How to spread the nets over nuisance data?

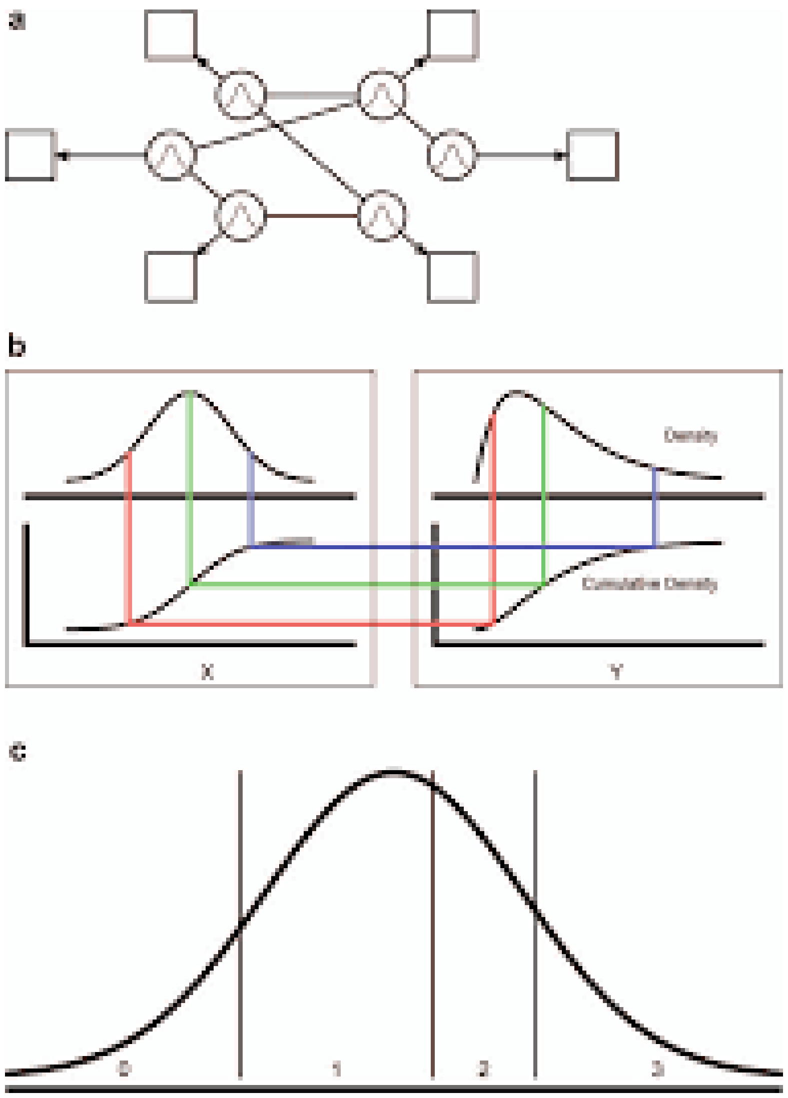
Introduction to Network Analysis, L3

Edita Chvojková

edita.chvojkova@mail.muni.cz

Today's lecture

- **non-normal continuous data**
  - transform variables first (`huge.npn` in `huge`)
- **ordinal data**
  - use Spearman correlations as input
- **binary data**
  - IsingModel
- **personalized networks & time series**
  - why are personalized networks so hot?



Binary data: The Ising model

The Ising model is a model from physics, which describes magnetisation.

• the model is based on pairwise interactions between neighbouring variables in a network

• can model binary data (-1 / 1 or 0 / 1) - two neighbouring nodes want to be in the same state

• joint estimation (e.g., maximum likelihood) is hard, because of the partition function.

• luckily, the Ising model can be estimated **in a series of logistic regressions!**

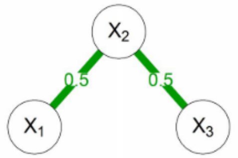
Binary data: The Ising model in `bootnet`

Regularized – IsingFit

- Claudia van Borkulo's eLasso algorithm uses a series of piecewise regressions
- One by one, each variable features as a dependent variable in a logistic regression, with all the other variables as independent
- This gives a regression equation, which is optimised using a penalisation function (lasso)
- When a variable X is included in the prediction function for Y, then we say that X is in the neighbourhood of Y
- When two variables are estimated to be in each other's neighbourhood, we connect them with an edge

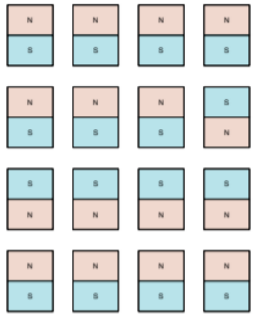
Unregularized – IsingSampler

- many different estimation methods
- don't do it

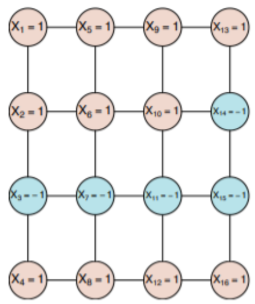


$$\boldsymbol{\Omega} = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0 \end{bmatrix}, \boldsymbol{\tau} = \begin{bmatrix} -0.1 \\ -0.1 \\ -0.1 \end{bmatrix}$$

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
			$Z = 10.4426$	



(a)

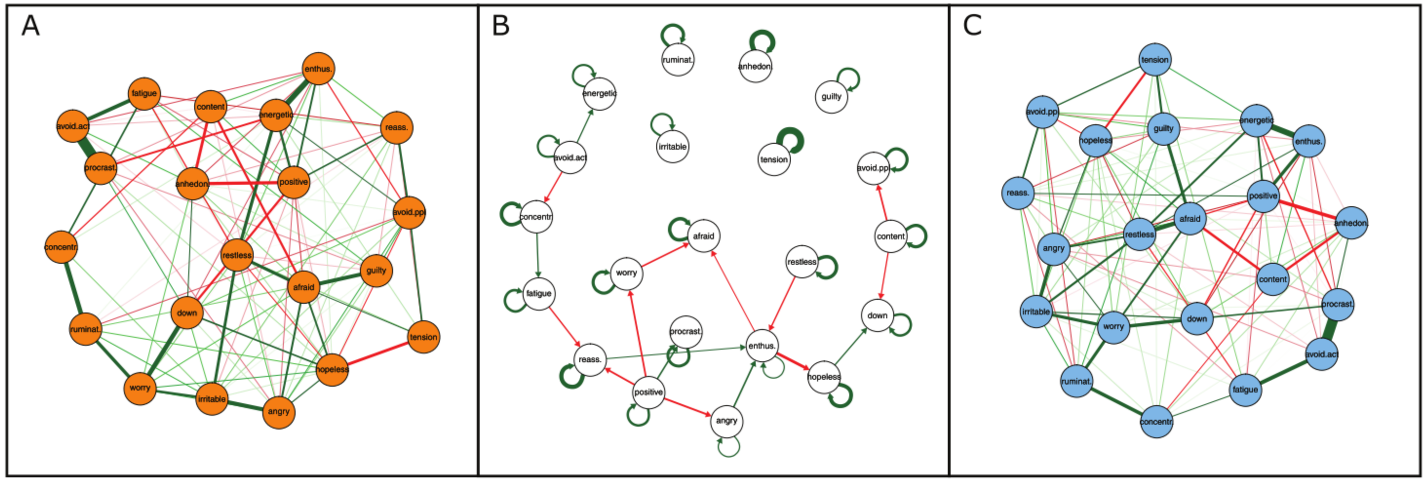


(a)

