Markov Random Fields

Network Analysis 2014

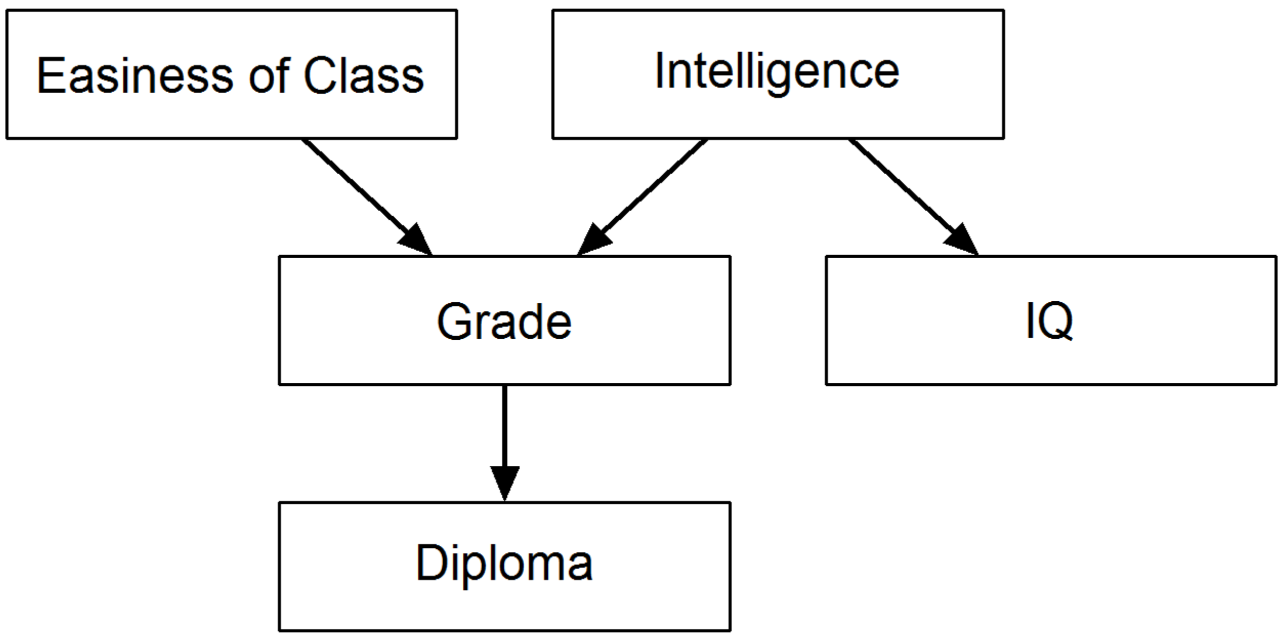
Sacha Epskamp

Disclaimer

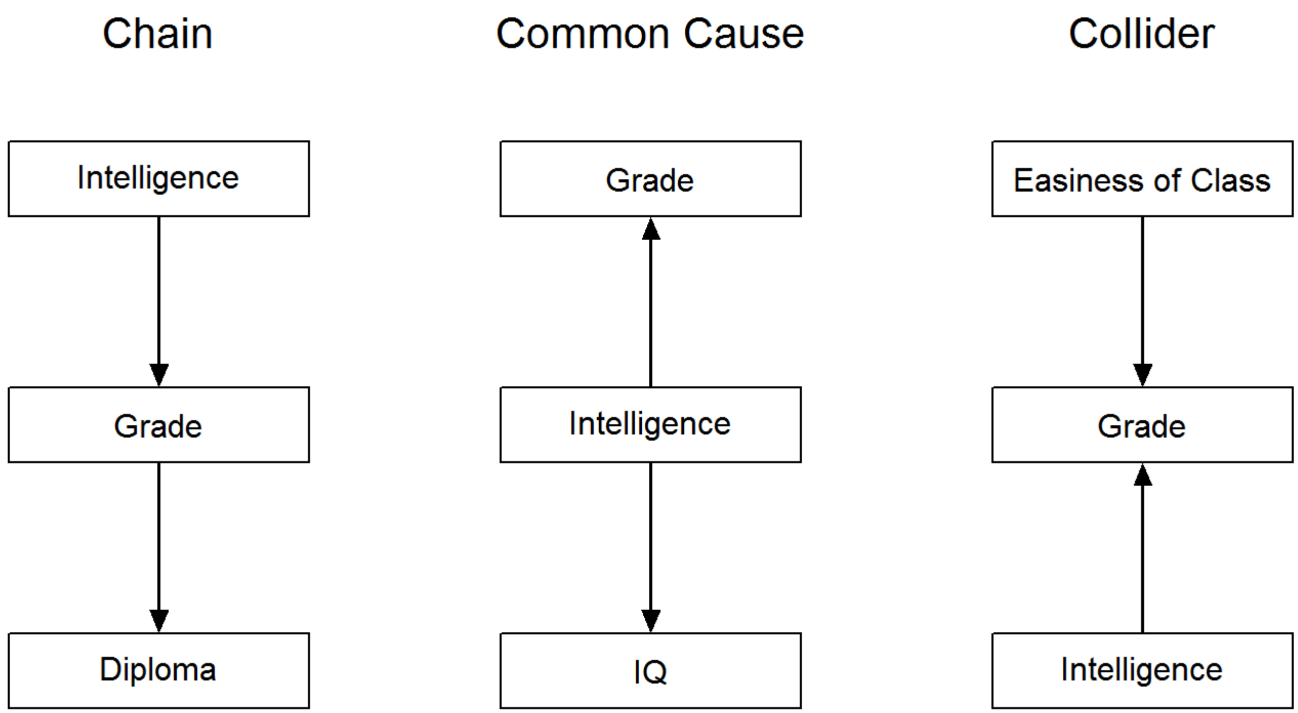
- These sheets contain mathematics to illustrate properties of Markov Random Fields (undirected networks)
  - As well as for reference
- Without mathematical statistics some concepts might be hard to follow
  - And that is ok!
  - I will not ask you to repeat the mathematics The main thing you should take from this lecture is:
  - Understand what Markov Random Fields are
  - Understand what independence relations they imply
  - Understand how the edge weights should be interpreted
  - Understand how they are estimated
    - As we will see on Thursday, the R codes are simple!!!
- Be there on Thursday!

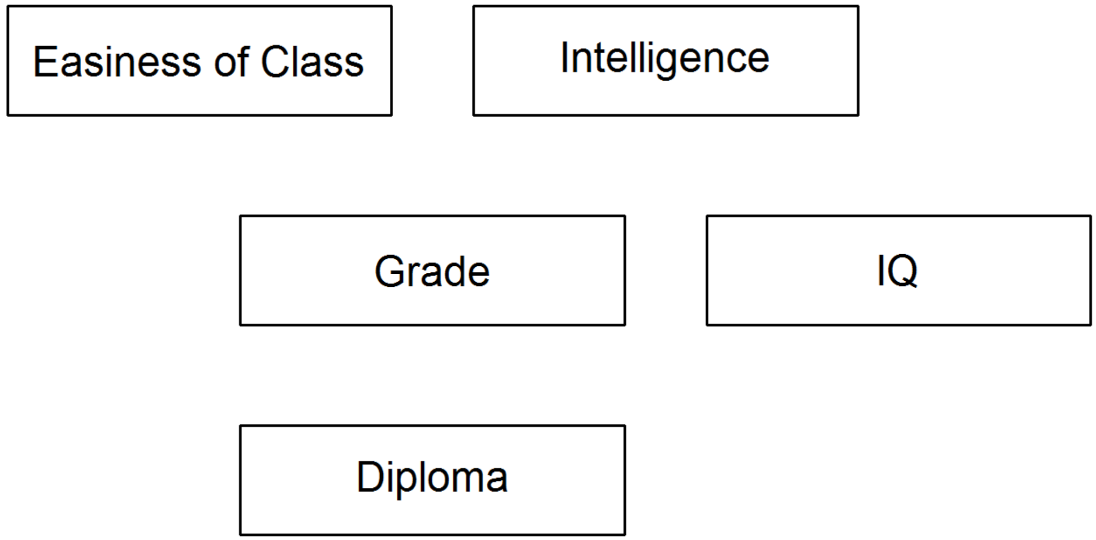
Recap

Causality

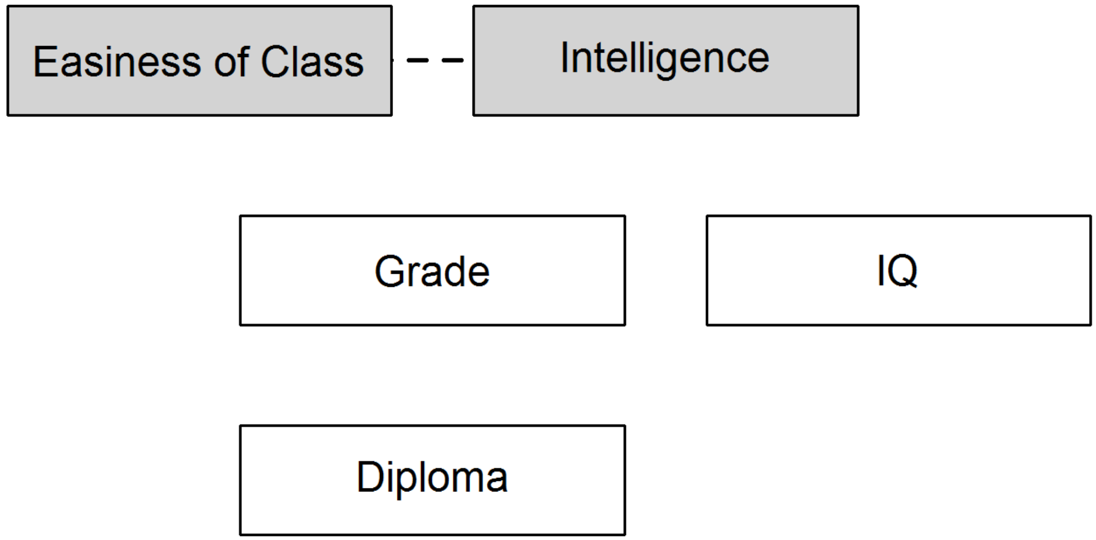


Causal structures



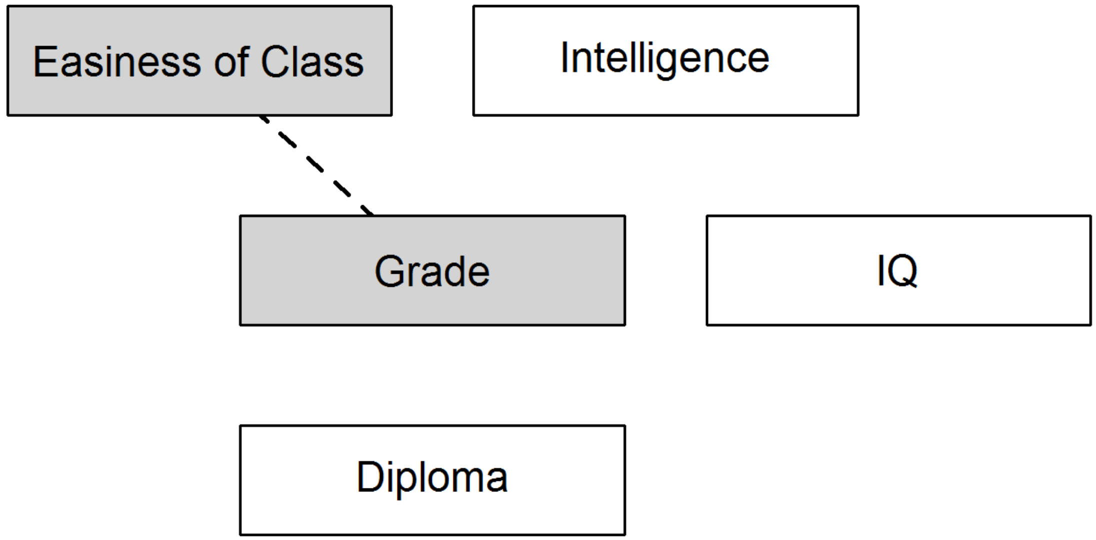


What if we don't know the structure?



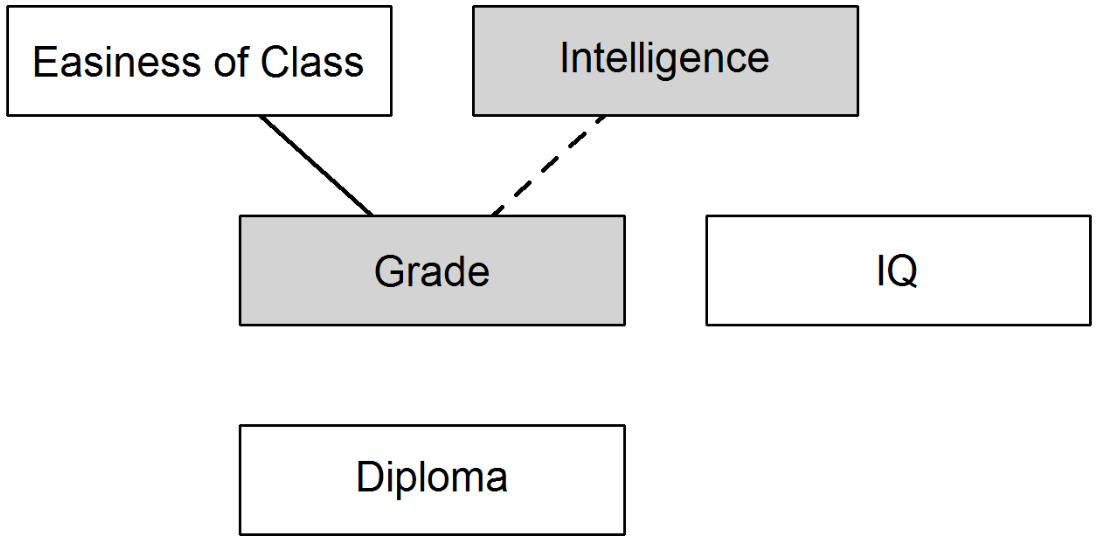
Are the two nodes independent given *any* set of other nodes (including the empty set)?

- Yes! They are independent to begin with!
- Draw no edge between Easiness of Class and Intelligence



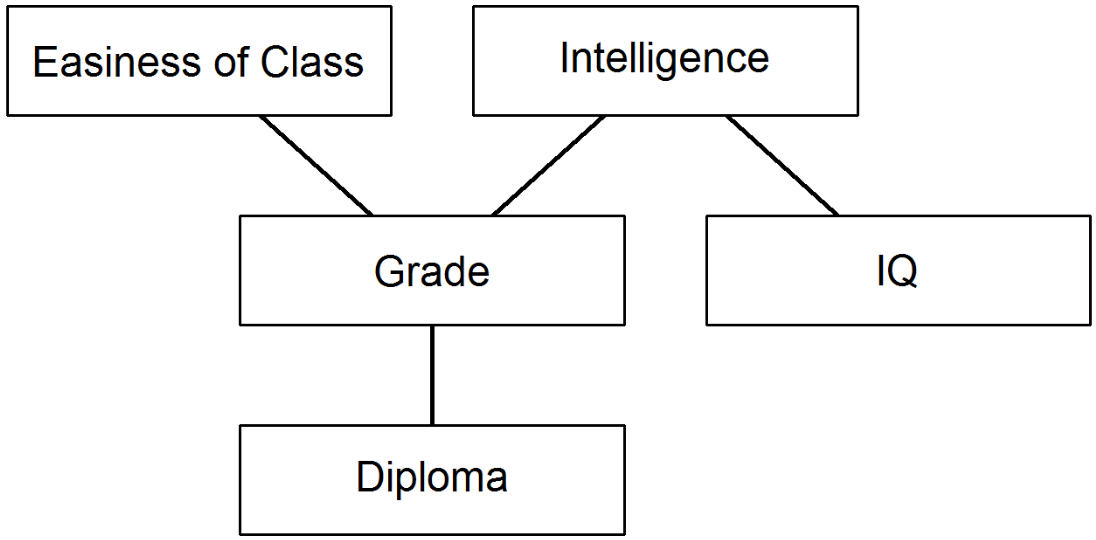
Are the two nodes independent given *any* set of other nodes (including the empty set)?

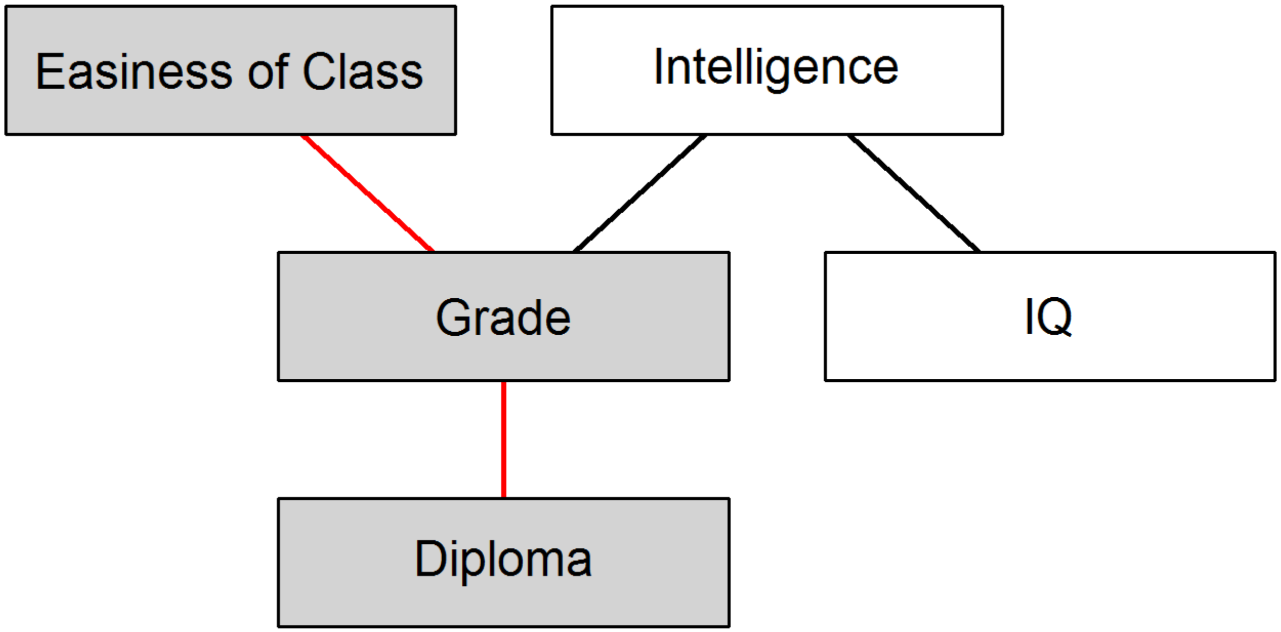
- No!
- Draw an edge between Easiness of Class and Grade



Are the two nodes independent given *any* set of other nodes (including the empty set)?

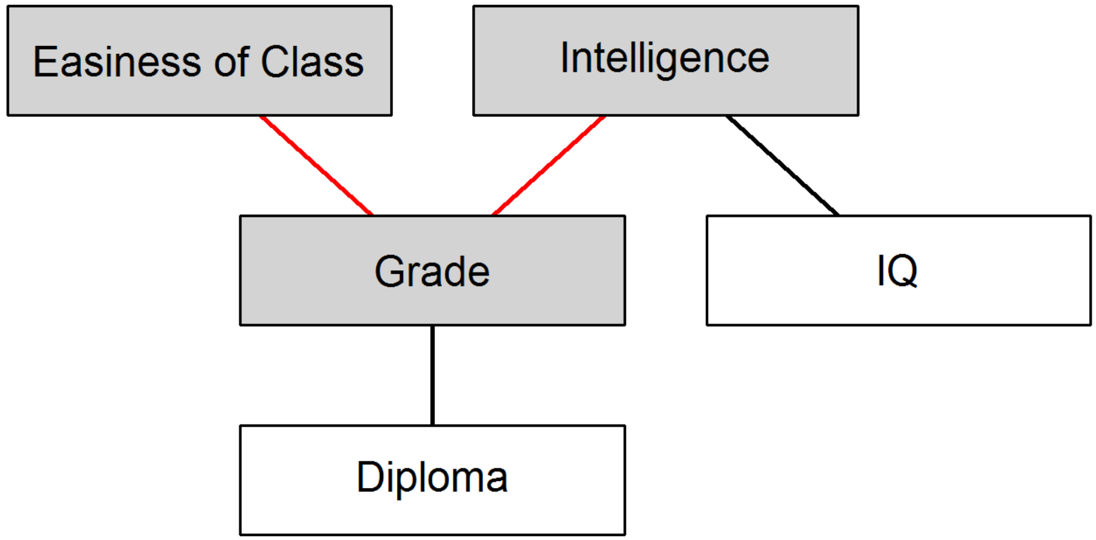
- No!
- Draw an edge between Grade and Intelligence


Easiness of Class
Intelligence
Grade
IQ
Diploma



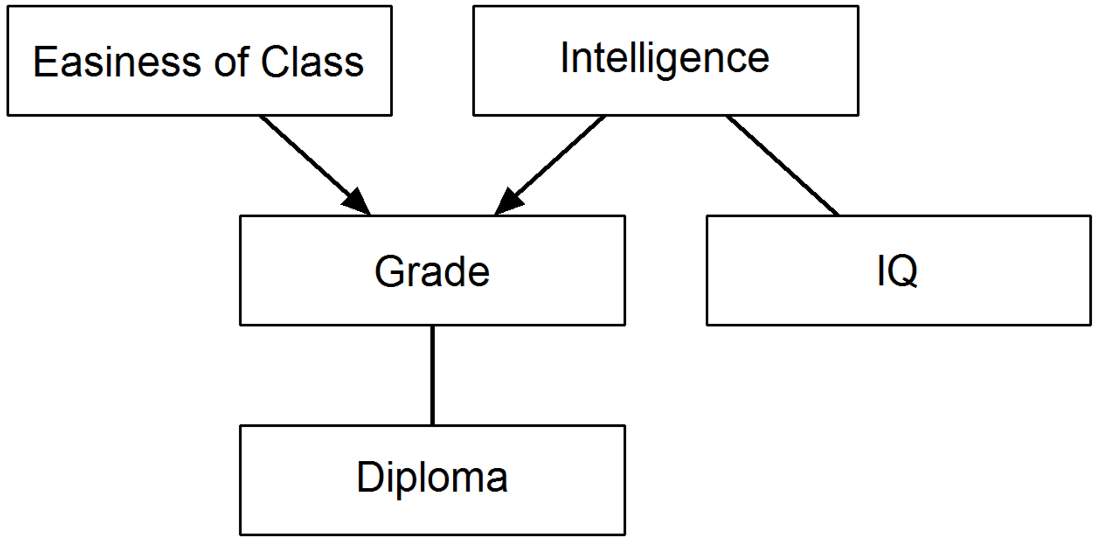
Is the middle node in the set that separated the other two nodes?

- Yes!
- No nothing

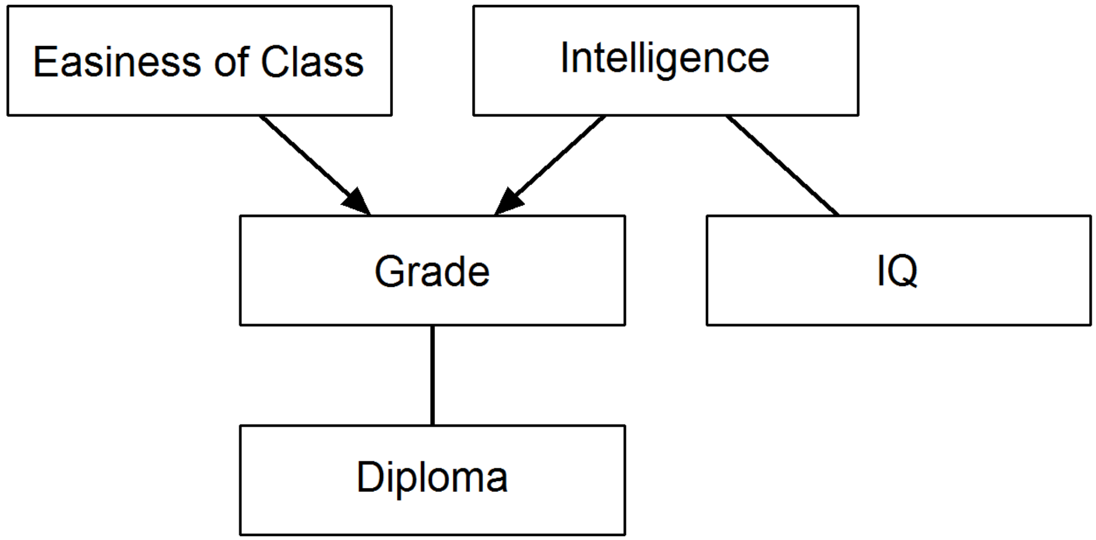


Is the middle node in the set that separated the other two nodes?

