Transforming Mutations into Models: Inferring Causal Networks from Experimental Data

Jolanda Kossakowski

University of Amsterdam
J.J.Kossakowski@uva.nl
http://jolandakossakowski.eu

IMPS 2017

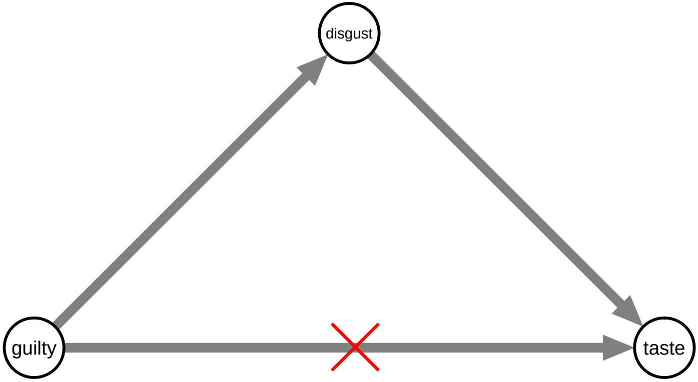
In search for the holy grail

In psychology, we often want to know what the parents of variables are.

"a causal relation is a relation between two variables where, when changing one variable, we expect to observe a change in the other variable"

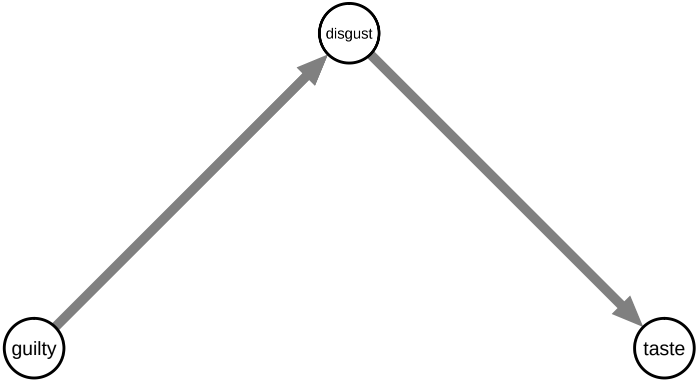
In search for the holy grail

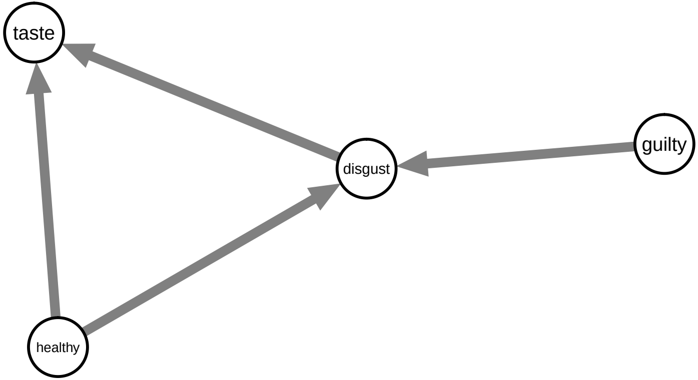
If there are multiple generations of parents and children in a network, we want to determine a node's true parent(s)



In search for the holy grail

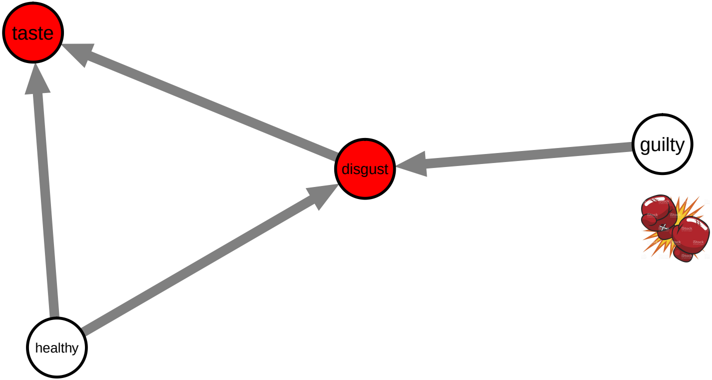
If there are multiple generations of parents and children in a network, we want to determine a node's true parent(s)

