

Statistical inference links data and theory in network science

10th SINM edition

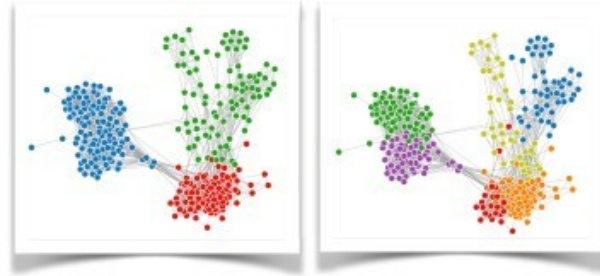
Leto Peel
Maastricht University
 @PiratePeel

 l.peel@maastrichtuniversity.nl

SINM 1st edition

Opportunities:

1. Model selection



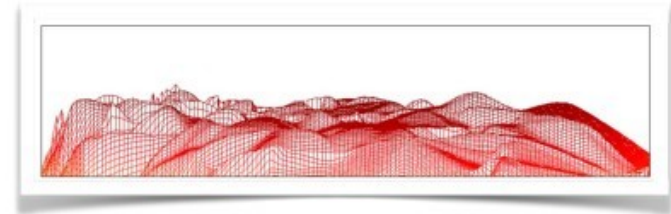
2. Tradeoffs between general and specific models



3. Dynamics



4. Computational challenges



Trade-offs between general and specific models



general ←————→ specific

Trade-offs between general and specific models



general ←————→ specific

More likely to get a “high-impact” paper
Good to raise awareness of Network Science
Doesn't really solve any actual problems

Trade-offs between general and specific models




general

More likely to get a “high-impact” paper
Good to raise awareness of Network Science
Doesn't really solve any actual problems



specific

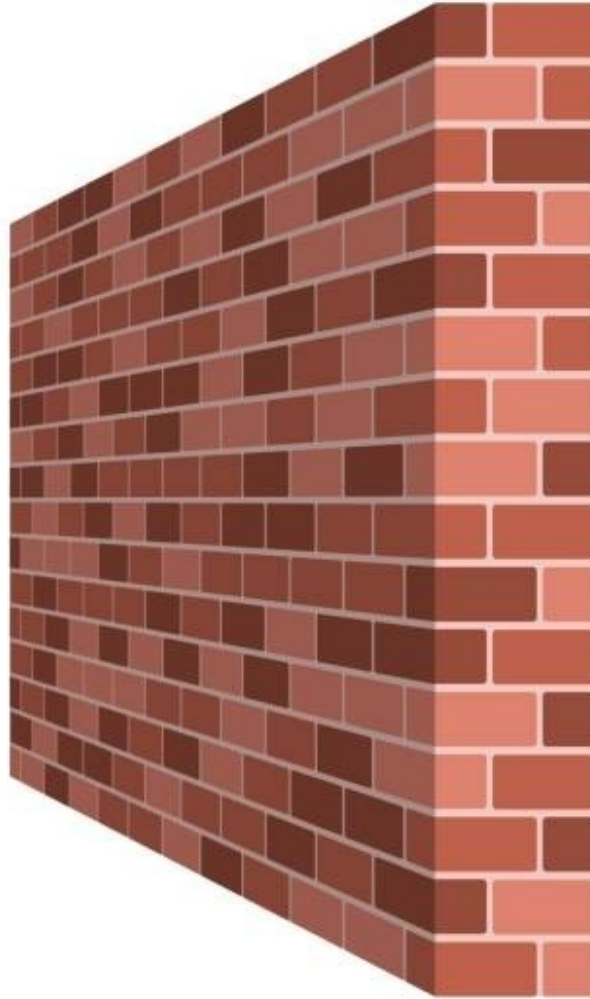
We need to be going more in this direction
This is where the real heroes will be

Network  Science



Network science allows us to
analyse systems as a whole!