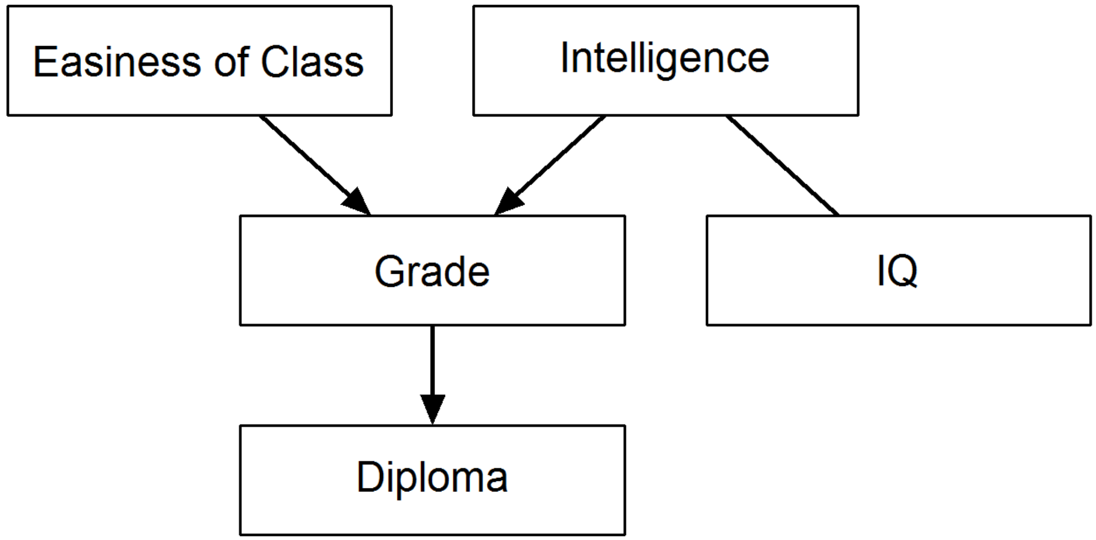
- No!
- Grade is a collider between Easiness of Class and Intelligence



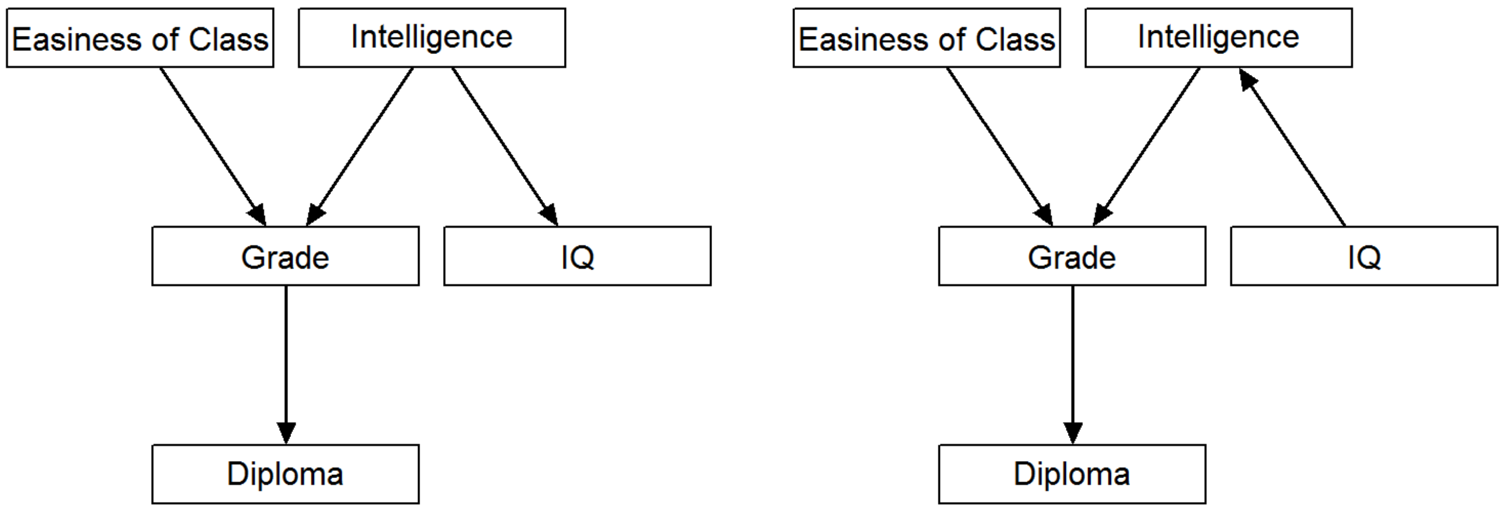
- Do we now know the direction of the edge between Intelligence and IQ?
  - No!



- Do we now know the direction of the edge between Grade and Diploma?
    - Yes! Grade was not a collider!


Easiness of Class
Intelligence
Grade
IQ
Diploma

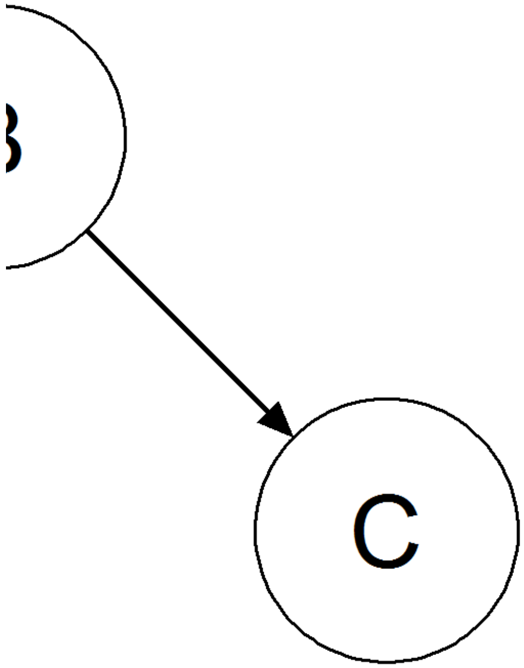
Equivalent models



Outline

- Testing Directed Networks
    - Structural Equation Modeling
    - Problems with directed networks
- Undirected networks / Markov Random Fields
- Undirected Networks as generating structures
- Important Markov Random Fields
    - Ising Model
    - Gaussian Random Field
- Examples
    - Radicalization
    - Quality of Life

Testing Directed Networks


C

Determinants of Radicalization of Islamic Youth in the Netherlands: Personal Uncertainty, Perceived Injustice, and Perceived Group Threat

Bertjan Doosje*
University of Amsterdam

Annemarie Loseman and Kees van den Bos
Utrecht University

In this study among Dutch Muslim youth (N = 131), we focus on the process of radicalization. We hypothesize that this process is driven by three main factors: (a) personal uncertainty, (b) perceived injustice, and (c) perceived group threat. Using structural equation modeling, we demonstrate that personal uncertainty, perceived injustice, and group-threat factors are important determinants of a radical belief system (e.g., perceived superiority of Muslims, perceived illegitimacy of Dutch authorities, perceived distance to others, and a feeling of being disconnected from society). This radical belief system in turn predicts attitudes toward violence by other Muslims, which is a determinant of own violent intentions. Results are discussed in terms of the role of individual and group-based determinants of radicalization.

Doosje, B., Loseman, A., & Bos, K. (2013). Determinants of radicalization of Islamic youth in the Netherlands: Personal uncertainty, perceived injustice, and perceived group threat. *Journal of Social Issues, 69(3)*, 586-604.

Table 1. The Means, Standard Deviations, and Inter-Correlations of All the Constructs

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Identification	4.56	0.85	–	−.19*	.08	−.25*	.42*	.07	.08	−.06	−.28*	.09	−.17	−.25*	−.04	−.07
2. Ind. Rel. Depri.	2.39	0.81		–	.49*	.36*	.23*	.50*	.21*	.50*	.25*	.12	.17	.21*	.12	.09
3. Col. Rel. Depri.	3.31	0.92			–	.11	.54*	.62*	.26*	.38*	.21	.31*	.18*	.09	.20*	.10
4. Int. Anxiety	−0.20	0.17				–	.01	.15	.19*	.21*	.35*	.08	.22*	.26*	.18*	.14
5. Symbolic Threat	3.46	0.76					–	.64*	.21*	.24*	.07	.39*	.01	.04	.17	−.01
6. Realistic Threat	3.10	0.88						–	.27*	.34*	.16	.35*	.19*	.14	.26*	.16
7. Per. Em. Uncertain.	2.84	0.67							–	.10	.08	.29*	.18	.00	.30*	.14
8. Perc. Proc. Injustice	2.38	0.68								–	.15	.01	.03	.23*	.04	.06
9. Perc. Illegitimacy	2.37	0.02									–	.22*	.17*	.35*	.35*	.24*
10. Perc. Ingr. Super.	3.26	0.93										–	.34*	.08	.53*	.30*
11. Distance	2.32	0.66											–	.08	.44*	.39*
12. Disconnected	2.79	0.96												–	.24*	.00
13. Moslim Violence	2.89	1.06													–	.47*
14. Violent Intentions	2.08	0.91														–

Note. 2 = *Individual Relative Deprivation*, 3 = *Collective Relative Deprivation*, 4 = *Intergroup Anxiety*, 5 = *Symbolic Threat*, 6 = *Realistic Threat*, 7 = *Personal Emotional Uncertainty*, 8 = *Perceived Procedural Injustice*, 9 = *Perceived Illegitimacy*, 10 = *Perceived Ingroup Superiority*. $*p < .05$.

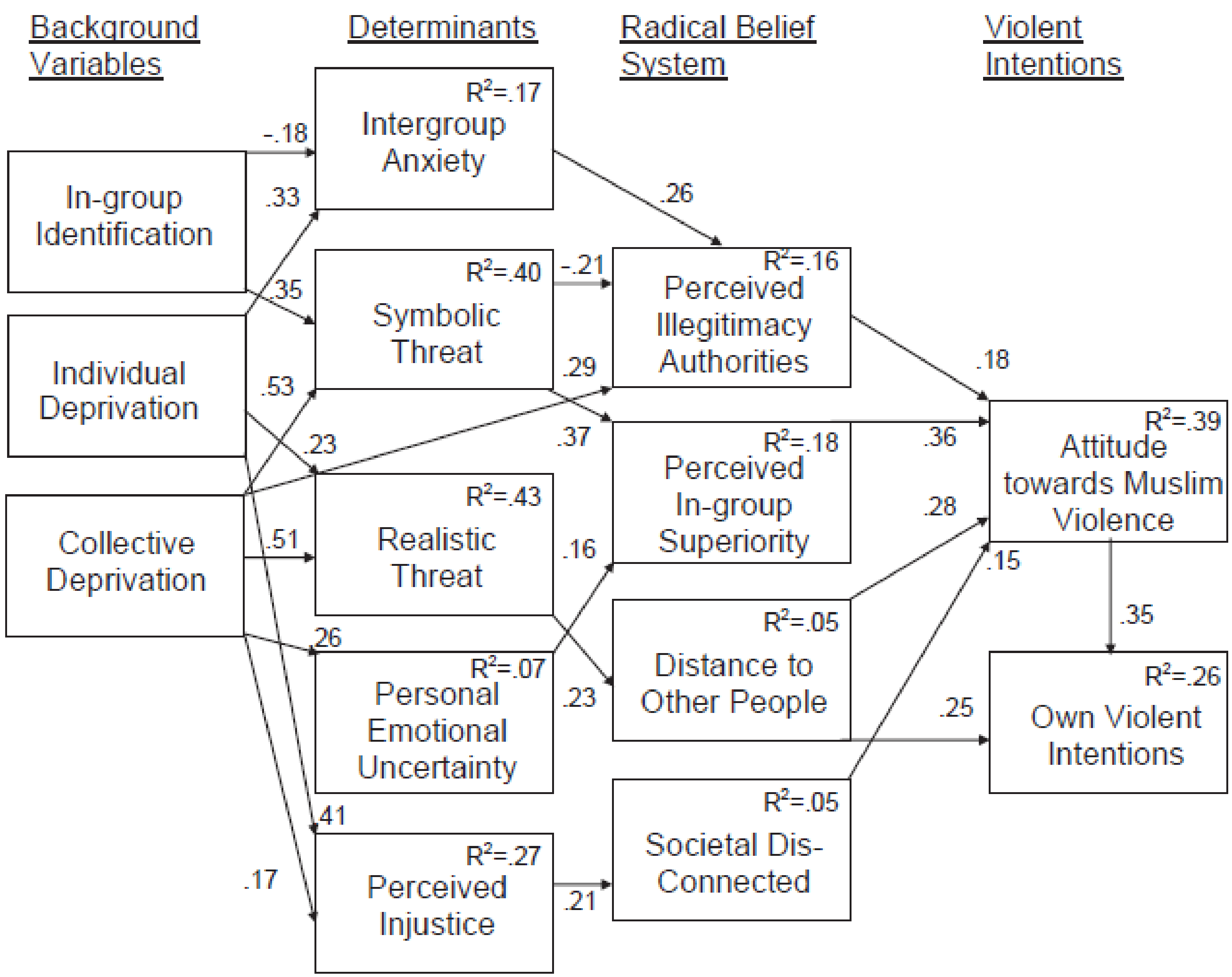
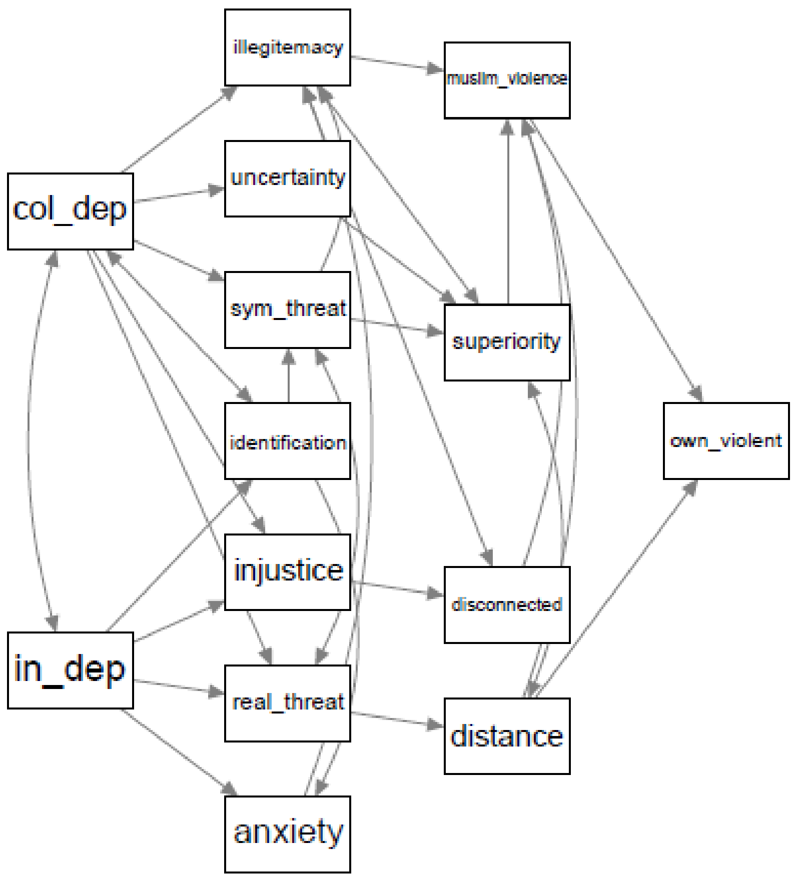


Fig. 1. Final structural equation model. All paths are significant. R^2 = % variance explained.

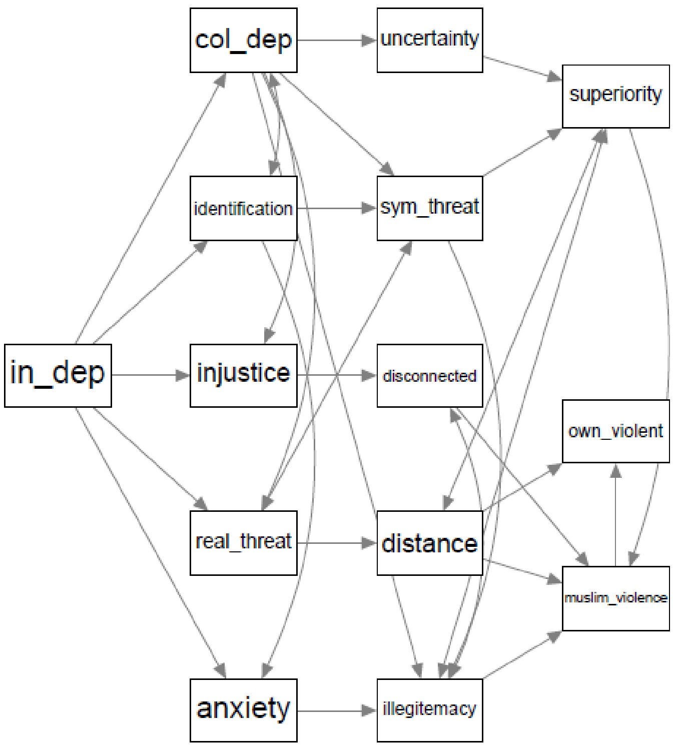
The hypothesized model had a good fit: Chi-square (65) = 76.58, p = .154, CFI = .98, NFI = .87, GFI = .93, SRMR = .082, and RMSEA = .037. La Grange Multiplier Test suggested including two direct paths: from collective deprivation to perceived illegitimacy of Dutch authorities, and from perceived distance to own violent intentions. When we included these paths, the fit became better. Our final model is presented in Figure 1. It has a very good fit: Chi-square (62) = 58.13, p = .650, CFI = 1.00, NFI = .90, GFI = .94, SRMR = .070, and RMSEA = .000. All paths included in the model are significant. We discuss this model in steps from left to right.

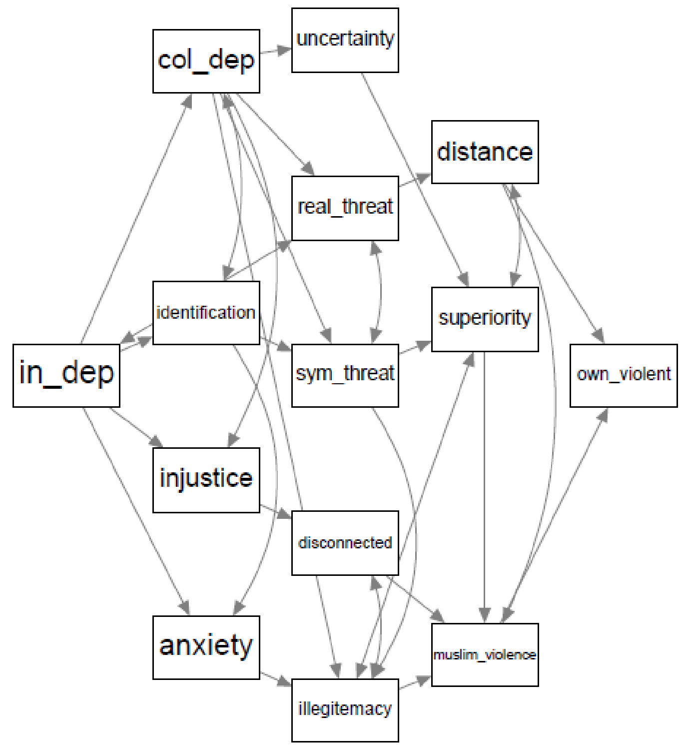
Equivalent model



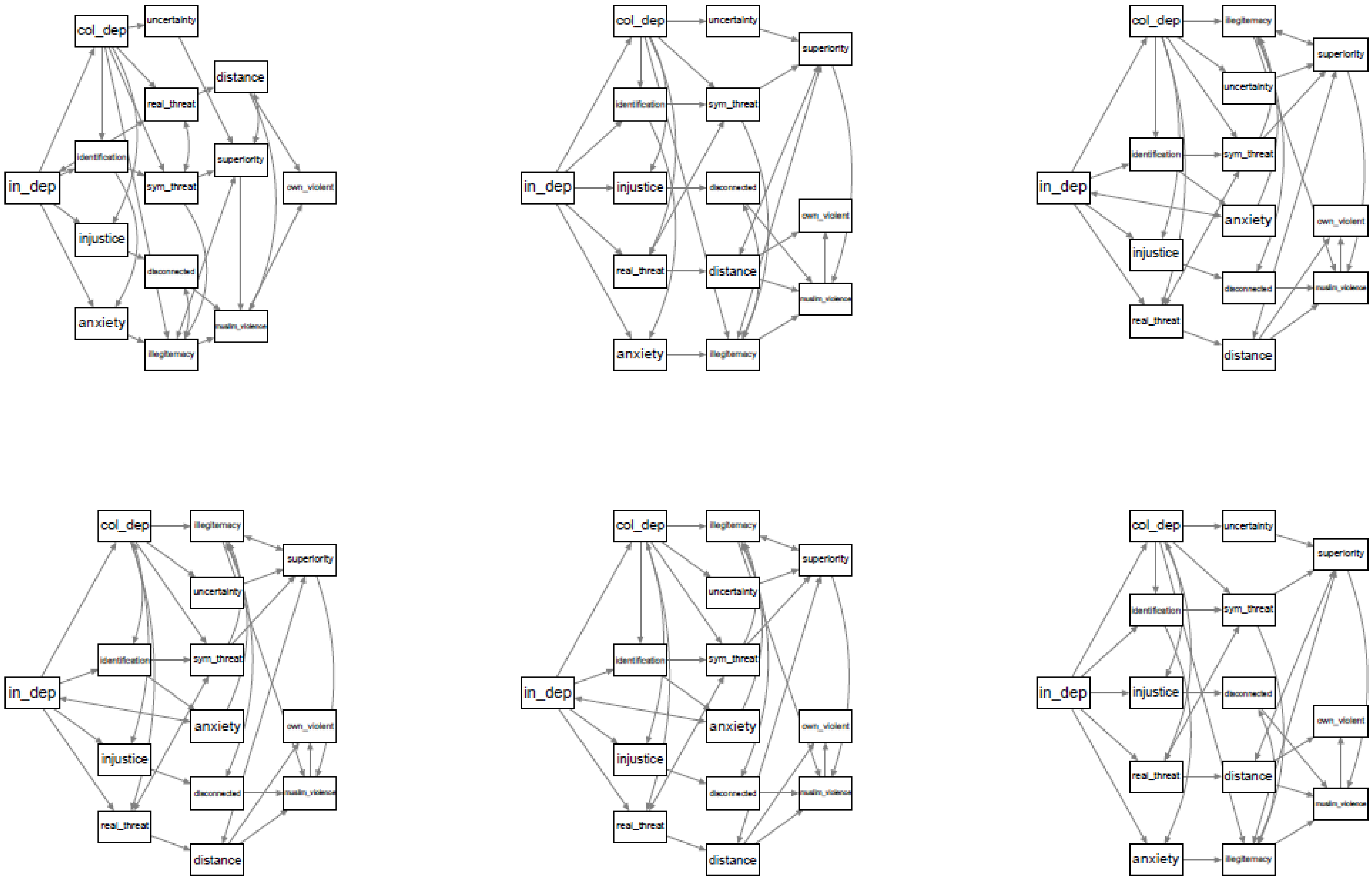
MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. Psychological bulletin, 114(1), 185.

