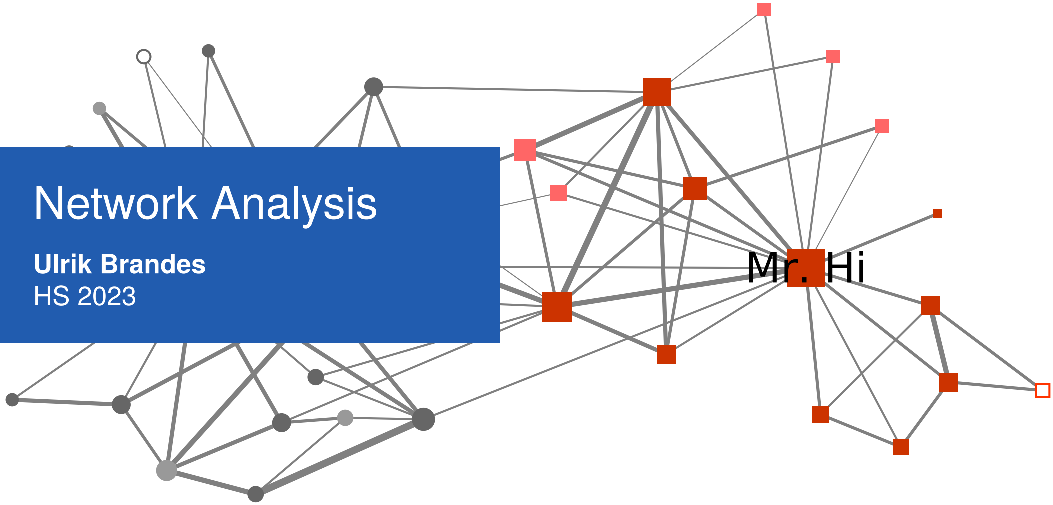
ETH zürich
Department of Humanities, Social and Political Sciences
Social Networks Lab

Network Analysis

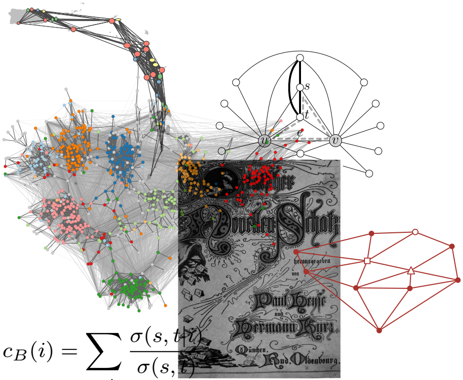
Ulrik Brandes

HS 2023



who are we?

Professorship Social Networks

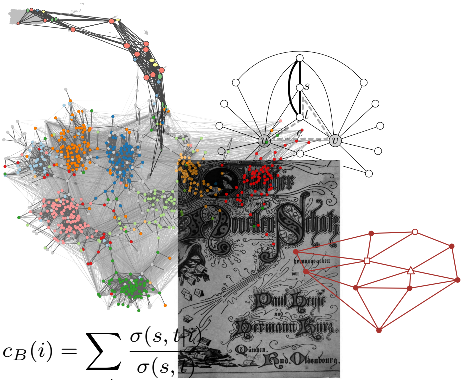


$$c_B(i) = \sum_{s,t} \frac{\sigma(s,t|i)}{\sigma(s,t)}$$

- Prof. Dr. Ulrik Brandes
- Denise Weber (WEP J 17)
- Dr. Julian Müller
- Seyedrouzbeh Hasheminezhad
- Gordana Marmulla
- Meher Chaitanya Pindiprolu
- Hadi Sotudeh
- Annina Stahl
- Hongjin Wu
- Wei Zhang

who are we?

