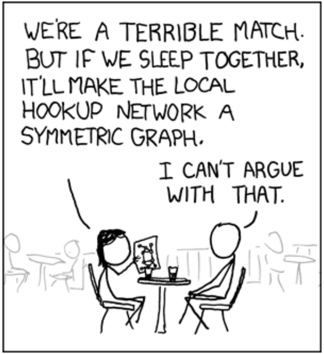
Social Networks, First Principles, and Graph Modification

Ulrik Brandes

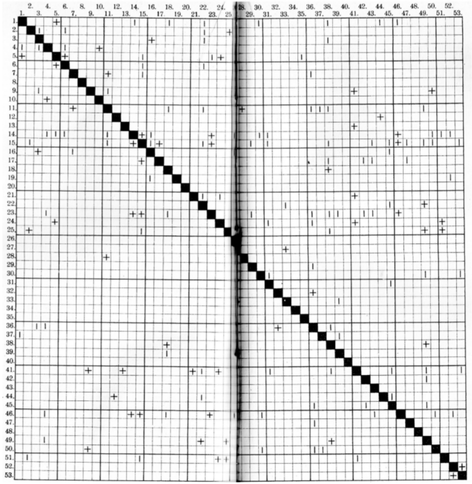
Social Networks Lab, ETH Zürich



xkcd no. 403

Workshop on Graph Modification
23-24 January 2020, Bergen

ETH zürich

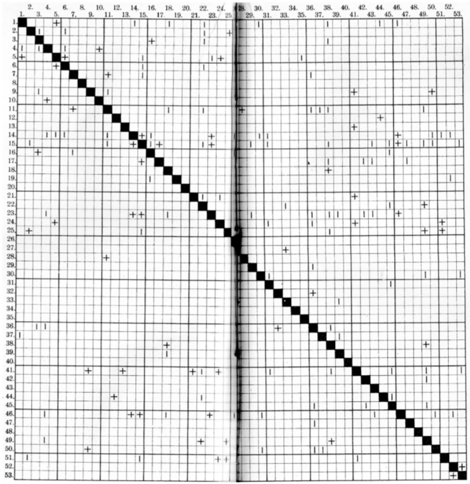

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53.
1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53.

Delitsch's Volksschulklasse

Johannes Delitsch (1900)
Über Schülerfreundschaften in einer Volksschulklasse
Zeitschrift für Kinderforschung 5:150–163

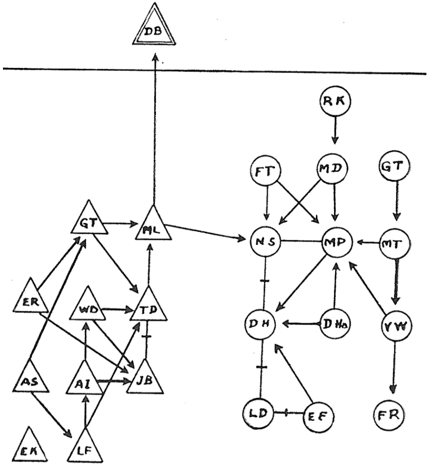
participant observation

- 53 boys in 4th grade
- school year 1880/81
- class room, school yard
- including staged situations



sociograms

Jacob L. Moreno: *Who Shall Survive?* Beacon House, 1953 [1934]



"A process of charting has been devised by the sociometrists, the sociogram, which is more than merely a method of presentation. It is first of all a method of exploration."

graph theory

Harary & Norman: Graph Theory as a Mathematical Model in the Social Sciences
Ann Arbor: Institute for Social Research, 1953

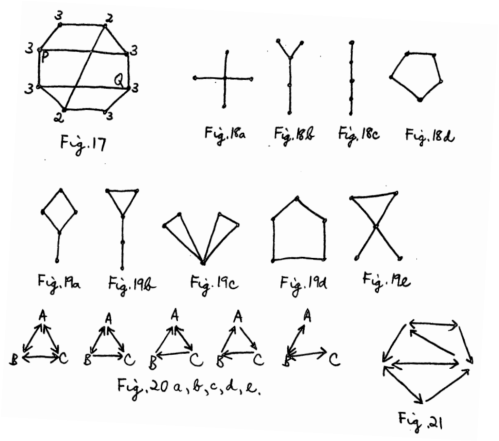
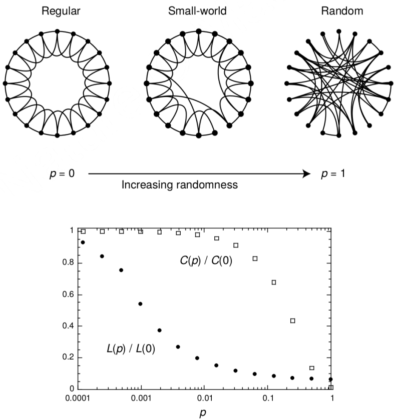


figure from:
Group structures and communication networks in terms of the mathematical theory of graphs
(Harary & Norman, 1952)

complex networks

Watts & Strogatz (1998). Collective dynamics of 'small world' networks, *Nature* 393:440–442