More equivalent models

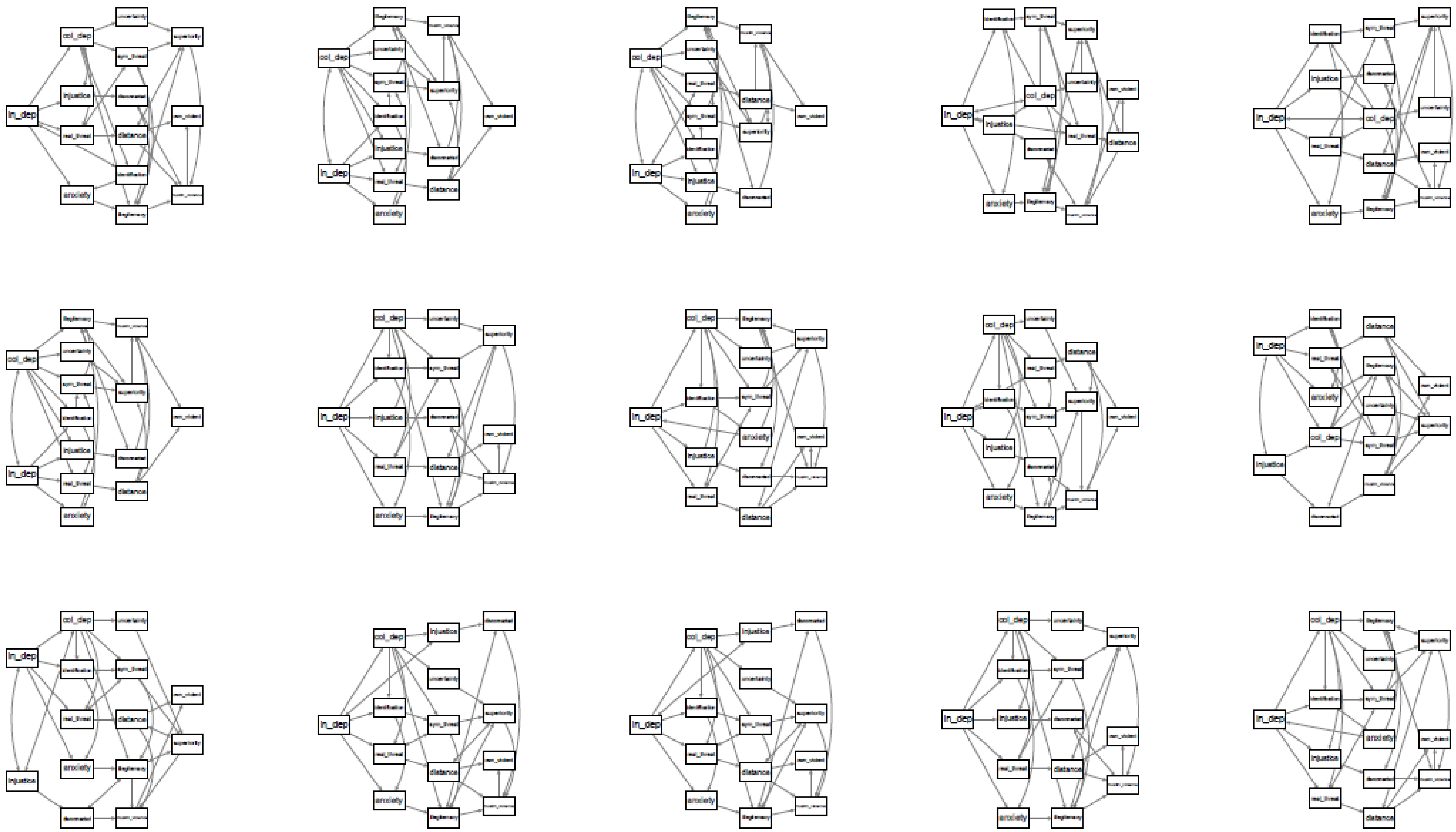




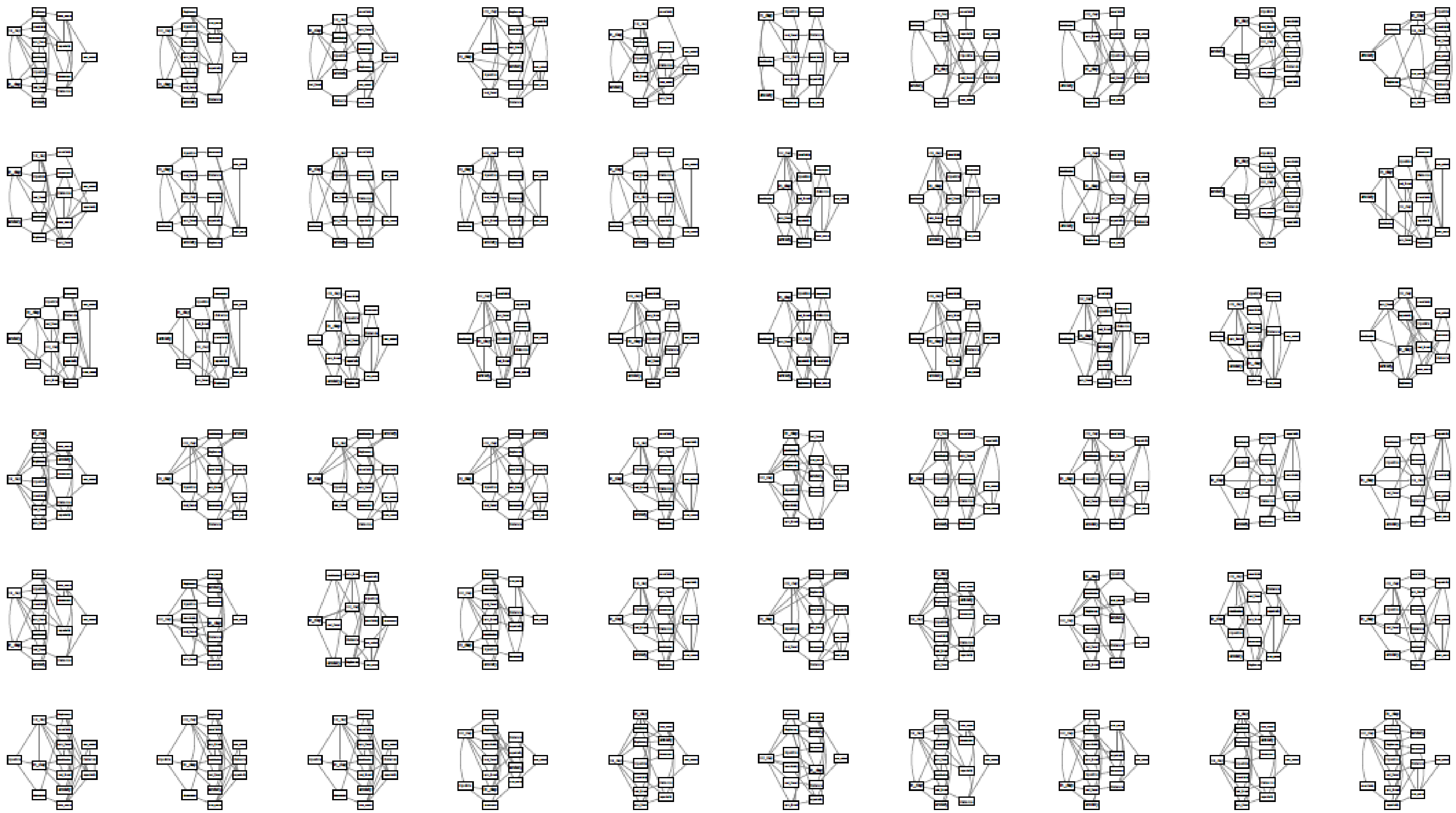
More equivalent models



More equivalent models



More equivalent models



More equivalent models

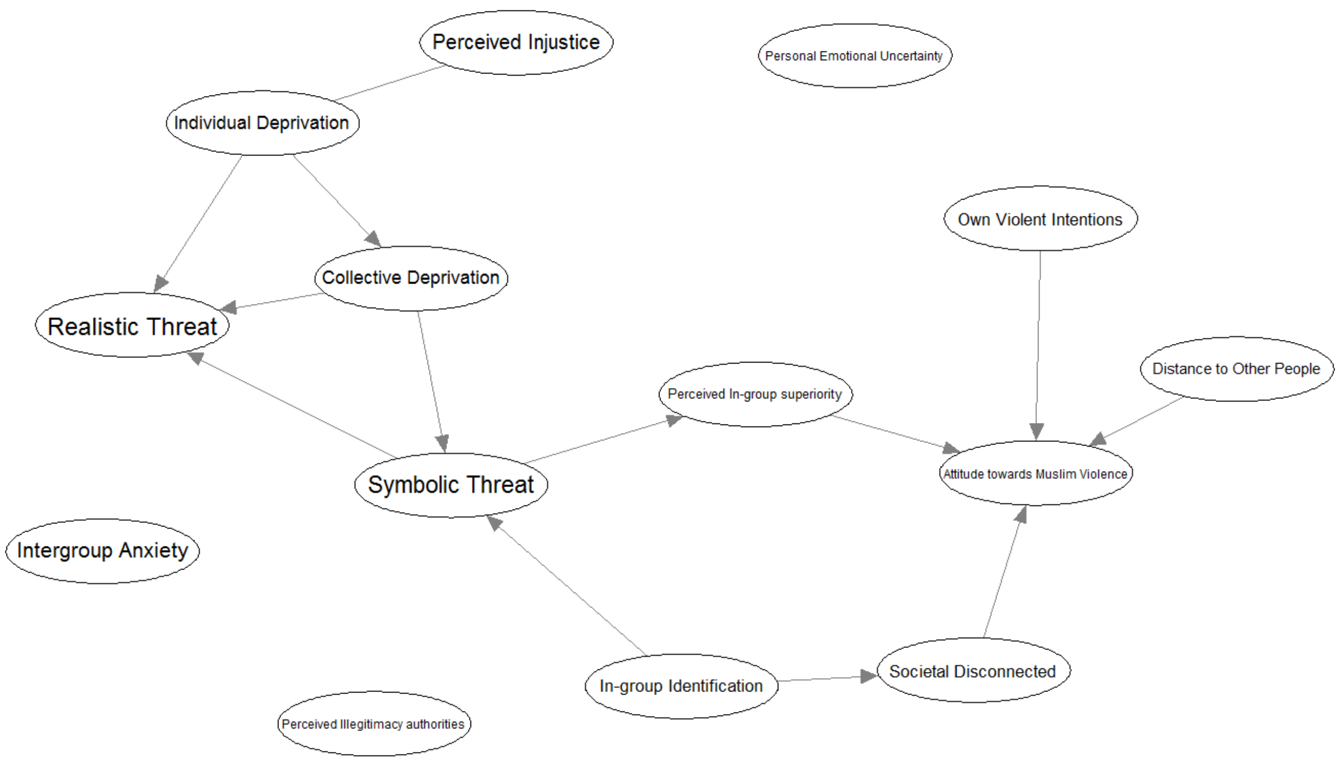
What does pcalg come up with?

```
Data <- read.csv(text = '
In-group Identification;identification; 4.56; 0.85
Individual Deprivation;in_dep; 2.39; 0.81
Collective Deprivation;col_dep; 3.31; 0.92
Intergroup Anxiety;anxiety; -0.20; 0.17
Symbolic Threat;sym_threat; 3.46; 0.76
Realistic Threat;real_threat; 3.10; 0.88
Personal Emotional Uncertainty;uncertainty; 2.84; 0.67
Perceived Injustice;injustice; 2.38; 0.68
Perceived Illegitimacy authorities;illegitemacy; 2.37; 0.02
Perceived In-group superiority;superiority; 3.26; 0.93
Distance to Other People;distance; 2.32; 0.66
Societal Disconnected;disconnected; 2.79; 0.96
Attitude towards Muslim Violence;muslim_violence; 2.89; 1.06
Own Violent Intentions;own_violent; 2.08; 0.91
', sep = ";", header = FALSE)
names(Data) <- c("name","var","mean", "sd")
```

```
library("lavaan")
corMat <- getCov('
1 -.19 .08  -.25 .42 .07 .08  -.06  -.28 .09  -.17  -.25  -.04  -.07
1 .49 .36 .23 .50 .21 .50 .25 .12 .17 .21 .12 .09
1 .11 .54 .62 .26 .38 .21 .31 .18 .09 .20 .10
1 .01 .15 .19 .21 .35 .08 .22 .26 .18 .14
1 .64 .21 .24 .07 .39 .01 .04 .17  -.01
1 .27 .34 .16 .35 .19 .14 .26 .16
1 .10 .08 .29 .18 .00 .30 .14
1 .15 .01 .03 .23 .04 .06
1 .22 .17 .35 .35 .24
1 .34 .08 .53 .30
1 .08 .44 .39
1 .24 .00
1 .47
1
',lower=FALSE)
covmat <- cor2cov(corMat, Data$sd)
colnames(covmat) <- rownames(covmat) <- Data$var
```

```
library("pcalg")
pc.fit <- pc(suffStat = list(C = covmat, n = 131),
             indepTest = gaussCItest,
             alpha=0.05, labels = as.character(Data$name))
```

```
library("qgraph")
qgraph(pc.fit, vsize = 15, vsize2 = 5, shape = "ellipse", repulsion = 0.7)
```



Does it fit?

```
Mod <-'
muslim_violence ~ own_violent + distance + disconnected + superiority
disconnected ~ identification
superiority ~ sym_threat
sym_threat ~ col_dep + identification
col_dep ~ in_dep
real_threat ~ in_dep + sym_threat
in_dep ~~ injustice
'


fit <- sem(Mod, sample.cov = covmat, sample.nobs = 131)
```

Does it fit?

```
round(fitMeasures(fit)[c('chisq','df', 'pvalue','cfi','nfi','rmsea','rmsea.ci.lower','rmsea.ci.upper')],
 2)
```

```
##          chisq             df         pvalue            cfi            nfi
##          80.52          39.00           0.00           0.89           0.81
##          rmsea rmsea.ci.lower rmsea.ci.upper
##           0.09           0.06           0.12
```

- Not really…

Directed Networks

- Modifying a causal network until it fits can lead to over-fitting
  - Many equivalent models
  - If your goal is to test a specific hypothesis (as was the goal of Doosje et al.) then SEM is good. For more exploratory research, different methods should be used.
- Causal search algorithms on the other hand can lack power
  - In my experience rarely an interpretable model comes out
- Both methods assume the true model is a DAG!
  - The arrows could be misleading, they imply a specific effect of **doing** something even though we only know what happens if we **condition** on something
- Equivalent models are also a huge problem for network analysis

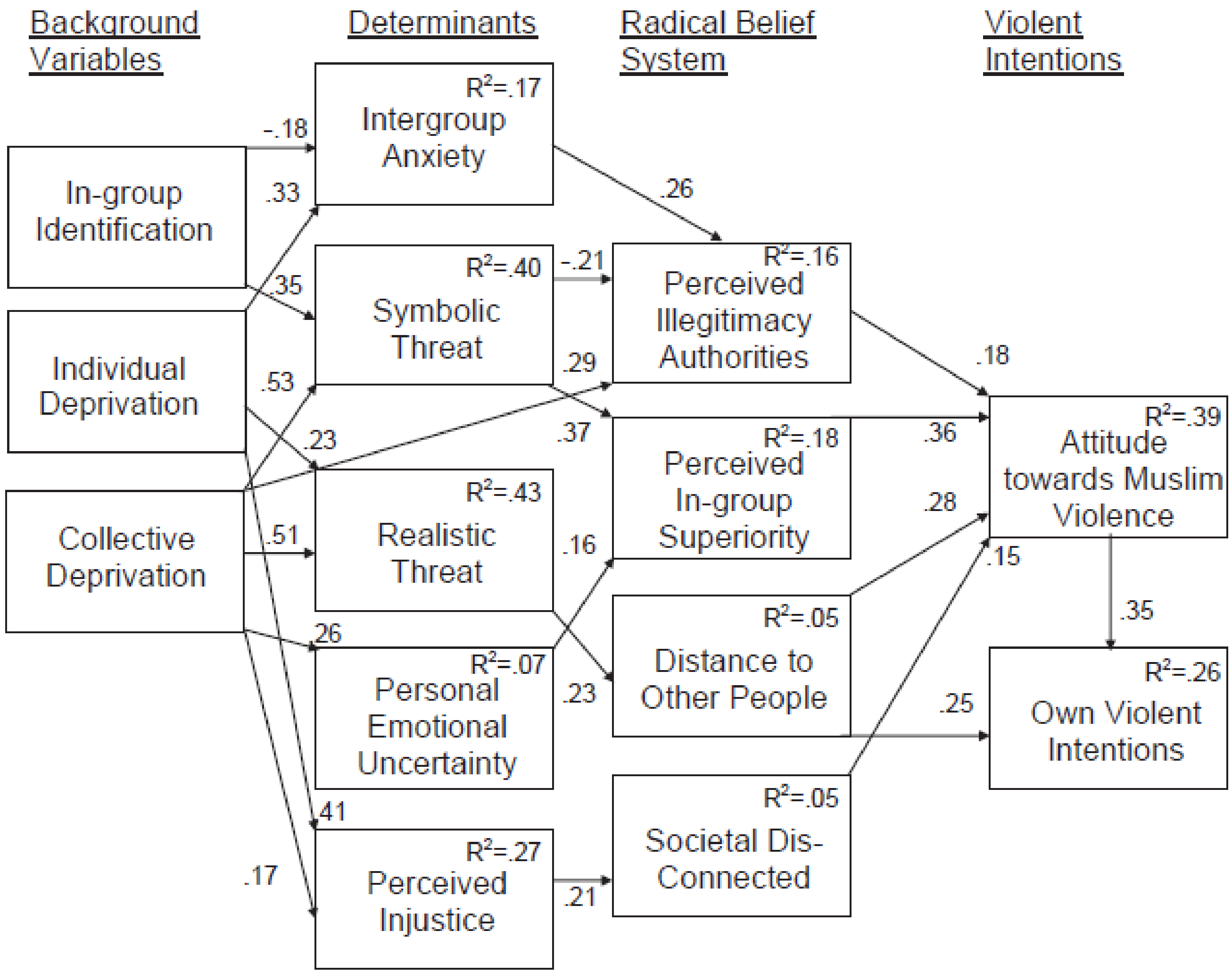
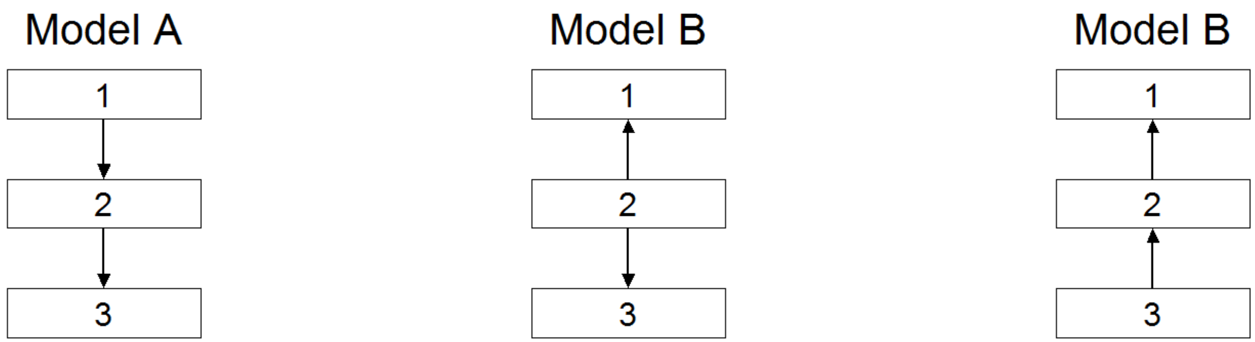
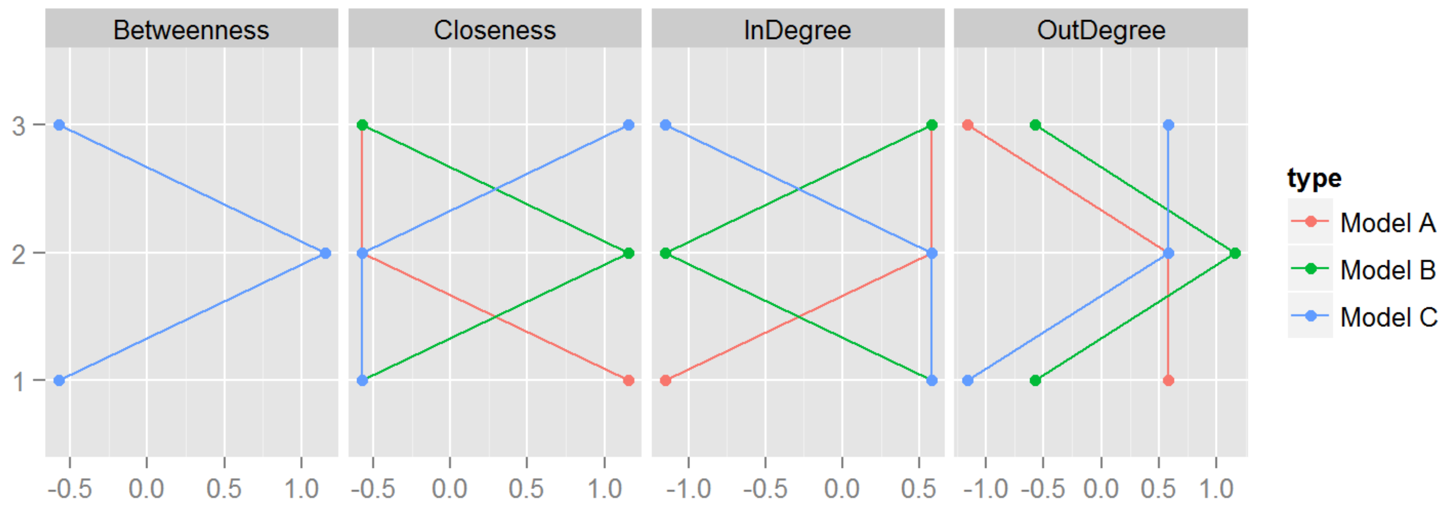


Fig. 1. Final structural equation model. All paths are significant. R^2 = % variance explained.

- Would, theoretically, Do(Attitude towards Muslim Violence) really not influence any of the predictors?

Equivalent Models and Centrality

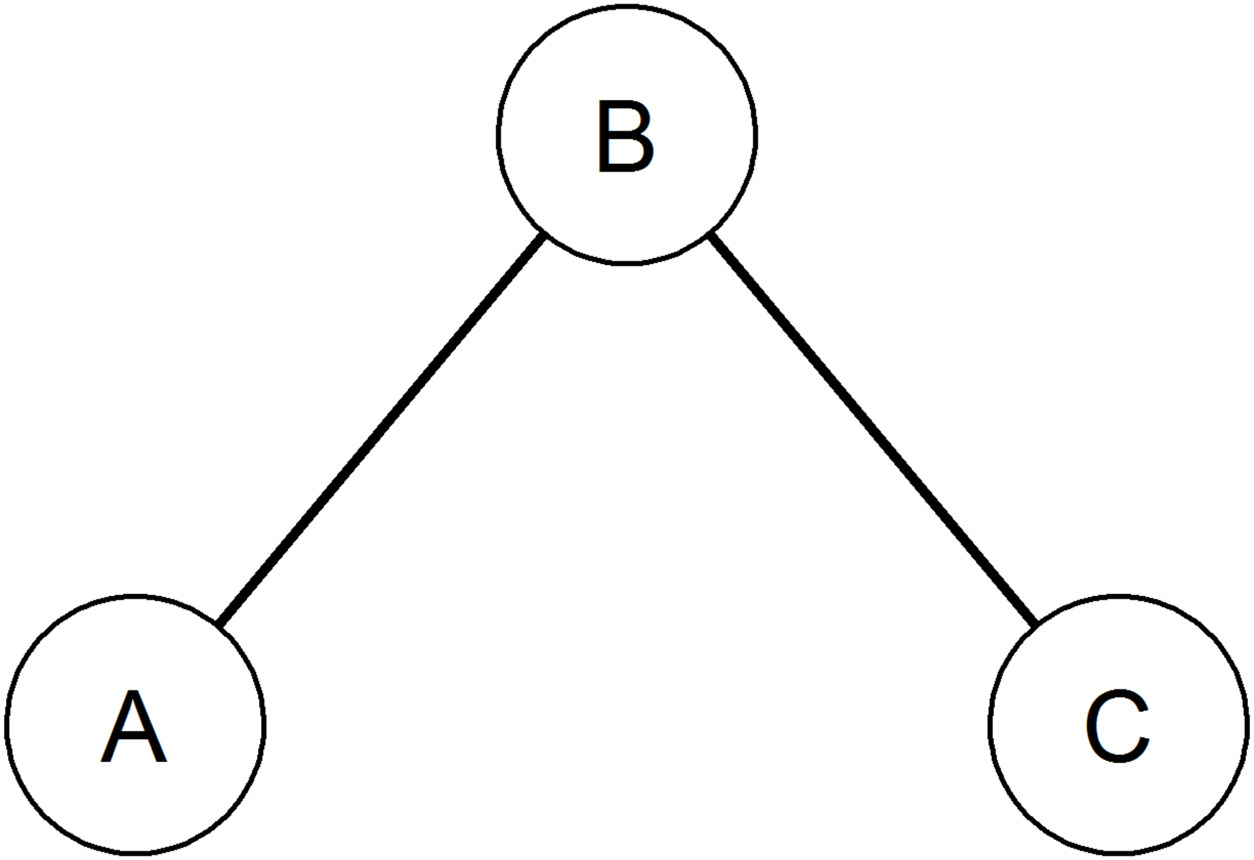




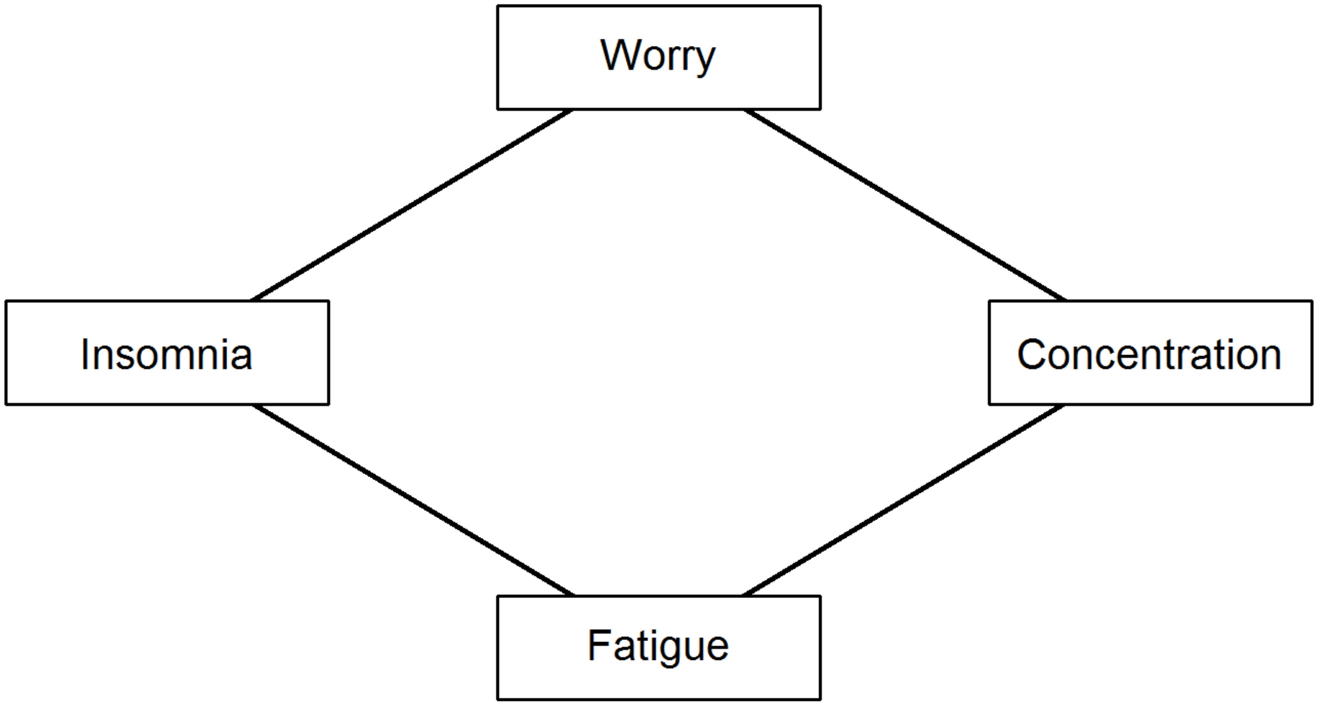
Markov Random Fields

Markov Random Fields

- A pairwise Markov Random Field (MRF) is an undirected network
- Two nodes are connected if they are not independent conditional on all other nodes.
- More importantly, two nodes are NOT connected if they are independent conditioned on all nodes:
- $X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}^{-(i,j)} = \boldsymbol{x}^{-(i,j)} \iff (i,j) \notin E$
- A node separates two nodes if it on all paths from one node to another
- No equivalent models!
  - Clear saturated model is a fully connected network
- Naturally cyclic!