Professorship Social Networks (algo@sn)



$$c_B(i) = \sum_{s,t} \frac{\sigma(s,t|i)}{\sigma(s,t)}$$

- Prof. Dr. Ulrik Brandes
- Denise Weber (WEP J 17)
- Dr. Julian Müller
- Seyedrouzbeh Hasheminezhad
- Gordana Marmulla
- Meher Chaitanya Pindiprolu
- Hadi Sotudeh
- Annina Stahl
- Hongjin Wu
- Wei Zhang

Social Networks Lab
jointly with Prof. Dr. Christoph Stadtfeld (zing@sn)

organization and outlook
Science in Perspective

wednesdays, 18:15–20:00, CAB G 11
3 credits: mid-term (30min; early November, online) and final exam (60min; late December, written exam)
slides, lecture notes, references, software via course moodle

organization and outlook
Science in Perspective

wednesdays, 18:15–20:00, CAB G 11
3 credits: mid-term (30min; early November, online) and final exam (60min; late December, written exam)
slides, lecture notes, references, software via course moodle

seminar (each semester)
network science applied in selected domain with teams, own contribution, mini-conference

theses and projects (on demand):
method development, implementation, replication, or your own network study

companion lectures
Introduction to Social Networks (spring, Stadtfeld): empirical research
Network Modeling (fall, Stadtfeld): statistical network models
Network Clustering (spring, Müller): clustering in and of networks

The Spectacles of Measurement (spring, Brandes): data, quantification, meaning
Soccer Analytics (spring, Brandes): rankings, match analysis, transfers, fans & media

who are you?

enrolled (as of earlier today)	**290**
computer science, computational science	183
mathematics, statistics & data science	66
life sciences, medicine & engineering	11
physics	10
cyber security	9
mobility & others	6
doctoral	5

BSc : MSc : PhD approx. 4 : 3 : 0.25

why are you here?

how did you learn about this course?

how did you learn about this course?

Mark S. Granovetter (1973). The strength of weak ties.
American Journal of Sociology 78(6):1360–1380

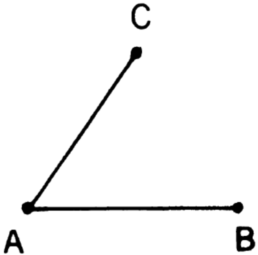


FIG. 1.—Forbidden triad

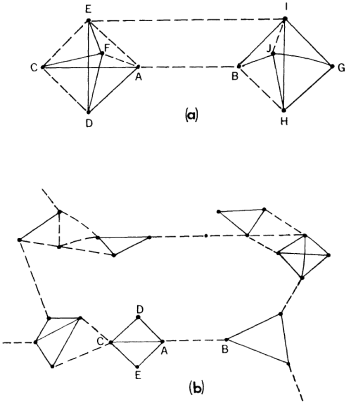
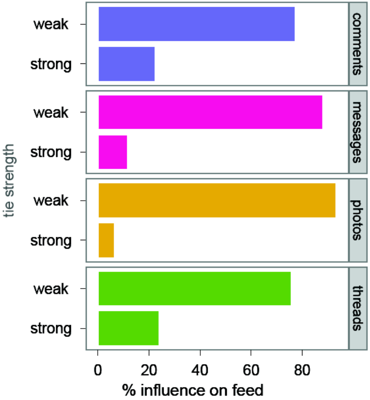


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; — — — = weak tie.

information bubble?

Bakshy, Rosenn, Marlow & Adamic (2012). The role of social networks in information diffusion. *Proc. WWW2012*, pp. 519–528



Meta: facebook, instagram, whatsapp, threads, . . .

3 000 mio. active users
(3rd among most-visited sites on the Web)

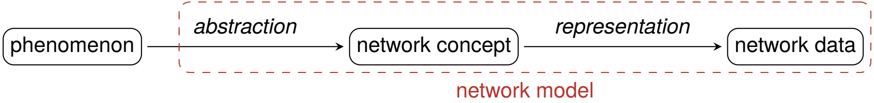
70 000 employees
110 000 mio. US$ revenue (98% advertising)
25 000 mio. US$ net income

course content overview

1. **introduction**
   - empirical research
   - dyadic data
   - network representations
   - software tool
   - friendship paradox
2. **network structure**
   - density
   - degree sequences
   - scale-free networks
   - core-periphery structure
   - small-world networks
   - micro structure and censuses
3. **centrality**
   - degree and domination
   - walks and shortest paths
   - dependency and duality
   - ranking
4. **roles**
   - role equivalence
   - role formation
5. **cohesion**
   - cliques
   - density
   - clustering
6. **influence**