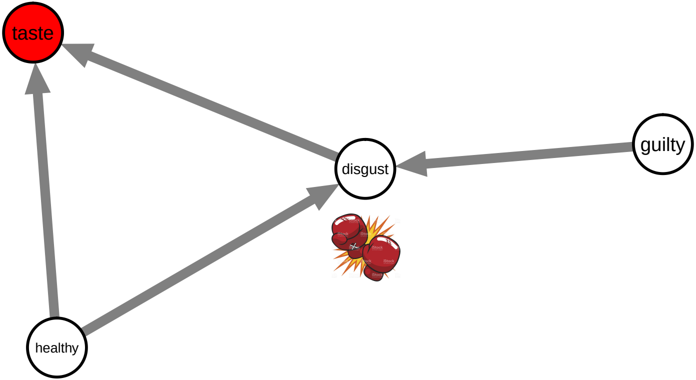



taste
disgust
guilty
healthy

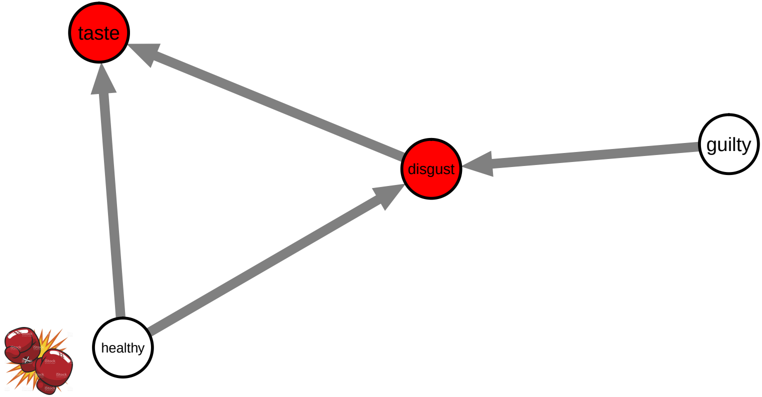
From observational to experimental data

Experimental data



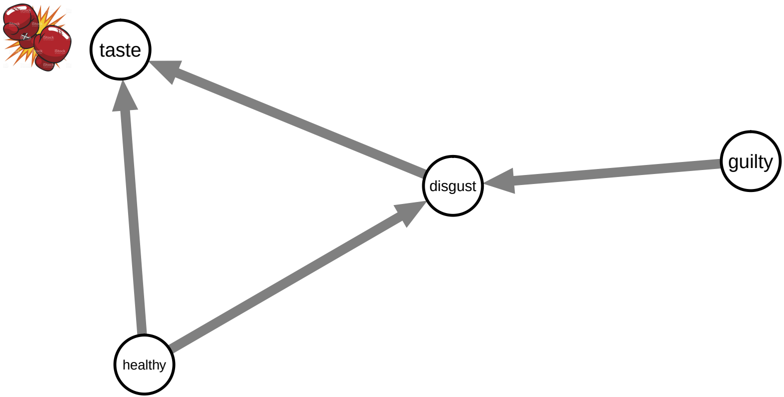


taste
disgust
guilty
healthy


taste
disgust
guilty
healthy

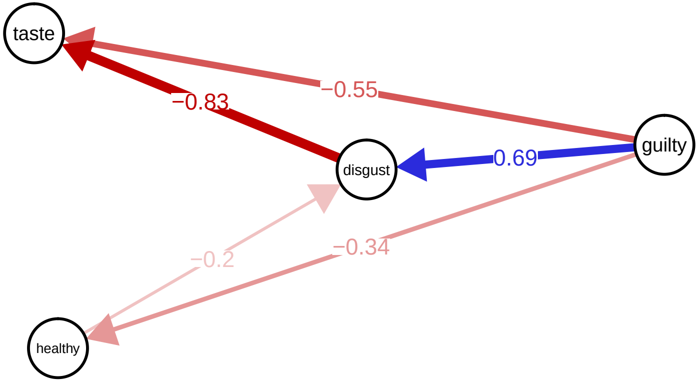
From observational to experimental data

Experimental data



Turning experimental data into a network

A high conditional correlation occurs when the effect shown in the observational data is also shown in the experimental data.