Personalized networks & time-series

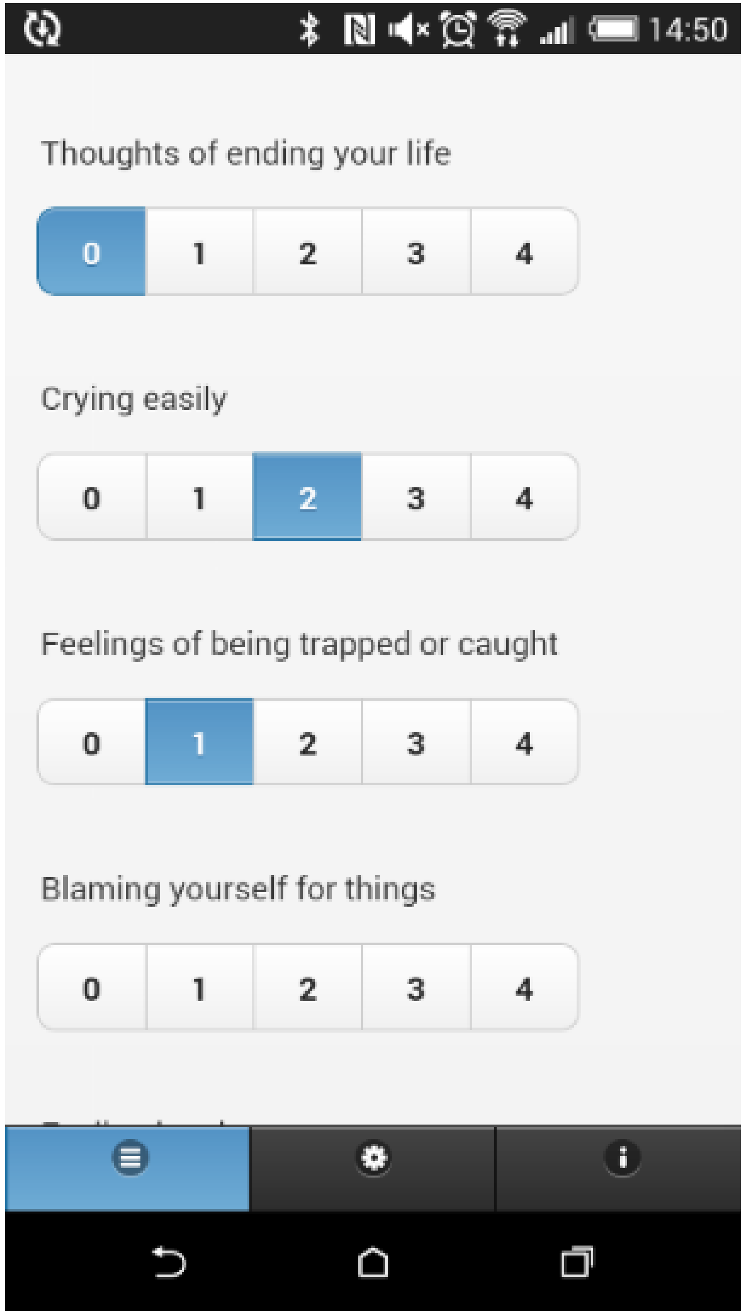
- N=1 graphical VAR
- N > 1 multilevel VAR
- panel data – a strange thing;

the mysterious latent trait makes its return!



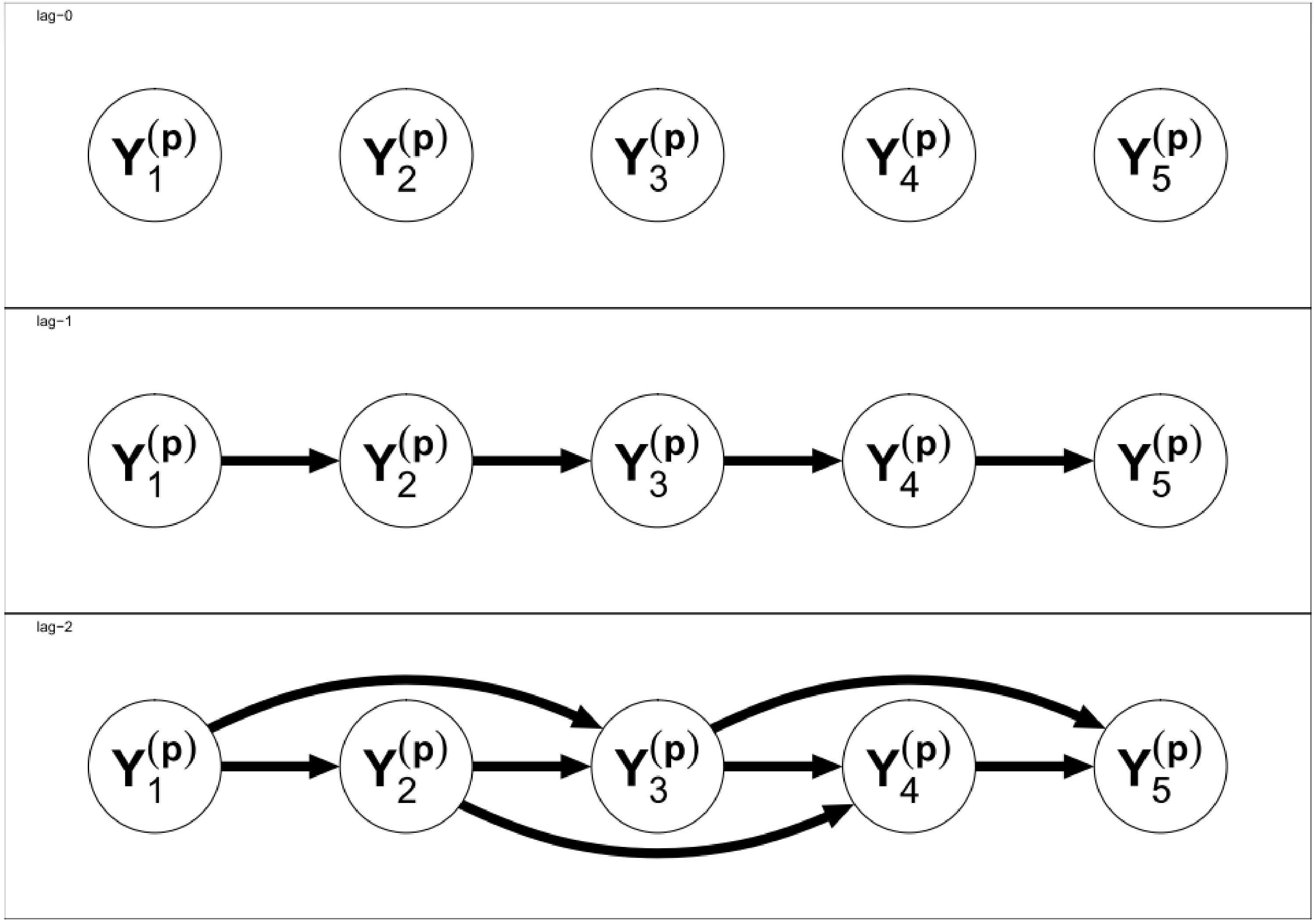
N = 1 ~ personalized networks

- Graphical VAR
- **One person** measured **several times** in a **short period**
- Cases cannot reasonably be assumed to be independent
  - Knowing someone's level of fatigue at a time point helps predict his or her level of fatigue at the next time point.
- Likelihood not easy to compute without two assumptions:
  - The time-series **factorize according to a graph**
  - The **model does not change over time**



Molenaar on ergodicity – read it!