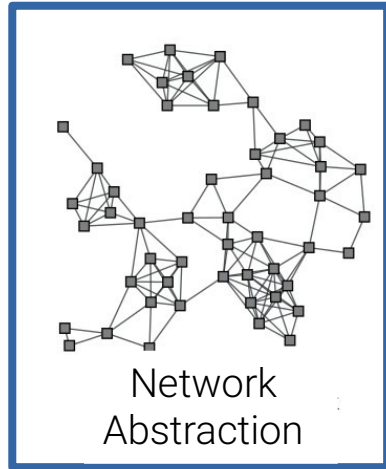
THEORY



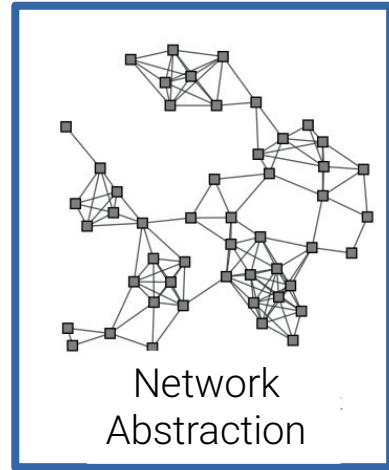
APPLICATION



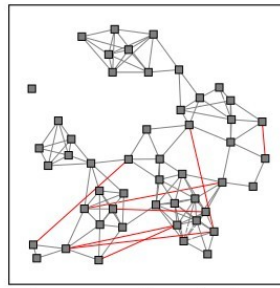
*Dramatic oversimplification



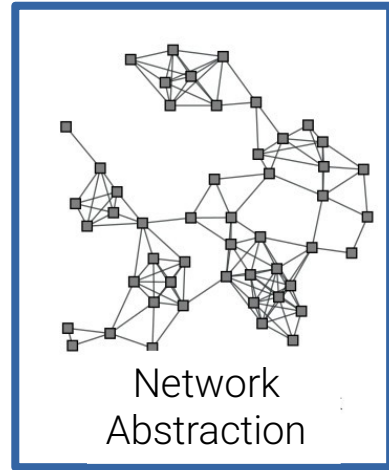
What we'd like to know



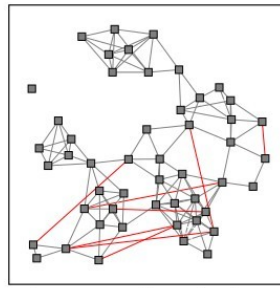
Measurement



What we'd like to know

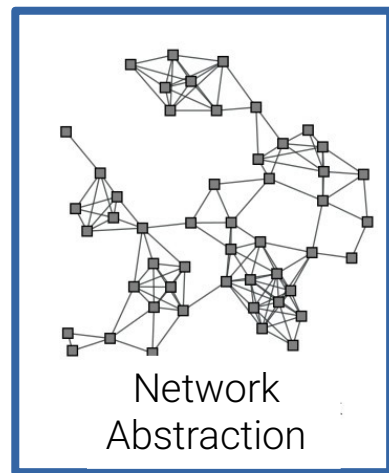


Measurement



Errors and Omissions

What we'd like to know

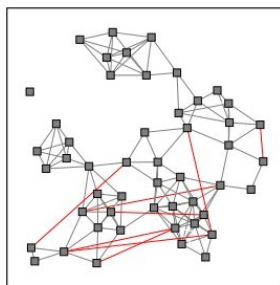


What we'd like to know

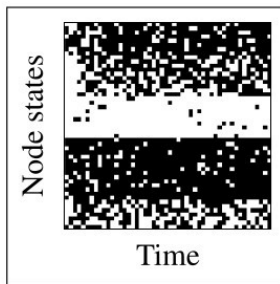
Measurement

Measurement

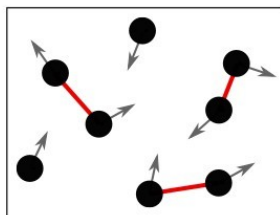
Measurement



Errors and omissions



Dynamics



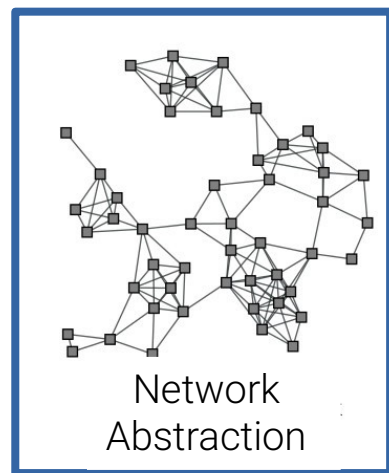
Proximity events

Errors and Omissions

Indirect observations

Thresholds and approximations

What we observe

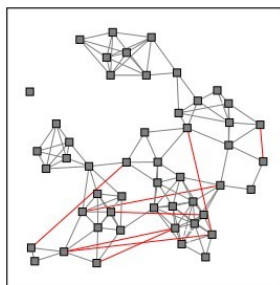


What we'd like to know

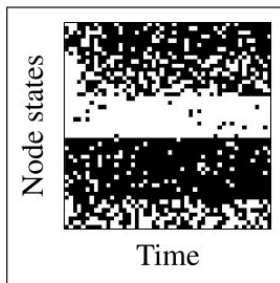
Measurement

Measurement

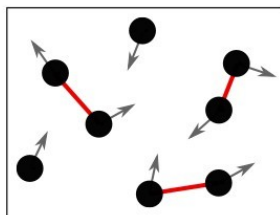
Measurement



Errors and omissions



Dynamics



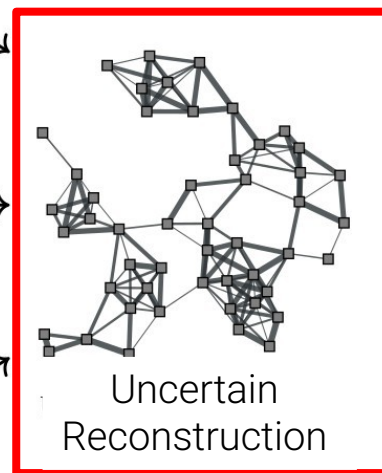
Proximity events

What we observe

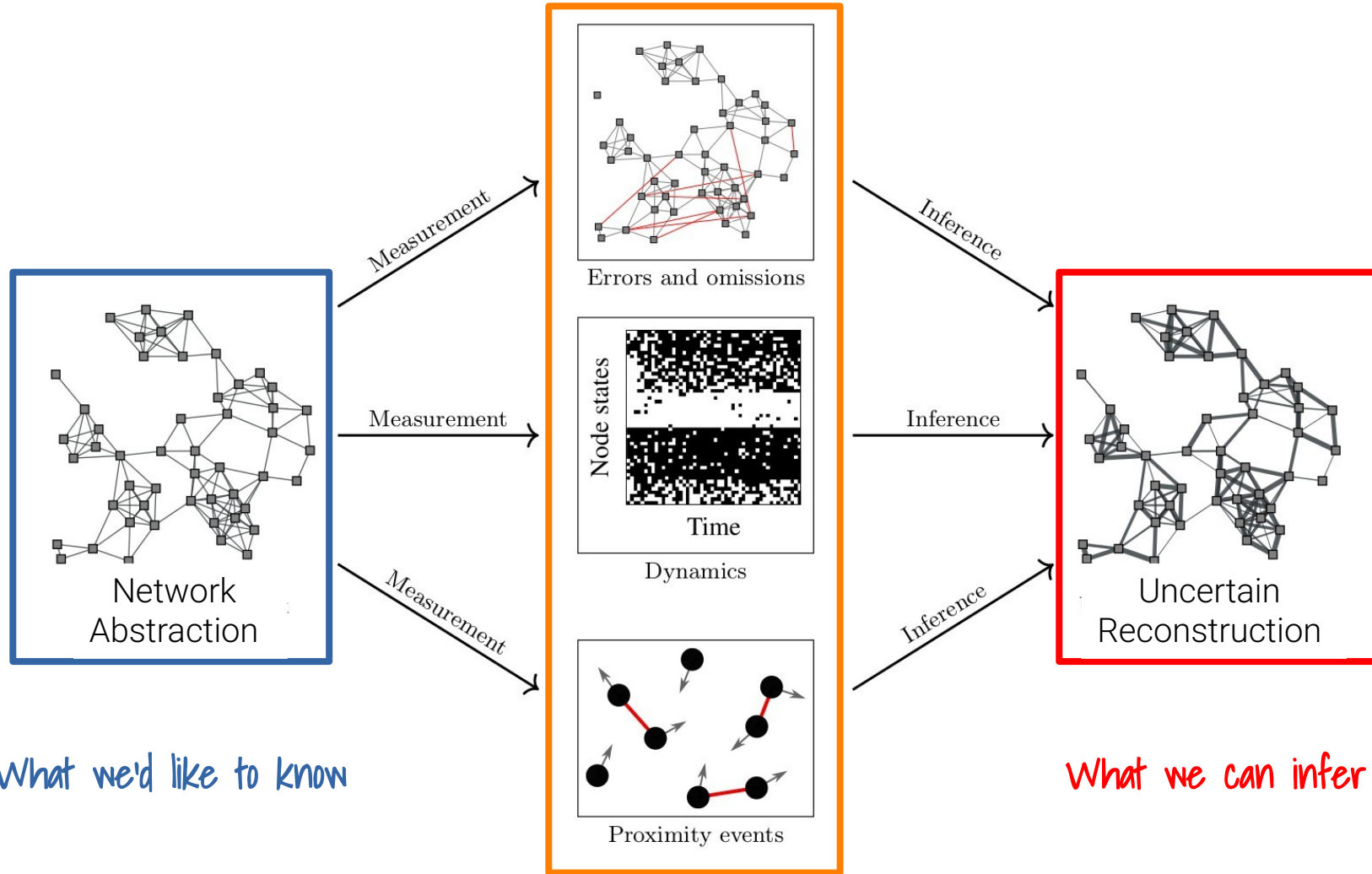
Inference

Inference

Inference



What we can infer



Three Zachary Karate club trophy winners
enter a Zoom...

Three Zachary Karate club trophy winners
enter a Zoom...



Three Zachary Karate club trophy winners
enter a Zoom...



Three Zachary Karate club trophy winners
enter a Zoom...



Three Zachary Karate club trophy winners
enter a Zoom...



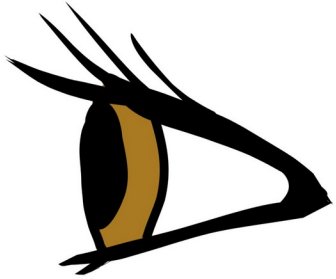
Three Zachary Karate club trophy winners
enter a Zoom...



This scene never actually happened, its a reconstruction!

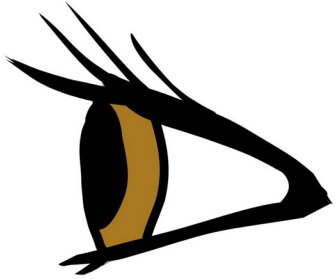
Network science in 3 easy steps...

1. Observations/ measurements



Network science in 3 easy steps...

1. Observations/
measurements

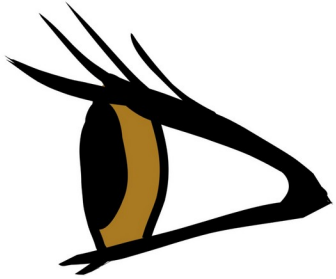


2. Network representation



Network science in 3 easy steps...

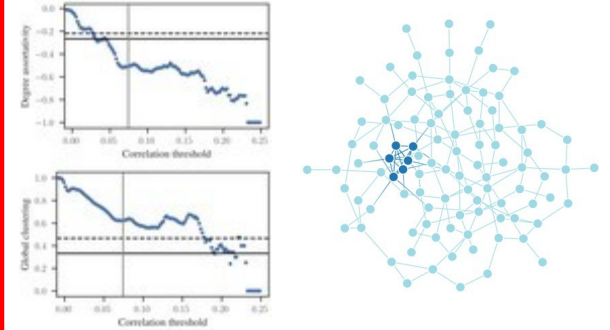
1. Observations/ measurements



2. Network representation

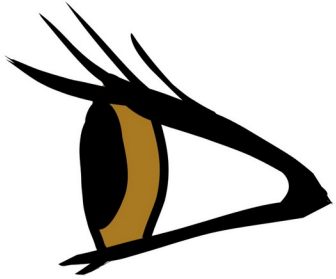


3. Network analysis



Network science in 3 easy steps...

1. Observations/ measurements

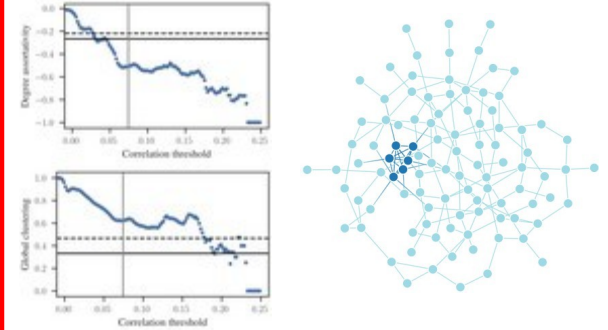


Obscured
quality of data

2. Network representation

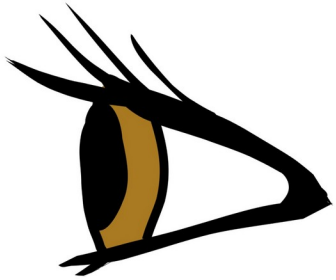


3. Network analysis



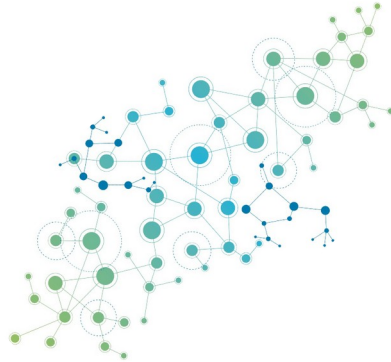
Network science in 3 easy steps...

1. Observations/ measurements



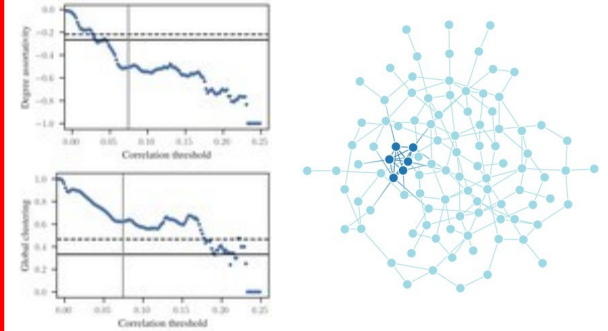
Obscured
quality of data

2. Network representation



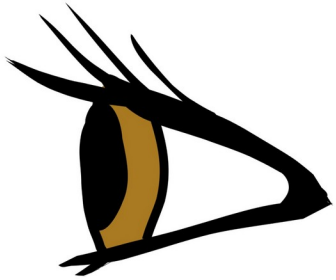
Choice of
representation

3. Network analysis



Network science in 3 easy steps...

1. Observations/ measurements



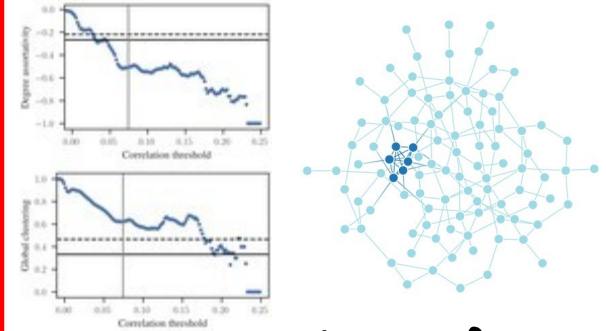
Obscured
quality of data

2. Network representation



Choice of
representation

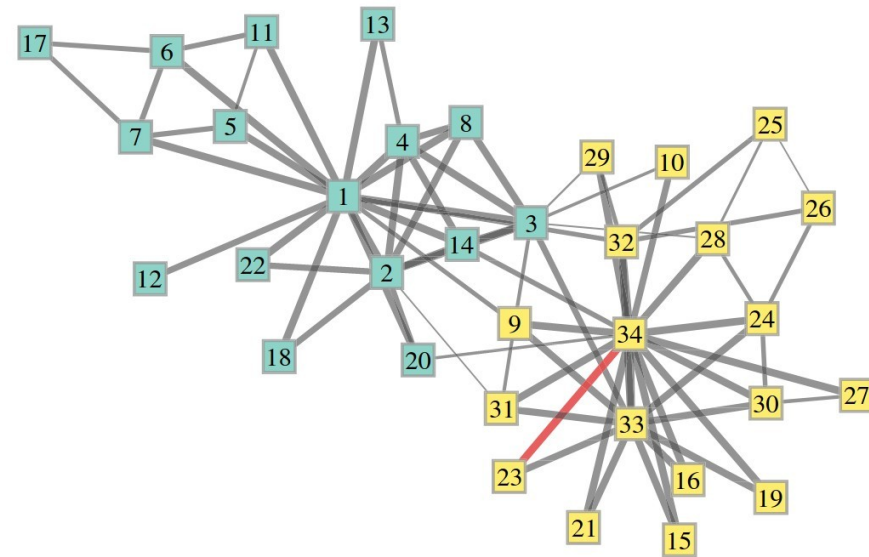
3. Network analysis



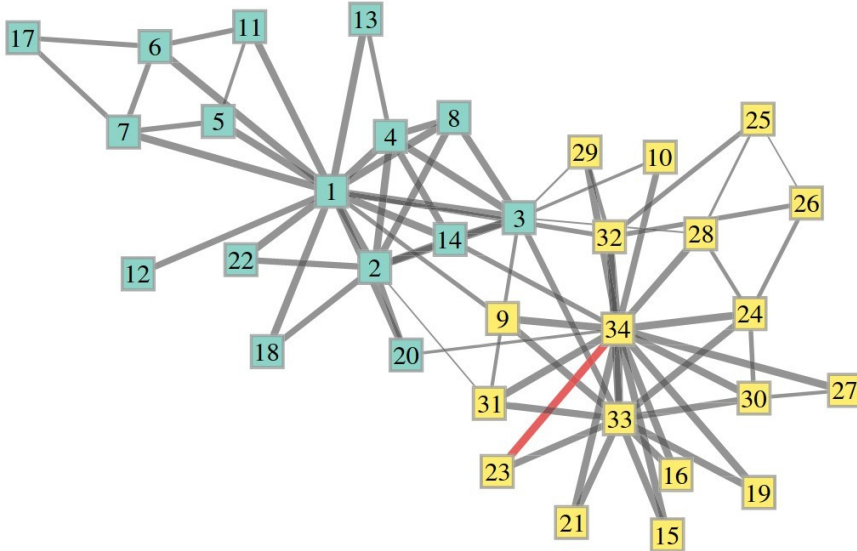
Suitability of
the methods

1. Obscured quality of data

Zachary's Karate Club

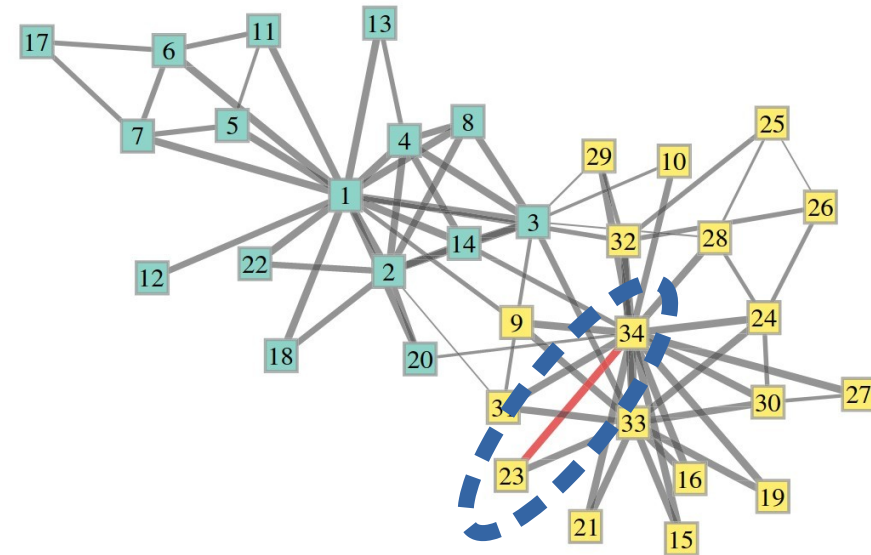


Karate Club

[illegible]

