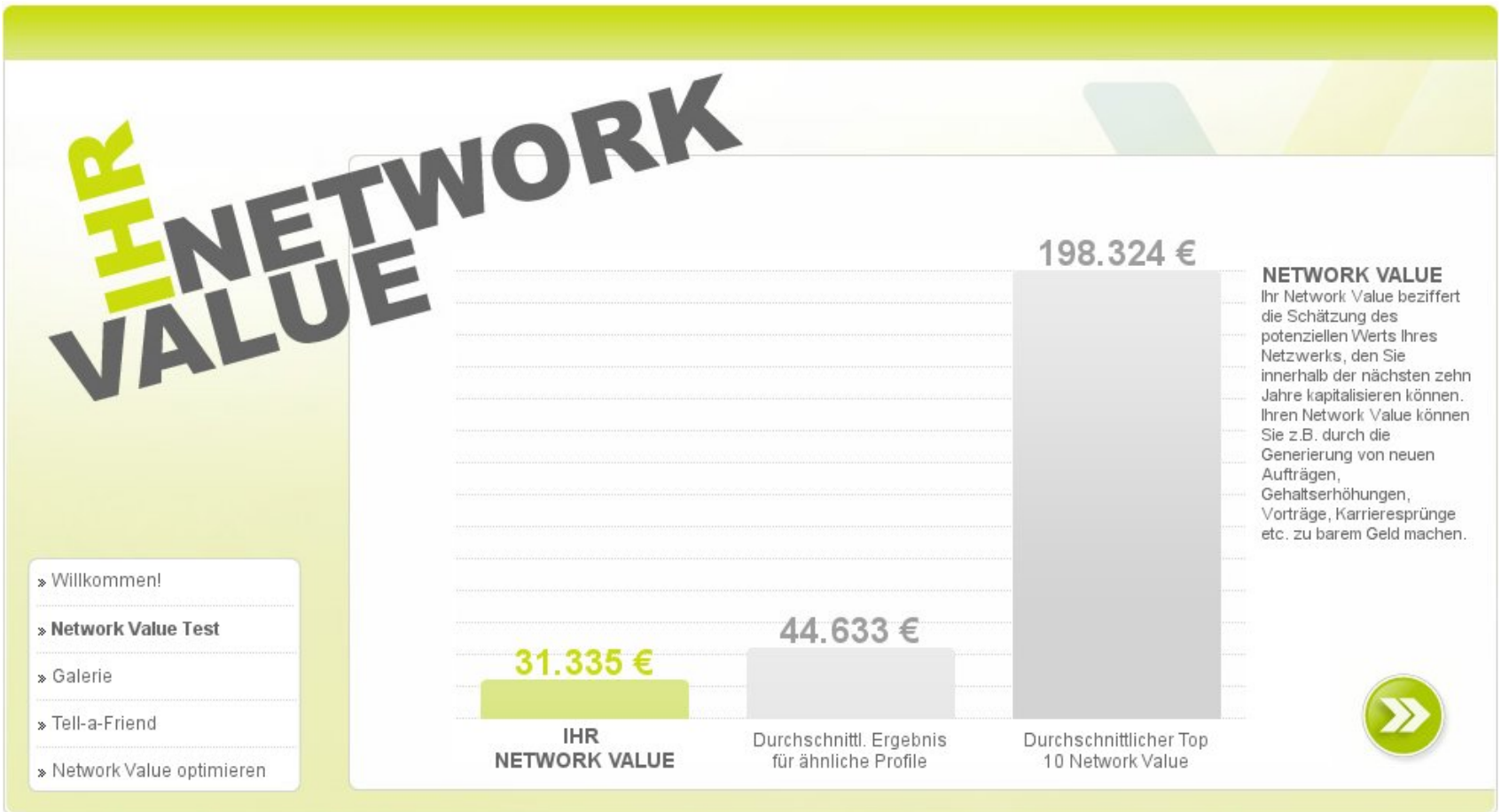
Robert Hilbrich

Department of Computer Science
Humboldt Universität Berlin

November 27, 2007

. . . taken from `http://mynetworkvalue.com/` (XING AG)

. . . and here my results:



How do they know that?