- Only under extremely unrealistic assumptions are parameters estimated from cross-sectional data the same as parameters you would estimate from time-series data
  - Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2(4*), 201-218.
- Many cross-sectional datasets, however, are not administered in the same way as a question in an time-series study.
  - E.g., “are you a person that on average is ...?” rather than “did you feel ... in the last hour?”
- Such results should definitely **not be interpreted as within-person**, but allow for a between-person interpretation, which can be very interesting!



We will use the lag-1 factorization and assume the variables to be centred!

Temporal network

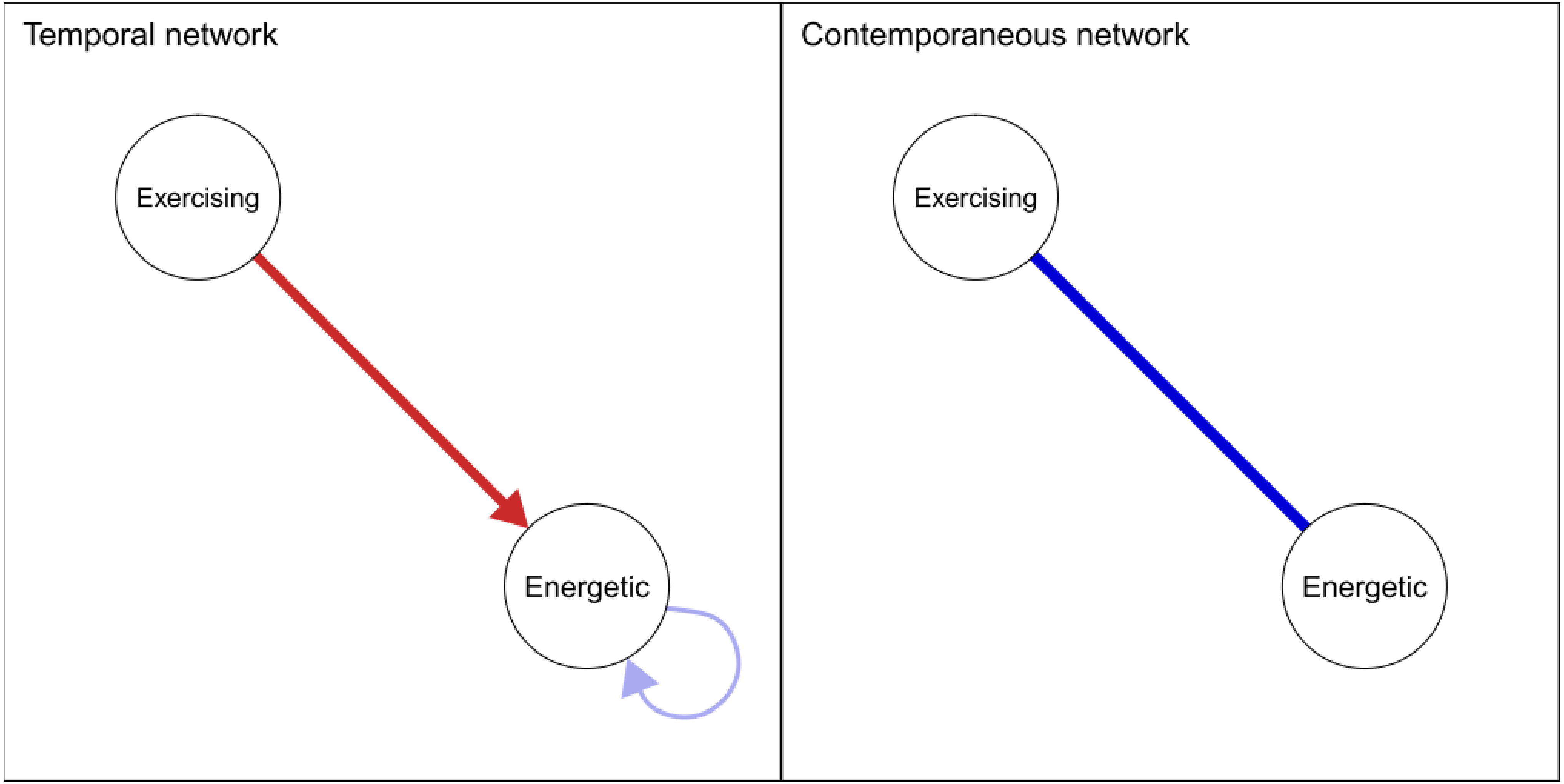
	relaxed	sad	nervous	concentration	tired	rumination	bodily.discomfort	time
31	5	6	4	5	6	4	4	2014-05-01 10:15:00
32	3	5	4	5	6	4	5	2014-05-01 13:15:00
33	NA	NA	NA	NA	NA	NA	NA	2014-05-01 16:15:00
34	3	5	4	5	6	4	4	2014-05-01 19:15:00
35	4	6	4	4	7	5	4	2014-05-01 22:15:00
36	4	3	3	5	5	2	3	2014-05-02 10:15:00
37	4	3	4	5	5	3	2	2014-05-02 13:15:00

- The temporal network shows that **one variable predicts another variable in the next measurement occasion**
- Granger causality – philosophical disclaimer
- Only temporal network from (graphical) VAR needed in predicting new responses

Contemporaneous network

	relaxed	sad	nervous	concentration	tired	rumination	bodily.discomfort	time
31	5	6	4	5	6	4	4	2014-05-01 10:15:00
32	3	5	4	5	6	4	5	2014-05-01 13:15:00
33	NA	NA	NA	NA	NA	NA	NA	2014-05-01 16:15:00
34	3	5	4	5	6	4	4	2014-05-01 19:15:00
35	4	6	4	4	7	5	4	2014-05-01 22:15:00
36	4	3	3	5	5	2	3	2014-05-02 10:15:00
37	4	3	4	5	5	3	2	2014-05-02 13:15:00

- The contemporaneous network shows that two variables predict one-another after taking temporal information into account
- Contains effects **faster than the time-window of measurement** (granularity)
  - Somatic arousal → anticipation of panic attack → anxiety
- The temporal network can be seen as a correction for dependent measurements in estimating the contemporaneous GGM


Temporal network
Exercising
Energetic
Contemporaneous network
Exercising
Energetic