Zachary's Karate Club

Individual Number																																				
	1	2	3	4	5	6	7	8	9	0	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3		
1	0	4	5	3	3	3	3	2	2	0	2	3	2	3	0	0	0	2	0	2	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	
2	4	0	6	3	0	0	0	4	0	0	0	0	0	5	0	0	0	1	0	2	0	2	0	0	0	0	0	0	0	0	2	0	0	0		
3	5	6	0	3	0	0	0	4	5	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	3	0	
4	3	3	3	0	0	0	0	3	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	3	0	0	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	3	0	0	0	0	0	5	0	0	0	3	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	3	0	0	0	2	5	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	2	4	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	4	3		
10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
11	2	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	3	5	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4	
17	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	
20	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	
22	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	4	0	2	0	0	0	
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	0	2	0	0	
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	7	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	2
28	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	0	0	0	0	0	0	0	0	4
29	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	4	0	0	0	0	0	3	2
31	0	2	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3
32	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7	0	0	2	0	0	0	4	4
33	0	0	2	0	0	0	0	0	3	0	0	0	0	0	3	3	0	0	1	0	3	0	0	5	0	0	0	0	0	4	3	4	0	5	0	0
34	0	0	0	0	0	0	0	4	2	0	0	0	3	2	4	0	0	2	1	1	0	3	0	0	0	2	4	2	2	3	4	5	0	0	0	0



Does this edge exist?

Assessing experimentally derived interactions in a small world

Debra S. Goldberg and Frederick P. Roth*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Edited by Lawrence A. Shepp, Rutgers, The State University of New Jersey–New Brunswick, Piscataway, NJ, and approved February 10, 2003 (received for review September 27, 2002)

Experimentally determined networks are susceptible to errors, yet important inferences can still be drawn from them. Many real networks have also been shown to have the small-world

negative errors (24, 25). Here we consider in detail a network of protein–protein interactions derived from high-throughput, error-prone yeast two-hybrid (Y2H) studies (26, 27). These data

Assessing experimentally derived interactions in a small world

Debra S. Goldberg and Frederick P. Roth*

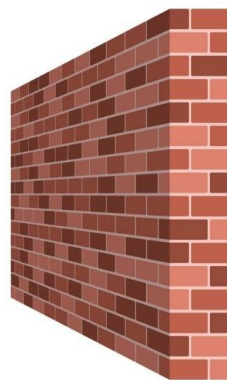
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Edited by Lawrence A. Shepp, Rutgers, The State University of New Jersey–New Brunswick, Piscataway, NJ, and approved February 10, 2003 (received for review September 27, 2002)

Experimentally determined networks are susceptible to errors, yet important inferences can still be drawn from them. Many real networks have also been shown to have the small-world

negative errors (24, 25). Here we consider in detail a network of protein–protein interactions derived from high-throughput, error-prone yeast two-hybrid (Y2H) studies (26, 27). These data

Link prediction

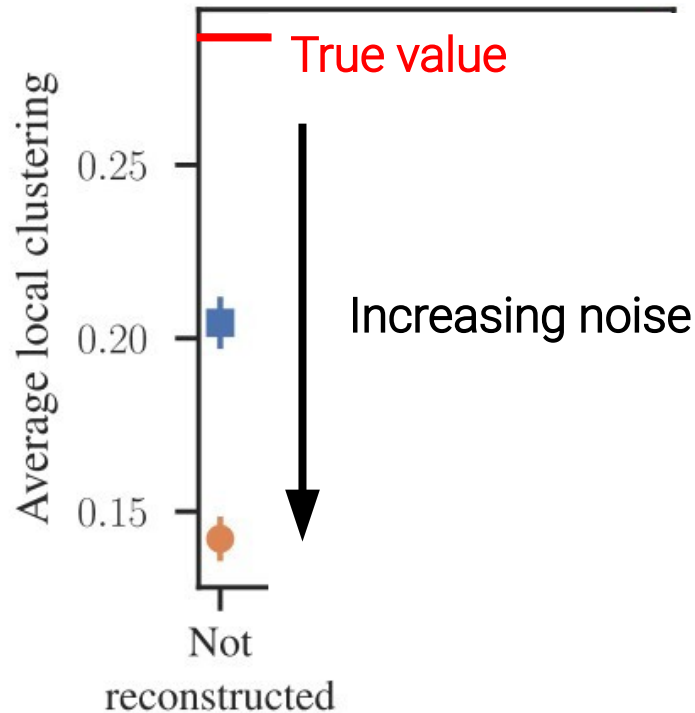


Rest of network science

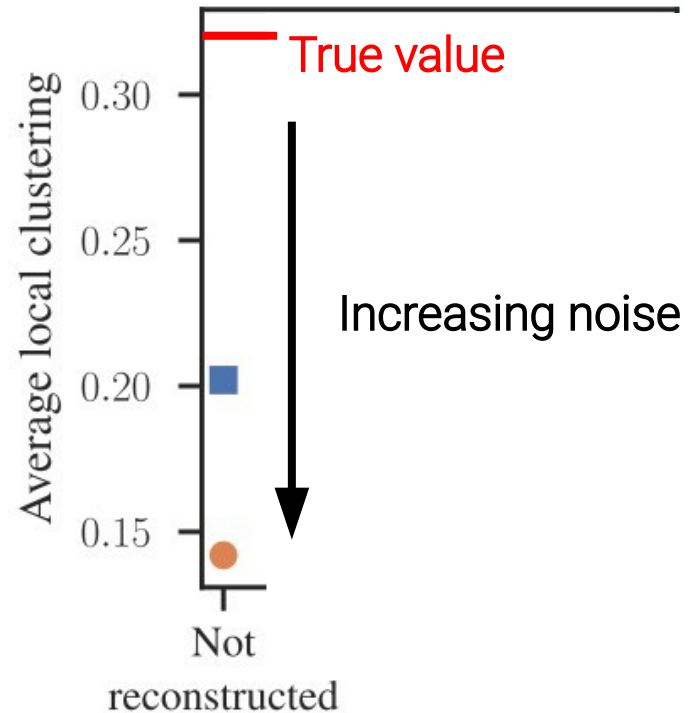
Errors in network data create systematic biases...

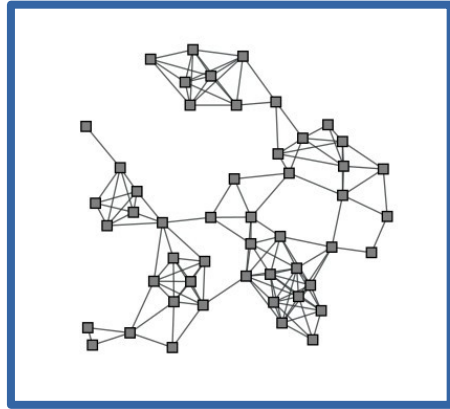
Errors in network data create systematic biases...

(a) High-school friendships

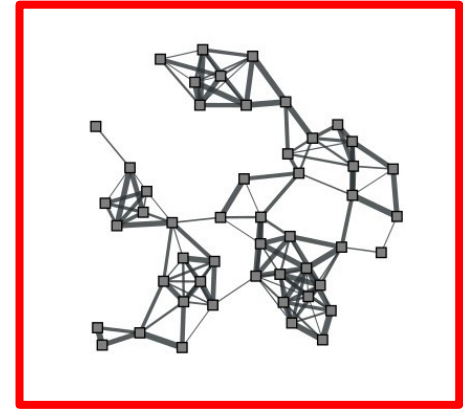


(b) Political blogs



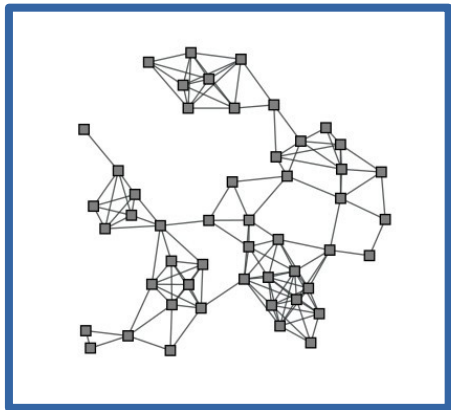


True Network



Reconstructed Network

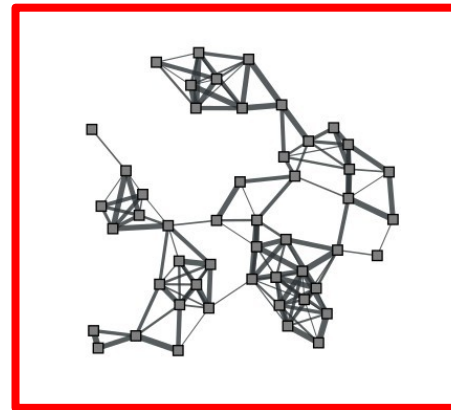
We don't know if the network represents the system



True Network

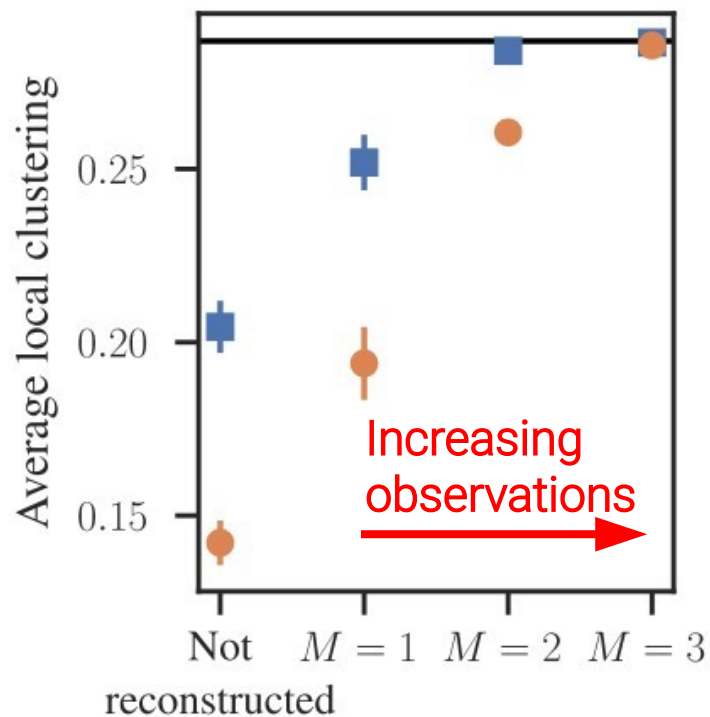
$$P(\mathbf{A}|\mathbf{D}) = \frac{P(\mathbf{D}|\mathbf{A})P(\mathbf{A})}{P(\mathbf{D})} .$$