Christakis & Fowler (2007). The spread of obesity in a large social network over 32 years, *New England Journal of Medicine* 357:370–379



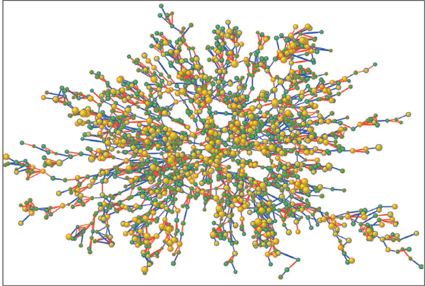
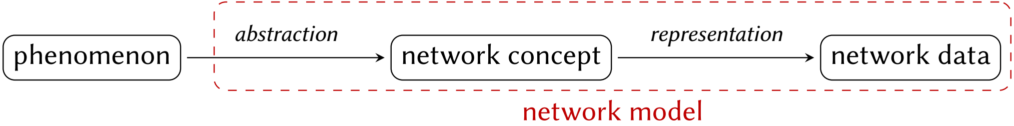


Figure 1. Largest Connected Subcomponent of the Social Network in the Framingham Heart Study in the Year 2000.
Each circle (node) represents one person in the data set. There are 2200 persons in this subcomponent of the social network. Circles with red borders denote women, and circles with blue borders denote men. The size of each circle is proportional to the person's body-mass index. The interior color of the circles indicates the person's obesity status: yellow denotes an obese person (body-mass index, ≥30) and green denotes a nonobese person. The colors of the ties between the nodes indicate the relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie.

network science

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



emerging mathematical discipline (a particular kind of data science)

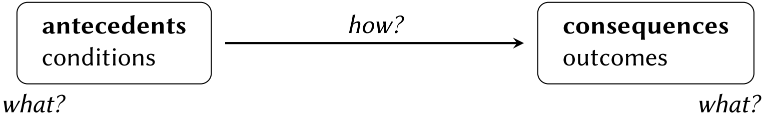
- mathematical network science
- computational network science
- applied network science

theory?

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

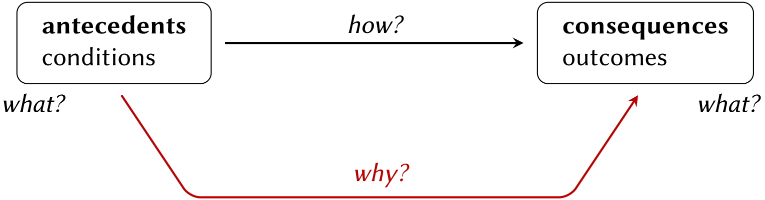
theory?

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



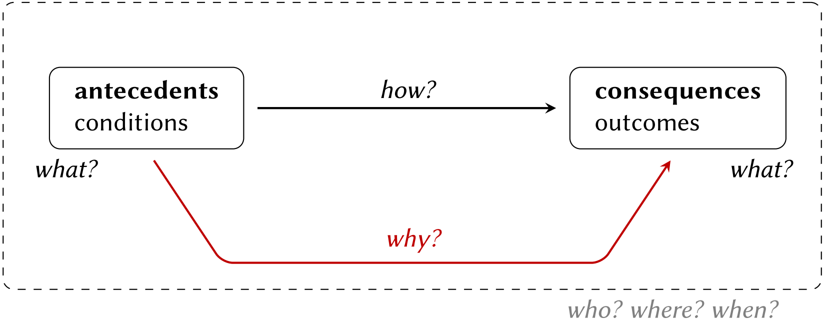
theory?

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



theory?

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



research design

networks as explanatory / mediating / dependent **variables**

research design

networks as explanatory / mediating / dependent **variables**

conditions → network → outcomes

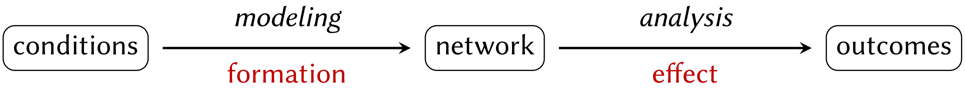
- conditions → network: *modeling* / formation
- network → outcomes: *analysis* / effect

longitudinal studies:

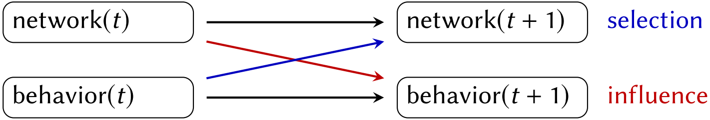
network(t) → network($t + 1$)