Research Area called Social Network Analysis (SNA)

. . . is a key technique in modern sociology, anthropology, sociolinguistics, geography, social psychology, communication studies, information science, organizational studies, economics, and biology as well as a popular topic of speculation and study.

How do they know that?

Research Area called Social Network Analysis (SNA)

. . . is a key technique in modern sociology, anthropology, sociolinguistics, geography, social psychology, communication studies, information science, organizational studies, economics, and biology as well as a popular topic of speculation and study.

Goal

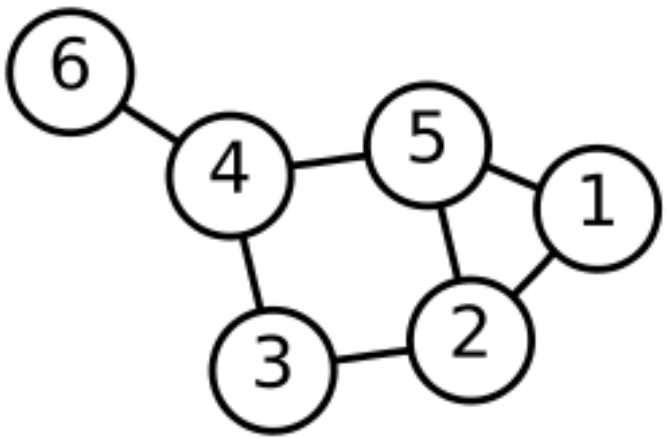
Analyze the structure of a social network to infer knowledge about an individual or a group.

Agenda

Graphs & Metrics

A graphical way to express relations.

What are Graphs?

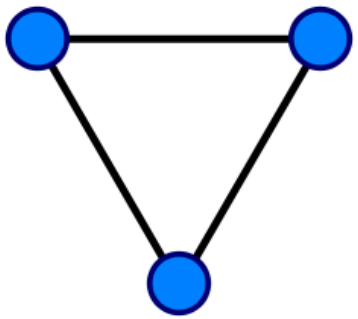


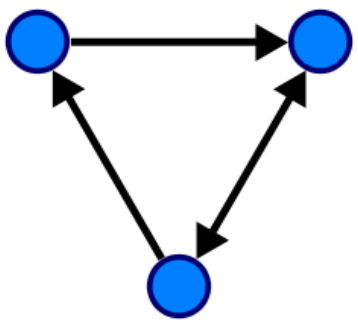
Graphs are ...

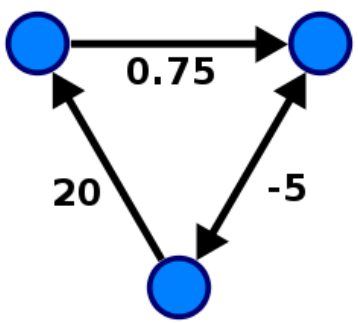
mathematical structures used to model pairwise relations between objects.

- nodes to represent objects ($\rightarrow$ *actors*)
- edges to express relations ($\rightarrow$ *communication paths*)

Graphs can be ...