- B separates A and C
- $A \perp\!\!\!\perp C \mid B$



- Worrying and fatigue separate Insomnia and Concentration

Markov Random Fields in Other Sciences

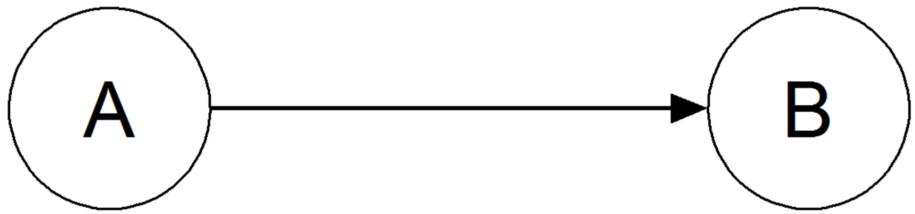
- MRF's have been used in many disciplines in Science
  - Physics
  - Machine Learning
  - Artificial Intelligence
  - Biometrics
  - Economics
  - Image processing (http://cs.stanford.edu/people/karpathy/visml/ising_example.html)
  - Neural Networks (http://nxxcxx.github.io/Neural-Network/)

Interpreting a Markov Random Field

The edges in a MRF can be interpreted in several ways:

- Predictive effects
- A representation of conditional independence relationships
- Pairwise interactions
- Genuine symmetric relationships between nodes
    - Ising Model

Predictive Effects



If this model is the generating model, does:

- A predict B?
  - Yes!
- B predict A?
  - Yes!
- A predict B just as well as B predict A?
  - Using linear or logistic regression, yes!

```r
# Generate data (binary):
A <- sample(c(0,1), 10000, replace = TRUE)
B <- 1 * (runif(10000) < ifelse(A==1, 0.8, 0.2))

# Predict A from B (logistic bregression):
AonB <- glm(A ~ B, family = "binomial")
coef(AonB)
```

```
## (Intercept)           B
##      -1.443       2.873
```

```r
# Predict B from A (logistic regression):
BonA <- glm(B ~ A, family = "binomial")
coef(BonA)
```

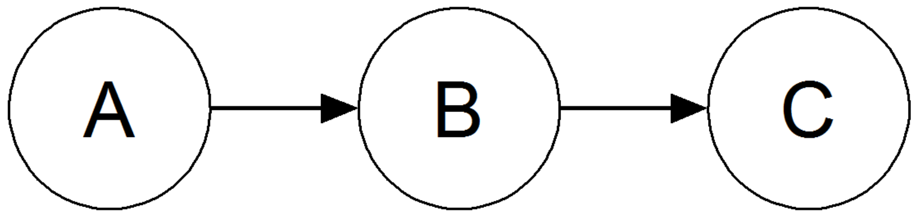
```
## (Intercept)           A
##      -1.440       2.873
```

- The logistic regression parameters are equal!

Predictive Effects

- A predicts B and B predicts A

Predictive Effects



If this model is the generating model, does:

- A predict C or vise versa?
  - Yes, they are correlated
- A predict C or vise versa when also taking B into account?
  - No!
- In a multiple (logistic) regression, C should not predict A when B is also taken as predictor

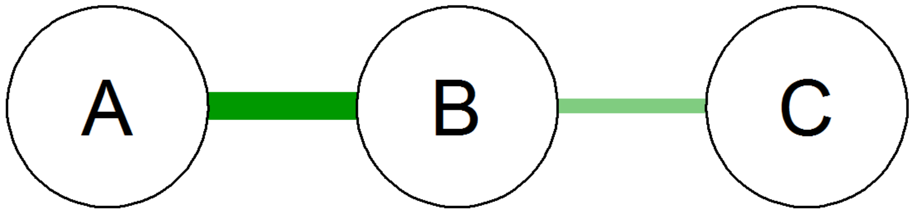
```r
# Generate data (Gaussian):
A <- rnorm(10000)
B <- A + rnorm(10000)
C <- B + 2*rnorm(10000)
```

```r
# Predict A from C:
AonC <- lm(A ~ C)
summary(AonC)
```

```
##
## Call:
## lm(formula = A ~ C)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.521 -0.617 -0.012  0.624  3.166
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01060    0.00914    1.16     0.25
## C            0.17084    0.00375   45.56   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.914 on 9998 degrees of freedom
## Multiple R-squared:  0.172,  Adjusted R-squared:  0.172
## F-statistic: 2.08e+03 on 1 and 9998 DF,  p-value: <2e-16
```

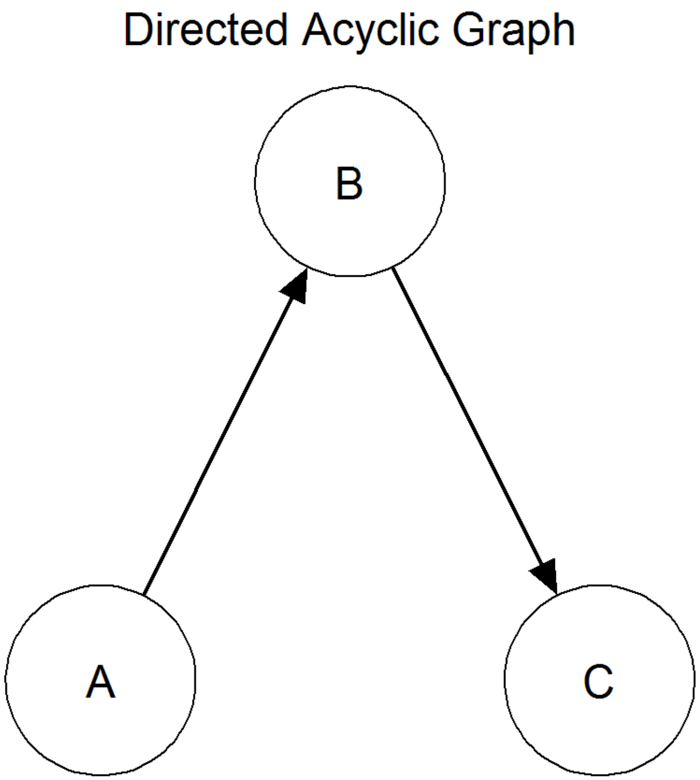
```r
# Predict A from B and C:
AonBC <- lm(A ~ B + C)
summary(AonBC)
```

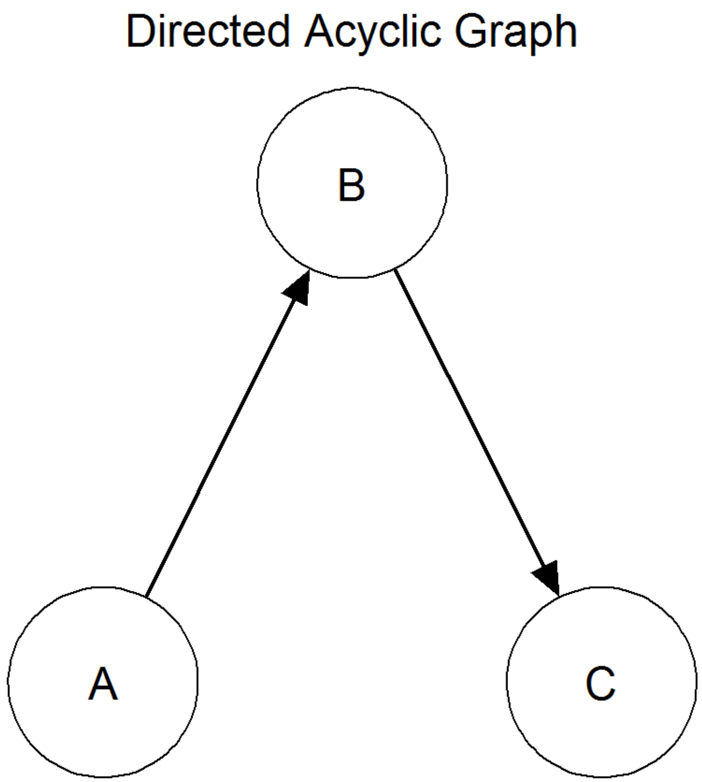
```
##
## Call:
## lm(formula = A ~ B + C)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9409 -0.4774 -0.0027  0.4785  2.9478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.00464    0.00711    0.65     0.51
## B            0.49992    0.00620   80.69   <2e-16 ***
## C            0.00417    0.00358    1.17     0.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.711 on 9997 degrees of freedom
## Multiple R-squared:  0.498,  Adjusted R-squared:  0.498
## F-statistic: 4.97e+03 on 2 and 9997 DF,  p-value: <2e-16
```

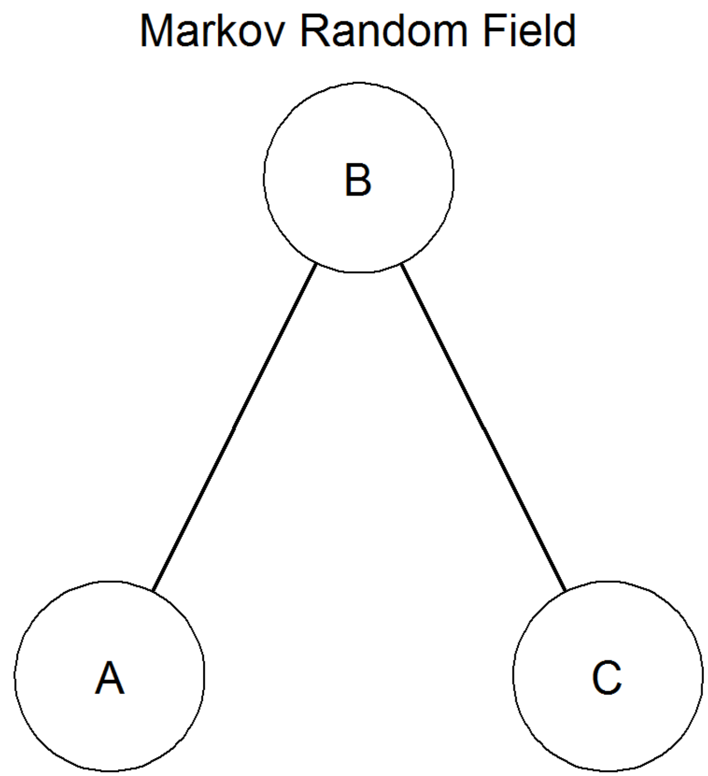


- A predicts B better than C predicts B
- The relationship between A and C is mediated by B