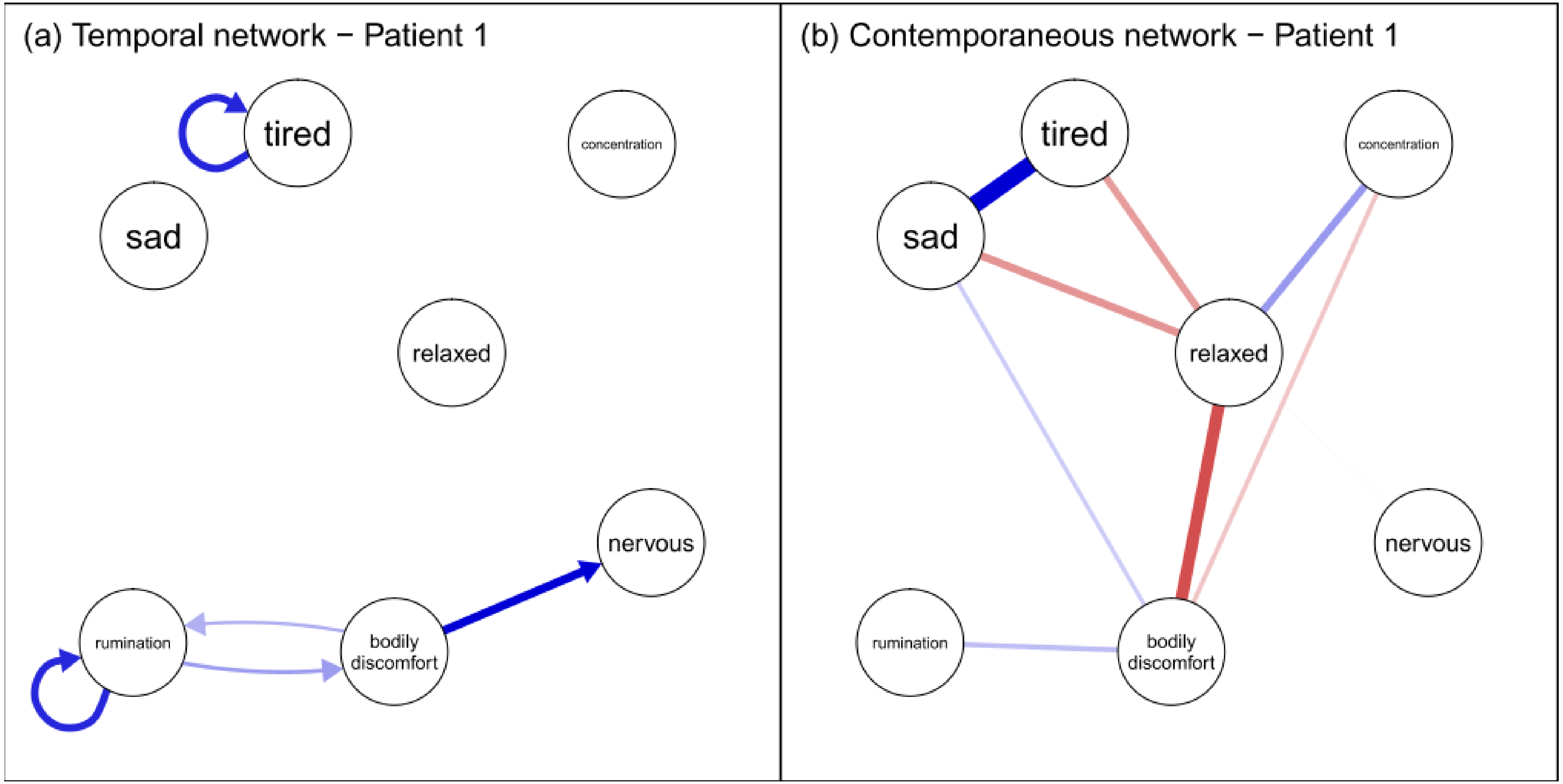
Estimation: 2 options

- Estimation without regularization
  - Stepwise non-regularized estimation implemented in the `psychonetrics` package (psychonetrics.org)
  - Inlcludes full information maximum likelihood (FIML) for missing data handling
- Model selection possibly using multiple regressions
  - LASSO estimation with EBIC model selection implemented in the `graphicalVAR` package
  - Missings can be handled using the Kahlman filter in the `impute.ts` package

Assumptions

- Stationarity
  - Plausible when data is obtained in short time-span, less plausible if data is obtained in longer time-span
- Stationary means
  - Can be assessed by regressing each time-series on time itself as predictor
  - Detrending is possible: for example one can remove a linear trend (see practical)
- Equidistant measurements
  - With multiple measurements per day, by default violated due to nights
  - Remove nights, or model nights as missing observations
  - Continuous time modelling poses promising ways to deal with non-equidistant measurements

Personalized Networks in Clinical Practice



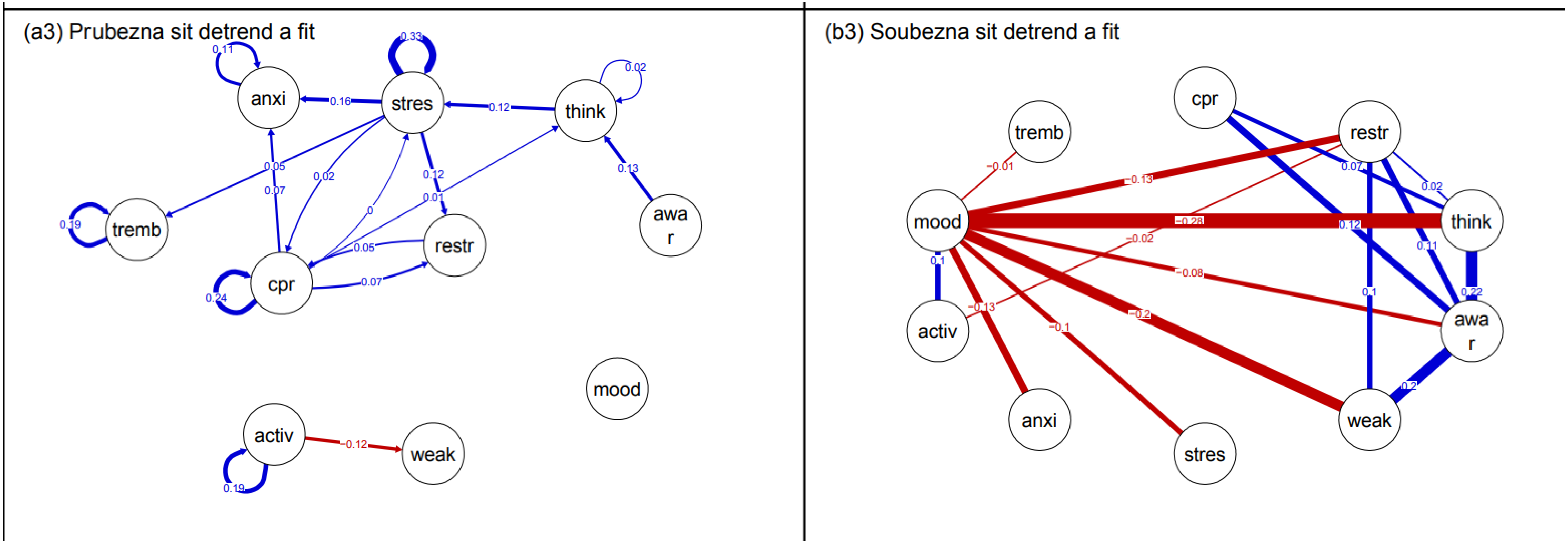
Contemporaneous network: conditional concentration given t − 1