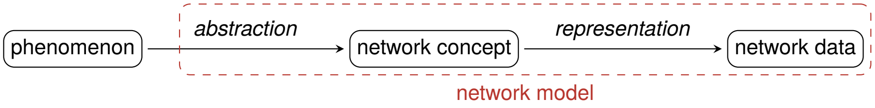
network science

B., Robins, McCranie & Wasserman (2013). What is Network Science? *Network Science* 1(1):1–15



network science

B., Robins, McCranie & Wasserman (2013). What is Network Science? *Network Science* 1(1):1–15



emerging mathematical discipline (compare to statistics)

- mathematical network science
- applied network science (social networks)
- computational network science

empirical research

pipeline

1. problem statement
2. theory
3. hypotheses
4. research design
5. data collection
6. exploration and analysis
7. interpretation and presentation

objectives

- description
- explanation
- prediction

empirical research

pipeline

1. problem statement
2. theory
3. hypotheses
4. research design
5. data collection
6. exploration and analysis
7. interpretation and presentation

objectives

- description (as prerequisite)
- explanation
- prediction

what is theory?

four questions to describe a relation

what?

antecedents
conditions

what?

consequences
outcomes

David A. Whetten (1989). What constitutes a theoretical contribution?
Academy of Management Review 14(4):490–495

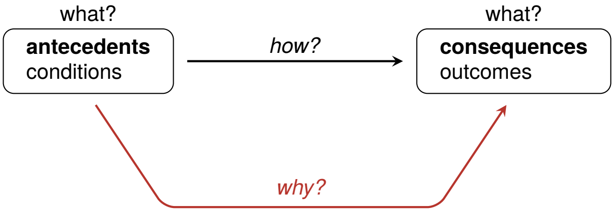
what is theory?

four questions to describe a relation

David A. Whetten (1989). What constitutes a theoretical contribution?
Academy of Management Review 14(4):490–495

what is theory?

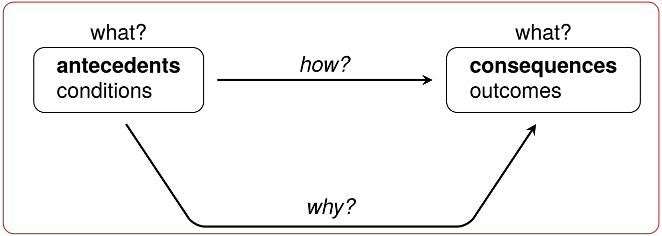
four questions to describe a relation



David A. Whetten (1989). What constitutes a theoretical contribution?
Academy of Management Review 14(4):490–495

what is theory?

four questions to describe a relation



who? where? when?

David A. Whetten (1989). What constitutes a theoretical contribution?
Academy of Management Review 14(4):490–495

hypotheses

theory based on the argument that
the utility that a given user derives from the good depends upon the number of other users who are in the same 'network' as is he or she

see the **model** in
Michael L. Katz & Carl Shapiro (1985). Network externalities, competition, and compatibility. *The American Economic Review* 75(3):424–440

hypotheses

theory based on the argument that
the utility that a given user derives from the good depends upon the number of other users who are in the same 'network' as is he or she

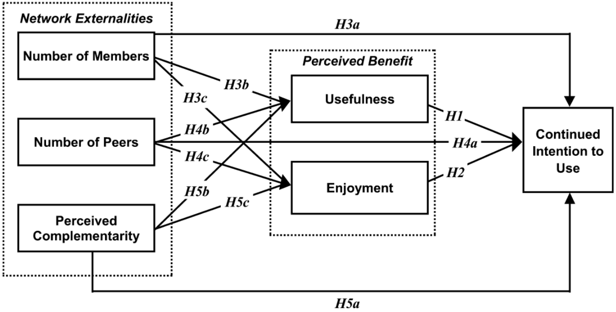
see the **model** in
Michael L. Katz & Carl Shapiro (1985). Network externalities, competition, and compatibility. *The American Economic Review* 75(3):424–440