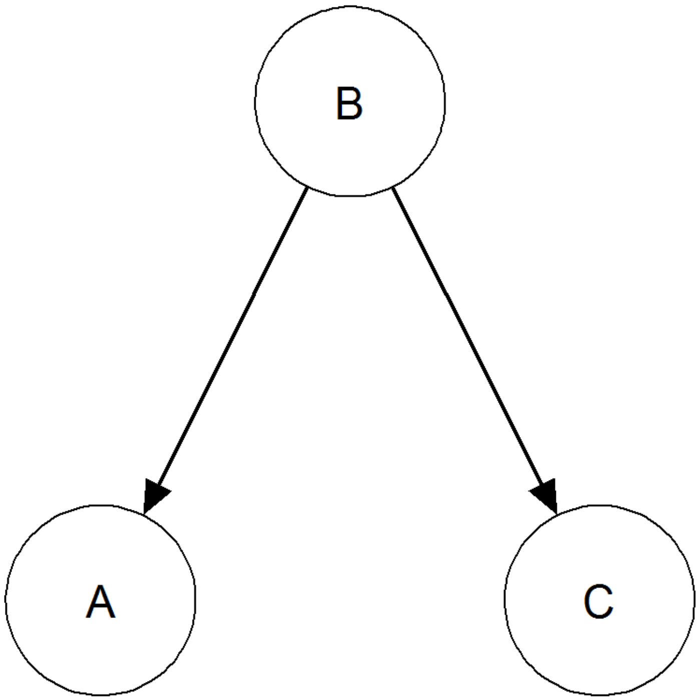
Conditional Independency Relationships



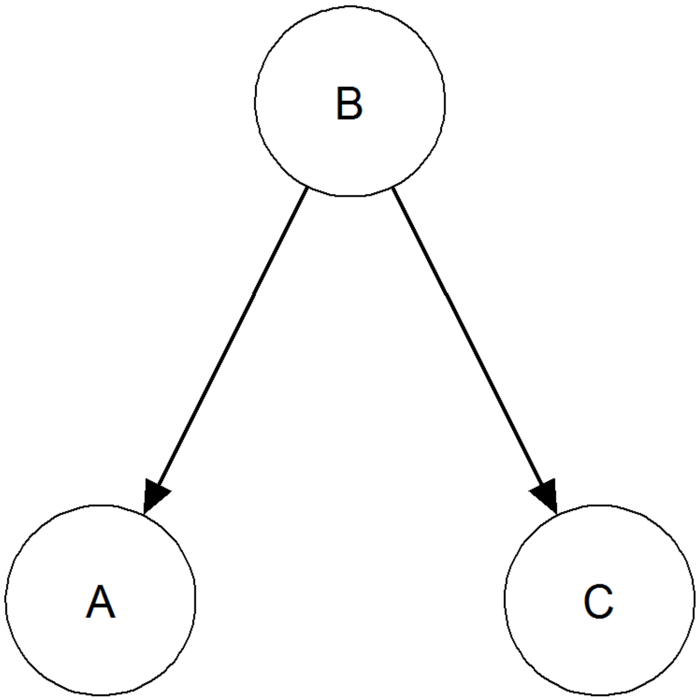

Directed Acyclic Graph
B
A
C


Markov Random Field
B
A
C

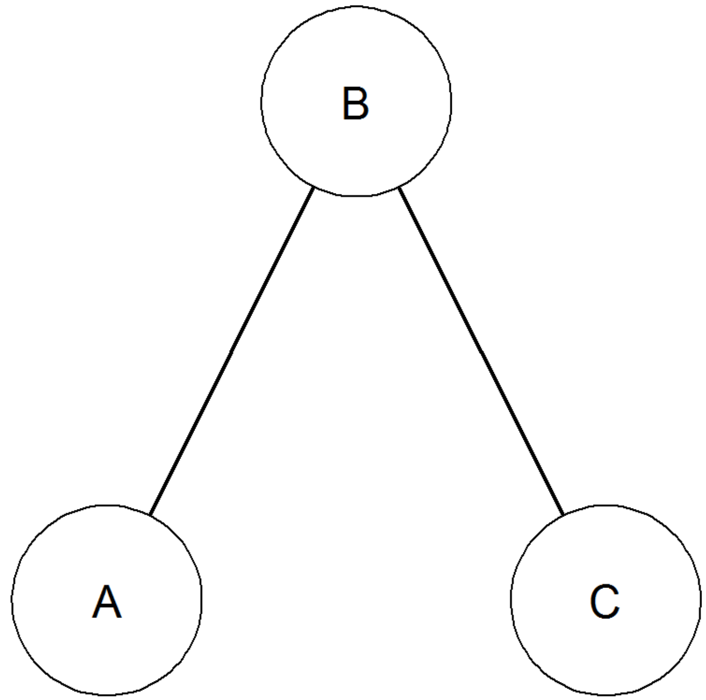
Directed Acyclic Graph



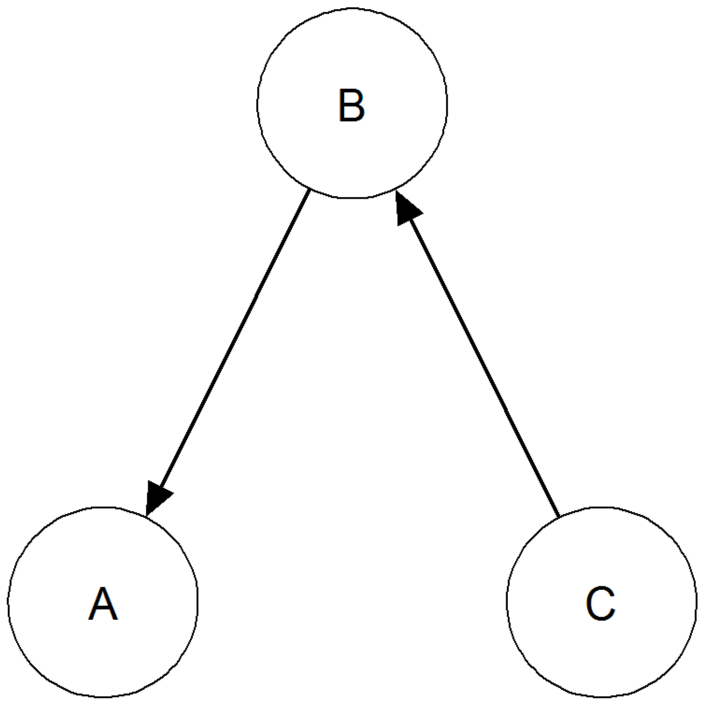
Directed Acyclic Graph



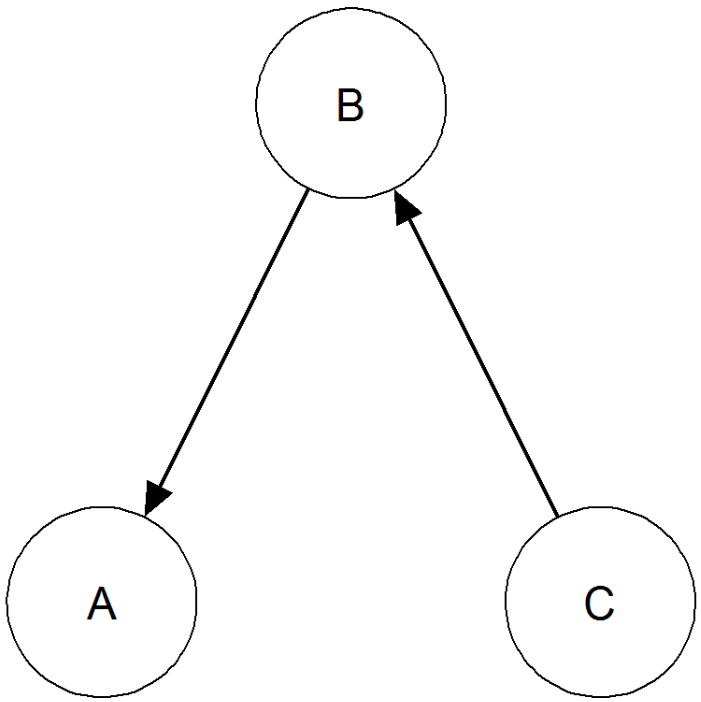
Markov Random Field



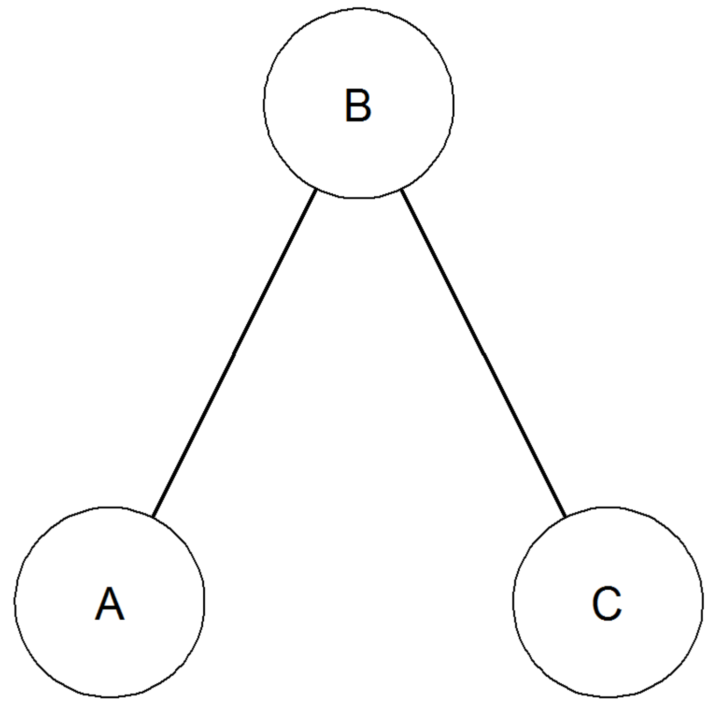
Directed Acyclic Graph



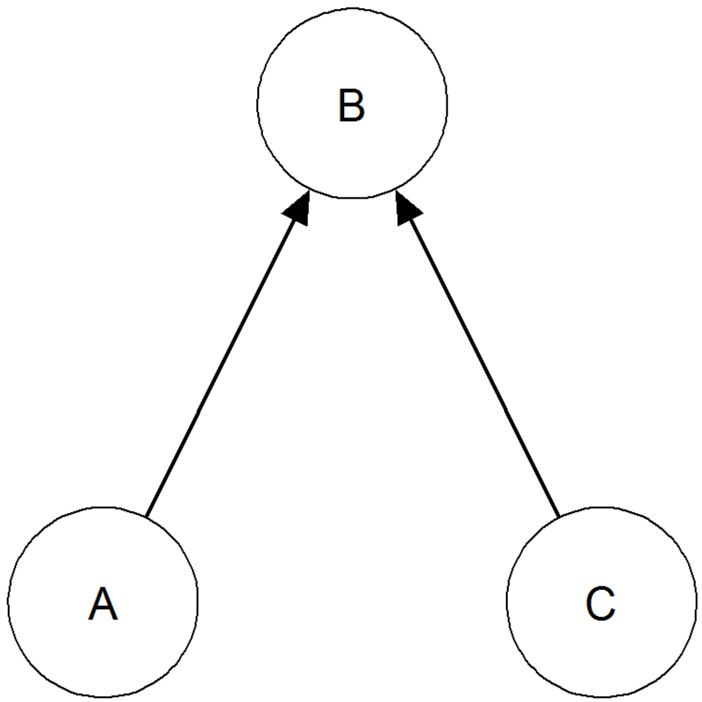
Directed Acyclic Graph

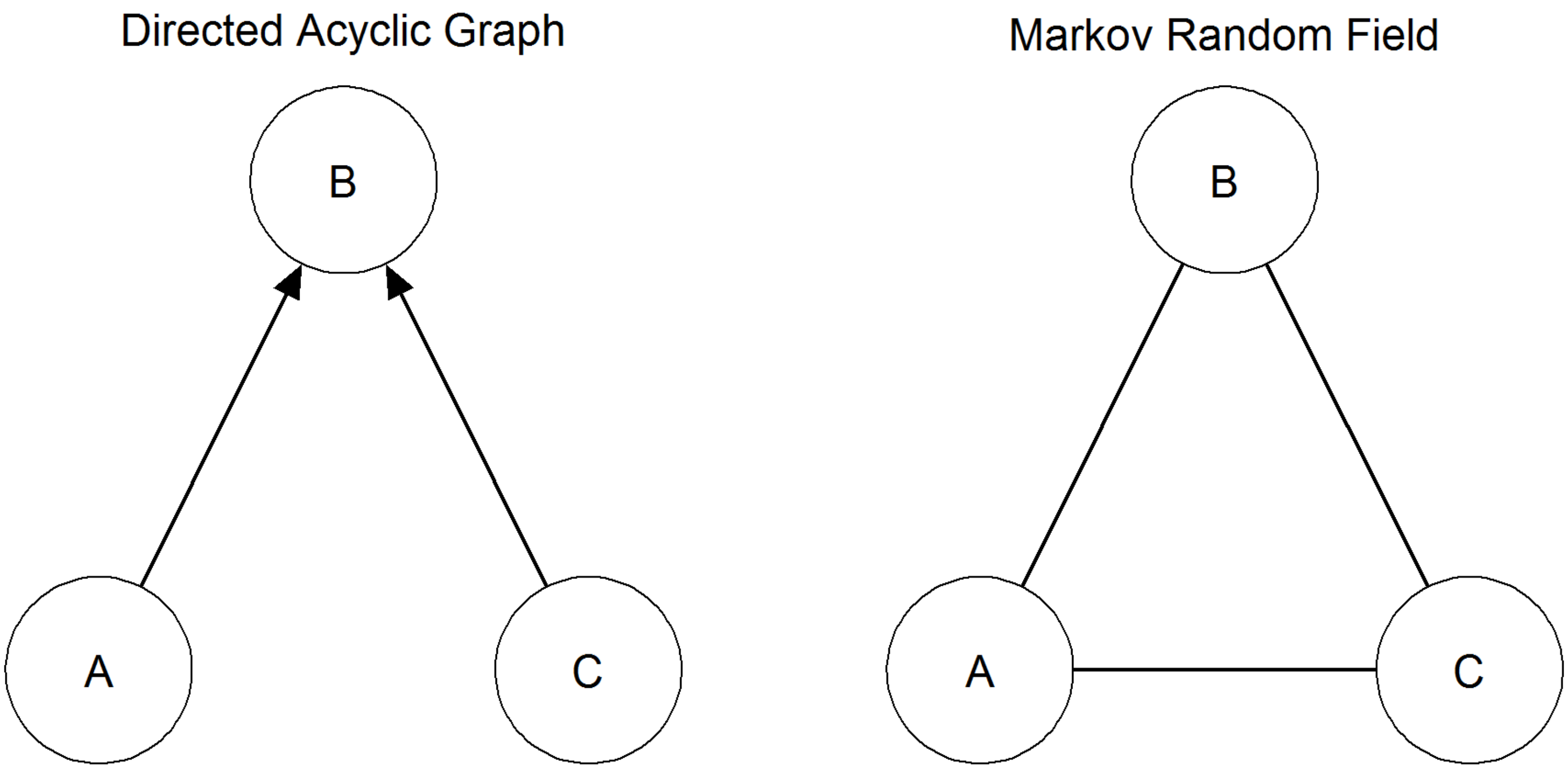


Markov Random Field



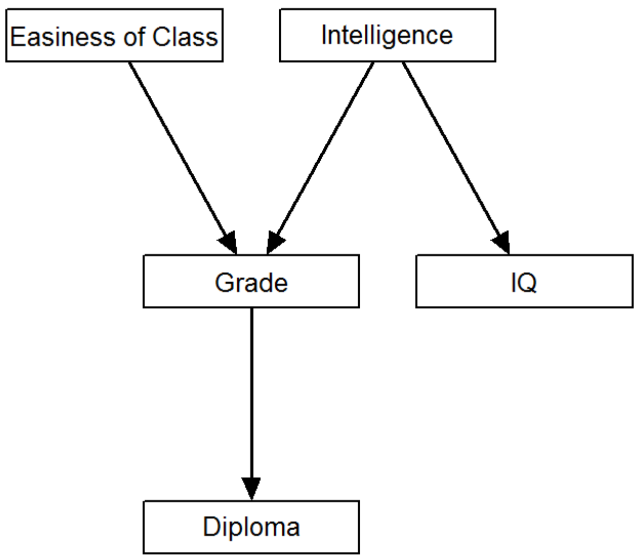
Directed Acyclic Graph



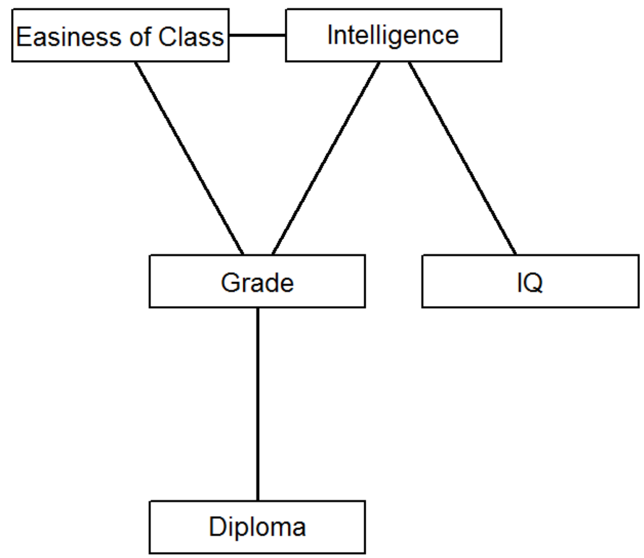


- A MRF can not represent the exact implied independence relationship of a collider structure
  - Three edges are needed instead of two
- However, exogenous variables are commonly modeled to be correlated anyway

Directed Acyclic Graph



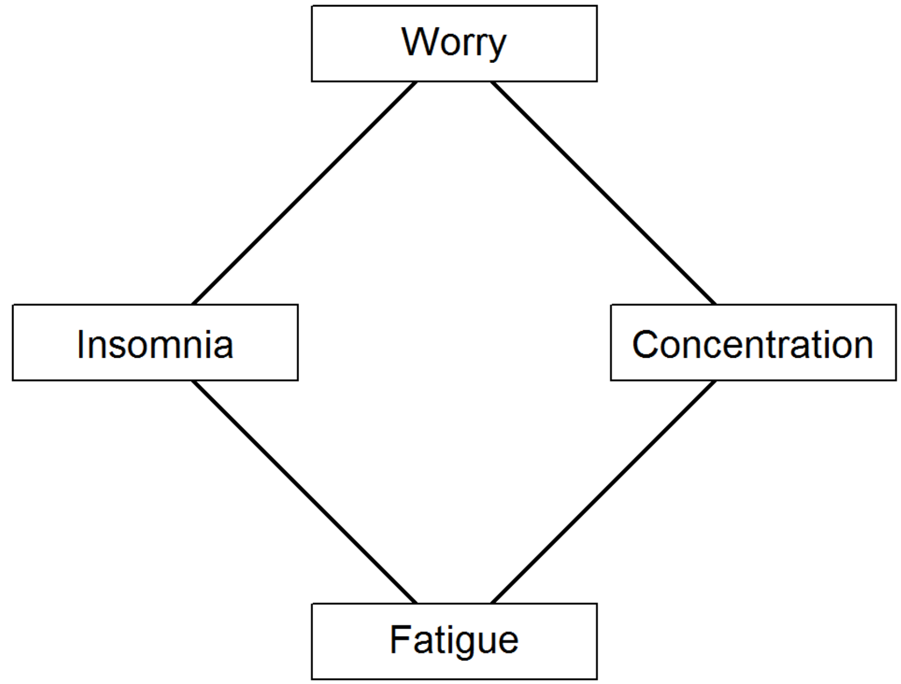
Markov Random Field



Directed Acyclic Graph

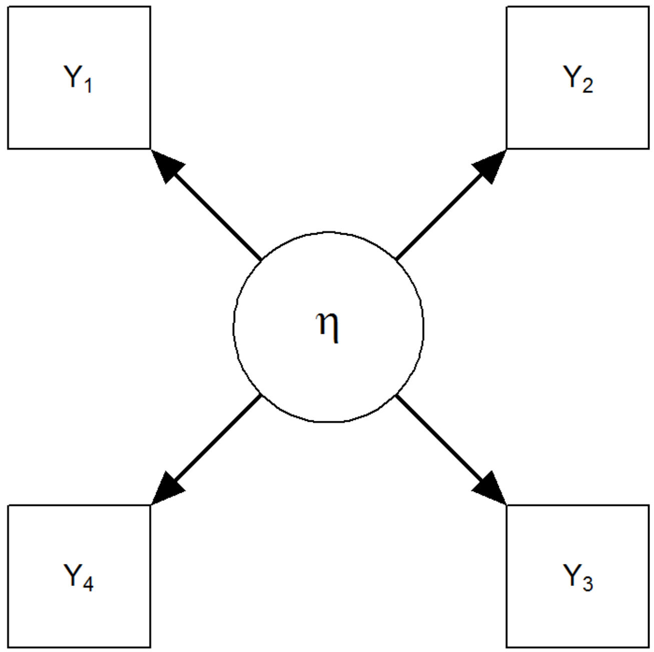
???

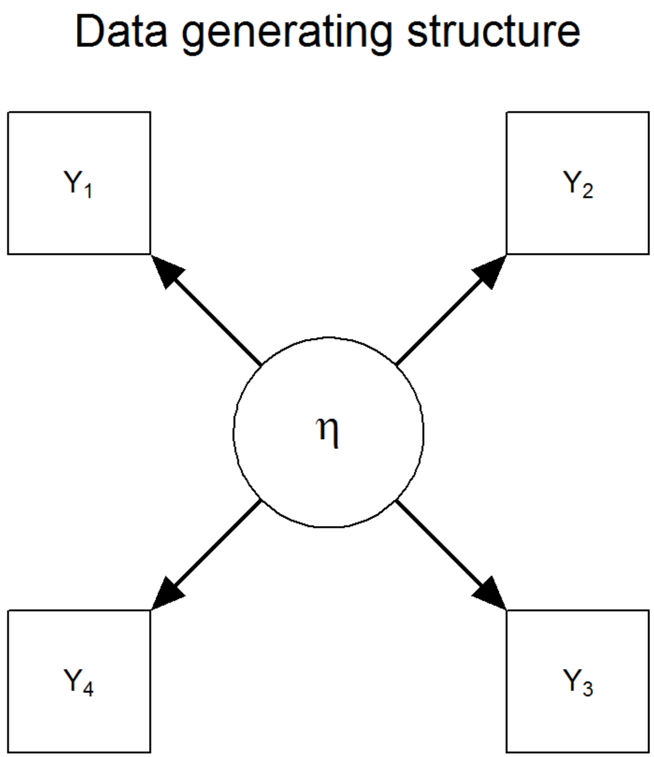
Markov Random Field



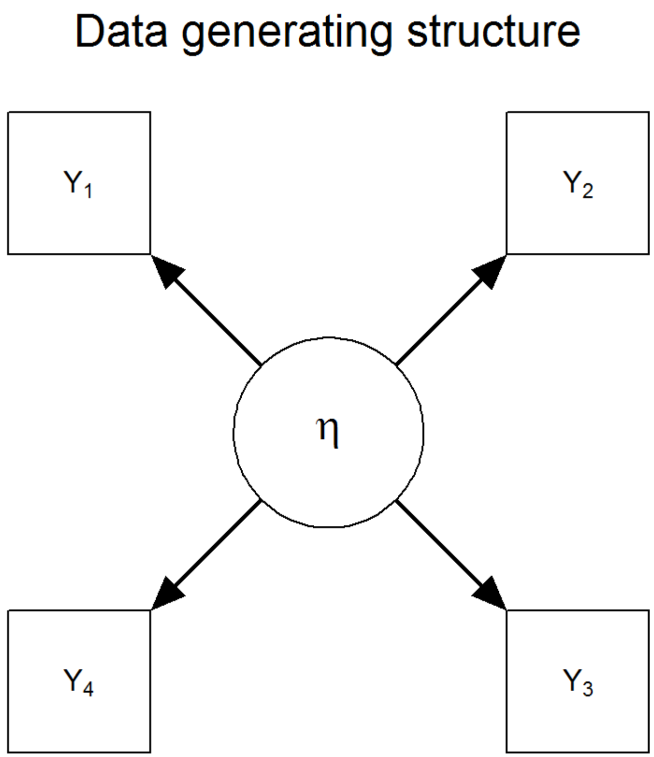
- MRF's can also represent independence relationships that DAG's can not.

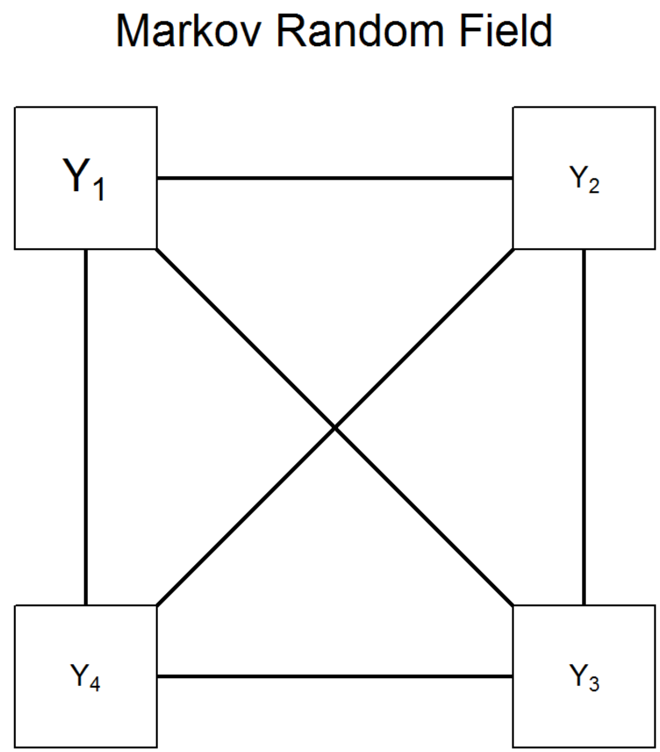
Data generating structure





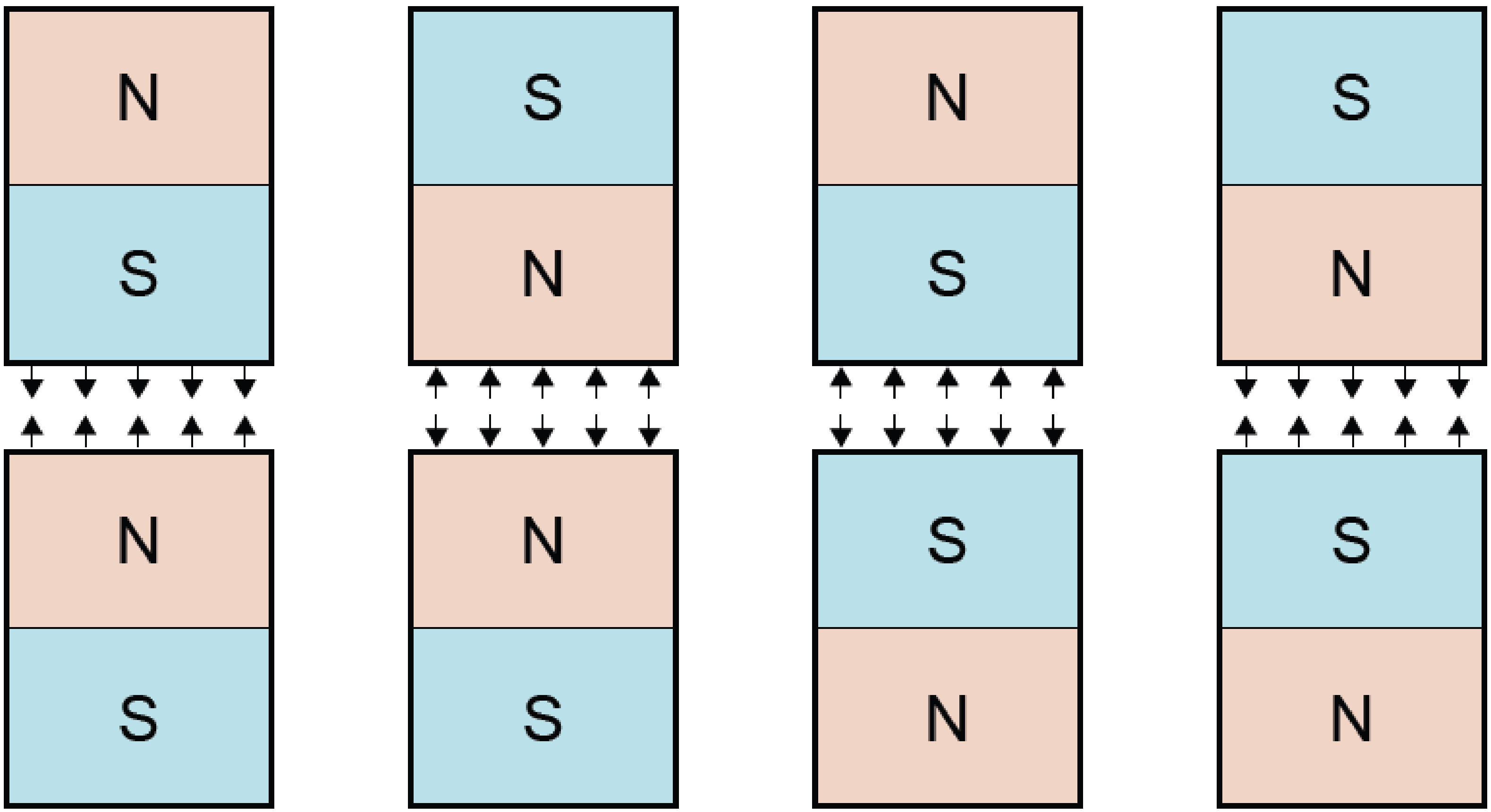
- If we could condition on η



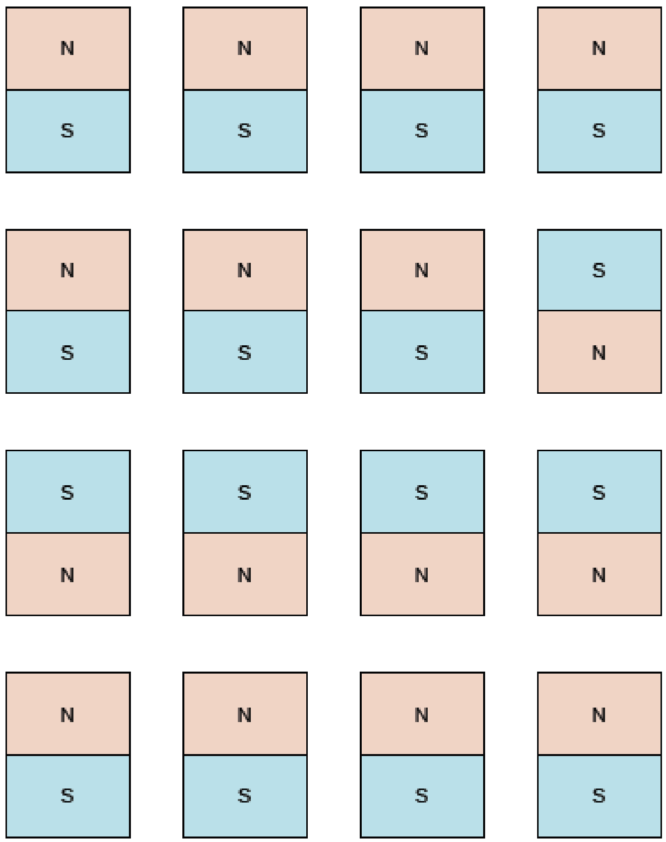


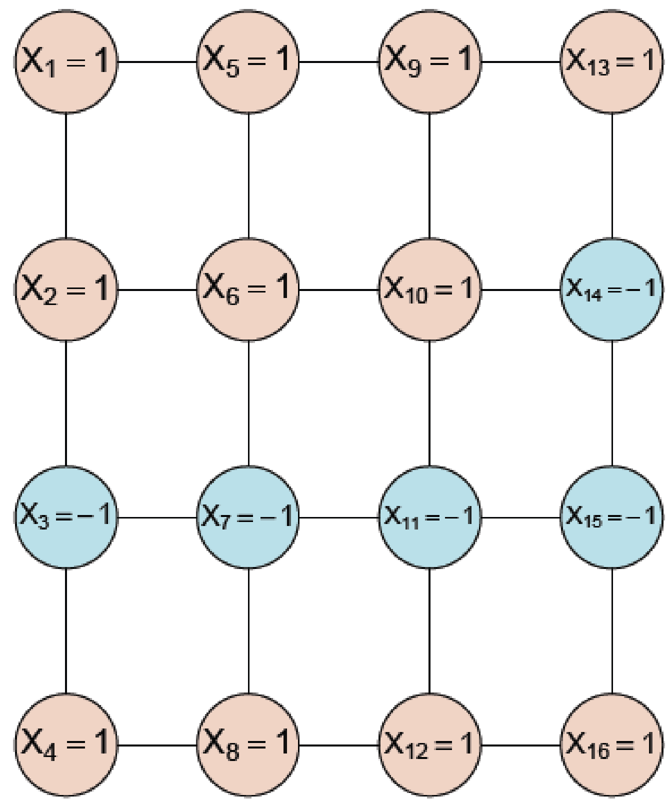
- If we could *not* condition on η
- **Equivalent models**
  - Data generated as a cluster of interacting components can fit a factor model perfectly!