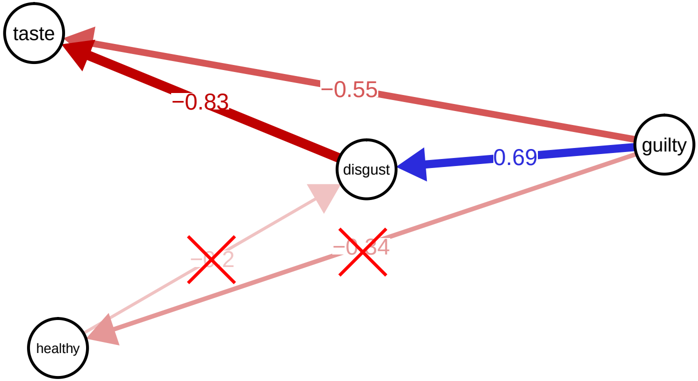


How do we determine if the relation between “guilty“ and “taste“ is a direct or indirect relation?

Current Methods for causal networks

Transitive Reduction

Transitive reduction prunes relations whose conditional correlation do not exceed some threshold:



$|0.20|$ & $|0.34| <$ threshold

Current Methods for causal networks

Transitive Reduction

- Transitive reduction algorithms start out with a graph that contains edges between variables whose conditional correlation exceed some threshold
- Edges are stepwise removed when alternative paths between those variables satisfy certain conditions
- DR-FFL[1] and TRANSWESD[2] are two methods that use transitive reduction
- DR-FFL and TRANSWESD vary in their usage of the conditional correlation when pruning relations from a network

[1] Pinna et al. (2010) From knockouts to networks: Establishing direct cause-effect relationships through graph analysis

[2] Klamt et al. (2010) TRANSWESD: inferring cellular networks with transitive reduction

Current Methods for causal networks

IC-algorithm

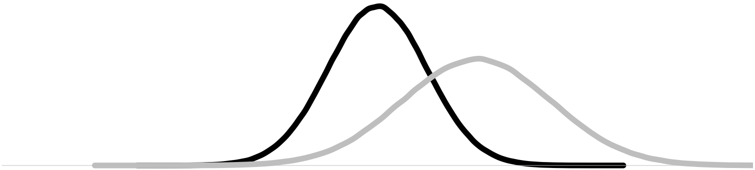
- Based on work by Pearl[3]
- Uses only observational data
- Sometimes finds unlogical relations

[3] Pearl. Causality

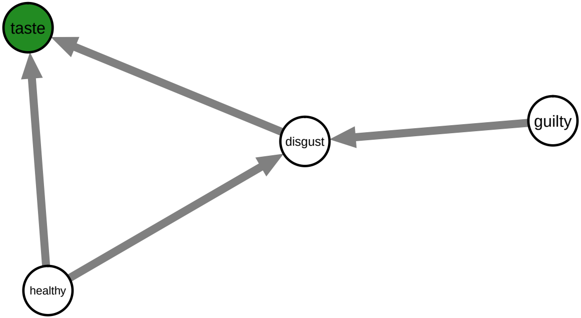
Current Methods for causal networks

Invariant Causal Prediction

- Invariant Causal Prediction[4] (ICP) regresses one *target* variable on each possible subset of remaining variables.
- Per regression, the regression coefficients and residual distribution are tested for equality ("invariance") across experiments.

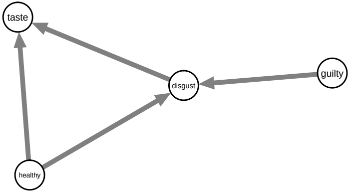


[4] Meinshausen, et al. (2016) Methods for causal inference from gene perturbation experiments and validation

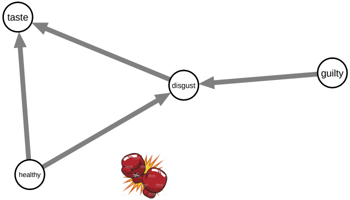


No parents	$\{\emptyset\}$
1 parent	{healthy} {disgust} {guilty}
2 parents	{healthy, disgust} {healthy, guilty} {disgust, guilty}
3 parents	{healthy, disgust, guilty}