Kuan-Yu Lin & Hsi-Peng Lu (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior* 27(3):1152–1161

hypothesis:
number of peers will have a positive effect on usefulness of a social network service

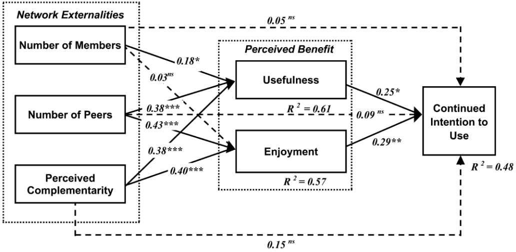
hypotheses

Kuan-Yu Lin & Hsi-Peng Lu (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior* 27(3):1152–1161

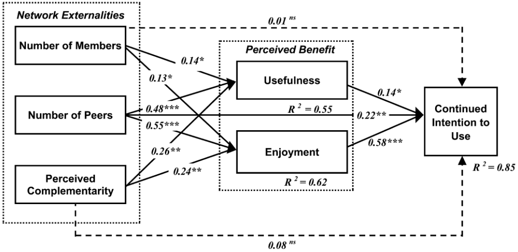


hypotheses

Kuan-Yu Lin & Hsi-Peng Lu (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior* 27(3):1152–1161



*** p < 0.001, ** p < 0.01, * p < 0.05, ns = not significant.



*** p < 0.001, ** p < 0.01, * p < 0.05, ns = not significant.

research design

networks as explanatory / intermediate / dependent variables

conditions —*formation* / modeling→ **network** —*effect* / analysis→ outcomes

research design

networks as explanatory / intermediate / dependent variables

$$\boxed{\text{conditions}} \xrightarrow[\text{modeling}]{\textit{formation}} \boxed{\text{network}} \xrightarrow[\text{analysis}]{\textit{effect}} \boxed{\text{outcomes}}$$

longitudinal studies:

$$\boxed{\text{network}(t)} \longrightarrow \boxed{\text{network}(t+1)}$$

research design

networks as explanatory / intermediate / dependent variables

$$\boxed{\text{conditions}} \xrightarrow[\text{modeling}]{\textit{formation}} \boxed{\text{network}} \xrightarrow[\text{analysis}]{\textit{effect}} \boxed{\text{outcomes}}$$

longitudinal studies: e.g., co-evolution of structure and behavior

$$\begin{tikzcd}
\boxed{\text{network}(t)} \arrow[r] \arrow[dr] & \boxed{\text{network}(t+1)} & \text{selection} \\
\boxed{\text{behavior}(t)} \arrow[r] \arrow[ur] & \boxed{\text{behavior}(t+1)} & \text{influence}
\end{tikzcd}$$

homophily as an outcome of social processes

McPherson, Smith-Lovin & Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–444

social selection

social influence

homophily as an outcome of social processes

McPherson, Smith-Lovin & Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–444

social selection

ties caused by attributes

social influence

homophily as an outcome of social processes

McPherson, Smith-Lovin & Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–444

social selection

ties caused by attributes

social influence

attributes caused by ties

data collection

Alice Altissimo (2016)
Combining egocentric network maps and narratives:
an applied analysis of qualitative network map interviews
Sociological Research Online 21(2):14

what is data?

data: **values** of variables
variable: mapping from domain (*units of observation*) to range (*potential values*)

$$x\colon \mathcal{S} \to \mathcal{X}$$

what is data?

data: values of variables
variable: mapping from domain (*units of observation*) to range (*potential values*)

$$x\colon \mathcal{S} \to \mathcal{X}$$

nominal
binary
ordinal
quantities

what is data?