Bayesian inference

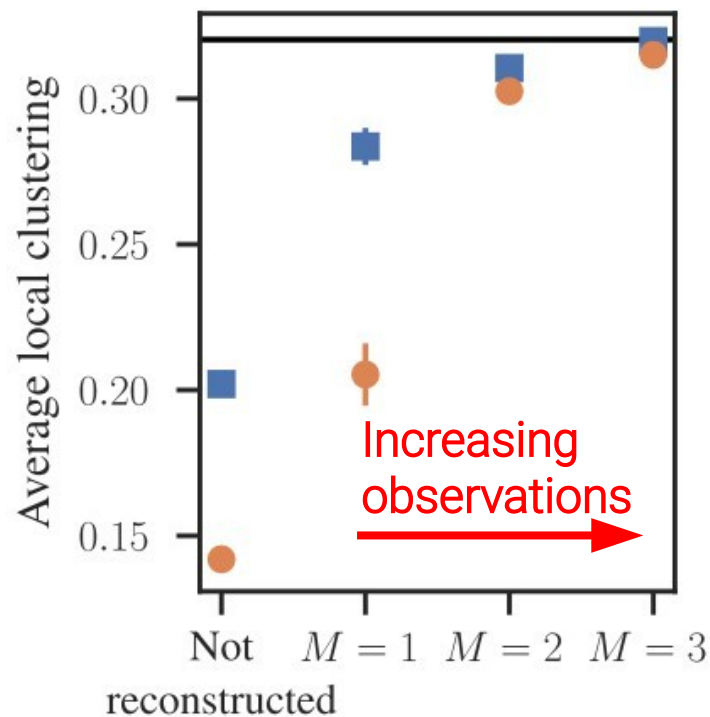


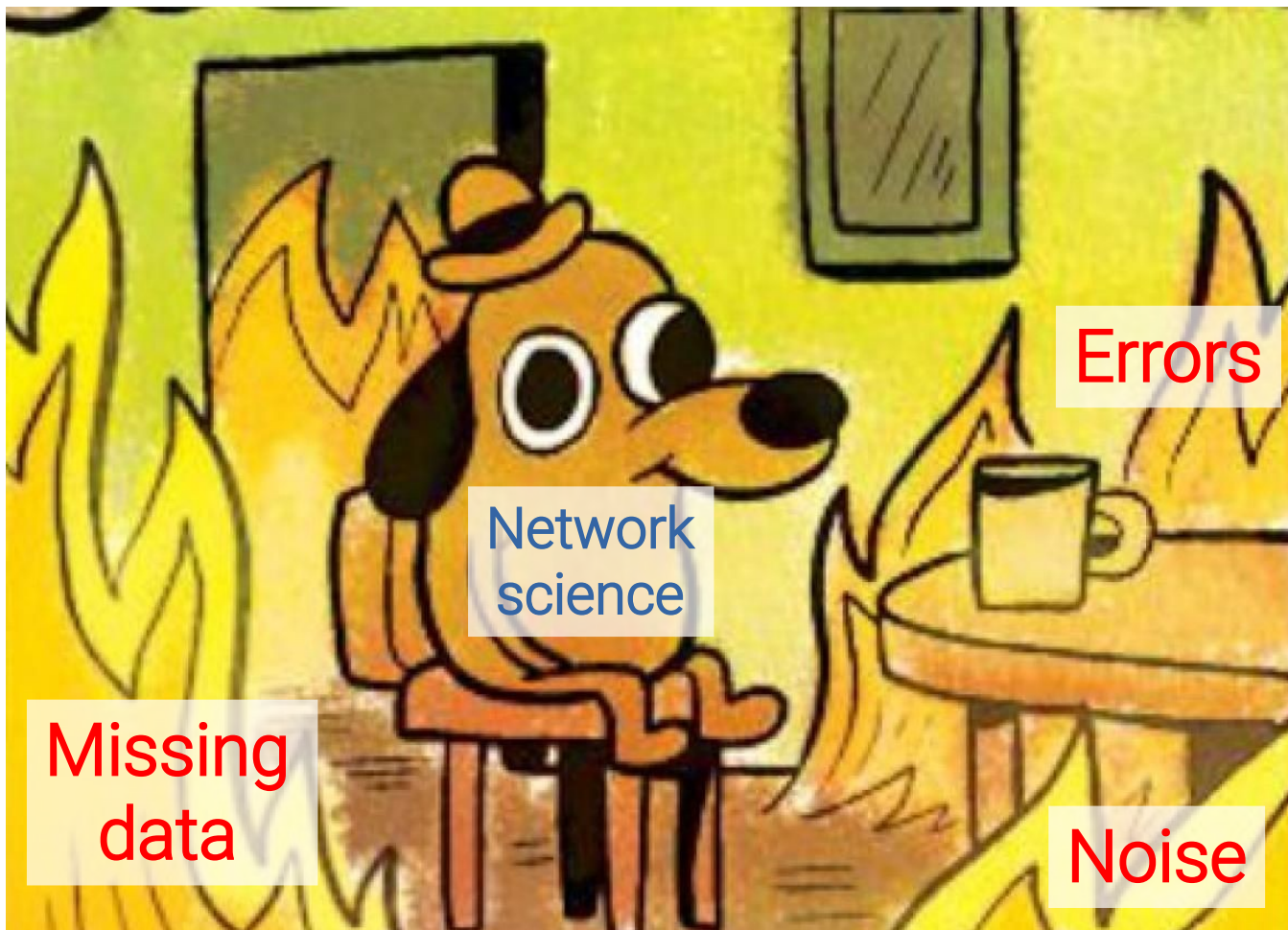
Reconstructed Network

(a) High-school friendships



(b) Political blogs

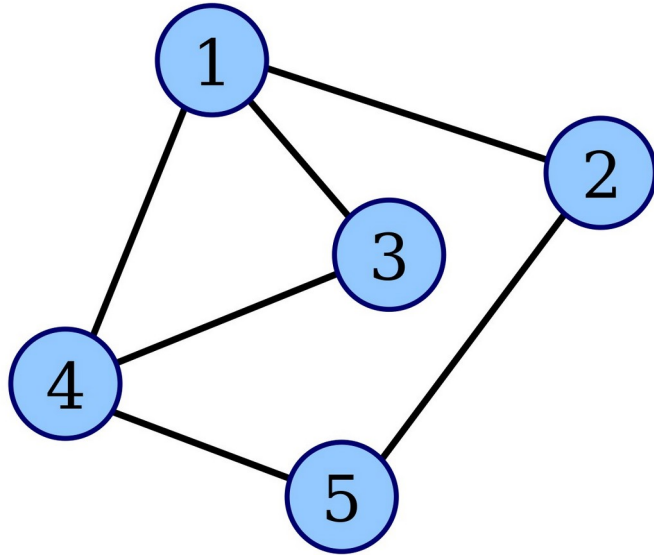




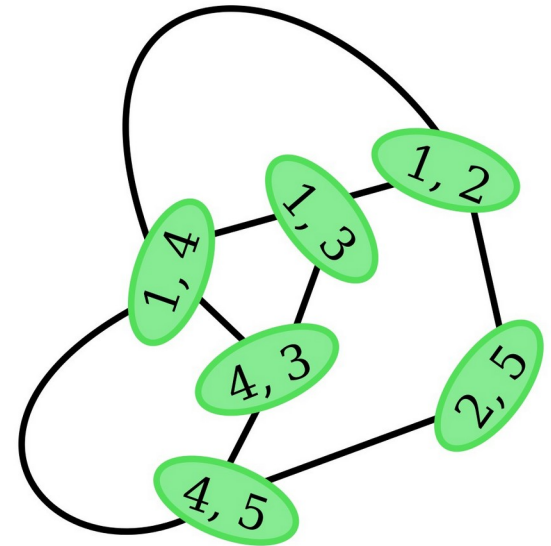
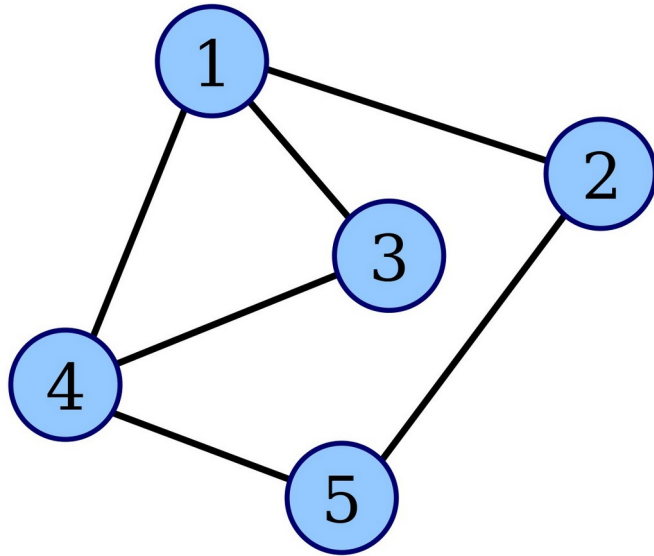
This is NOT fine

II. Choice of representation

What are the nodes and what are the edges?

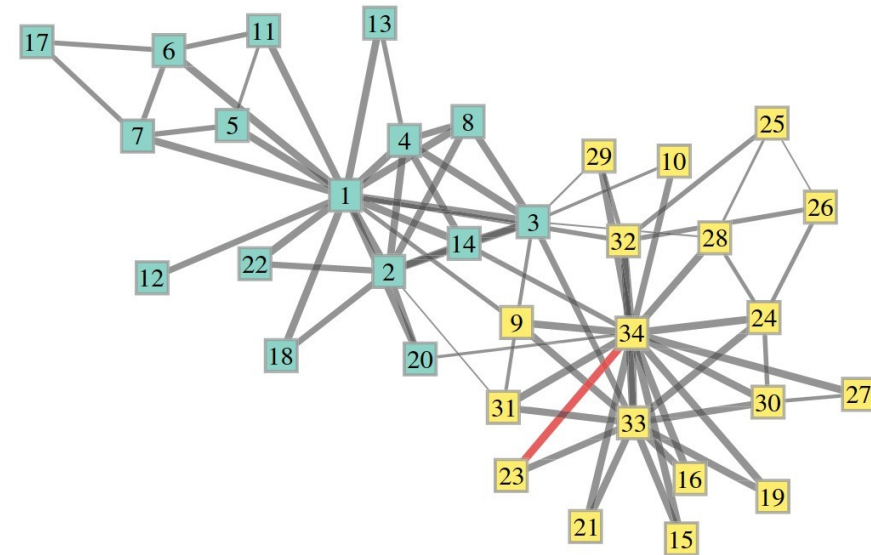


What are the nodes and what are the edges?