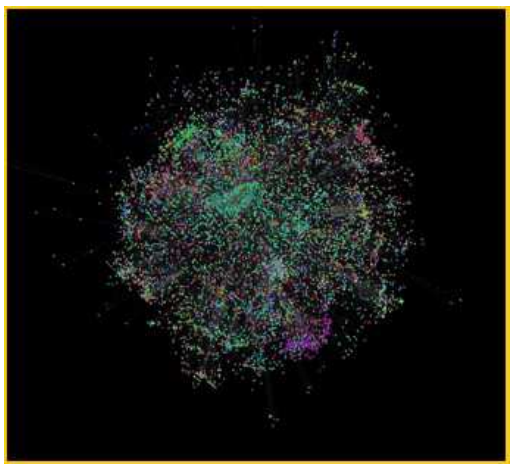






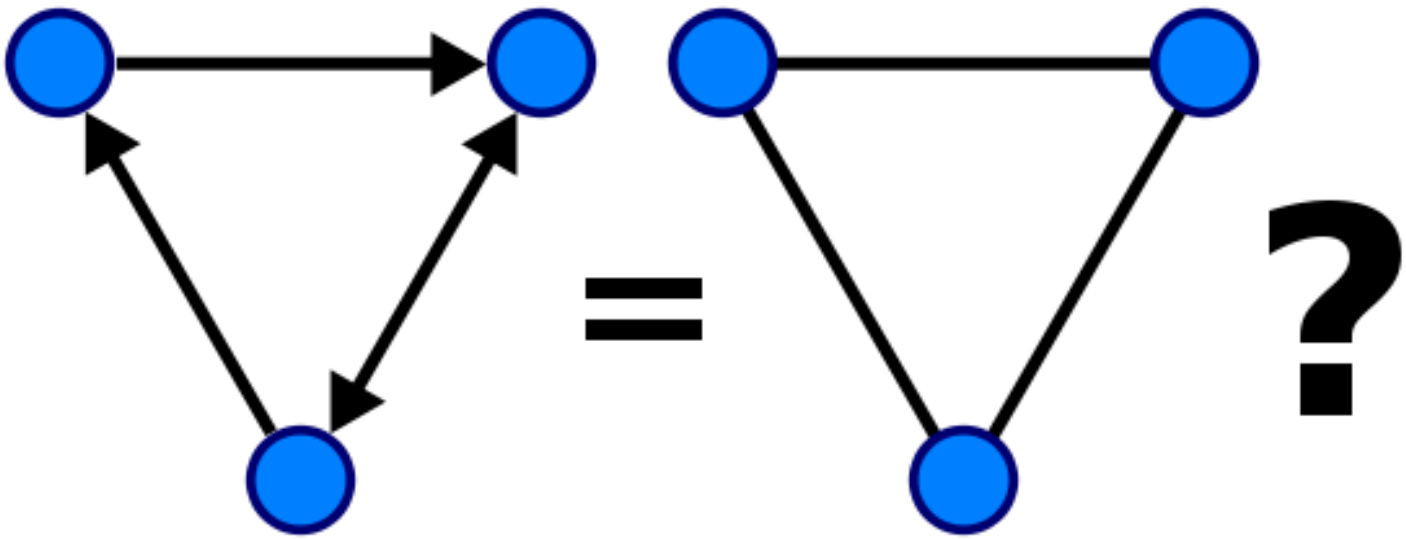
- undirected
  to represent (only) symmetric relations
- directed
  to represent asymmetric ($\rightarrow$ directed) and symmetric relations
- weighted
  to represent intensities, distances or costs of relations