Temporal network: regression coefficients between t − 1 and t

Pacient 1

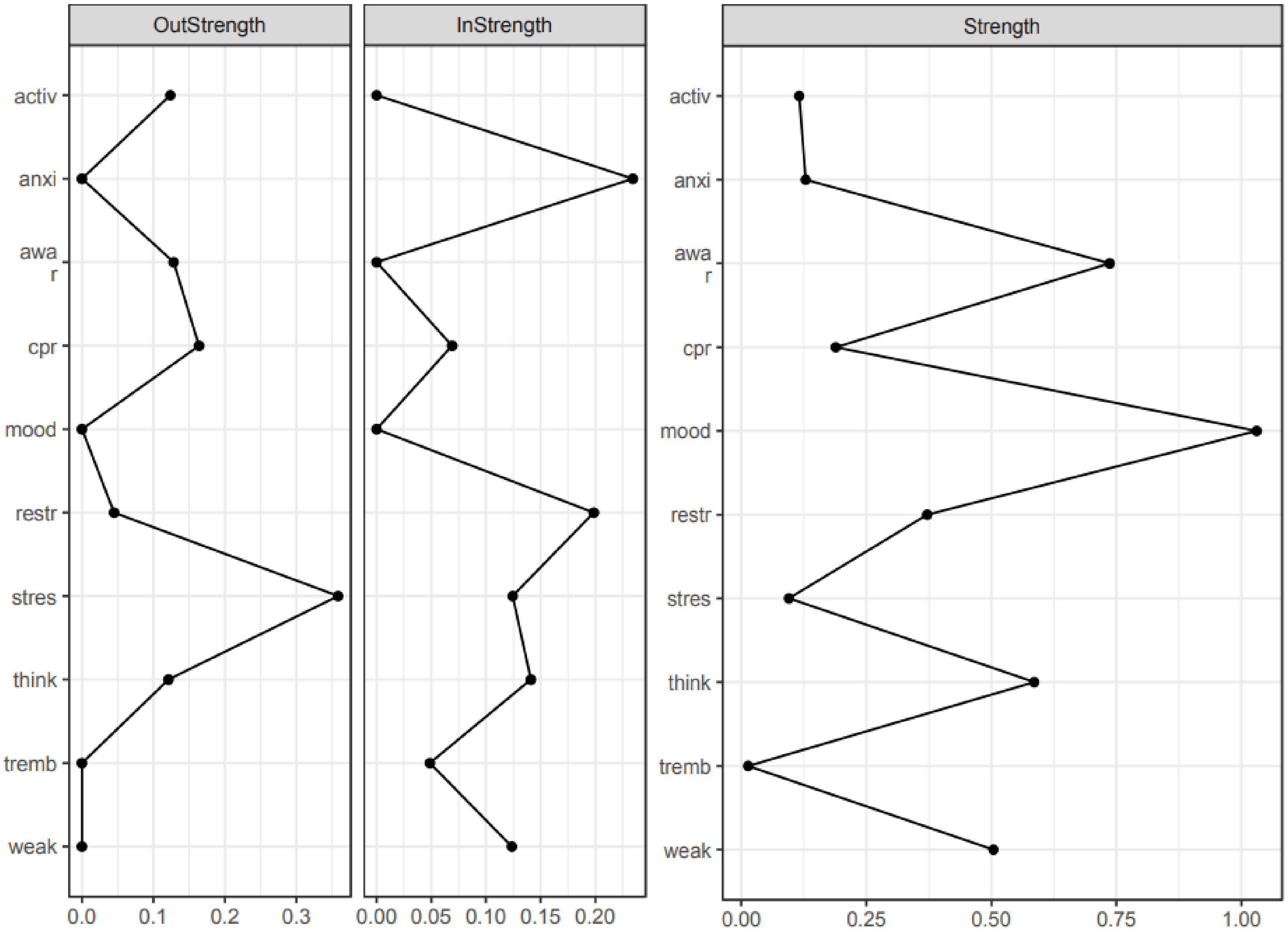
Muž, 42 let. Přijat na skupinovou terapii v pobytovém stacionáři Psychosomatické kliniky v Praze. V současné době na čekací listině. U pacienta se projevovaly následující symptomy: úzkost, nervozita, bušení srdce, ztuhlost, výboje na šíji, rozostřené vidění, náhlá ztráta energie a bolesti nohou, otoky kloubů (diagnostikována revmatoidní artritida, nyní v remisi). Za hlavní spouštěč byla v anamnéze označena pracovní i mimopracovní zátěž. V rozhovoru před spuštěním měření byl seznam symptomů aktualizován. Vzhledem k upravenému pracovnímu režimu se pracovní zátěž povedlo omezit. Situace u mimopracovní zátěže přetrvává v důsledku vážně nemocného blízkého člověka v rodině. Pacient je velice aktivní. Sportuje, věnuje se horské turistice, má dvě zaměstnání. Potíže se začaly projevovat po velké zátěži, která byla spojená s vlastnoruční rekonstrukcí rodinného domu. Významnou roli hrála také stresující a špatně organizovaná práce v jeho tehdejším zaměstnání.

Nováček, T. (2019). Personalizovaný síťový model pro medicínsky nevysvětlené symptomy. Nepublikovaná diplomová práce, Masarykova Univerzita.



Pozn.: crp = mám pocit tlaku v oblasti hrudníku, restr = moje tělesné obtíže mi zabraňují dělat to, co chci, think = aktuálně myslím na své tělesné obtíže, awa.r = právě v tomto okamžiku si všímám, kde v těle co cítím, weak = cítím se zeslábly, tremb = třesou se mi ruce / nohy, stres = jsem v napětí, anxi = mám z něčeho obavy, activ = právě se věnuji aktivitě, která mě naplňuje, mood = celkově se cítím takto: (vizualizace grafickými znaky s mírou zamračení – úsměvu), hand = cítím bolest v dlaních / chodidlech, neck = Cítím výboje na šíji, help.r = aktuálně potřebuji kontakt s jinými lidmi, Coll = aktuálně jsem s kolegou, Home = aktuálně jsem doma.

Nováček, T. (2019). Personalizovaný síťový model pro medicínsky nevysvětlené symptomy. Nepublikovaná diplomová práce, Masarykova Univerzita.