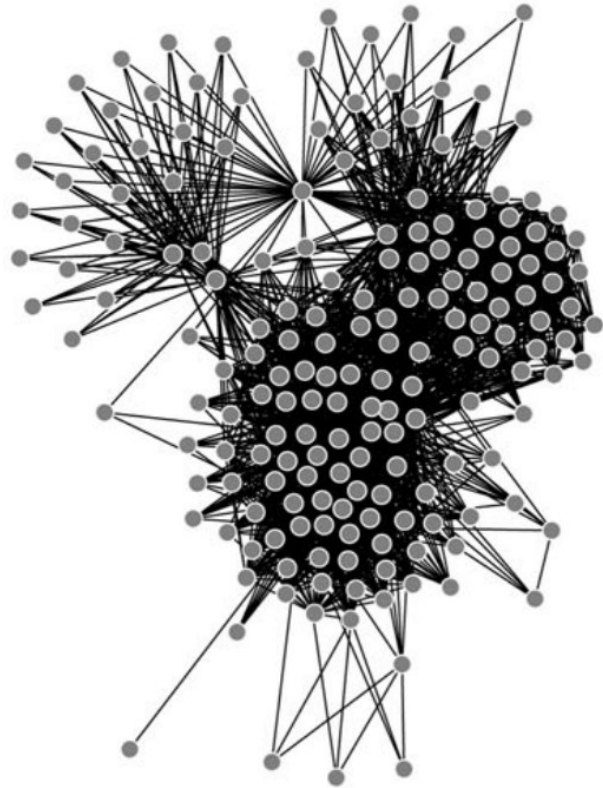


Zachary's Karate Club

What about
the raw data?

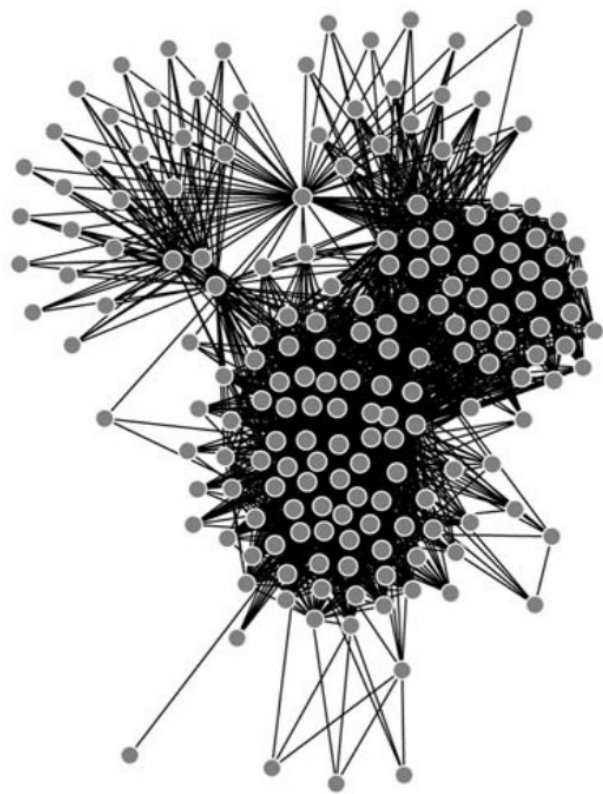


How does the network generate data?

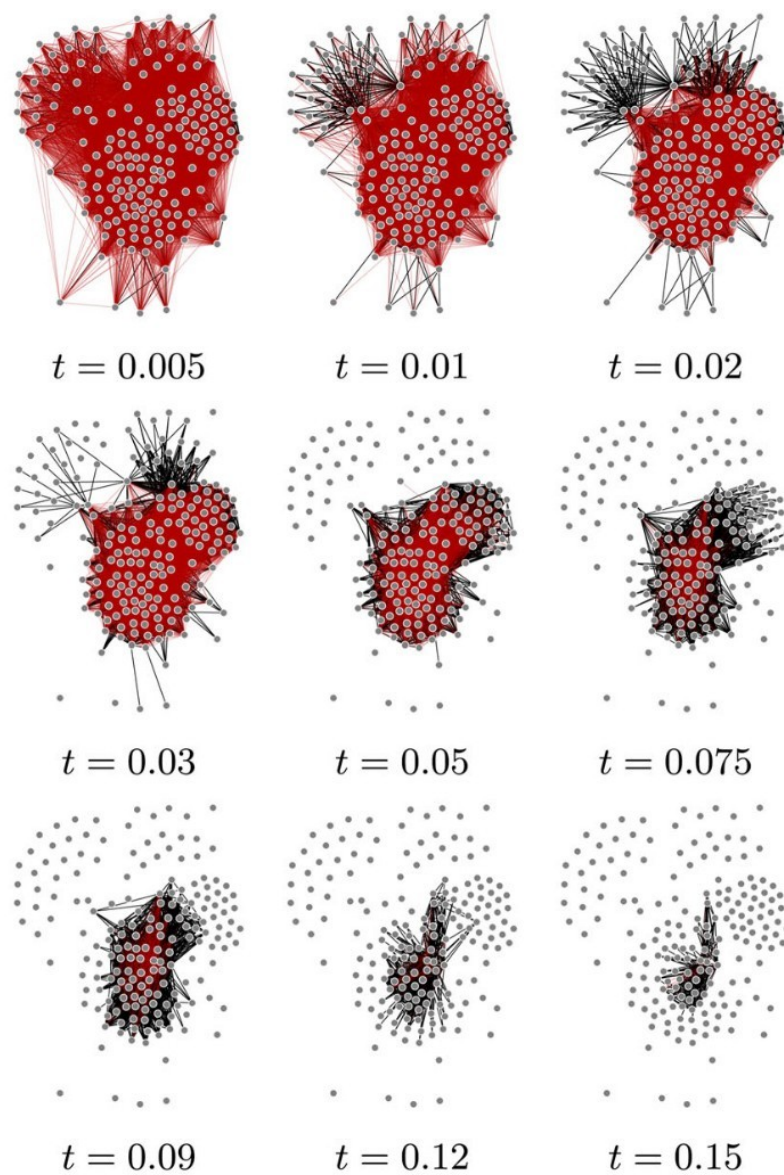


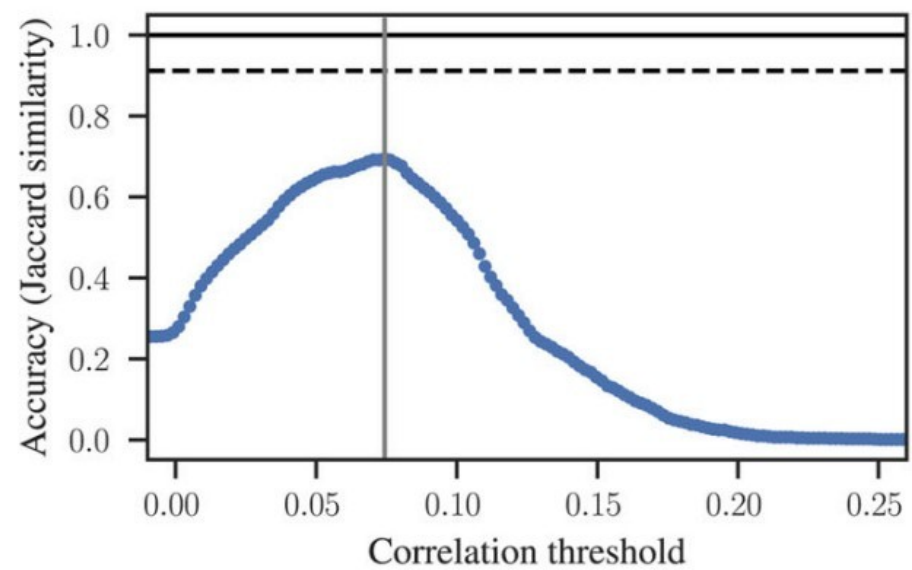
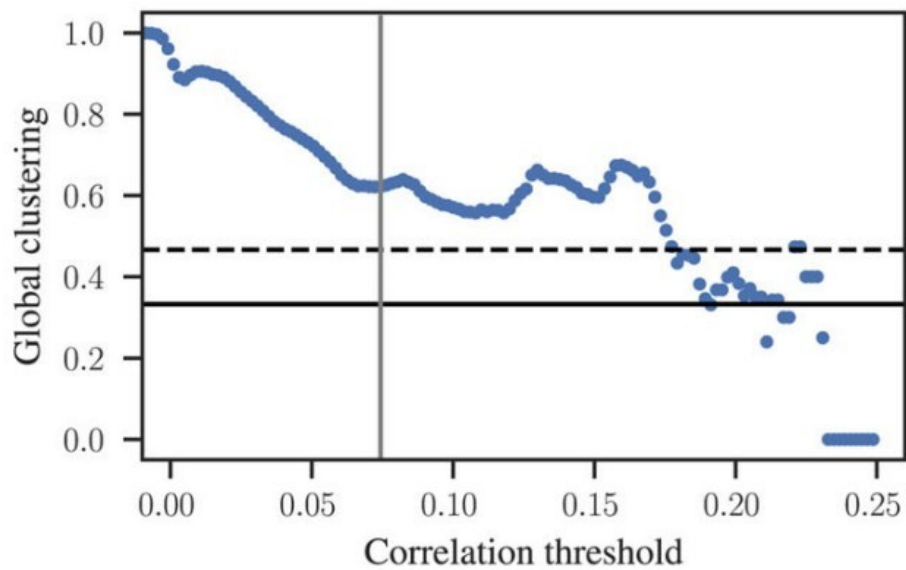
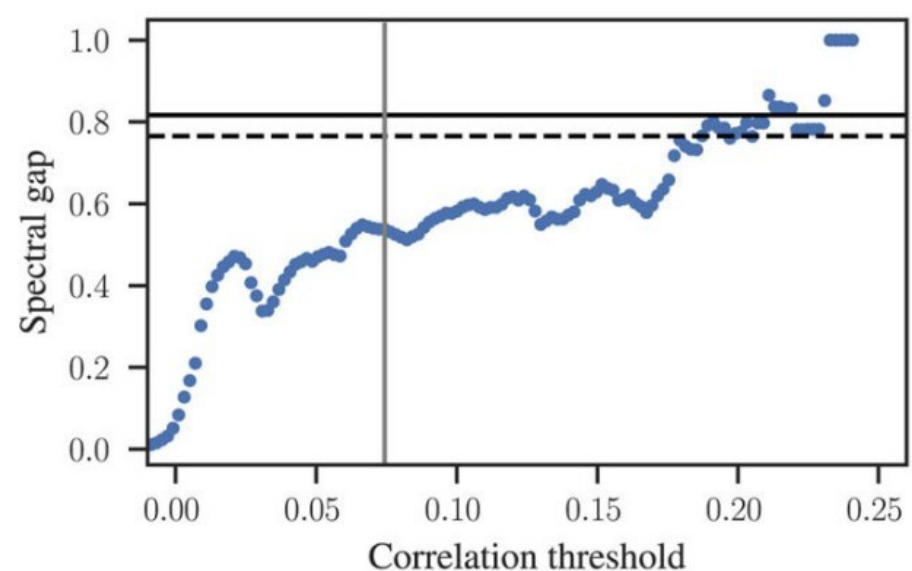
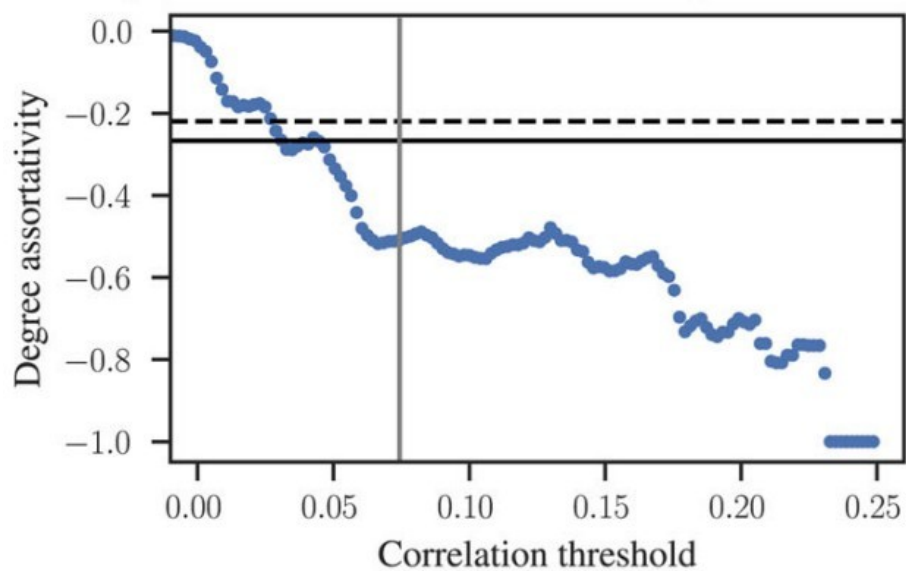
(a) True network