research design

networks as explanatory / mediating / dependent **variables**



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



social network analysis

old school

centrality

$$c_D(i) = \deg(i)$$

$$c_C(i) = \left(\sum_t dist(i,t)\right)^{-1}$$

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

$$c_E(i) = \frac{1}{\lambda_1} \sum_{i \in N(i)} c_E(j)$$

$$\vdots$$

roles

$$i \sim j \implies N(i) = N(j)$$

$$i \approx j \implies [N(i)]_{\approx} = [N(j)]_{\approx}$$

$$i \equiv j \iff i = \alpha(j)$$

$$\vdots$$

cohesion

dense subgraphs

iterated cuts

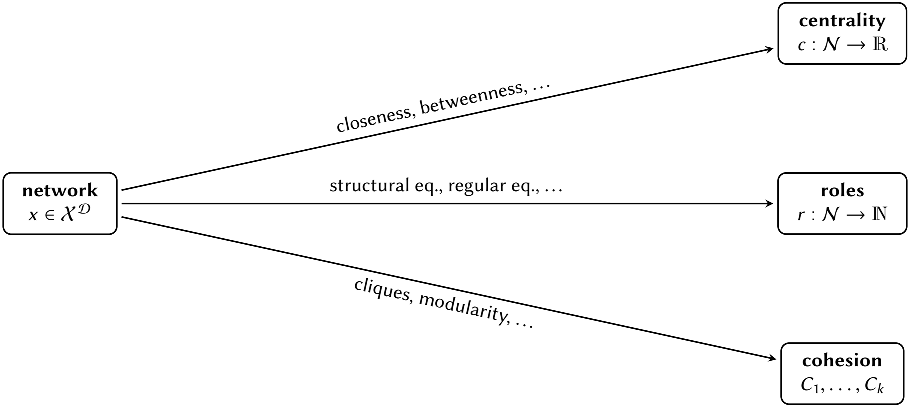
spectral partitioning

$$Q(C) = \sum_C \left[\frac{m(C)}{m} - \left(\frac{\sum_{i \in C} \deg(i)}{2m} \right)^2 \right]$$

$$\vdots$$

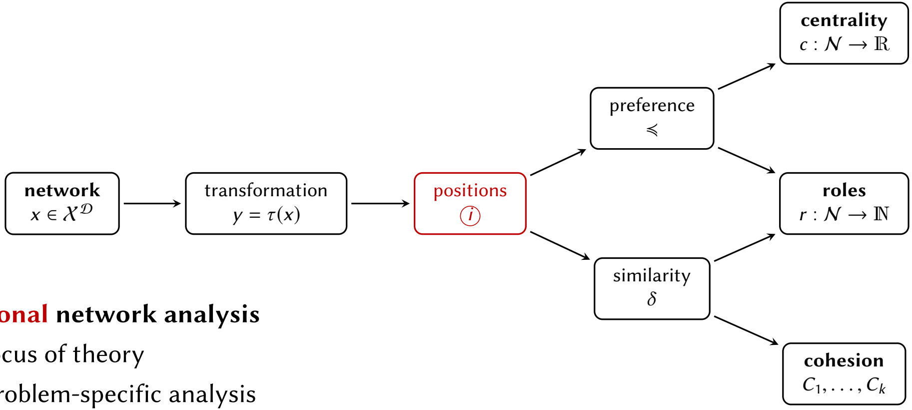
analytic categories

choose or refine method from a list of options



analytic pipeline

agenda: break down choices

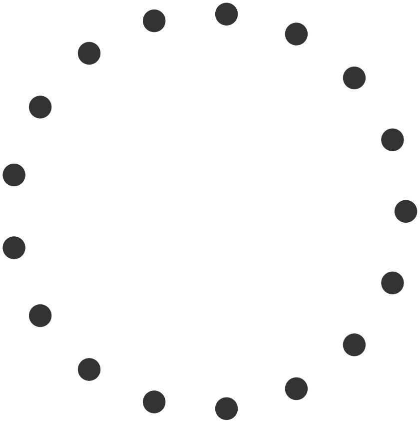


positional network analysis

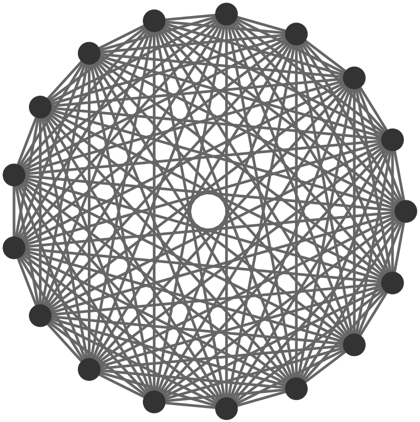
- locus of theory
- problem-specific analysis
- small-scale hypothesis testing
- research program

B. (2016). Network positions. *Methodological Innovations* 9:2059799116630650

isolates

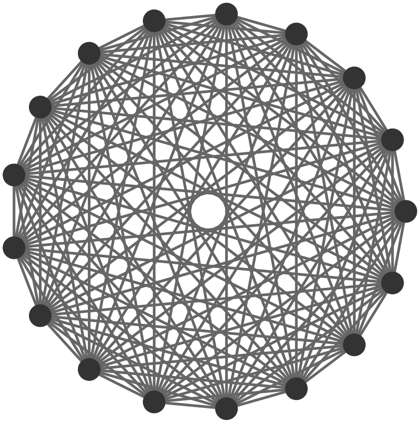


isolates vs. cliques



Luce & Perry (1949).
A method of matrix analysis of group structure.
Psychometrika 14:95–116