data: values of variables
variable: mapping from domain (*units of observation*) to range (*potential values*)

$$x\colon \mathcal{S} \to \mathcal{X}$$

nominal
binary
ordinal
quantities

tuples
distributions
sequences
time intervals
networks

structure of range

what is data?

data: values of variables
variable: mapping from domain (*units of observation*) to range (*potential values*)

$$x\colon \mathcal{S} \to \mathcal{X}$$

dyadic
time series
vector fields
spatial data
multi-level
networks

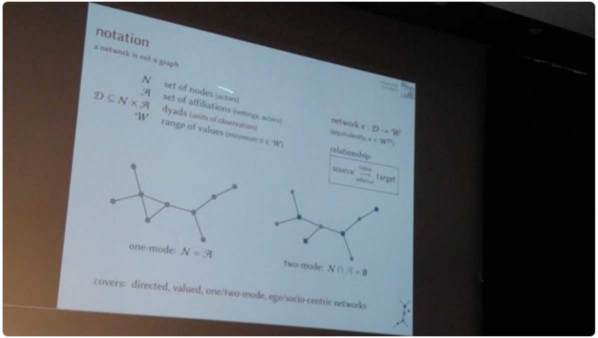
structure of domain

networks are data on intersecting dyads

or rather the answer to the big (fat) data problem?

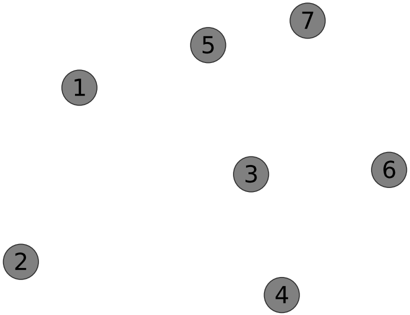
NetSciX @NetSciX2016 · Jan 13

U. Brandes: Network is not a graph it's data (on a diet) #networkscience #datascience



data with unstructured domains

$x : \mathcal{P} \to \mathcal{X}$ where $\mathcal{P} \subseteq \{i_1, i_2, i_3, \ldots\}$ population

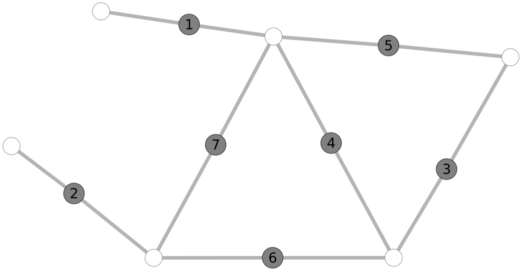


data with structured domains

$x : \mathcal{T} \to \mathcal{X}$ where $\mathcal{T} \subseteq \langle t_0, t_1, t_2, \ldots \rangle$ time series

data with structured domains

$x : \mathcal{D} \to \mathcal{X}$ where $\mathcal{D} \subseteq \mathcal{N} \times \mathcal{N}$ network



network data

network variable $x : \mathcal{D} \rightarrow \mathcal{X}$

$\mathcal{D} \subseteq \mathcal{N} \times \mathcal{A}$ dyadic domain

network data

network variable $x : \mathcal{D} \to \mathcal{X}$

$\mathcal{D} \subseteq \mathcal{N} \times \mathcal{A}$ dyadic domain

$\mathcal{N} = \mathcal{A}$ one-mode

$\mathcal{N} \cap \mathcal{A} = \emptyset$ two-mode

network data

network variable $x : \mathcal{D} \to \mathcal{X}$

$\mathcal{D} \subseteq \mathcal{N} \times \mathcal{A}$ dyadic domain

$\mathcal{N} = \mathcal{A}$ one-mode

$\mathcal{N} \cap \mathcal{A} = \emptyset$ two-mode

networks represented as graphs, matrices, tupels, ...

caveat: structural zeros $(i, j) \notin \mathcal{D}$ vs. absent ties $x_{ij} = 0$

network data

data science perspective: incidence-structured domain (data on overlapping dyads)