Examples



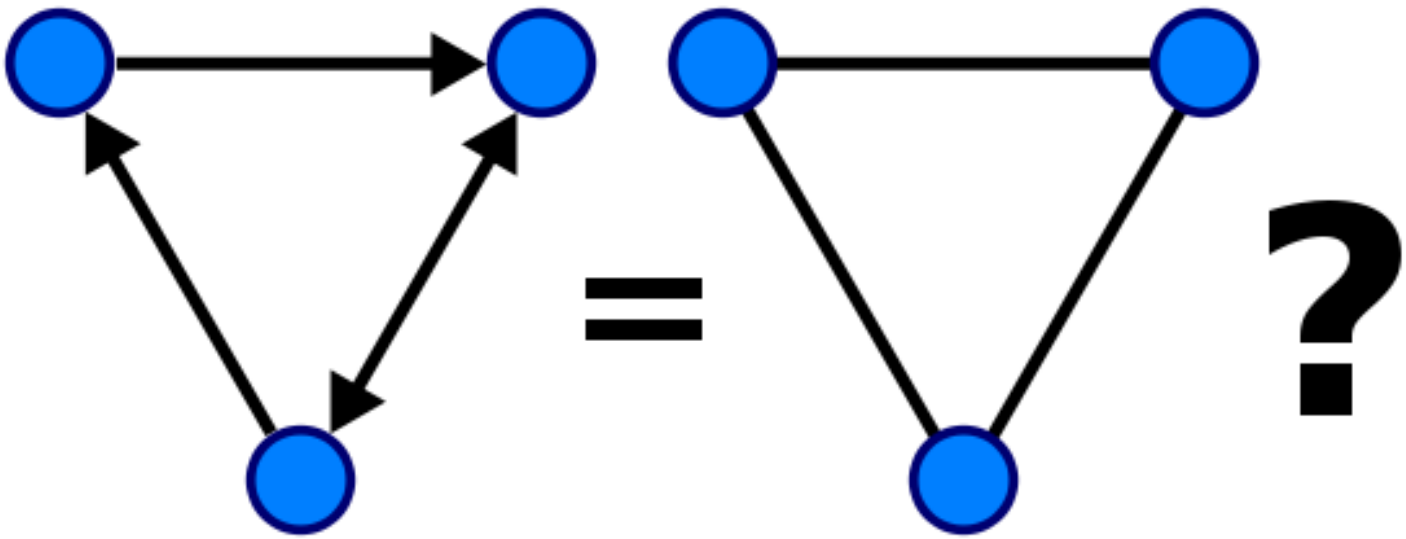
- `http://mit.edu/networks`
  Research with a 250 million node graph (call logs)
- Vizster - Visualizing Online Social Networks
  "... exploration of the community structure of friendster, tribe, ..."
  `http://www.cs.berkeley.edu/~jheer/vizster/early_design/`
- ...

What are Graphs Metrics?



In SNA, we need to compare Graphs with other Graphs

What are Graphs Metrics?



In SNA, we need to compare Graphs with other Graphs
Need measures for Graphs $\longrightarrow$ Graph Metrics!

Graph Metrics

Graph Metrics

. . . are of properties of graphs to compare

- Graphs with other Graphs
- Nodes with other Nodes

Graph Metrics

Graph Metrics

. . . are of properties of graphs to compare

- Graphs with other Graphs
- Nodes with other Nodes

Simple examples

- # of nodes per graph
- # of edges per node

Graph Metrics in SNA

Social Network Analysis (SNA) needs/deserves special metrics.

Using SNA metrics in . . .

- static graphs
  graph properties at a given point in time ($\rightarrow$ snapshot)
- dynamic graphs
  graph properties observed over a period of time ($\rightarrow$ evolution)

Graph Metrics in SNA

Social Network Analysis (SNA) needs/deserves special metrics.

Using SNA metrics in . . .

- static graphs
  graph properties at a given point in time ($\rightarrow$ snapshot)
- dynamic graphs
  graph properties observed over a period of time ($\rightarrow$ evolution)

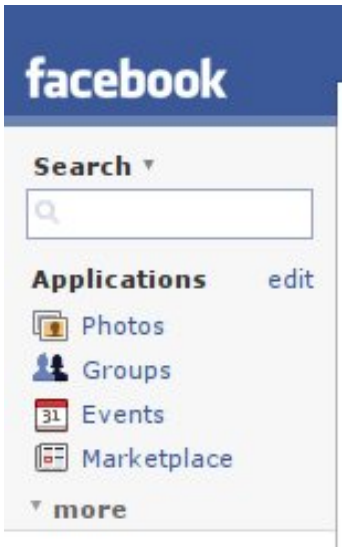
Some popular SNA metrics

- *Density*: $\frac{\#\text{ of edges}}{\#\text{ of possible edges}}$
- *Characteristic Path Length (CPL)*:
  relative connectedness of a social network
- . . .