isolates vs. cliques

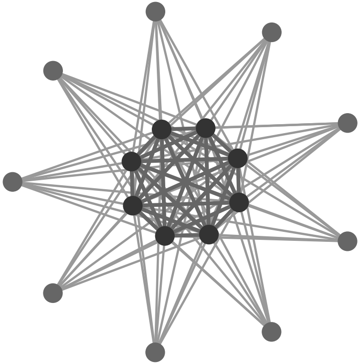


Luce & Perry (1949).
A method of matrix analysis of group structure.
Psychometrika 14:95–116

density: (normalized) distance

$m = \lvert E \rvert$	edges = edits
$\frac{2m}{n} = \langle \deg \rangle$	average degree
$\frac{m}{\binom{n}{2}} = \frac{\langle \deg \rangle}{n-1}$	both normalized

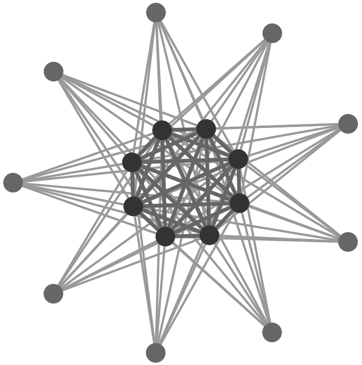
ideal core-periphery structure



split graph: ex. $V = C \uplus P$ s.t.

- $G[C]$ clique
- $G[P]$ isolates

ideal core-periphery structure



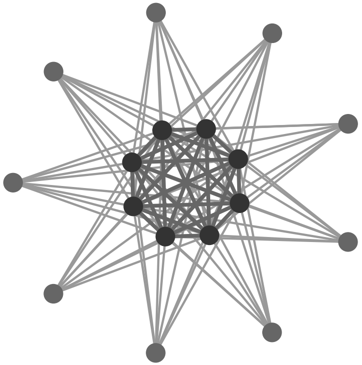
degree sequence $d_1 \geq \ldots \geq d_n$
corrected Durfee number $h(G) = \max\{i \,:\, d_i \geq i - 1\}$

$$\sum_{i=1}^{h} d_i = h(h-1) + \sum_{j=h+1}^{n} d_j$$

split graph: ex. $V = C \uplus P$ s.t.

- $G[C]$ clique
- $G[P]$ isolates

ideal core-periphery structure



degree sequence $d_1 \geq \ldots \geq d_n$
corrected Durfee number $h(G) = \max\{i \,:\, d_i \geq i-1\}$

$$\sum_{i=1}^{h} d_i = h(h-1) + \sum_{j=h+1}^{n} d_j$$

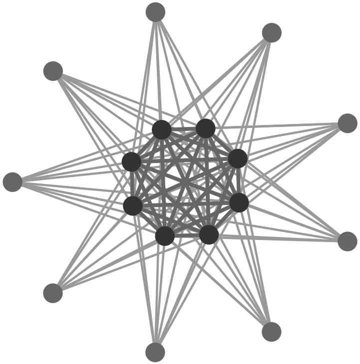
splittance: split edit distance (Hammer & Simeone, 1981)

$$\frac{1}{2}\left(h(h-1) + \sum_{j=h+1}^{n} d_j - \sum_{i=1}^{h} d_i \right)$$

split graph: ex. $V = C \uplus P$ s.t.

- $G[C]$ clique
- $G[P]$ isolates

ideal core-periphery structure



split graph: ex. $V = C \uplus P$ s.t.

- $G[C]$ clique
- $G[P]$ isolates

degree sequence $d_1 \geq \ldots \geq d_n$
corrected Durfee number $h(G) = \max\{i \,:\, d_i \geq i-1\}$

$$\sum_{i=1}^{h} d_i = h(h-1) + \sum_{j=h+1}^{n} d_j$$

splittance: split edit distance (Hammer & Simeone, 1981)

$$\frac{1}{2}\left(h(h-1) + \sum_{j=h+1}^{n} d_j - \sum_{i=1}^{h} d_i \right)$$

$\mathcal{NP}$-hard for multiple cores/peripheries, other densities
Bruckner, Hüffner & Komusiewicz (2015)
B., Holm & Karrenbauer (2016)

the world trade network has a core-periphery structure

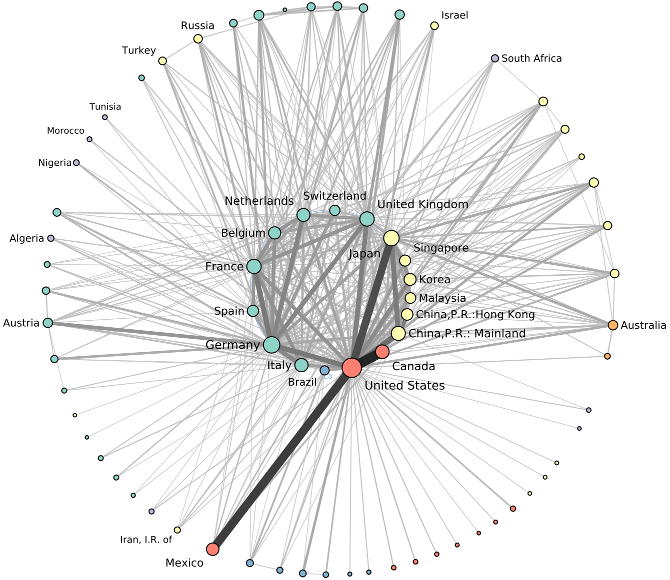
international trade

- in the year 2000
- only if ≥ 10 billion US$

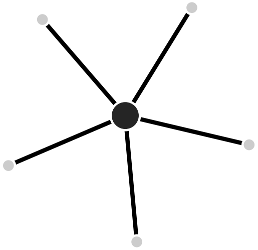
- 17 countries in core
- 16 edges missing in core