- nodes $\mathcal{N}$ ("social actors")

attributes
monadic variables

$x \in \mathcal{X}^{\mathcal{N}}$
$x : \mathcal{N} \to \mathcal{X}$

node	attribute	attribute
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3

independence assumption

network data

data science perspective: incidence-structured domain (data on overlapping dyads)

- nodes $\mathcal{N}$ ("social actors") and affiliatons $\mathcal{A}$ ("social settings")
- dyads $\mathcal{D} \subseteq \mathcal{N} \times \mathcal{N}$ ("one-mode") or $\mathcal{D} \subseteq \mathcal{N} \times \mathcal{A}$ ("two-mode")

attributes
monadic variables

$x \in \mathcal{X}^{\mathcal{N}}$
$x : \mathcal{N} \to \mathcal{X}$

node	attribute	attribute
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3

independence assumption

relations
dyadic variables

$x \in \mathcal{X}^{\mathcal{D}}$
$x : \mathcal{D} \to \mathcal{X}$

	nodes / affiliations		
relation	1	2	3
nodes 1	$\cdot$	x_{12}	x_{13}
2	x_{21}	$\cdot$	x_{23}
3	x_{31}	x_{32}	$\cdot$

interest in dependencies

network data

data science perspective: incidence-structured domain (data on overlapping dyads)

- nodes $\mathcal{N}$ ("social actors") and affiliatons $\mathcal{A}$ ("social settings")
- dyads $\mathcal{D} \subseteq \mathcal{N} \times \mathcal{N}$ ("one-mode") or $\mathcal{D} \subseteq \mathcal{N} \times \mathcal{A}$ ("two-mode")

attributes
monadic variables

$x \in \mathcal{X}^{\mathcal{N}}$
$x : \mathcal{N} \to \mathcal{X}$

node	attribute	attribute
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3

independence assumption

relations
dyadic variables

$x \in \mathcal{X}^{\mathcal{D}}$
$x : \mathcal{D} \to \mathcal{X}$

	nodes / affiliations		
relation	1	2	3
nodes 1	$\cdot$	x_{12}	x_{13}
2	x_{21}	$\cdot$	x_{23}
3	x_{31}	x_{32}	$\cdot$

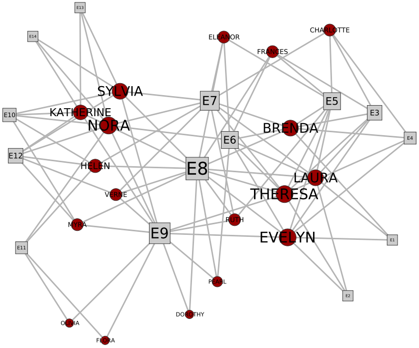
interest in dependencies

NB: $\mathcal{X}$ can be any data type

types of networks: two-mode network

e.g., presence of individuals at events

$\mathcal{N} = \{\text{BRENDA}, \text{CHARLOTTE}, \ldots\}$
$\mathcal{A} = \{\text{E1}, \text{E2}, \ldots\}$

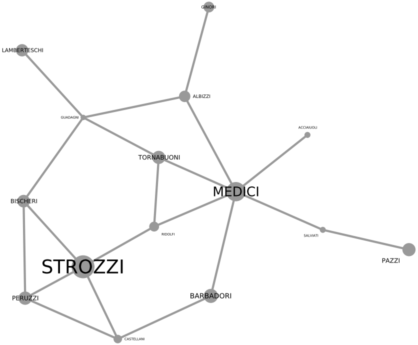


	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
MYRA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0