Graphs for Social Networks

Modelling a web-based social network using graphs

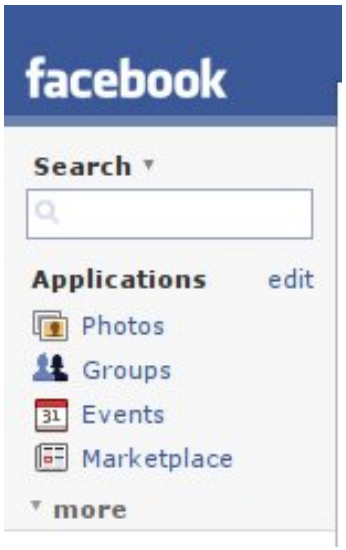
Modelling a Social Network



Now you got all the tools

- to build a model ($\rightarrow$ graphs) and
- analyze it ($\rightarrow$ graph metrics)

Modelling a Social Network



Now you got all the tools

- to build a model ($\rightarrow$ graphs) and
- analyze it ($\rightarrow$ graph metrics)

Questions

What can you infer about real life from these models?
To what extent do your metrics reflect real life?

Critical Remarks

Example

Intuitive metric: Popularity of a node $\rightsquigarrow$ # edges per node

Critical Remarks

Example

Intuitive metric: Popularity of a node $\backsim$ # edges per node

How about "A. Schmidt" with 192 344 friends in XING?
What if "A. Schmidt" also had a Gold Account at StayFriends?

Critical Remarks

Example

Intuitive metric: Popularity of a node $\backsim$ # edges per node

How about "A. Schmidt" with 192 344 friends in XING?
What if "A. Schmidt" also had a Gold Account at StayFriends?

Critical Comments

- web-based social networks reflect an *artificial* social network
- being part of a social network attaches a *label* (Facebook vs. XING)
- joining a community may have *purpose* or *strategy* behind it

Application Areas

Who is using graphs and metrics of social networks?

static models: Reputation management

...taken from `http://amazon.com`

main idea: reciprocity

"the more other people value my opinions, the more weight has my opinion on other topics."

dynamic models: Find terrorist activity

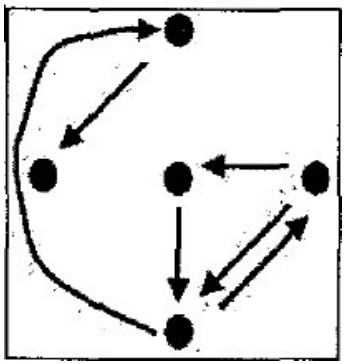
Idea

- Given a directed graph representing communication paths
- Goal: Find secretive / terrorist activity in cells in the graph
- Method: Find *communication patterns* that uncommon for human social networks

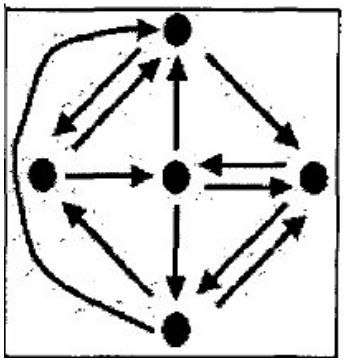
. . . DOD / DARPA sponsored research (!!)

Evolution of *Human Social Networks* vs. *Secretive Activity*

Human Social Networks



$t = t_0$

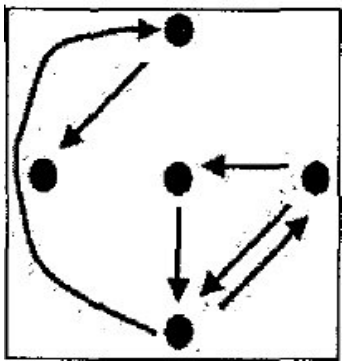


$t = t_1$

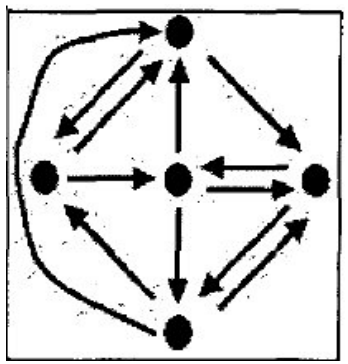
- no central mediator
- geodesic assumption
- redundancy assumption