The Ising Model

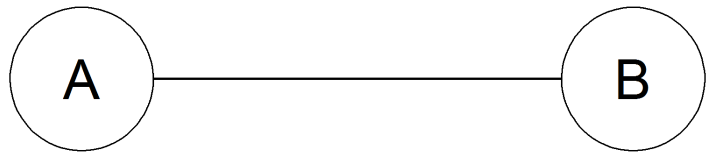


The Ising Model





Markov Random Fields as Generating Structure



- In the Ising model, we could hold a magnet in some way—$\mathrm{Do}(A)$— which can cause adjacent nodes to "flip" with the same probability if we conditioned on A
  - $\Pr(B \mid \mathrm{Do}(A)) = \Pr(B \mid A)$
  - $\Pr(A \mid \mathrm{Do}(B)) = \Pr(A \mid B)$
- Symmetric relationship that can not be represented in a DAG
- Real relationship that occurs in physics

Parameterizing Markov Random Fields

The Ising Model Probability distribution

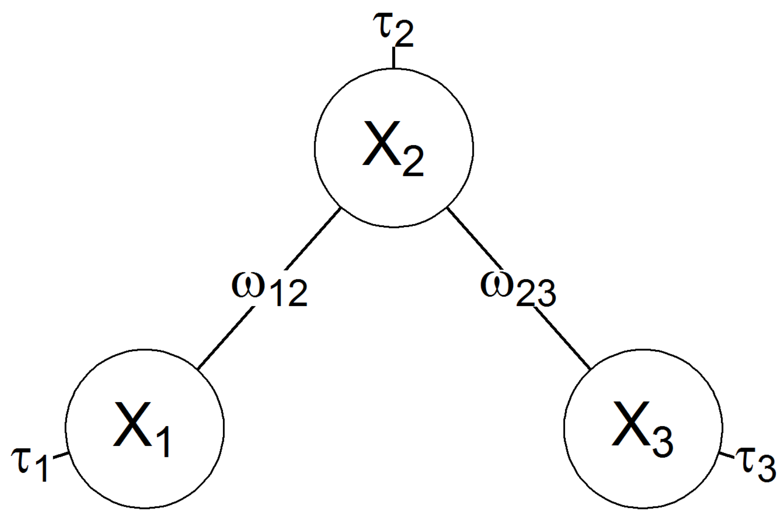
$$\Pr\left(\boldsymbol{X} = \boldsymbol{x}\right) = \frac{1}{Z} \exp\left(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j\right)$$

- All X variables can typically take the values -1 and 1
- τ_i is called the *threshold* parameter and denotes the tendency for node i to be in some state
- ω_{ij} is called the *network* parameter and denotes the preference for nodes i and j to be in the same state
  - Edge weights
- Z is a normalizing constant (partition function) and takes the sum over all possible configurations of $\boldsymbol{X}$:
  - $Z = \sum_{\boldsymbol{x}} \exp\left(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j\right)$

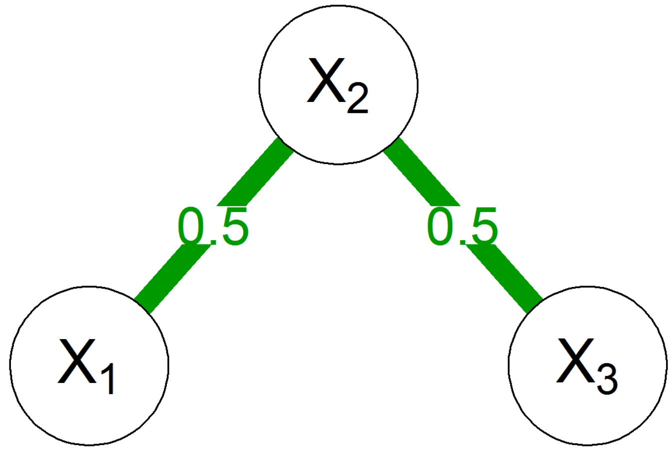
In matrix form the Ising Model probability distribution is proportional to:

$$\Pr\left(\boldsymbol{X} = \boldsymbol{x}\right) \propto \exp\left(\boldsymbol{\tau}^\top \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Omega}\boldsymbol{x}\right)$$

- $\boldsymbol{x}$ is a vector of binary variables (-1 or 1)
- $\boldsymbol{\tau}$ is a vector containing threshold parameters
- $\boldsymbol{\Omega}$ is a matrix containing the network parameters and an arbitrary diagonal
  - Weights matrix that encodes a network
    - $\omega_{ij} = 0$ means that there is no edge between nodes i and j
    - Positive weights are comparable in strength to negative weights



$$\boldsymbol{\Omega} = \begin{bmatrix} 0 & \omega_{12} & 0 \\ \omega_{12} & 0 & \omega_{23} \\ 0 & \omega_{23} & 0 \end{bmatrix}, \boldsymbol{\tau} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}$$



$$\boldsymbol{\Omega} = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0 \end{bmatrix}, \boldsymbol{\tau} = \begin{bmatrix} -0.1 \\ -0.1 \\ -0.1 \end{bmatrix}$$

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z} \exp\left(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j \right)$$

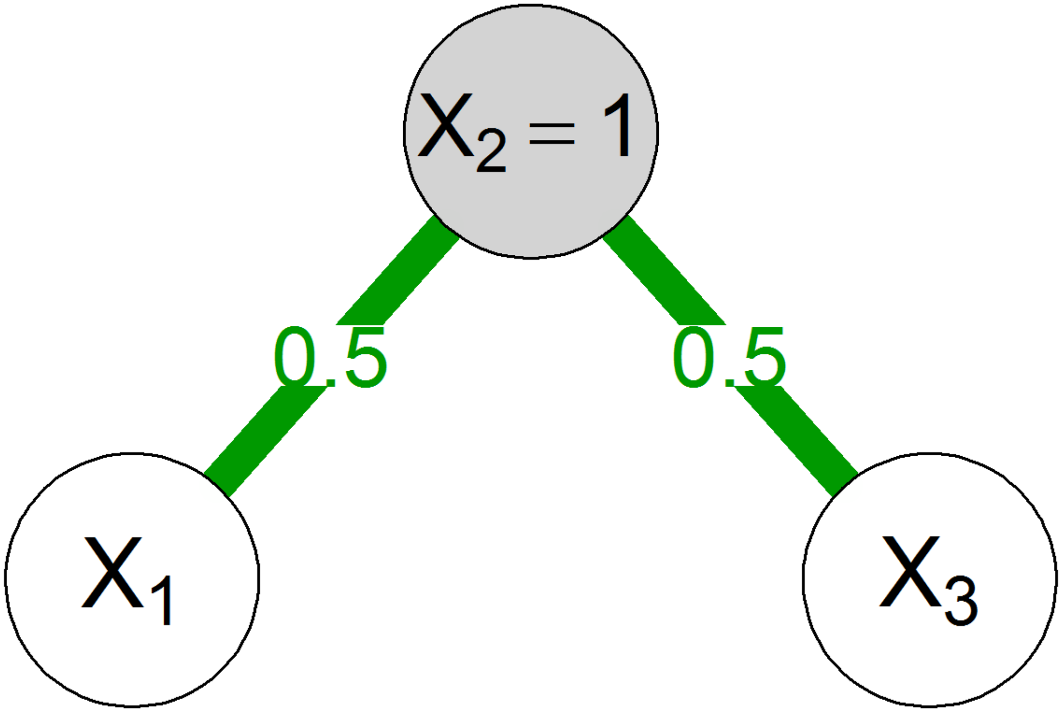
We can compute the unnormalized probability that all nodes are 1:

$$\exp(-0.1 + -0.1 + -0.1 + 0.5 + 0.5) = 2.0138$$

- We will call this the *potential* for the nodes to be in this state
- Summing the potential of every possible state gives the normalizing constant Z
- Which can then be used to compute the probabilities

x_1	x_2	x_3	Potential	Probability
-1	-1	-1	3.6693	0.3514
1	-1	-1	1.1052	0.1058
-1	1	-1	0.4066	0.0389
1	1	-1	0.9048	0.0866
-1	-1	1	1.1052	0.1058
1	-1	1	0.3329	0.0319
-1	1	1	0.9048	0.0866
1	1	1	2.0138	0.1928
			$Z = 10.4426$	

What if we observe $X_2 = 1$?



What if we observe $X_2 = 1$?

x_1	x_2	x_3	Potential	Probability
-1	-1	-1	3.6693	0.3514
1	-1	-1	1.1052	0.1058
-1	1	-1	0.4066	0.0389
1	1	-1	0.9048	0.0866
-1	-1	1	1.1052	0.1058
1	-1	1	0.3329	0.0319
-1	1	1	0.9048	0.0866
1	1	1	2.0138	0.1928
			$Z = 10.4426$	

What if we observe $X_2 = 1$?

x_1	x_2	x_3	Potential	Probability
-1	1	-1	0.4066	0.0961
1	1	-1	0.9048	0.2139
-1	1	1	0.9048	0.2139
1	1	1	2.0138	0.4761

$$Z = 4.229997$$

- $(0.2139 + 0.4761) * (0.4761 + 0.2139) = 0.4761$
- $\Pr(X_1 = 1, X_3 = 1 \mid X_2 = 1) = \Pr(X_1 = 1 \mid X_2 = 1)\Pr(X_3 = 1 \mid X_2 = 1)$
- X_1 and X_3 are conditionally independent given X_2!

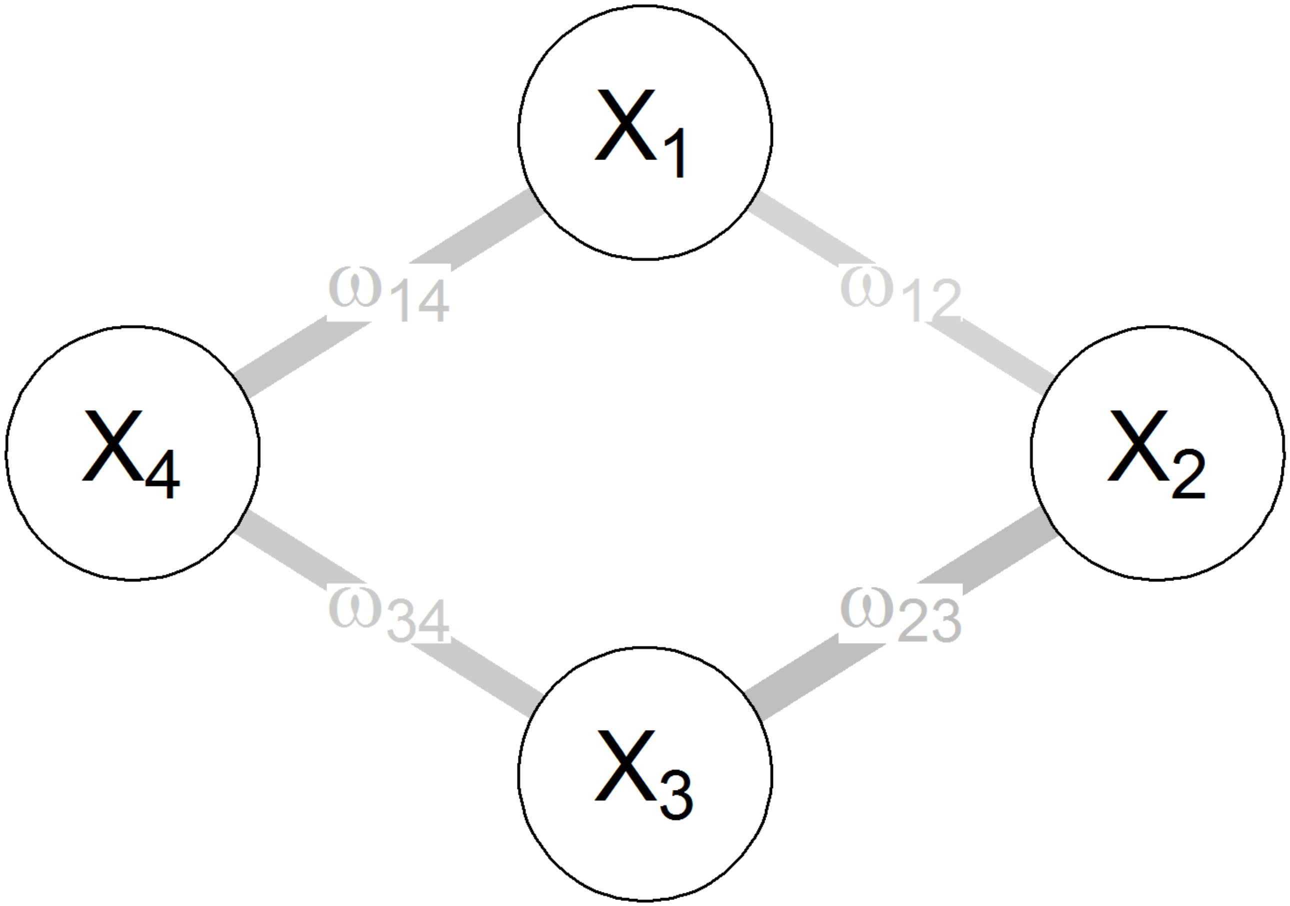
Conditional Distribution for the Ising Model

The conditional distribution for node i given that we observe **all** other nodes is:

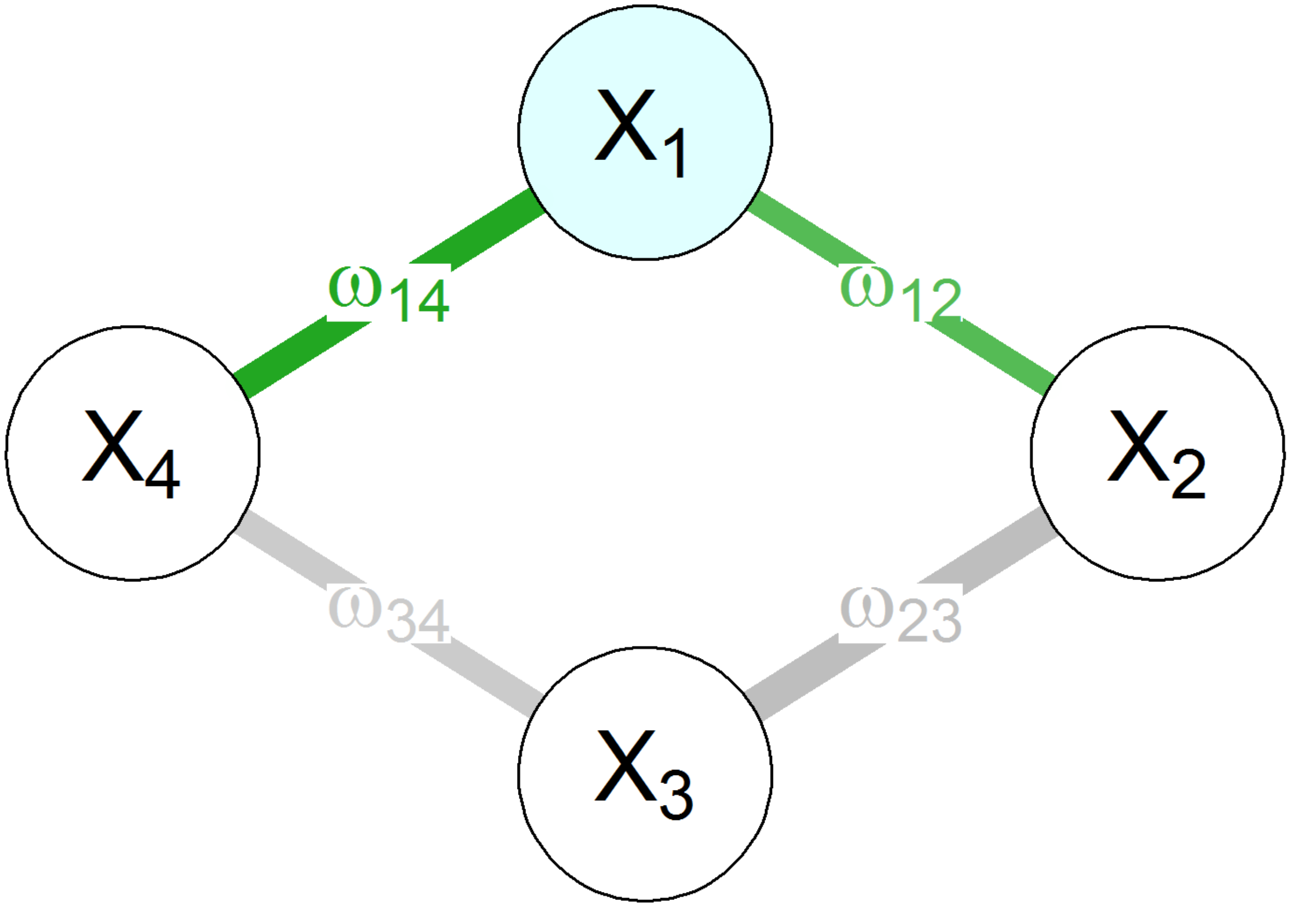
$$\Pr\left(X_i = x_i \mid \boldsymbol{X}^{-(i)} = \boldsymbol{x}^{-(i)}\right) \propto \exp\left(\left(\tau_i + \sum_{j, j \neq i} \omega_{ij} x_j\right) x_i\right)$$

- This is a *multiple logistic regression* model!
- The most common model to predict the value of one binary variable (dependent variable) given a set of other variables (independent variables)
- The Ising model is a combination of predictive models. Edges represent how strongly one node predicts another
    - A path means that the predictive strength of one node on another is *mediated*

Combination of Logistic models

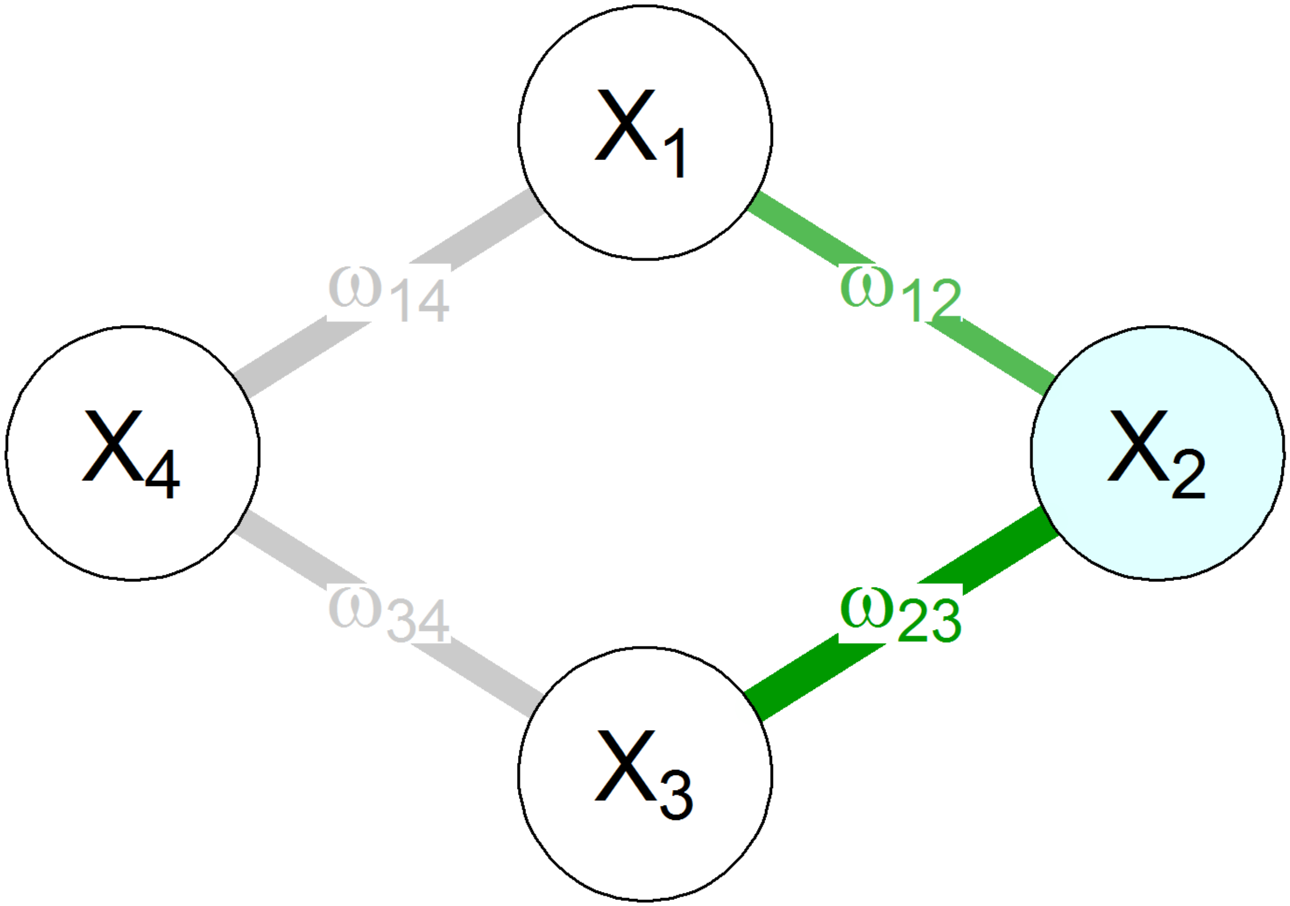


Ising Model as Combination of Logistic models



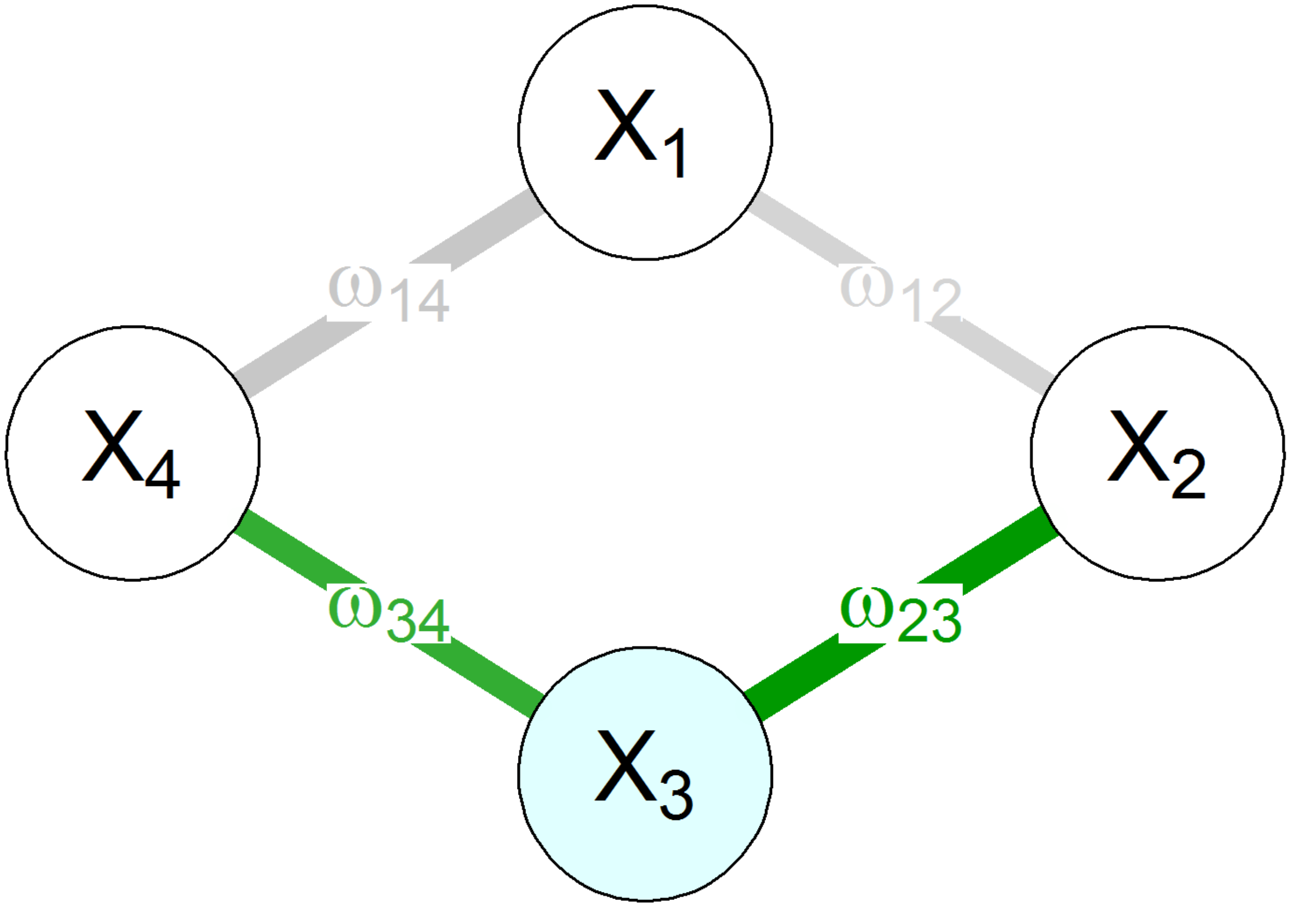
$$\Pr\left(X_1 = 1\right) \propto \exp\left(\tau_1 + \omega_{12} x_2 + \omega_{14} x_4\right)$$

Ising Model as Combination of Logistic models



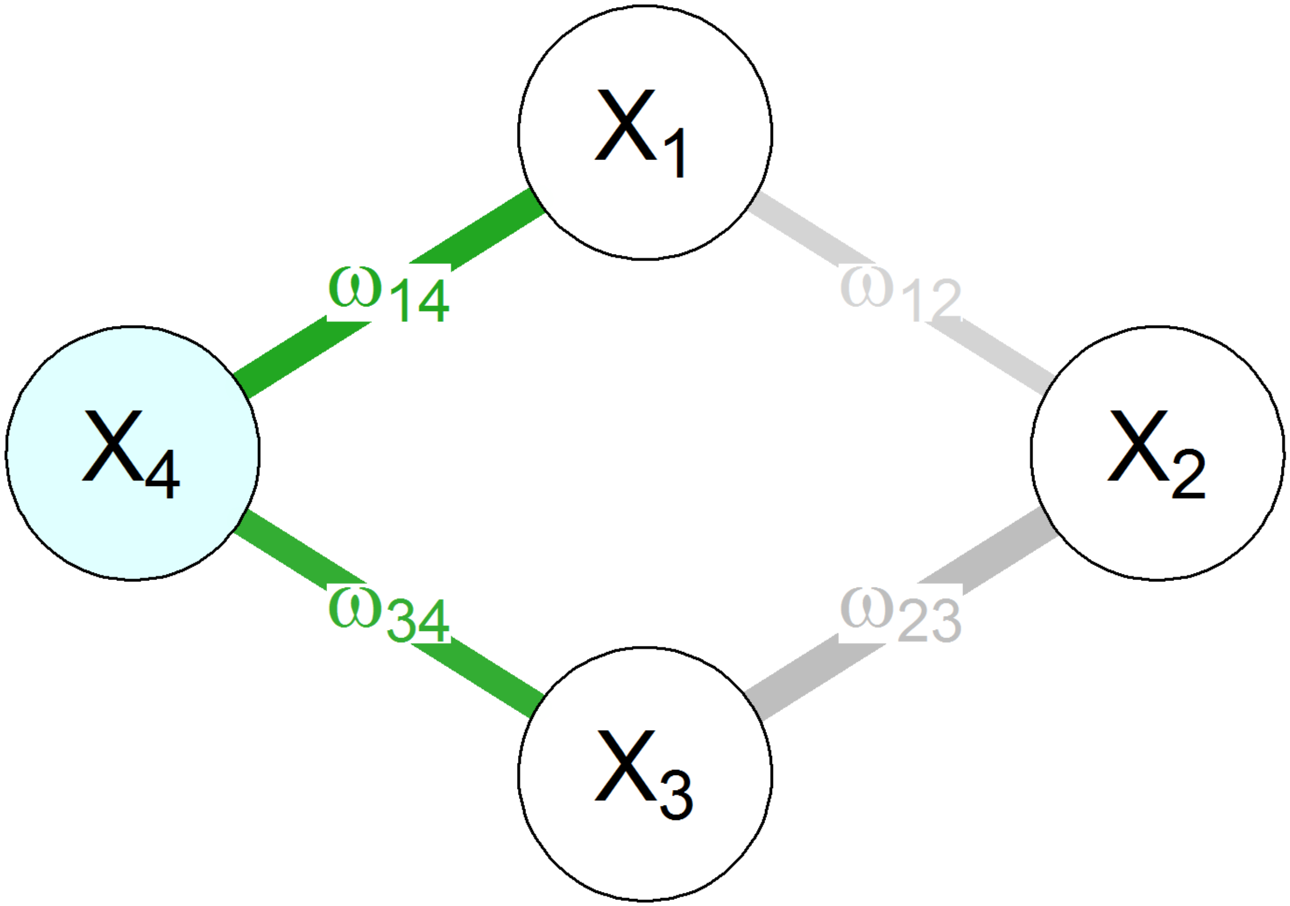
$$\Pr\left(X_2 = 1\right) \propto \exp\left(\tau_2 + \omega_{12} x_1 + \omega_{23} x_3\right)$$

Ising Model as Combination of Logistic models



$$\Pr\left(X_3 = 1\right) \propto \exp\left(\tau_3 + \omega_{23} x_2 + \omega_{34} x_4\right)$$

Ising Model as Combination of Logistic models



$$\Pr\left(X_4 = 1\right) \propto \exp\left(\tau_4 + \omega_{14} x_1 + \omega_{34} x_3\right)$$

Continous Data

If $\boldsymbol{x}$ is not binary but assumed Gaussian we can use a multivariate Gaussian distribution:

$$f(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

- $\boldsymbol{\mu}$ is a vector that encodes the means
- $\boldsymbol{\Sigma}$ is the variance-covariance matrix
- Now we can rearrange:
  - $f(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$
  - $f(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})\right)$
  - $f(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}\right)$

Gaussian Random Field

Reparameterizing $\boldsymbol{\tau} = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Omega} = -\boldsymbol{\Sigma}^{-1}$ we obtain the following expression for the Multivariate Normal Distribution:

$$f(\boldsymbol{x}) \propto \exp\left(\boldsymbol{\tau}^\top \boldsymbol{x} + \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{\Omega} \boldsymbol{x}\right)$$

- Exactly the same form as the Ising Model!!!
- Except:
  - $\boldsymbol{x}$ is now continuous
  - The normalizing constant is different
- The multivariate normal distribution encodes a network
  - This network is called a *Gaussian Random Field* (GRF), *Concentration Graph*, *Gaussian Graphical Model* or *Partial Correlation Network*.