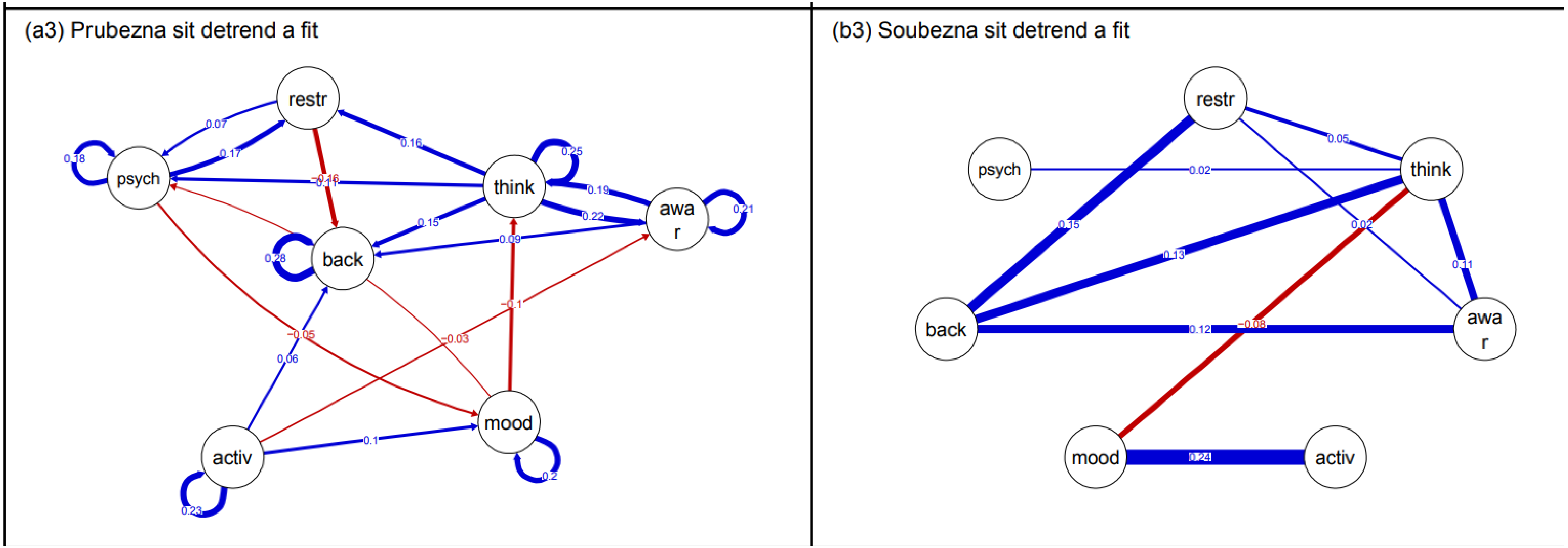


Nováček, T. (2019). Personalizovaný síťový model pro medicínsky nevysvětlené symptomy. Nepublikovaná diplomová práce, Masarykova Univerzita.

Pacient 2

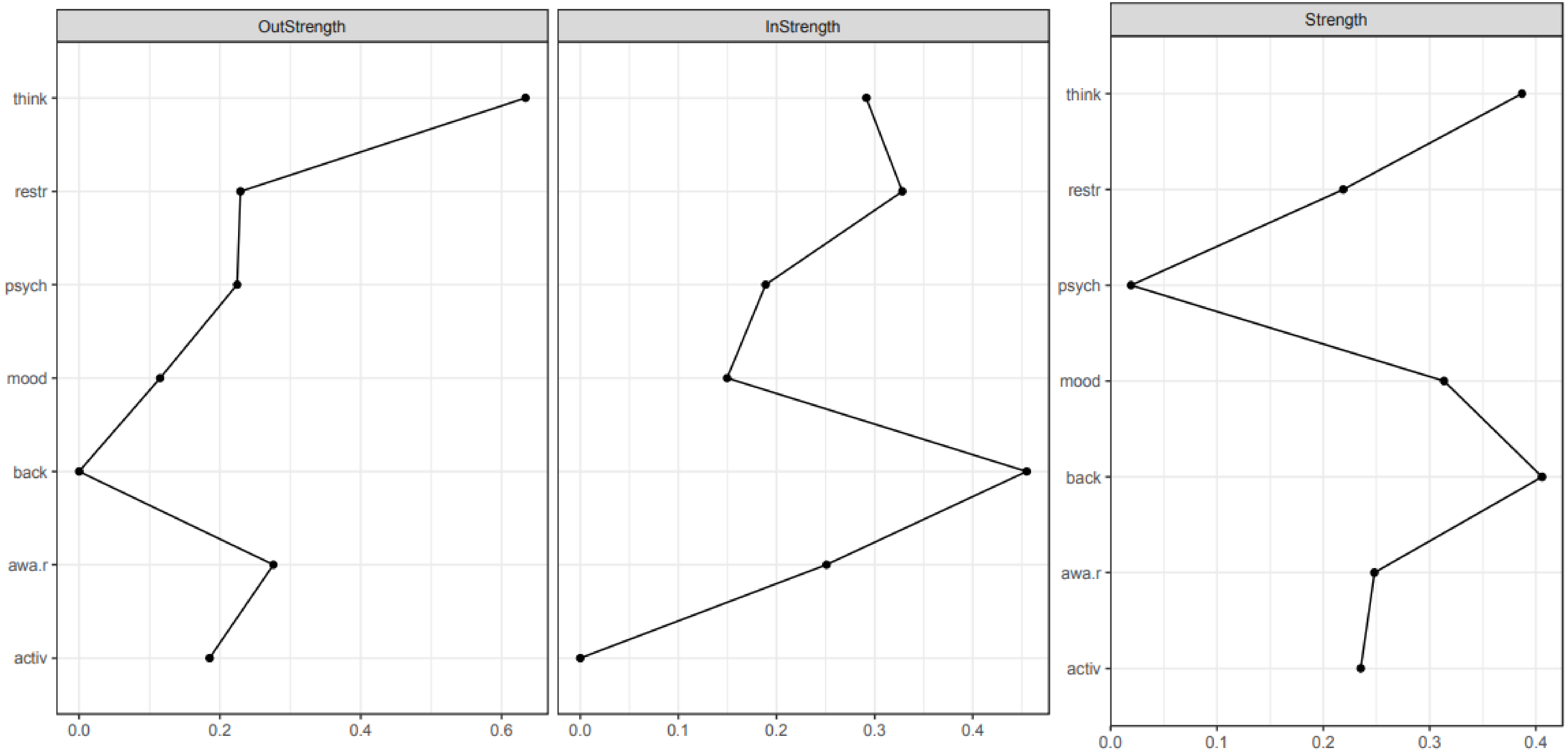
Žena, 51 let. Přijata na skupinovou terapii do pobytového stacionáře Psychosomatické kliniky v Praze. V době výzkumu byla na čekací listině. Týden po ukončení sběru dat nastoupila do terapie. Mezi klíčové symptomy patří bolesti zad a svalů jako reakce na zátěžové situace. Před započetím sběru dat byl s pacientkou uskutečněn rozhovor pro aktualizaci symptomů a možných spouštěčů. Pacientka při rozhovoru měla silné bolesti zad, nicméně dokázala si nastavit sezení v křesle, abychom dokázali vše potřebné zvládnout.

Nováček, T. (2019). Personalizovaný síťový model pro medicínsky nevysvětlené symptomy. Nepublikovaná diplomová práce, Masarykova Univerzita.