Comparing methods



IC-algorithm

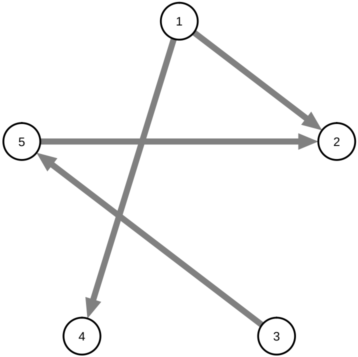


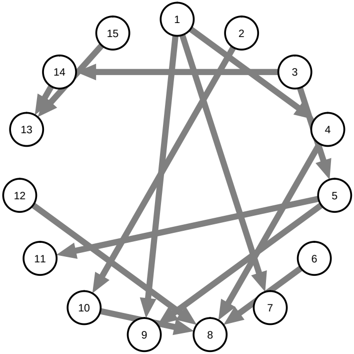
DR-FFL/TRANSWESD/ICP

Comparing methods

	Pro	Con
DR-FFL	Creates single-subject network	Causal network is always unweighted
TRANSWESD	Creates between-subject network	Computation time is long for reasonably sized networks
IC	Computationally fast	Uses solely observational data
ICP	Can combine data from different sources	Tests all possible subsets for each node

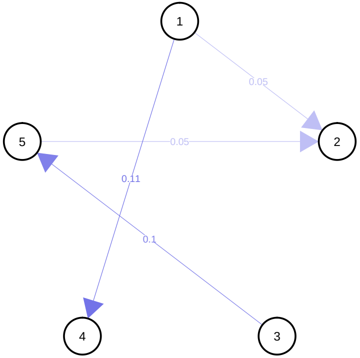
Simulation study



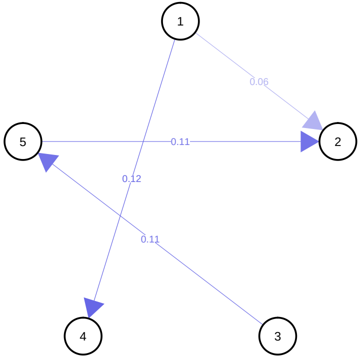


$$V = \{5, 15\}$$
$$N = \{500, 1000, 5000\}$$
$$\overline{P} = \{1, 5\}$$

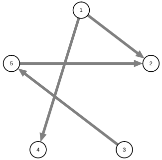
Simulation results



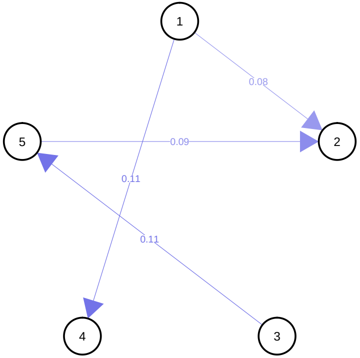
DR-FFL



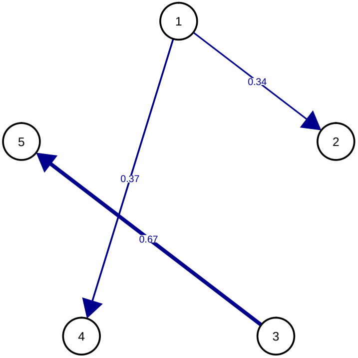
TRANSWESD



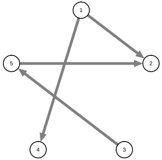
Simulation results

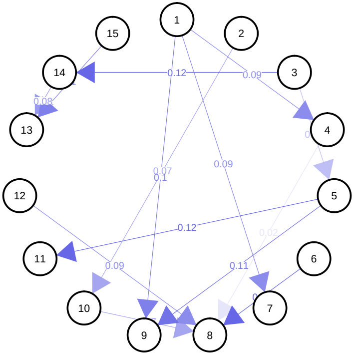


IC-algorithm

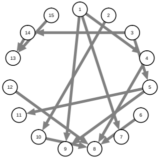


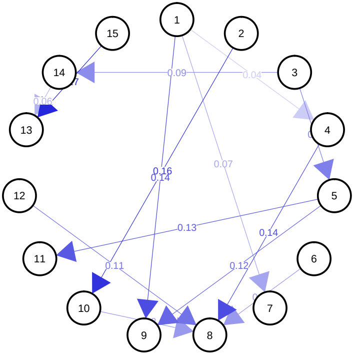
ICP





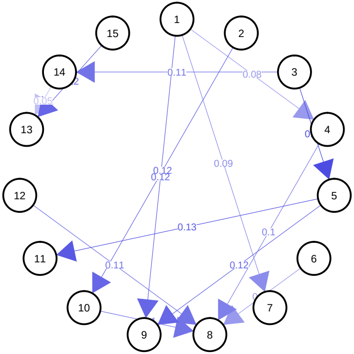
DR-FFL