Gaussian Random Field

- The negative inverse covariance matrix of the multivariate normal distribution, also called *precision matrix* encodes a network
- Through mathematical magic, standardized elements of the negative inverse precision matrix are equal *partial correlation coefficients* between nodes conditioned on all other nodes
  - $\rho_{ij} = \omega_{ij} / \sqrt{\omega_{ii}\omega_{jj}} = \mathrm{Cor}\left(X_i, X_j \mid \boldsymbol{X}^{-(i,j)} = \boldsymbol{x}^{-(i,j)}\right)$
- Typically these partial correlation coefficients are used in the weights matrix of the corresponding network structure
  - There is no edge between nodes i and j if $\rho_{ij} = 0$
  - Which clearly corresponds to the Markov property that two nodes are then independent conditioned on all other nodes in the network

Conditional Distribution for the Gaussian Random Field

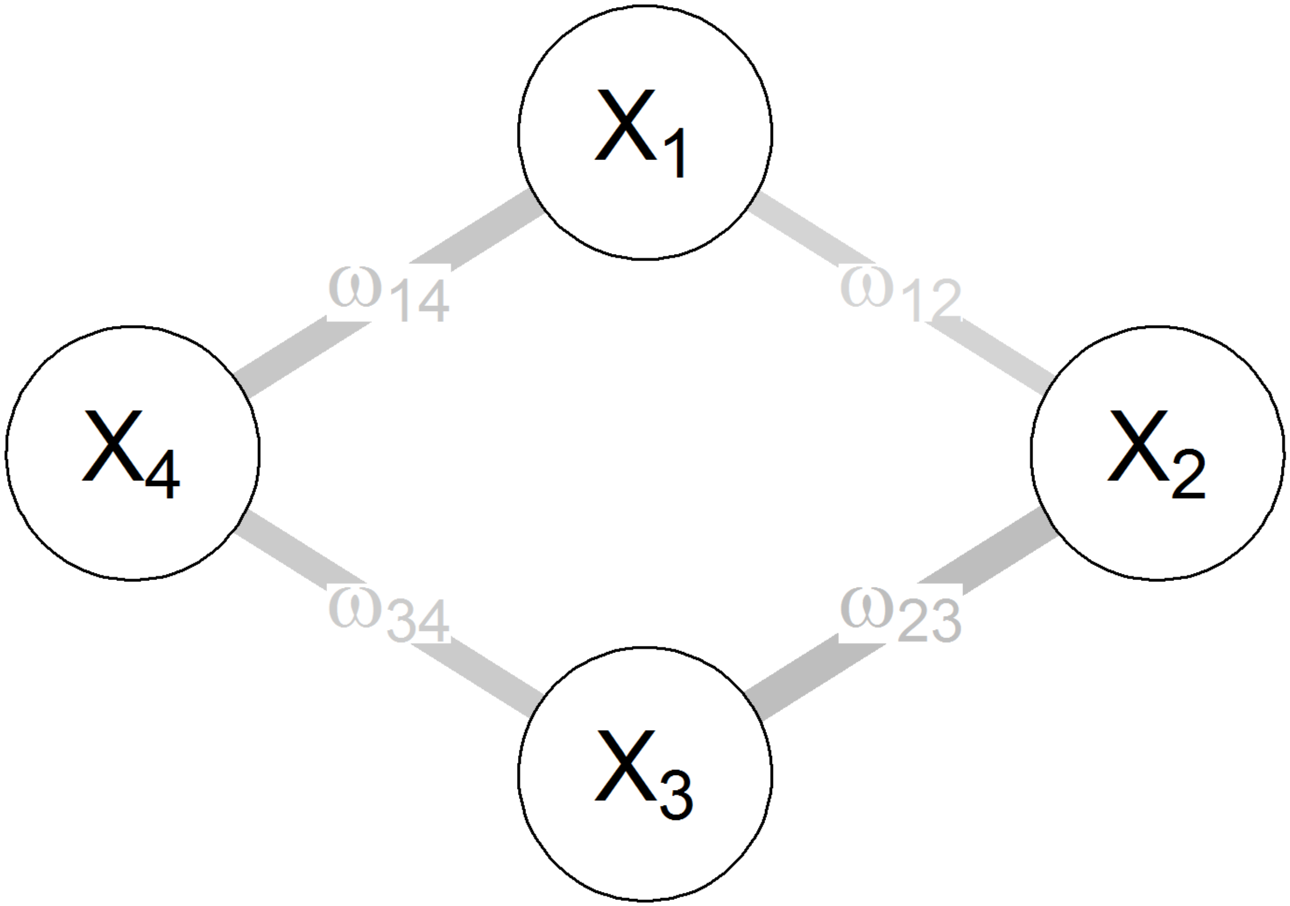
If we observe the value of every node except node i we obtain the following (assuming all variables are standardized):

$$X_i \mid \boldsymbol{X}^{-(i)} \sim N\left(\sum_{j \neq i} \frac{\omega_{ij}}{\theta_i^2}\, x_j, \theta_i^2\right)$$

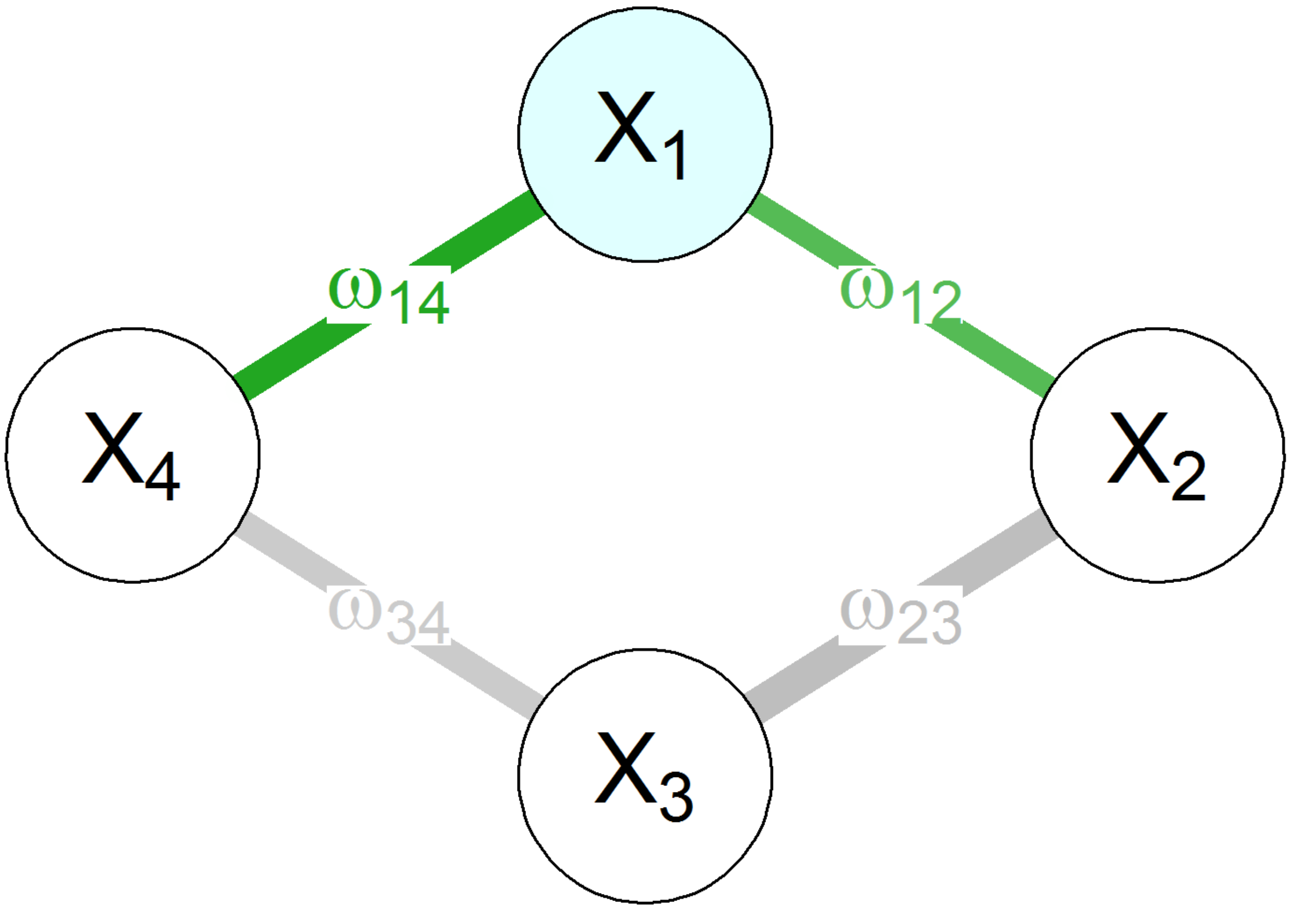
In which θ_i^2 is the ith diagonal element of the precision matrix, also called the residual variance.

- This is a multiple linear regression model!
- Similar to the Ising model, the Gaussian Random Field shows how well nodes predict each other!

Combination of Linear models

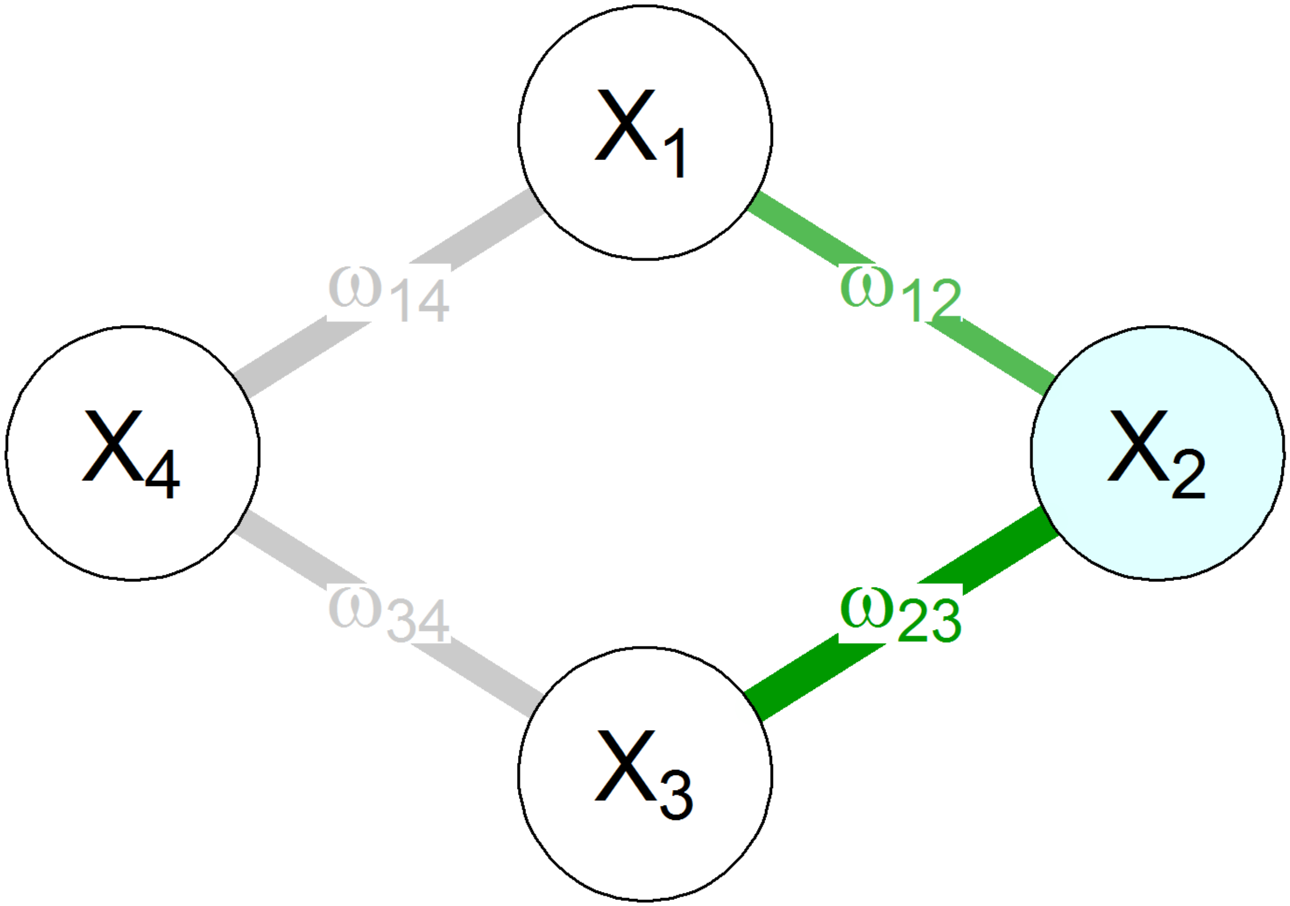


Ising Model as Combination of Linear models



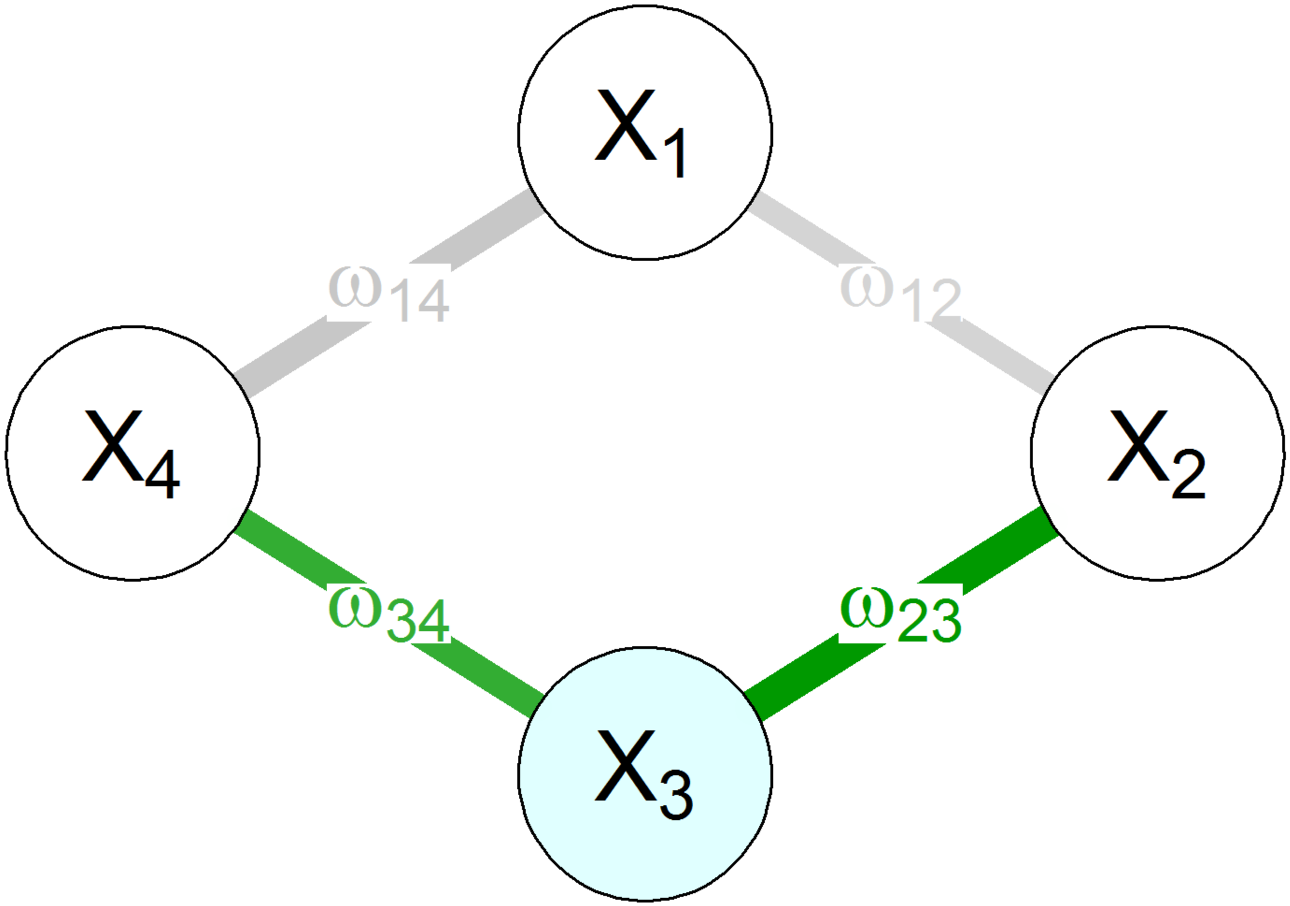
$$X_1 = \tau_1 + \omega_{12}\theta_1^{-2}x_2 + \omega_{14}\theta_1^{-2}x_4 + \varepsilon_1$$

Ising Model as Combination of Linear models



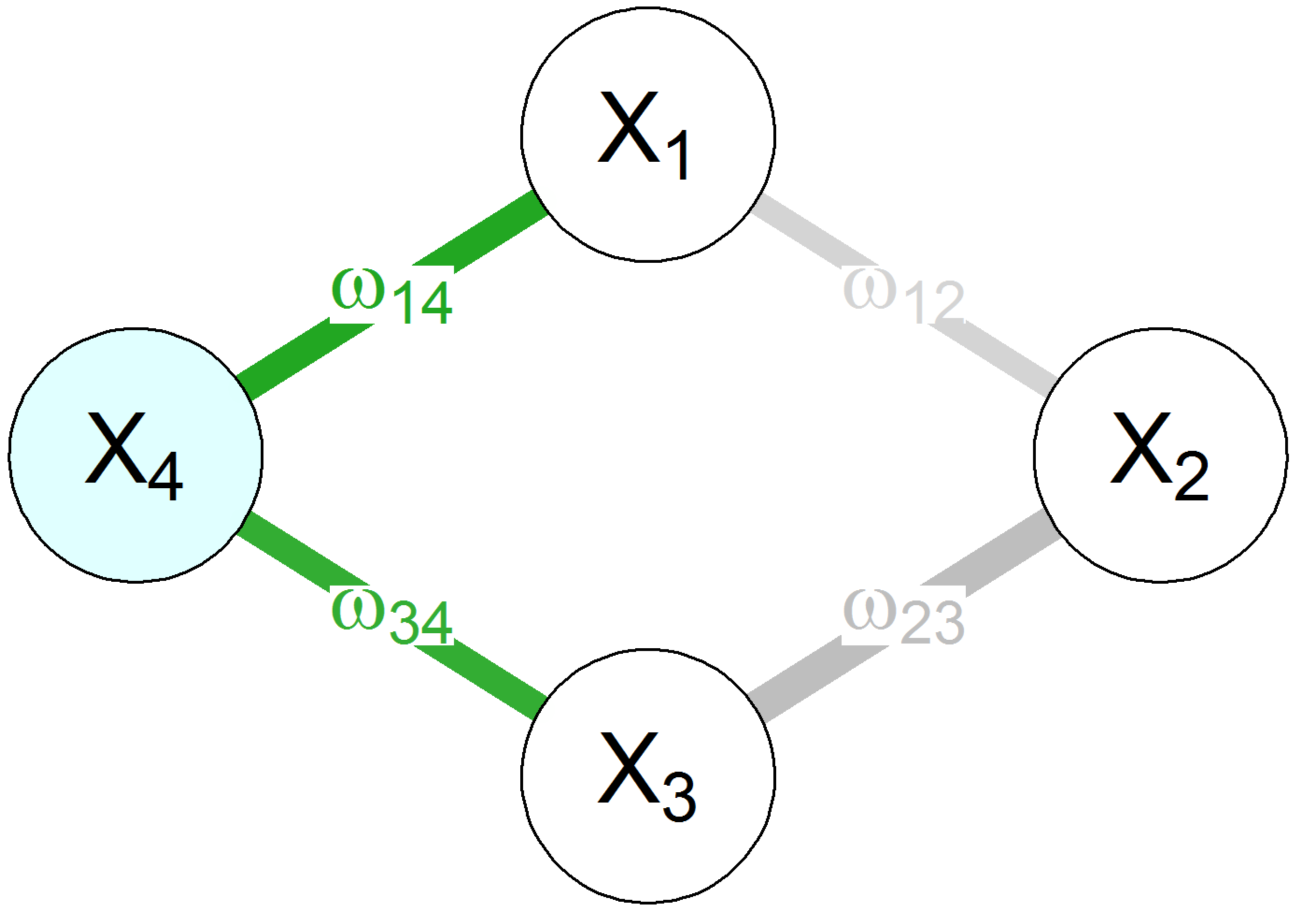
$$X_2 = \tau_2 + \omega_{12}\theta_2^{-2}x_2 + \omega_{23}\theta_2^{-2}x_3 + \varepsilon_2$$

Ising Model as Combination of Linear models



$$X_3 = \tau_3 + \omega_{23}\theta_3^{-2}x_3 + \omega_{34}\theta_3^{-2}x_4 + \varepsilon_3$$

Ising Model as Combination of Linear models



$$X_4 = \tau_4 + \omega_{34}\theta_4^{-2}x_3 + \omega_{14}\theta_4^{-2}x_1 + \varepsilon_4$$

Markov Random Fields

- **Ising Model**
  - All nodes are binary
    - -1 or 1
  - Combination of multiple logistic regression models
- **Gaussian Random Field**
  - All nodes assumed normally distributed
  - Graph structure directly related to the inverse variance-covariance matrix
  - Graph usually standardized to *partial correlations*
  - Combination of multiple linear regression models