- 56 countries in periphery
- 23 edges appear in periphery

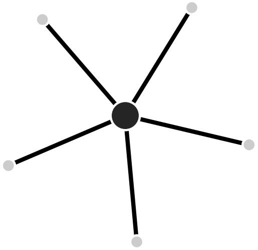
- 9% editing needed



ideal centrality ranking



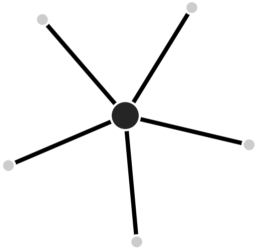
ideal centrality ranking



A person located in the center of a star
is universally assumed
to be structurally more central
than any other person
in any other position
in any other network of similar size.

Freeman (*Social Networks* 1979)

ideal centrality ranking



A person located in the center of a star
is universally assumed
to be structurally more central
than any other person
in any other position
in any other network of similar size.

Freeman (*Social Networks* 1979)

axiomatic: Sabidussi (*Psychometrika* 1966)

- invariance under graph isomorphisms
- edge addition and switching increase centrality

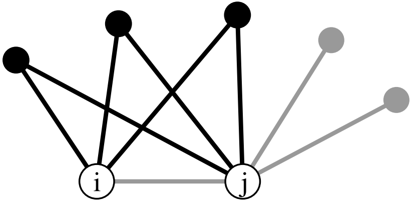
Boldi & Vigna (*Internet Mathematics* 2014)

conceptual:

- flow processes
  Borgatti (*Social Networks* 2005)
- radial and medial positions
  Borgatti & Everett (*Social Networks* 2006)

the essence of centrality

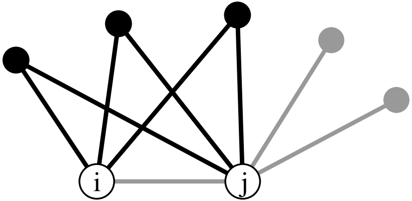
better relations $\implies$ higher centrality



$$N(i) \subseteq N[j] \implies c(i) \leq c(j)$$

the essence of centrality

better relations $\implies$ higher centrality



$$N(i) \subseteq N[j] \implies \tau(i,t) \leq \tau(j,t) \implies c(i) \leq c(j)$$

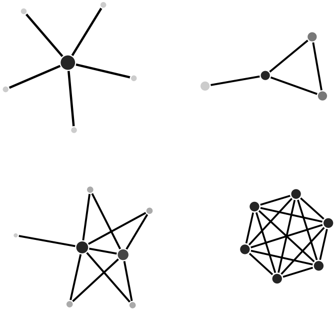
Schoch & B. (*European Journal of Applied Mathematics* 2016)

path algebra $\tau(s,t) = \bigoplus_{s \rightarrow^* t} x_{sv_1} \odot \cdots \odot x_{v_{k-1}t}$

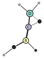
$\odot, \oplus$ *decreasing* semiring $\implies$ neighborhood-inclusion respected

ideal ranking by centrality

Schoch, Valente & B. (*Social Networks* 2017)

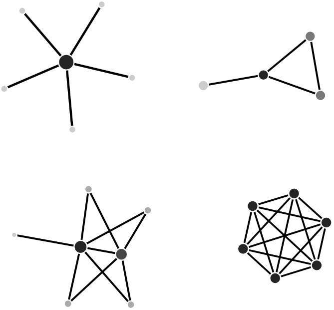


completely ranked by
neighborhood-inclusion



ideal ranking by centrality

Schoch, Valente & B. (*Social Networks* 2017)



completely ranked by
neighborhood-inclusion

threshhold graphs

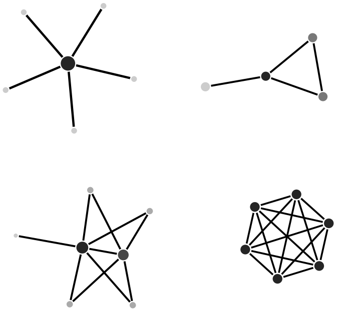
$$\sum_{i=1}^{1} d_i = 1(1-1) + \sum_{j=1+1}^{n} d_j$$

$$\vdots$$

$$\sum_{i=1}^{h} d_i = h(h-1) + \sum_{j=h+1}^{n} d_j$$

ideal ranking by centrality

Schoch, Valente & B. (*Social Networks* 2017)



completely ranked by
neighborhood-inclusion
majorization gap, edge rotation

threshhold graphs

$$\sum_{i=1}^{1} d_i = 1(1-1) + \sum_{j=1+1}^{n} d_j$$

$$\vdots$$

$$\sum_{i=1}^{h} d_i = h(h-1) + \sum_{j=h+1}^{n} d_j$$

threshold edit distance $\mathcal{NP}$-hard

Drange, Dregi, Lokshtanov & Sullivan (ESA 2015)

how far from a threshold graph?