Correlation "networks"

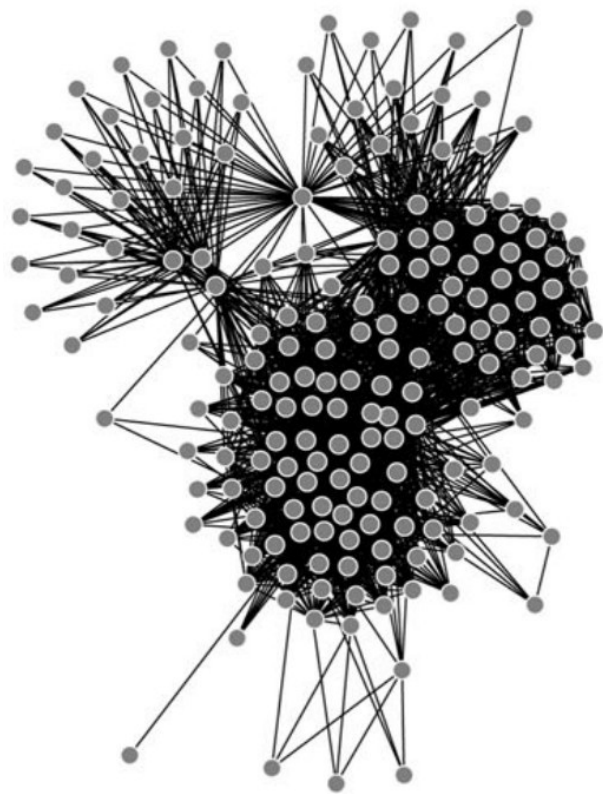


(a) True network

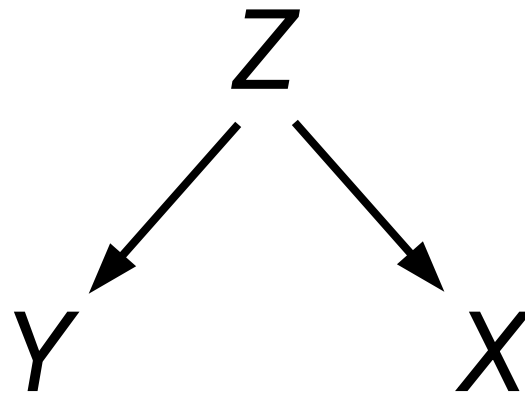




Correlation "networks"



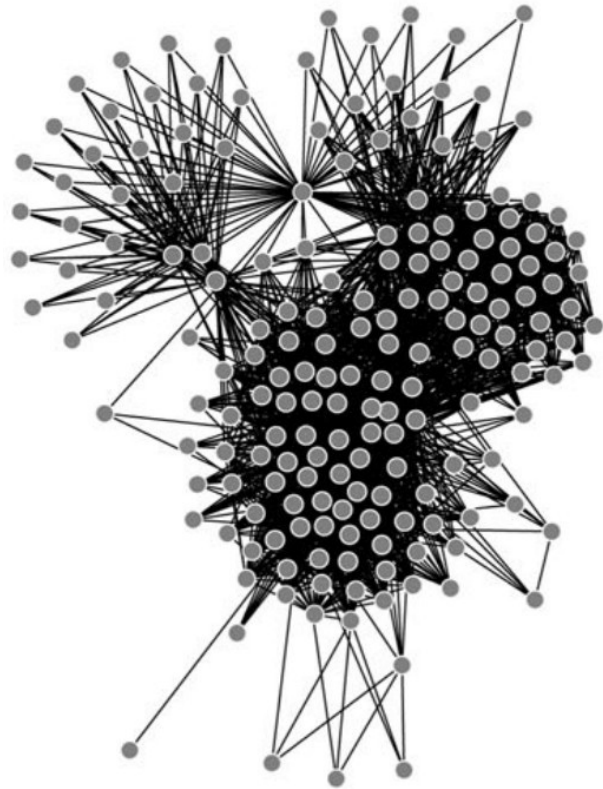
(a) True network



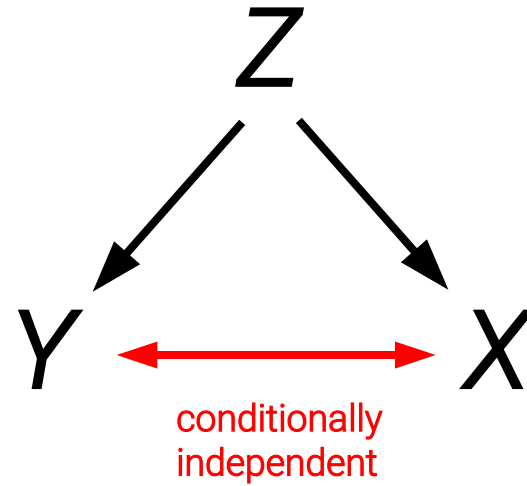
$$X(t) = Z(t) + \epsilon_1$$

$$Y(t) = Z(t) + \epsilon_2,$$

Correlation "networks"