Estimating Markov Random Fields

Constructing Markov Random Fields

- Because the Ising Model is a combination of multiple logistic models and the Gaussian Random field a combination of multiple linear models we can form MRF's in the following way:

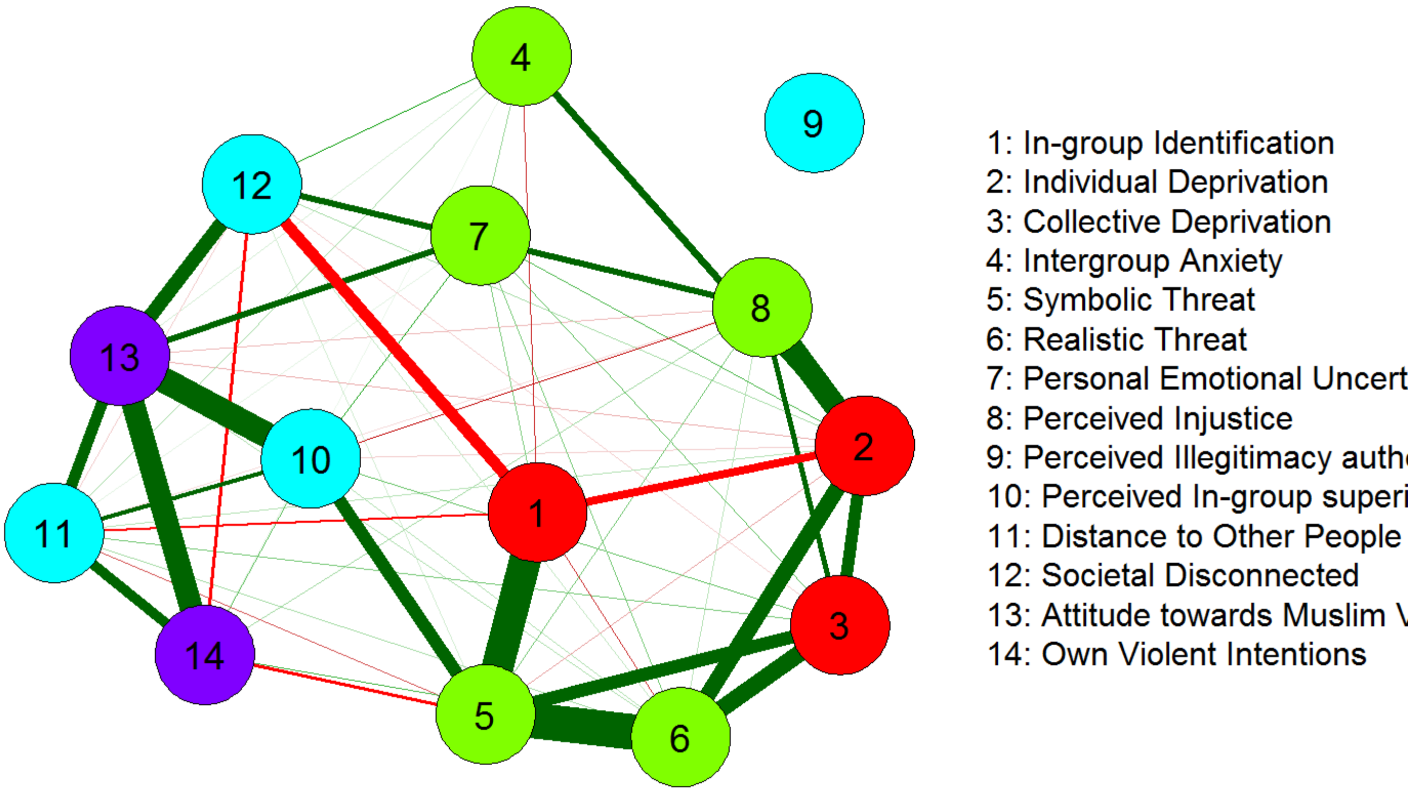
1. Predict each node from all other nodes in a multiple (logistic) regression
2. This gives you two parameters for each edge in the network. For example:
    - The regression parameter from predicting A from B and the regression parameter from B to A.
    - In logistic regression, these should be about equal. In linear regression, they can be standardized to be about equal
3. Take the mean of these two parameters as edge weight
    - If the regression parameter is 0 there is no edge!

Advanced MRF estimation

- The methods on the previous slide work but require a lot of data!
- To obtain a simpler interpretable and more stable network we will apply *regularization*
  - LASSO
- Also, the assumption of multivariate normal data can be relaxed in MRF's
- This will both be discussed on Thursday!

Examples

Radicalization

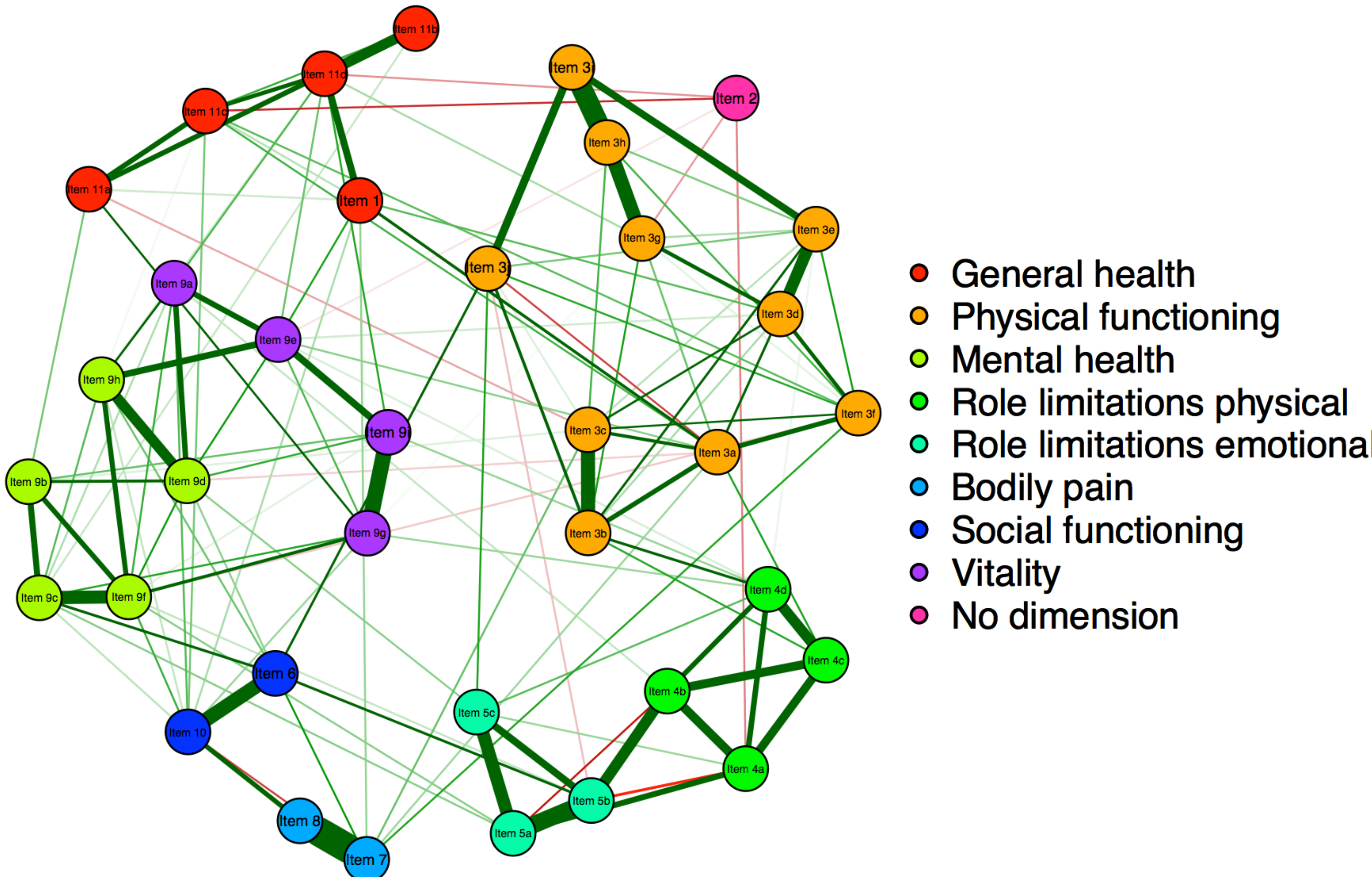


Radicalization

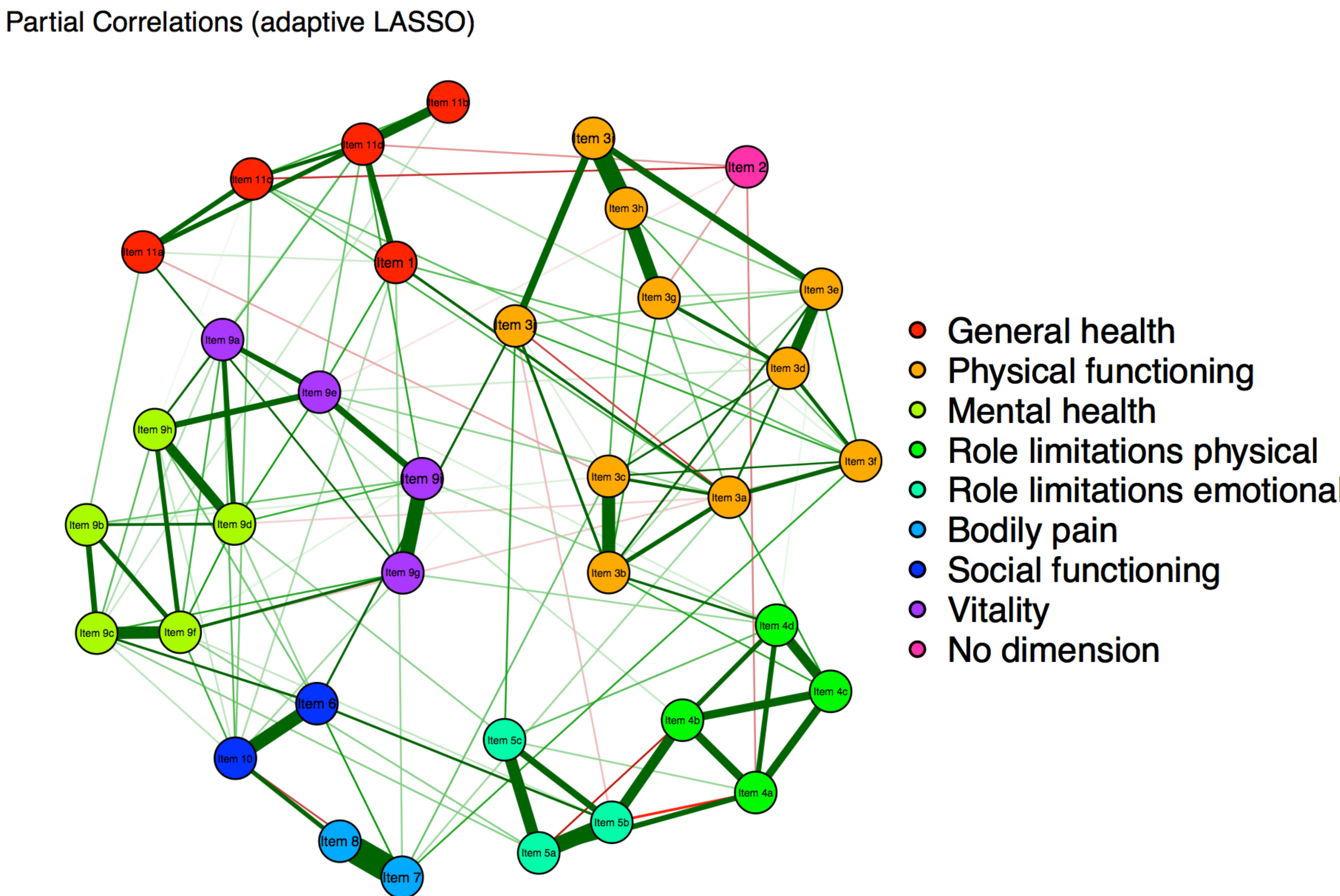
	Muslim Violence	Violent Intentions
In-group Identification	16.12	19.89
Individual Deprivation	20.10	23.87
Collective Deprivation	17.05	20.82
Intergroup Anxiety	Inf	Inf
Symbolic Threat	11.44	15.21
Realistic Threat	14.81	18.58
Personal Emotional Uncertainty	10.73	14.50
Perceived Injustice	24.83	28.60
Perceived Illegitimacy authorities	Inf	Inf
Perceived In-group superiority	3.16	6.93
Distance to Other People	5.70	8.81
Societal Disconnected	9.15	12.92
Attitude towards Muslim Violence	0.00	3.77
Own Violent Intentions	3.77	0.00

Quality of Life

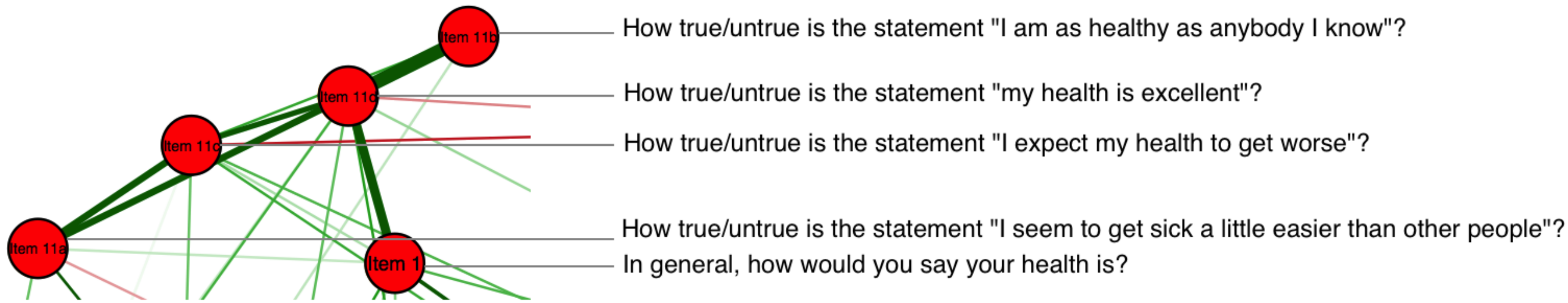
Partial Correlations (adaptive LASSO)



Quality of Life

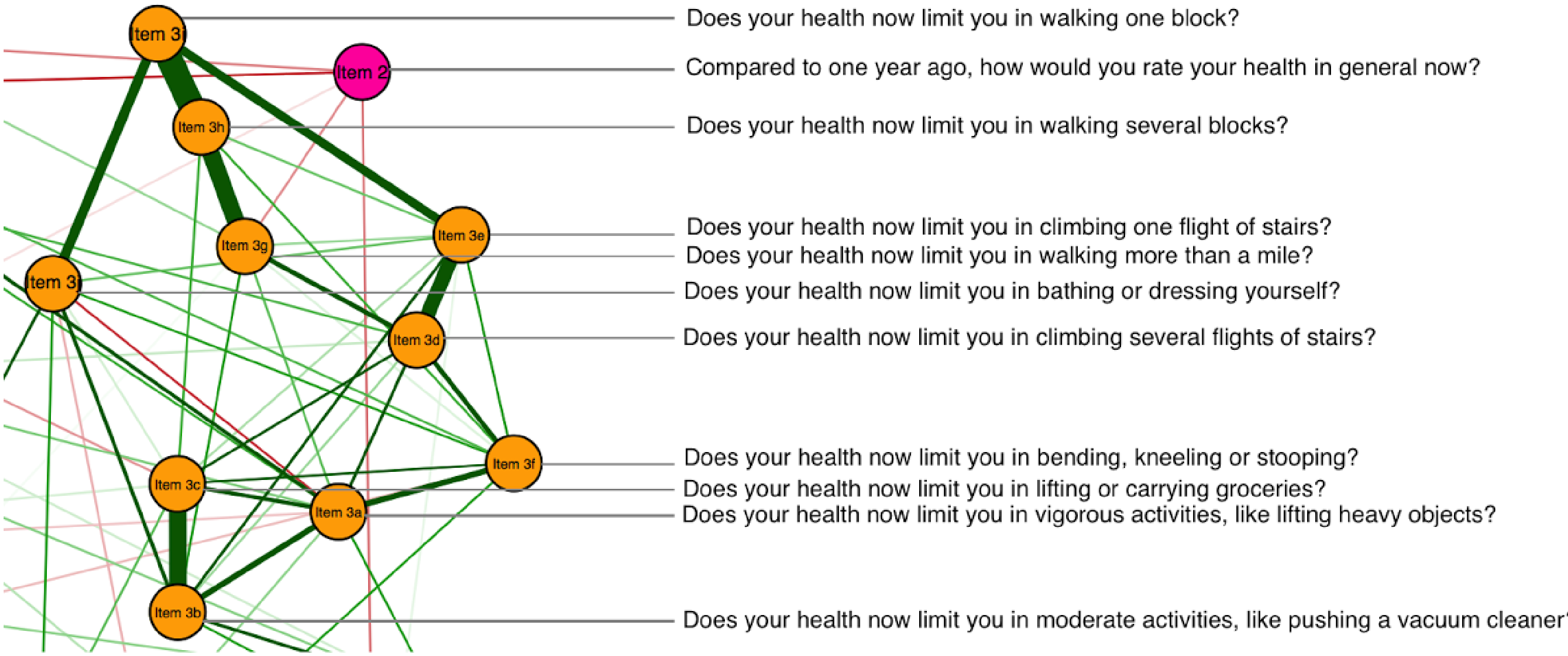


Quality of Life



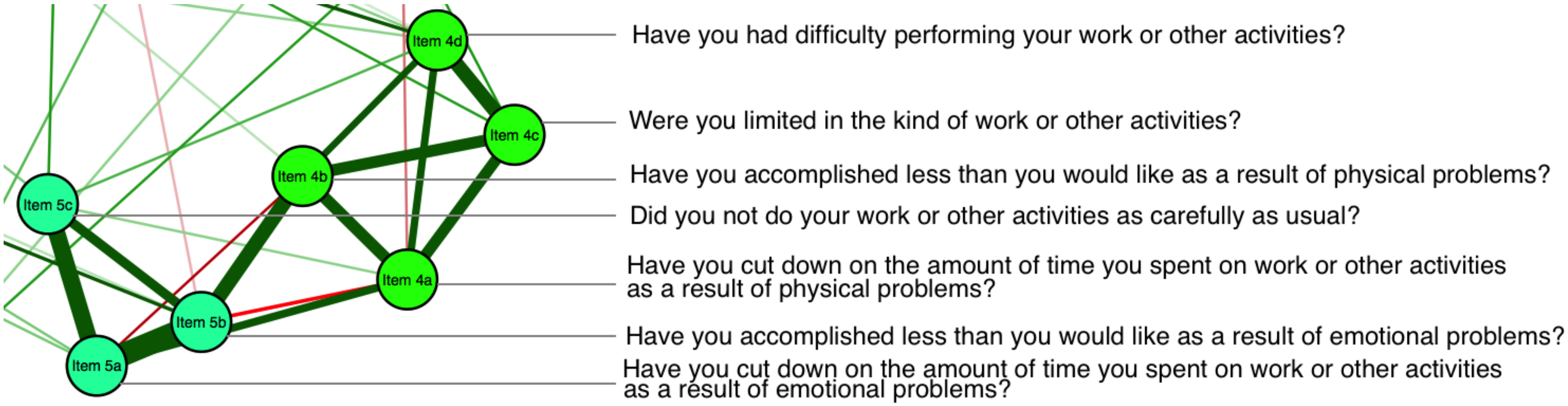
By Jolanda Kossakowski

Quality of Life



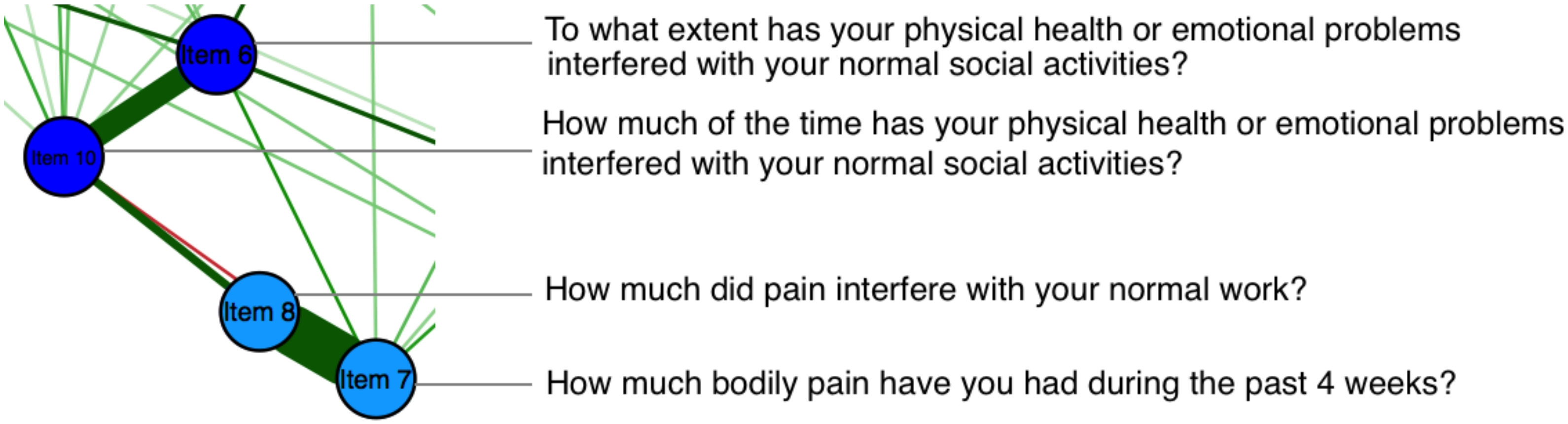
By Jolanda Kossakowski

Quality of Life



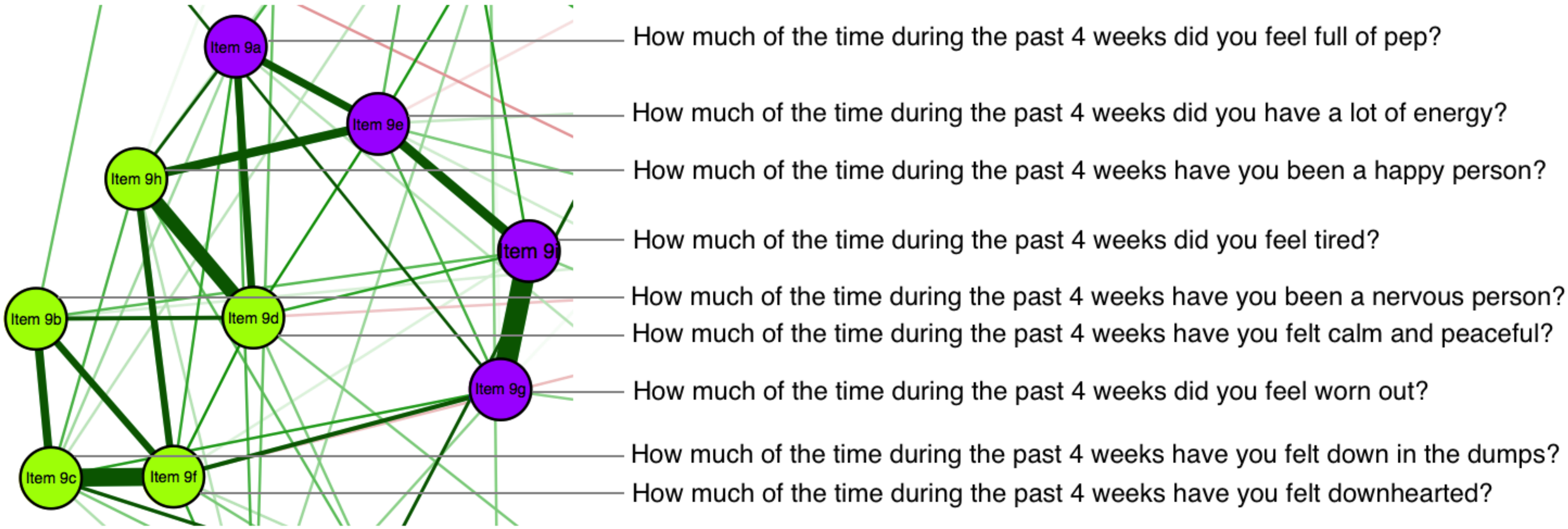
By Jolanda Kossakowski

Quality of Life



By Jolanda Kossakowski

Quality of Life



By Jolanda Kossakowski

Thursday:

- Estimating sparse networks
- Normality assumption
- Another example
- Cookbook
  - Most codes of the course!
- Be there!