Mahadev & Peled (1995). *Threshold Graphs and Related Topics.*

(corrected) conjugate sequences $\overline{d}_i = |\{d_j \geq i-1 : j = 1, \ldots, i-1\}| + |\{d_j \geq i : j = i+1, \ldots, n\}|$

	$\overline{d}_1$	$\overline{d}_2$	$\overline{d}_3$	$\overline{d}_4$	$\overline{d}_5$	$\overline{d}_6$	
d_1	×	●	●	●	●	●	5
d_2	●	×	●	●	●		4
d_3	●	●	×	●			3
d_4	●	●	●	×			3
d_5	●				·		1
d_6	●					·	1
d_7	●						1
	6	3	3	3	2	1	18

equivalent to Erdős-Gallai inequalities:

$$\sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} \overline{d}_i \qquad \text{for all } k = 1, \ldots, n-1$$

how far from a threshold graph?

Mahadev & Peled (1995). *Threshold Graphs and Related Topics.*

(corrected) conjugate sequences $\overline{d}_i = |\{d_j \geq i-1 : j = 1, \dots, i-1\}| + |\{d_j \geq i : j = i+1, \dots, n\}|$

	$\overline{d}_1$	$\overline{d}_2$	$\overline{d}_3$	$\overline{d}_4$	$\overline{d}_5$	$\overline{d}_6$	
d_1	×	●	●	●	●	●	5
d_2	●	×	●	●	●		4
d_3	●	●	×	●			3
d_4	●	●	●	×			3
d_5	●				·		1
d_6	●					·	1
d_7	●						1
	6	3	3	3	2	1	18

equivalent to Erdős-Gallai inequalities:

$$\sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} \overline{d}_i \qquad \text{for all } k = 1, \dots, n-1$$

$$h = \max\{i : d_i \geq i-1\}$$

splittance $\frac{1}{2}\left(\sum_{i=1}^{h} \overline{d}_i - \sum_{i=1}^{h} d_i\right)$

threshold gap $\frac{1}{2}\sum_{i=1}^{h} \left|\overline{d}_i - d_i\right|$

= **majorization gap**

≤ 2× edit distance

how far from a threshold graph?

Mahadev & Peled (1995). *Threshold Graphs and Related Topics.*

(corrected) conjugate sequences $\overline{d}_i = |\{d_j \geq i-1 : j = 1, \ldots, i-1\}| + |\{d_j \geq i : j = i+1, \ldots, n\}|$

	$\overline{d}_1$	$\overline{d}_2$	$\overline{d}_3$	$\overline{d}_4$	$\overline{d}_5$	$\overline{d}_6$	
d_1	×	●	●	●	●	●	5
d_2	●	×	●	●	●		4
d_3	●	●	×	●			3
d_4	●	●	●	×			3
d_5	●				·		1
d_6	●					·	1
d_7	●						1
	6	3	3	3	2	1	18

equivalent to Erdős-Gallai inequalities:

$$\sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} \overline{d}_i \qquad \text{for all } k = 1, \ldots, n-1$$

$$h = \max\{i : d_i \geq i-1\}$$

splittance $\frac{1}{2}\left(\sum_{i=1}^{h} \overline{d}_i - \sum_{i=1}^{h} d_i\right)$

threshold gap $\frac{1}{2}\sum_{i=1}^{h} \left|\overline{d}_i - d_i\right|$

$=$ **majorization gap**

$\leq 2\times$ edit distance

open: edge rotation distance

ETH zürich

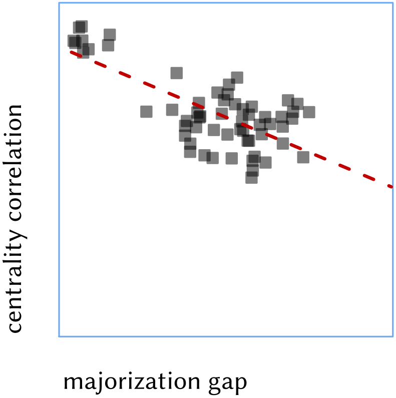
distance from threshold graph

Schoch, Valente & B. (2017). Correlations among centrality indices and a class of uniquely ranked graphs. *Social Networks* 50:46–54