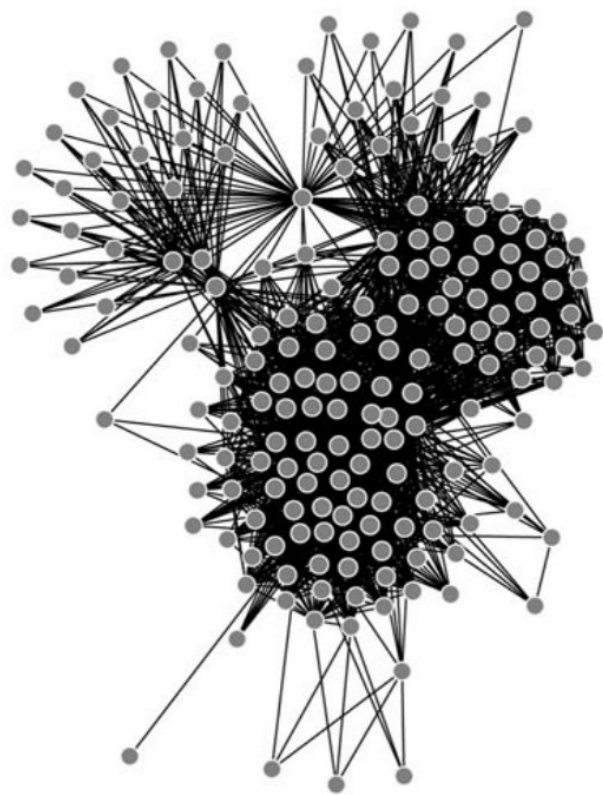


(a) True network

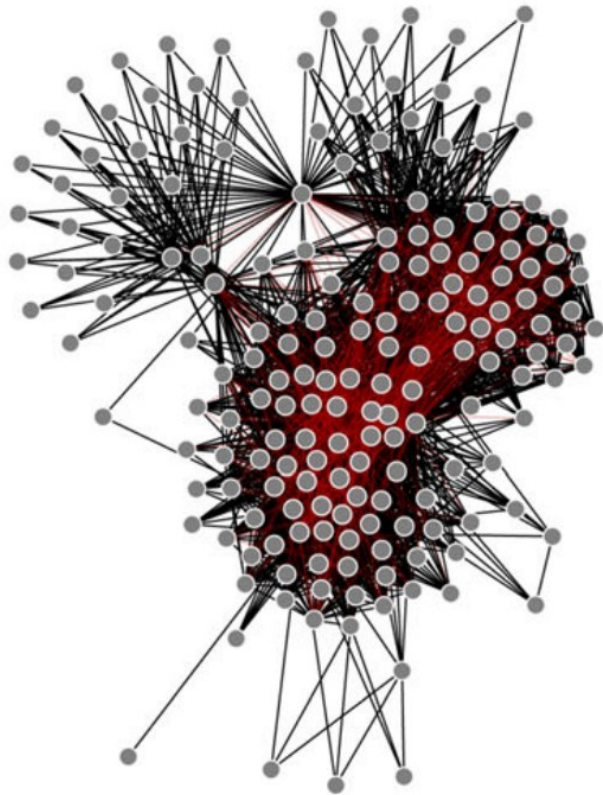


$$X(t) = Z(t) + \epsilon_1$$

$$Y(t) = Z(t) + \epsilon_2,$$

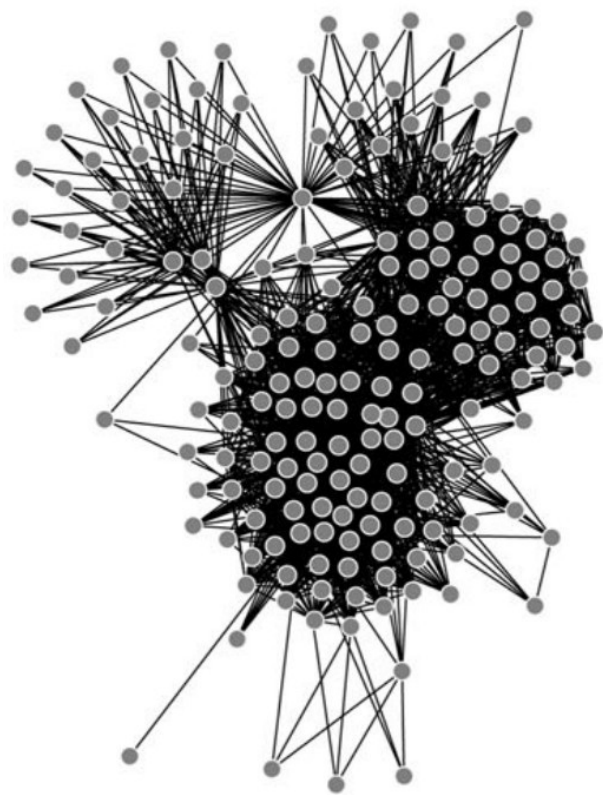


(a) True network

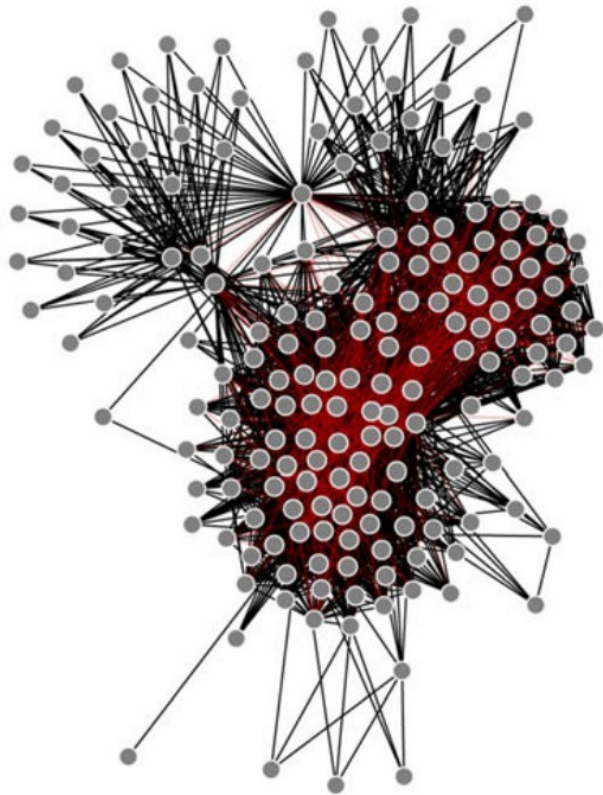


(b) Graphical LASSO

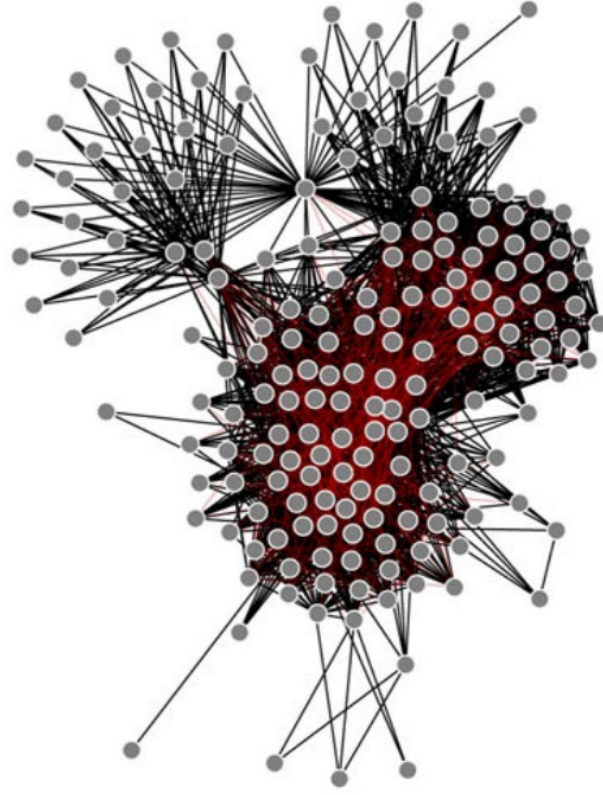
Friedman, J., Hastie, T. & Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432-441 (2008).



(a) True network



(b) Graphical LASSO



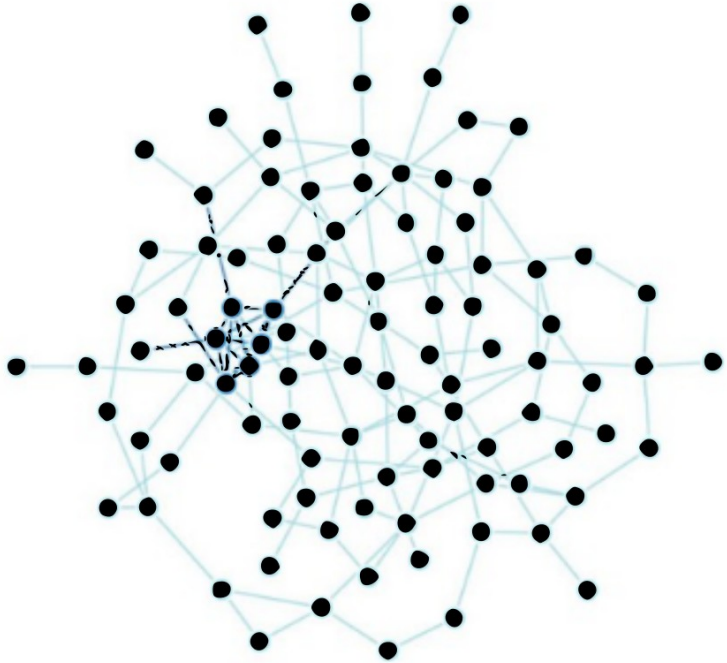
(c) Bayesian inference



"I see networks!"

III. Suitability of the methods

Summary descriptors used out of context



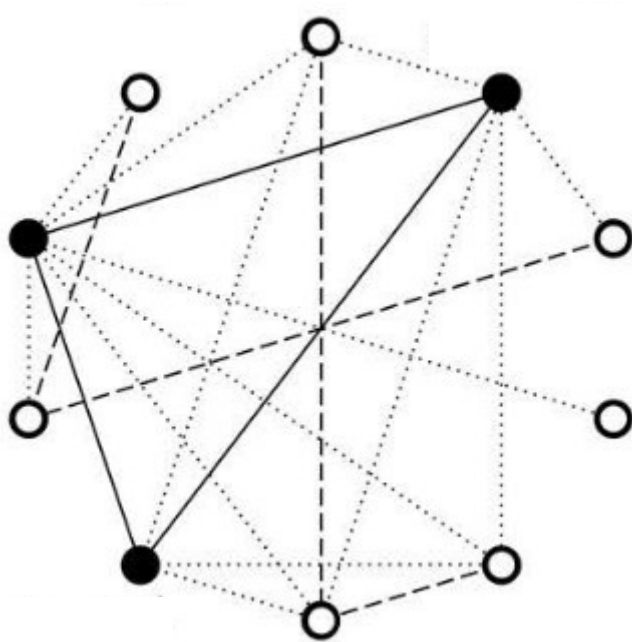
Shortest path of a correlation network?

Maximum modularity of a network?

What to vary, what to keep the same?

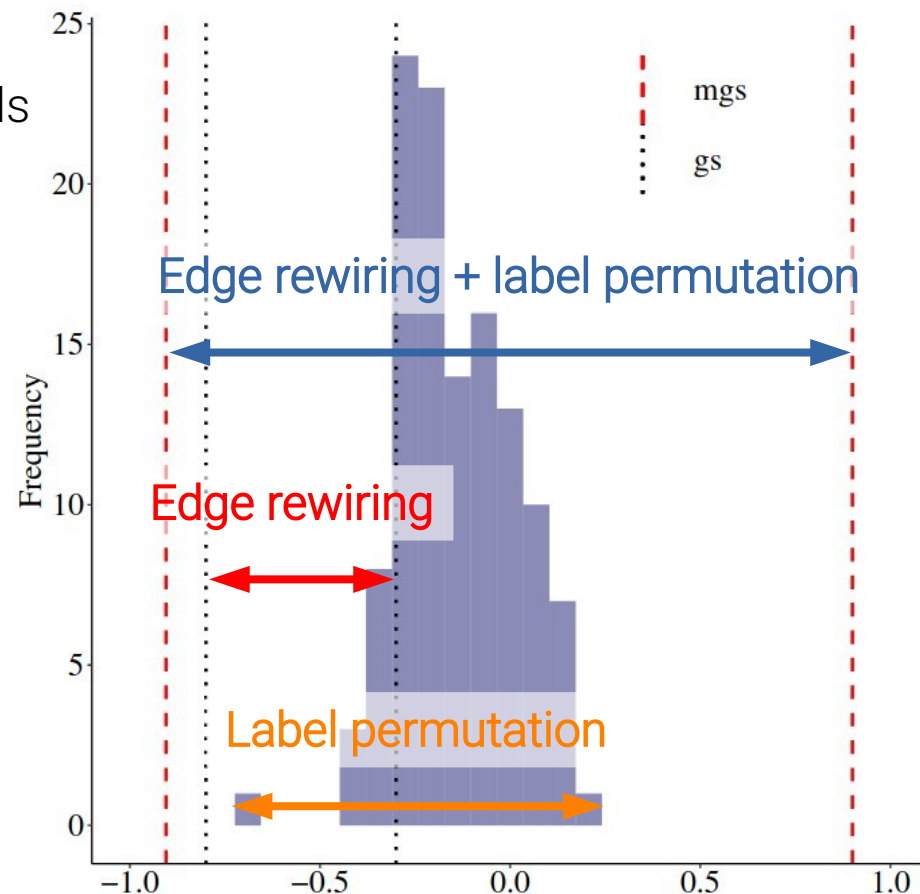
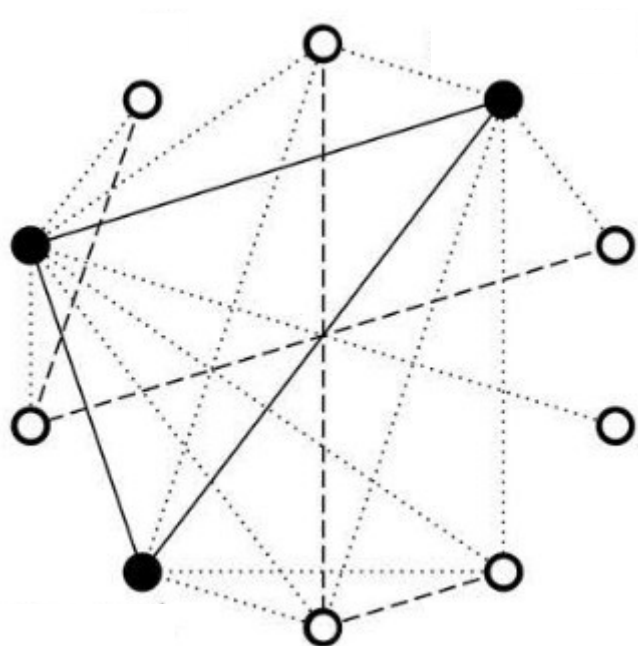
What to vary, what to keep the same?

Assortativity of node labels for different null models

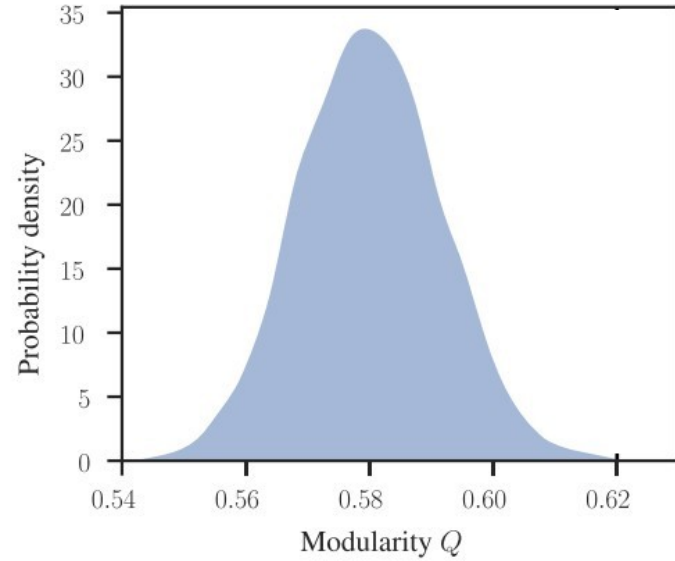


What to vary, what to keep the same?