types of networks: one-mode network

e.g., marriage relationships between families

$$\mathcal{N} = \{\text{ACCIAIUOLI}, \ldots\}$$
$$\mathcal{A} = \mathcal{N}$$



ACCIAIUOLI	·	0	0	0	0	0	0	0	1	0	0	0	0	0	0
ALBIZZI	0	·	0	0	0	1	1	0	1	0	0	0	0	0	0
BARBADORI	0	0	·	0	1	0	0	0	1	0	0	0	0	0	0
BISCHERI	0	0	0	·	0	0	1	0	0	0	1	0	0	1	0
CASTELLANI	0	0	1	0	·	0	0	0	0	0	1	0	0	1	0
GINORI	0	1	0	0	0	·	0	0	0	0	0	0	0	0	0
GUADAGNI	0	1	0	1	0	0	·	1	0	0	0	0	0	0	1
LAMBERTESCHI	0	0	0	0	0	0	1	·	0	0	0	0	0	0	0
MEDICI	1	1	1	0	0	0	0	0	·	0	0	1	1	0	1
PAZZI	0	0	0	0	0	0	0	0	0	·	0	0	1	0	0
PERUZZI	0	0	0	1	1	0	0	0	0	0	·	0	0	1	0
RIDOLFI	0	0	0	0	0	0	0	0	1	0	0	·	0	1	1
SALVIATI	0	0	0	0	0	0	0	0	1	1	0	0	·	0	0
STROZZI	0	0	0	1	1	0	0	0	0	0	1	1	0	·	0
TORNABUONI	0	0	0	0	0	0	1	0	1	0	0	1	0	0	·

data collection

measurement
- objectivity
- reliability
- validity

sources
- surveys
- observational
- secondary
- archival

management
- preprocessing
- transformation
- filtering
- security and privacy

storing network data

$$\begin{array}{c|c|} & \mathcal{A} \\ \hline \mathcal{N} & \mathcal{X} \\ \hline \end{array}$$

adjacency matrix

$\mathcal{N}$	$(\mathcal{A} \times \mathcal{X})^+$

adjacency list

$\mathcal{N}$	$\mathcal{A}$	$\mathcal{X}$

edge list

$\mathcal{A}$	$\mathcal{X}$

(ego networks)

further complications

- symmetry
- sparse data
- complex ranges
- multiple relations
- (multiple) node attributes

- cognitive social structures
- multilayer networks
- personal networks

representing network data

caution: unobserved and absent ties typically indistinguishable

- **matrix**

- **graph**
  given network $x : \mathcal{D} \to \mathcal{X}$ define $G(x) = (V, E; \omega)$ with

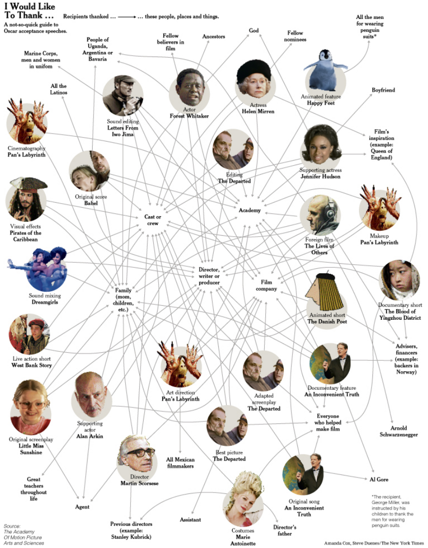
$$\begin{aligned} V &= \mathcal{N} \\ E &= \{(i,j) \in \mathcal{D} \, : \, x_{ij} \neq 0\} \\ \omega : E \to \mathcal{X}(x) &= x_{|E} \end{aligned}$$