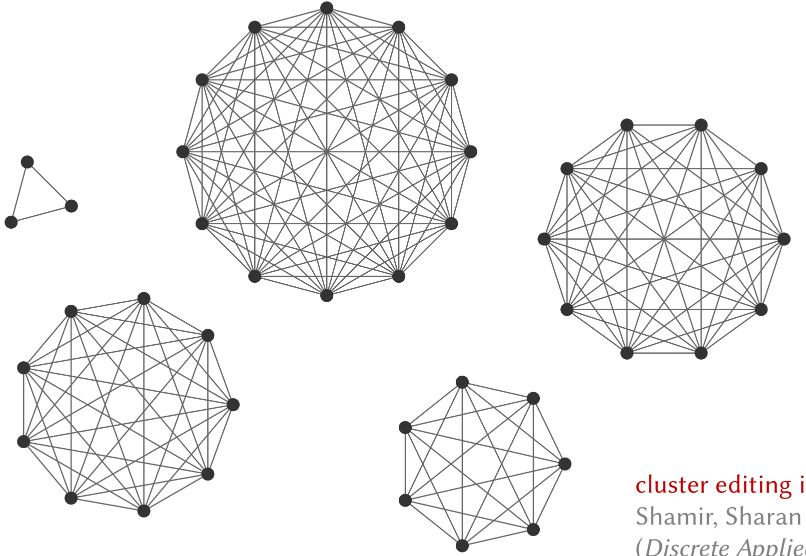
replication of empirical study
Valente et al. (*Connections* 2003)

core-periphery networks
yield high correlation of centralities

⟹ beware of generated data!



ideal community structures



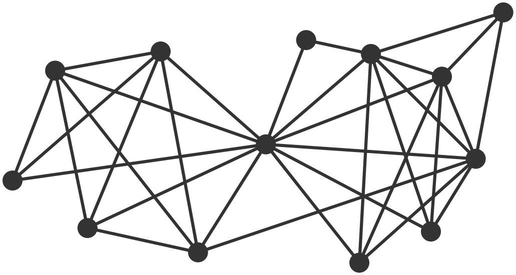
cluster editing is $\mathcal{NP}$-hard

Shamir, Sharan & Tsur

(*Discrete Applied Mathematics* 2004)

brokered communities

Nastos & Gao (2013). Familial groups in social networks. *Social Networks* 35(3):439–450



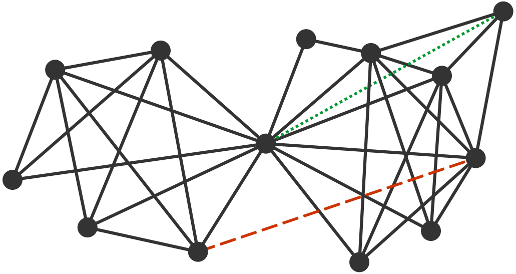
threshold graphs:

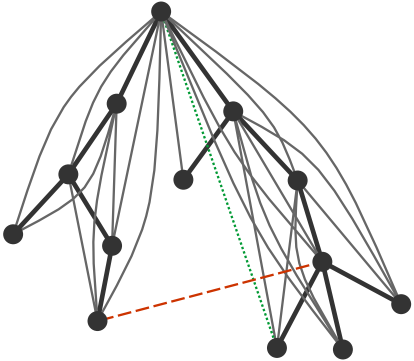
- add universal vertex (top of ranking)
- or add isolated vertex (bottom of ranking)

brokered communities

Nastos & Gao (2013). Familial groups in social networks. *Social Networks* 35(3):439–450

edit distance $\mathcal{NP}$-hard





quasi-threshold graphs:

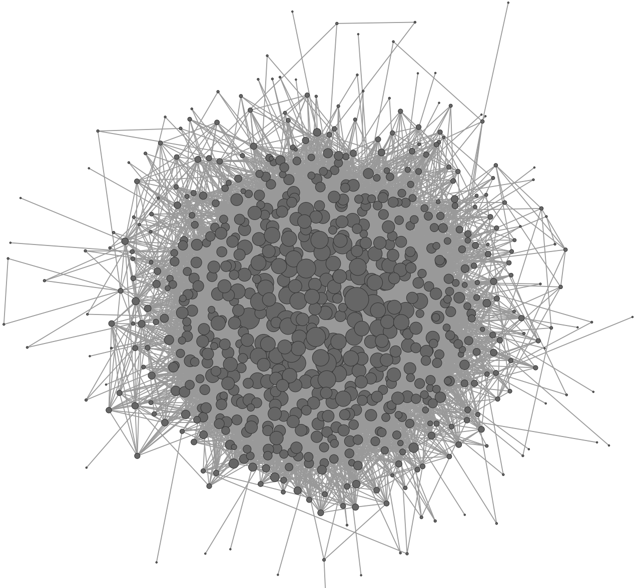
- add universal vertex (top of hierarchies)
- or add isolated vertex (new hierarchy)
- merge two quasi-threshold graphs (separate hierarchies)

facebook communities

B., Hamann, Strasser & Wagner (*Proc. ESA* 2015)