Assortativity of node labels for different null models

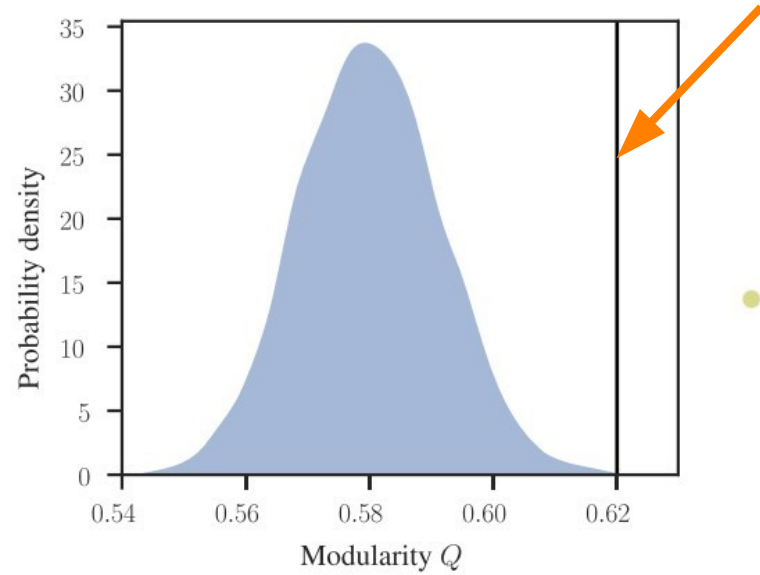


Null models and testing hypotheses

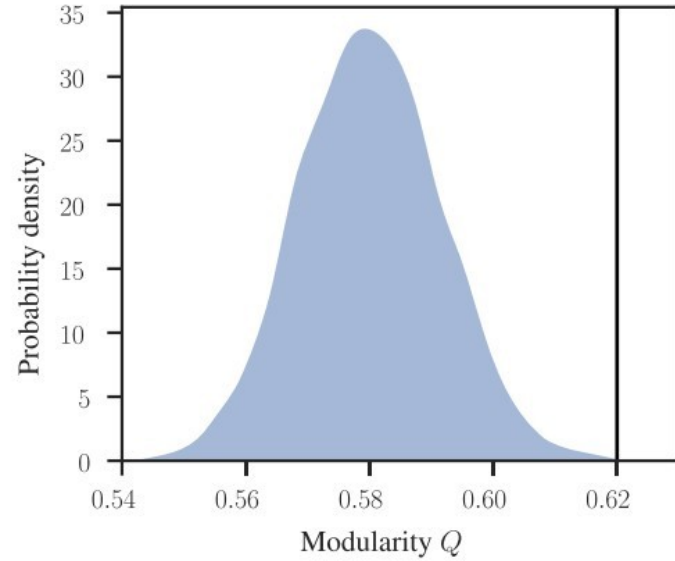


Null models and testing hypotheses

Reject the null hypothesis

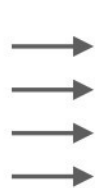
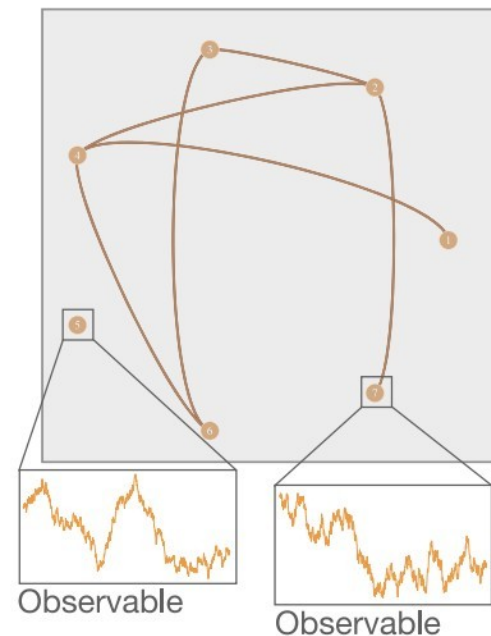


Rejecting the null hypothesis does not test
the alternative...



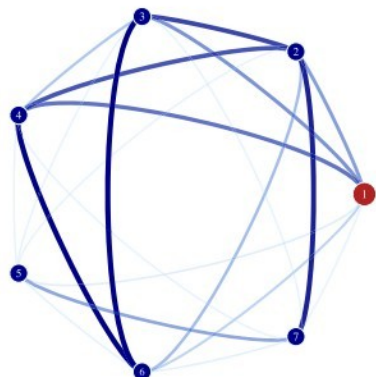
Accounting for reconstruction uncertainty

Original (unknown) Network



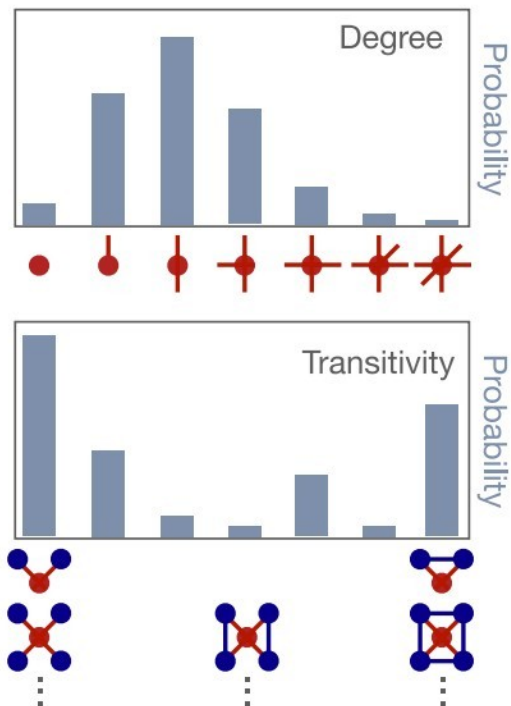
Network
Reconstruction
Method

Probabilistic Network Model



Low High
Probability of
link presence

Descriptors Inference





MODEL FREE

**IS JUST MODELLING
WITH YOUR EYES CLOSED**

IV. Outlook

Looking forwards...

Looking forwards...



Eat our own dog food. More focus on collaborations, less on individuals

Looking forwards...



Eat our own dog food. More focus on collaborations, less on individuals



Break down walls. Strengthen the link between theory and application.

Looking forwards...



Eat our own dog food. More focus on collaborations, less on individuals

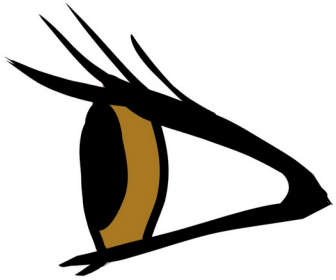


Break down walls. Strengthen the link between theory and application.



Better modelling. Generative models + statistical inference. Focus on more specific models. Solve real problems.

Observations/
measurements



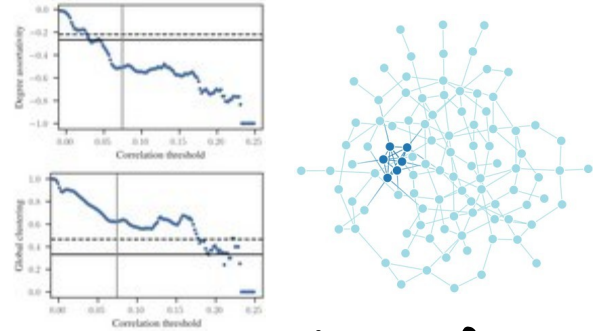
Obscured
quality of data

Network representation



Choice of
representation

Network analysis



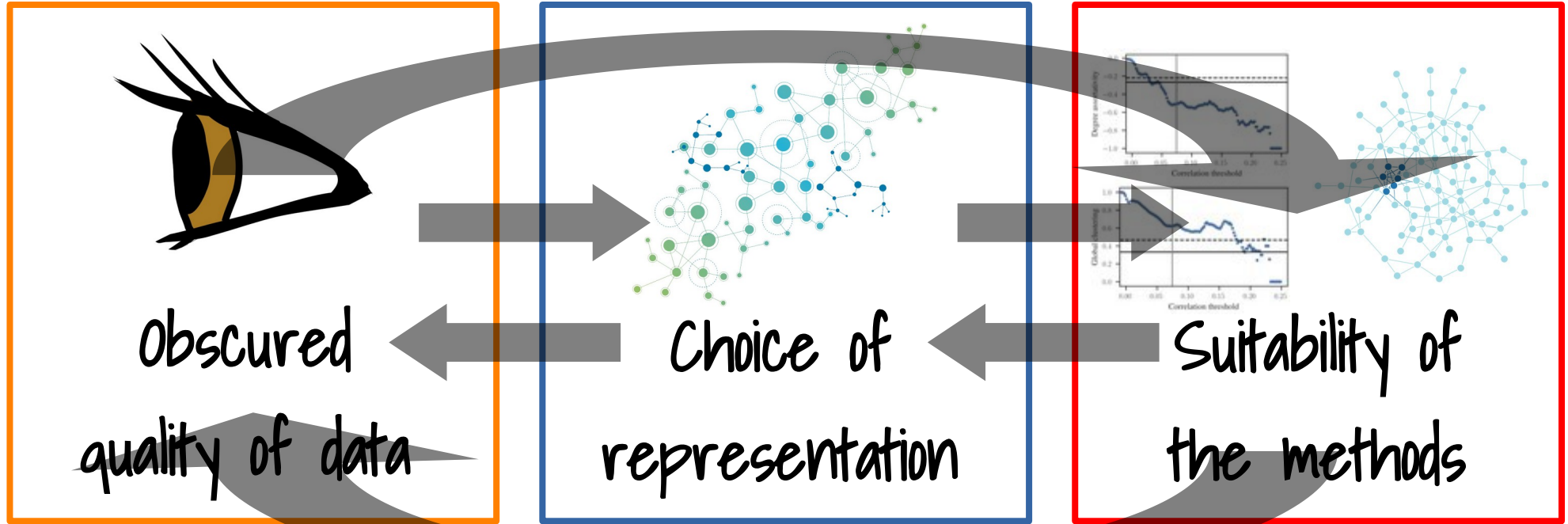
Suitability of
the methods

These steps are interdependent

Observations/
measurements

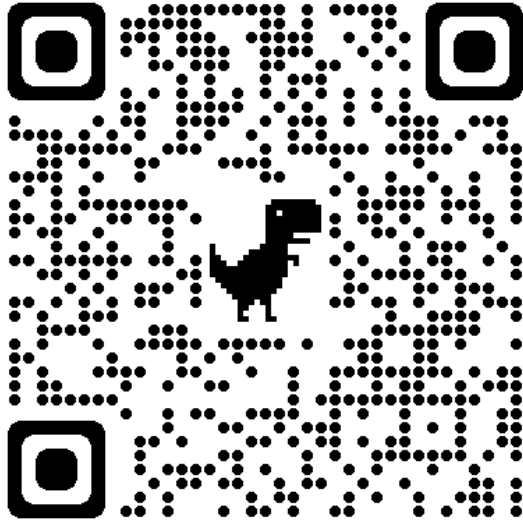
Network representation

Network analysis



Download the paper!

Peel, L., Peixoto, T.P. & De Domenico, M. Statistical inference links data and theory in network science. *Nat Commun* **13**, 6794 (2022).



Contact:



@PiratePeel



l.peel@maastrichtuniversity.nl

Joint work with...



Tiago P.
Peixoto



Manlio De
Domenico