  $x_{ij} = x_{ji}$ for all $(i,j) \in \mathcal{D} \implies$ undirected graph

- **(valued) relation**

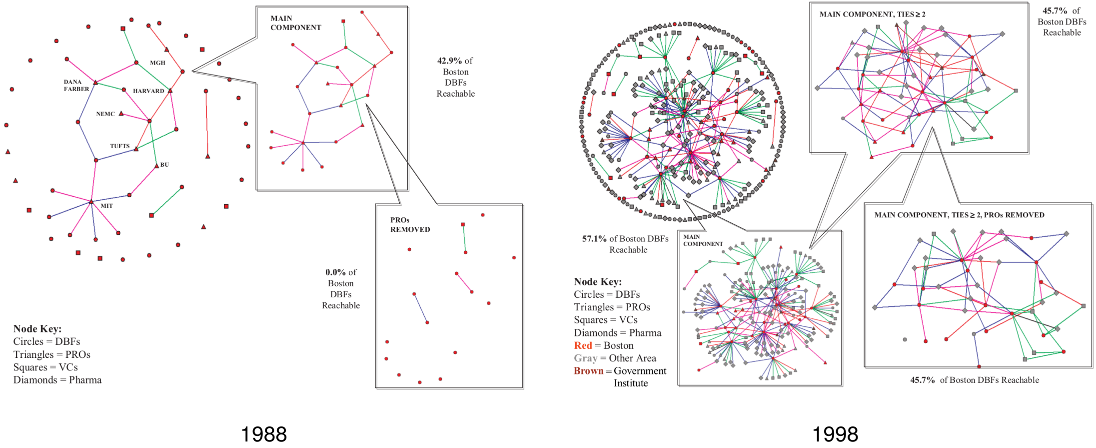
presentation

mind the purpose



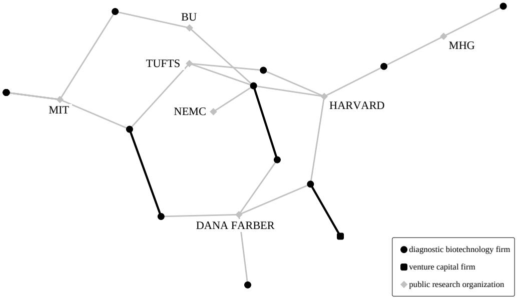
message in a figure

Owen-Smith & Powell (2004). Knowledge networks as channels and conduits:
The effects of spillovers in the Boston biotechnology community. *Organization Science* 15(1):5–21

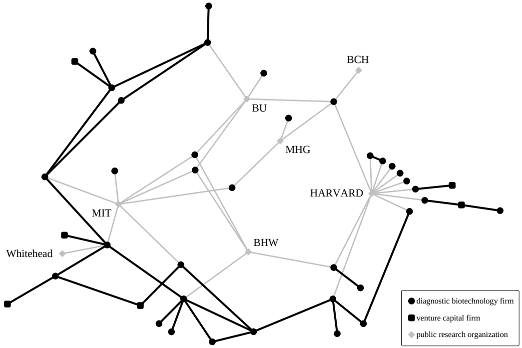


(same) message in a (different) figure

B. & Schneider (2009). Netzwerkbilder. Politiknetzwerke in Metaphern, Modellen und Visualisierungen. In Schneider, Janning, Leifeld & Malang: *Politiknetzwerke*. VS Verlag für Sozialwissenchaften, pp. 31–58



1988

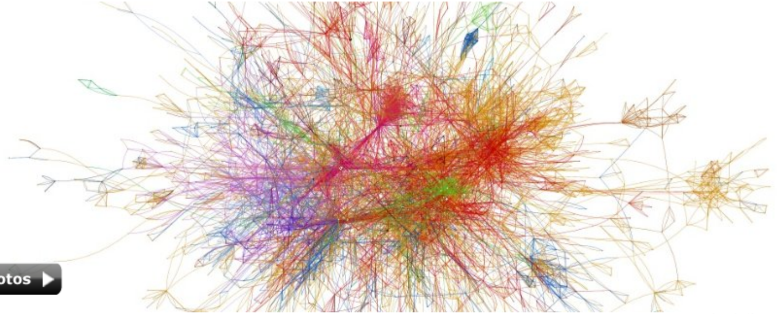


1998

10.07.2015 – 11:27 Uhr

Interaktive Grafik: So sind Deutschlands YouTuber vernetzt

Von Christian Stöcker, Christina Elmer und Maximilian Schäfer



gugelproductions.de

Mit wem ist LeFloid befreundet, wer sind die Kumpel von Gronkh und Sarazar, und wo versammeln sich die Bodybuilder? Eine interaktive Netzwerk-Grafik gibt faszinierende Einblicke in die deutsche YouTuber-Szene.

data and diagram: Bertram Gugel (2015)

presentation