facebook friendships

- CalTech in 2005



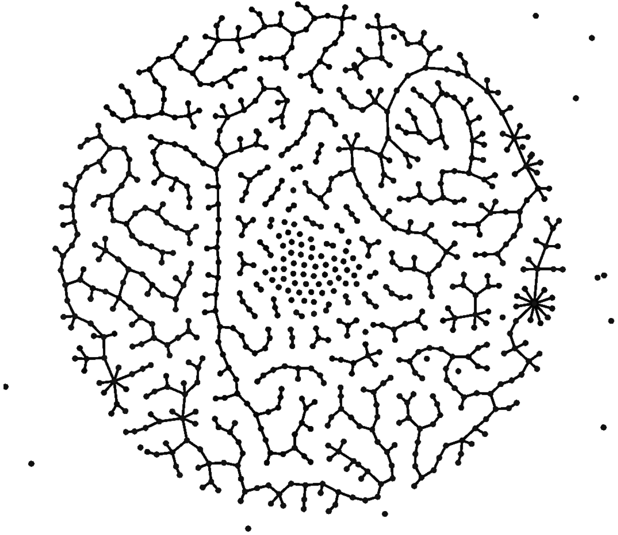
ETH zürich

facebook communities

B., Hamann, Strasser & Wagner (*Proc. ESA* 2015)

facebook friendships

- CalTech in 2005
- skeleton of nearby quasi-threshold graph

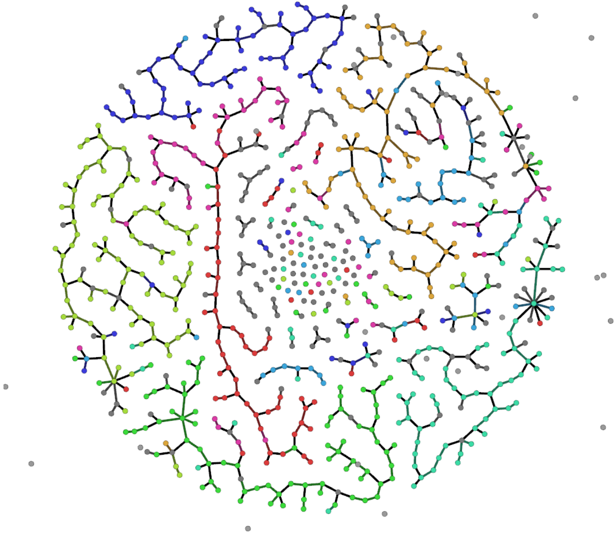


facebook communities at CalTech in 2005 lived under one roof

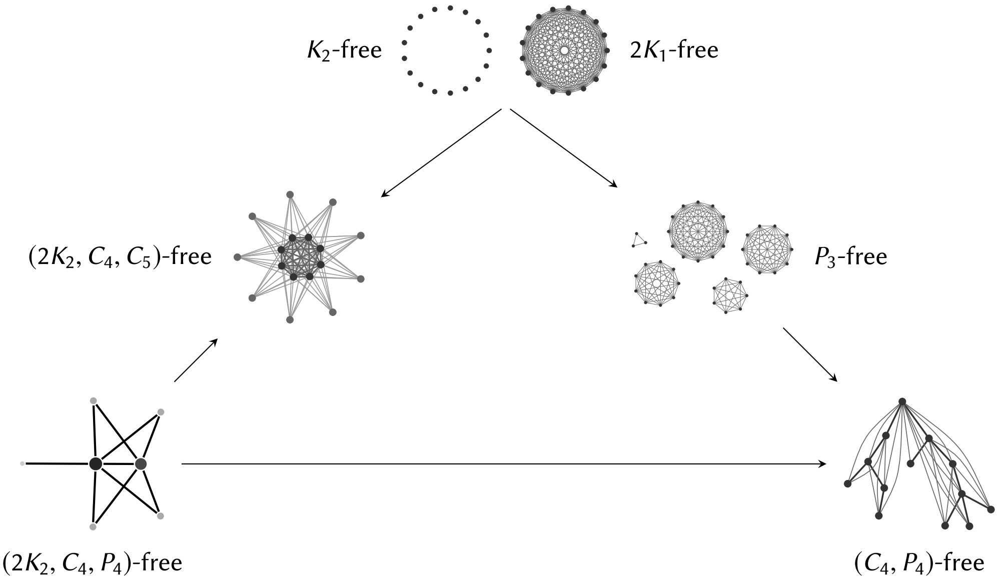
B., Hamann, Strasser & Wagner (*Proc. ESA* 2015)

facebook friendships

- CalTech in 2005
- skeleton of nearby quasi-threshold graph
- vertices colored according to dorm



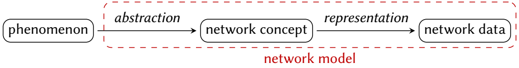
summary: ideal structures



conclusions

- **social networks:** network science applied in social domain

  observations on intersecting dyads
  dependencies yield structure
  delineated by substance, not methods



- **first principles:** ideal structure as reference point

  graph modification to assess deviations

  new models for measurement error?

- **generalizations:** positional approach suggests new problems

  graph modification effects on indirect relations (distance, connectivity, etc.)
  effects of homogeneity assumptions on graph modifications (extrinsic, automorphic. etc.)
  graph modification and non-binary dyad variables (multiplex, signed, valued, temporal, etc.)

  ⋮