Poznámka. *crp = mám pocit tlaku v oblasti hrudníku, restr = moje tělesné obtíže mi zabraňují dělat to, co chci, think = aktuálně myslím na své tělesné obtíže, awa.r = právě v tomto okamžiku si všímám, kde v těle co cítím, stres = jsem v napětí, anxi = mám z něčeho obavy, activ = právě se věnuji aktivitě, která mě naplňuje, mood = celkově se cítím takto: (vizualizace grafickými znaky s mírou zamračení – úsměvu), trape = bolí mě trapézový sval, back = bolí mě záda v oblastí beder, help.r = aktuálně potřebuji kontakt s jinými lidmi, psych = cítím se psychicky vyčerpaná, muscl = bolí mě svaly, nause = je mi nevolno , avoid = vyhýbám se fyzické aktivitě, protože aktuálně šetřím síly, fatiq = jsem unavená, Home = aktuálně jsem doma, Child = aktuálně jsem s vnučkou/dcerou.*

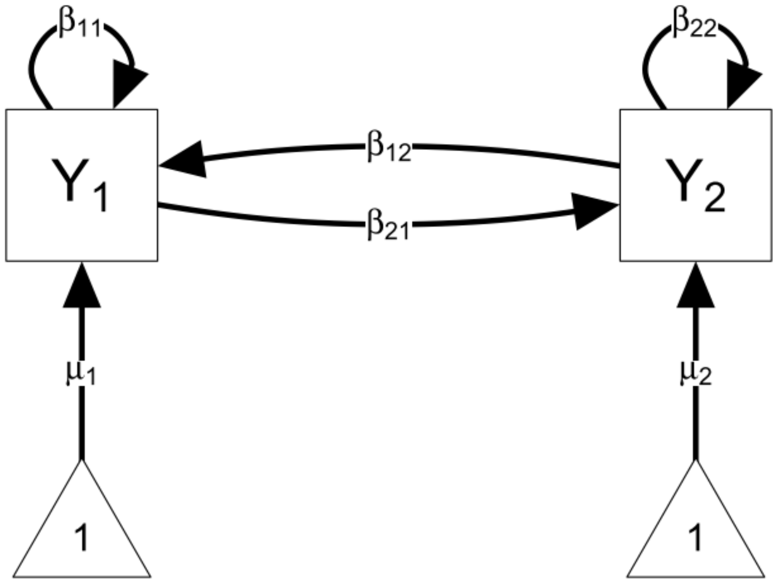
Nováček, T. (2019). Personalizovaný síťový model pro medicínsky nevysvětlené symptomy. Nepublikovaná diplomová práce, Masarykova Univerzita.



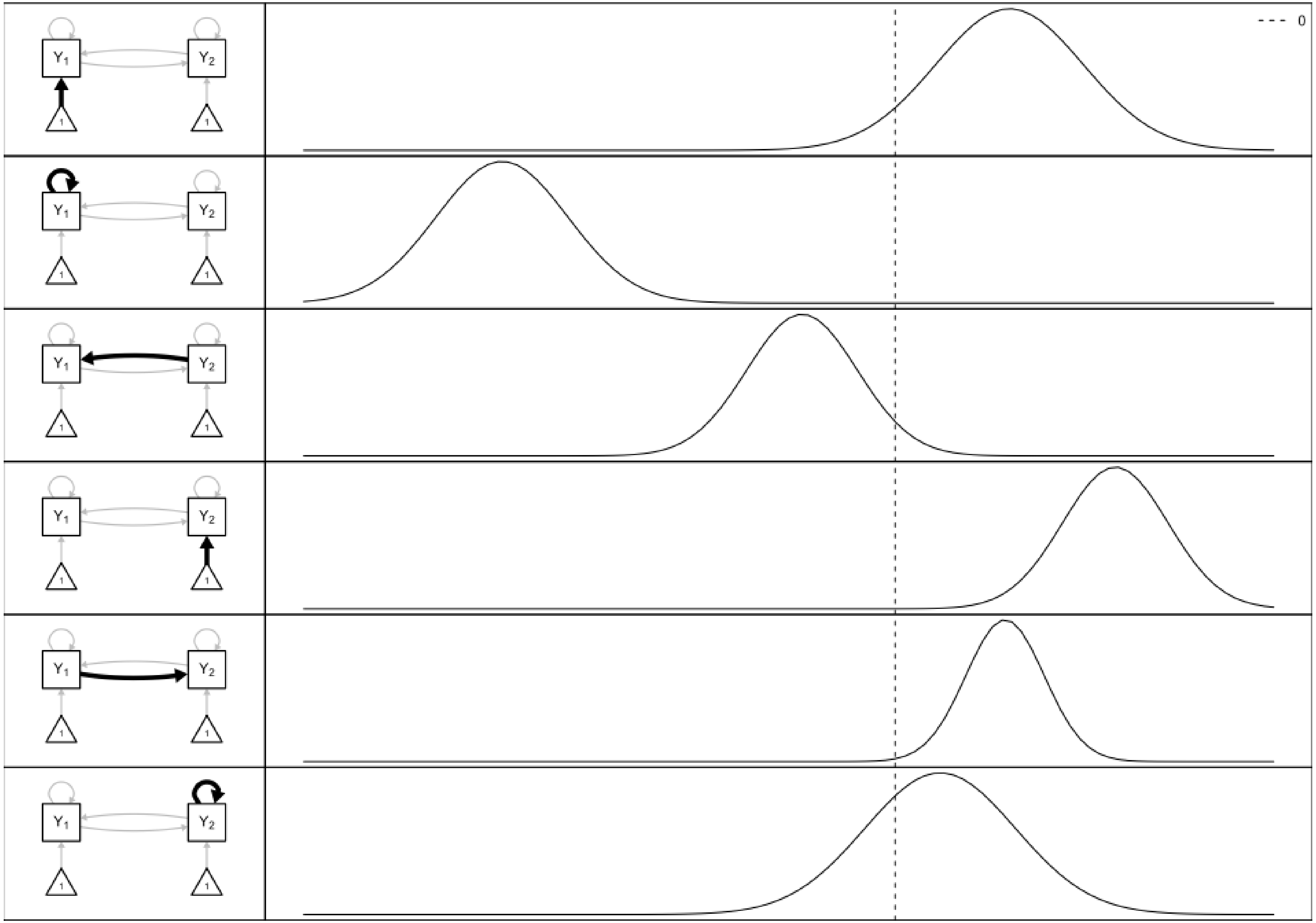
Nováček, T. (2019). Personalizovaný síťový model pro medicínsky nevysvětlené symptomy. Nepublikovaná diplomová práce, Masarykova Univerzita.