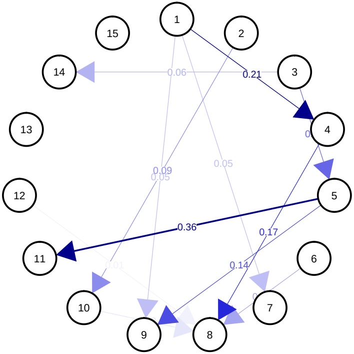


TRANSWESD

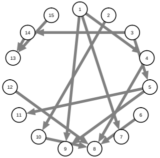
Simulation results



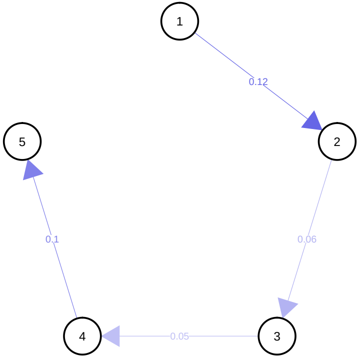
IC-algorithm



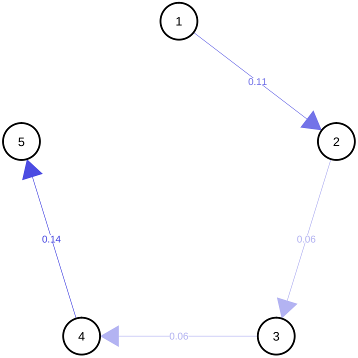
ICP



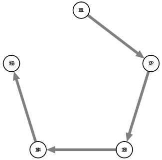
Simulation results



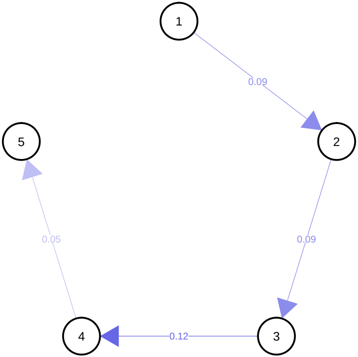
DR-FFL



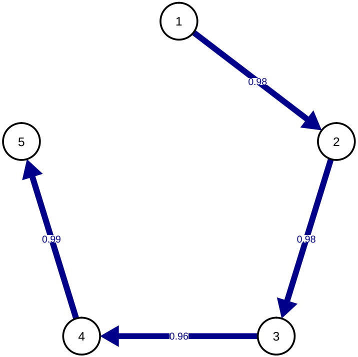
TRANSWESD



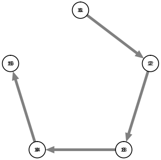
Simulation results



IC-algorithm



ICP



Conclusion

- Sensitivity was generally low in all conditions
- Sensitivity increased immensely for ICP when collider effect was removed

- Specificity was generally high in all conditions
- Higher thresholds for DR-FFL/TRANSWESD resulted in increased specificity

- IC-algorithm is "good" in finding reversed causal relations in comparison to the simulated network

The future

We are currently working on a model that combines the advantages of transitive reduction and invariant causal prediction.

Collaborators

Lourens Waldorp

Han van der Maas

Ria Hoekstra

Contact info:

`http://jolandakossakowski.eu`

`J.J.Kossakowski@uva.nl`