Evolution of *Human Social Networks* vs. *Secretive Activity*

Human Social Networks



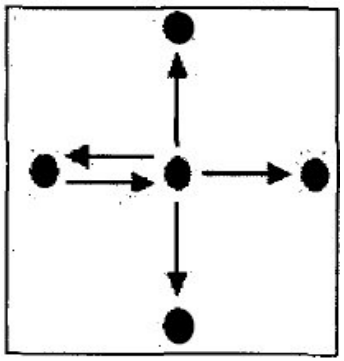
$t = t_0$



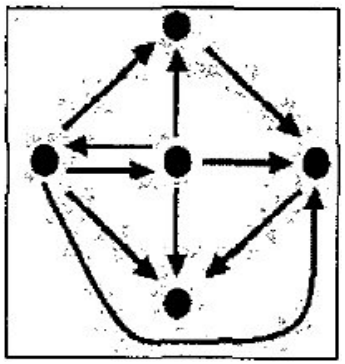
$t = t_1$

- no central mediator
- geodesic assumption
- redundancy assumption

Secretive Activity



$t = t_0$ (*sleeping*)



$t = t_1$ (*active*)

- *sleeping*: central mediator to reduce risk for leaks
- *active*: efficiency needs direct communication

Find secretive activity

Which metrics are used?

- mediated communication $\longrightarrow$ Characteristic Path Length$\uparrow$
- Density is lower compared to real live
  (reduce # possible leaks $\Rightarrow$ reduce # communication paths)

Find secretive activity

Which metrics are used?

- mediated communication $\longrightarrow$ Characteristic Path Length$\uparrow$
- Density is lower compared to real live
  (reduce # possible leaks $\Rightarrow$ reduce # communication paths)

Experiments

- on a generated graph do *pattern-search* for metrics
- Result: accuracy rate of 96%

Find secretive activity

Which metrics are used?

- mediated communication $\longrightarrow$ Characteristic Path Length$\uparrow$
- Density is lower compared to real live
  (reduce # possible leaks $\Rightarrow$ reduce # communication paths)

Experiments

- on a generated graph do *pattern-search* for metrics
- Result: accuracy rate of 96%

Remarks ...

- Found a pattern = Found Terror Cell?
- Pattern based on Leninist Cell-Communication Model
- However: AT&T develops similar approach to detect fraud!

Keynote

What should you take home?

Keynote

Should we use graph metrics?

Keynote

Should we use graph metrics?

Yes, because . . .

- AT&T, DOD(DARPA), . . .
- Data-Mining
  (fast, simple, cheap)
- Only need the Graph
- High accuracy rate

Keynote

Should we use graph metrics?

Yes, because ...

- AT&T, DOD(DARPA), ...
- Data-Mining
  (fast, simple, cheap)
- Only need the Graph
- High accuracy rate

No, because ...

- Based on Models
  $\longrightarrow$ Model Errors?
- Observation noise
  $\longrightarrow$ effect on metrics?
- False interpretations possible!
- How to build a complete Social Network?

Further Plans for the Paper

1. Further analysis of terror papers to show *weak spots and false assumptions*

2. Learn more about *Reputation Management*
3. Start it soon :)

Bibliography (. . . more on the WIKI)


Thayne Coffman, Seth Greenblatt, and Sherry Marcus.
Sensitivity of social network analysis metrics to observation noise.
In *Proc. IEEE Aerospace Conf.*, 2004.


Thayne Coffman and Sherry Marcus.
Dynamic classification of groups using social network analysis and hmms.
In *Proc. IEEE Aerospace Conf.*, 2004.


Corinna Cortes, Daryl Pregibon, and Chris Volinsky.
Communities of interest.
Lecture Notes in Computer Science, 2189:105–??, 2001.


John Scott.
Social Network Analysis, A Handbook, Second Edition.
SAGE Publications, 2000.

Questions or Comments?