N > 1: Multi-level VAR

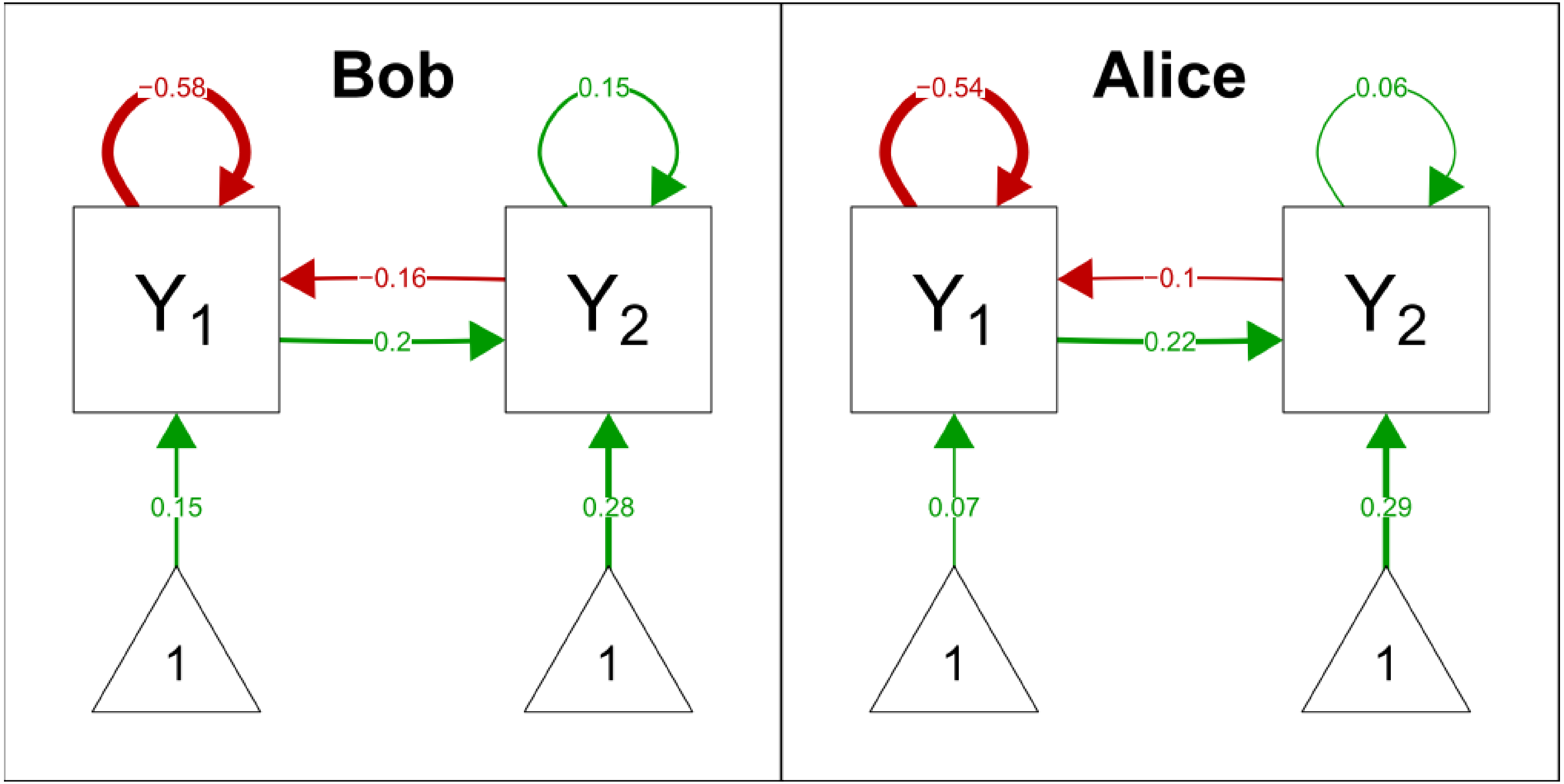
- Each subject is assumed to have their own temporal and contemporaneous VAR model
- Between-subject effects can be modeled in a level 2 model
- which gives 3 networks in total!



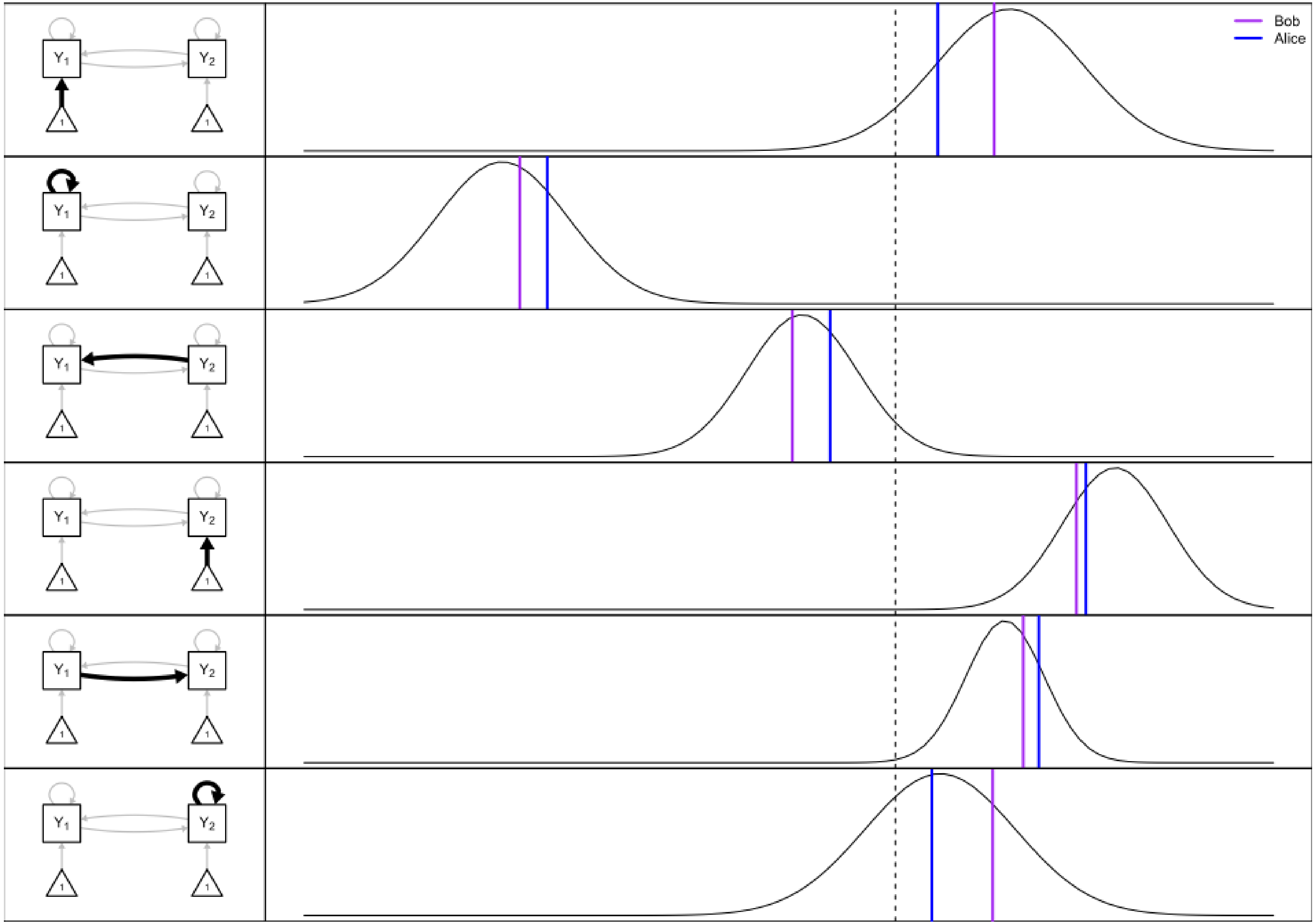
Each Parameter has a Distribution



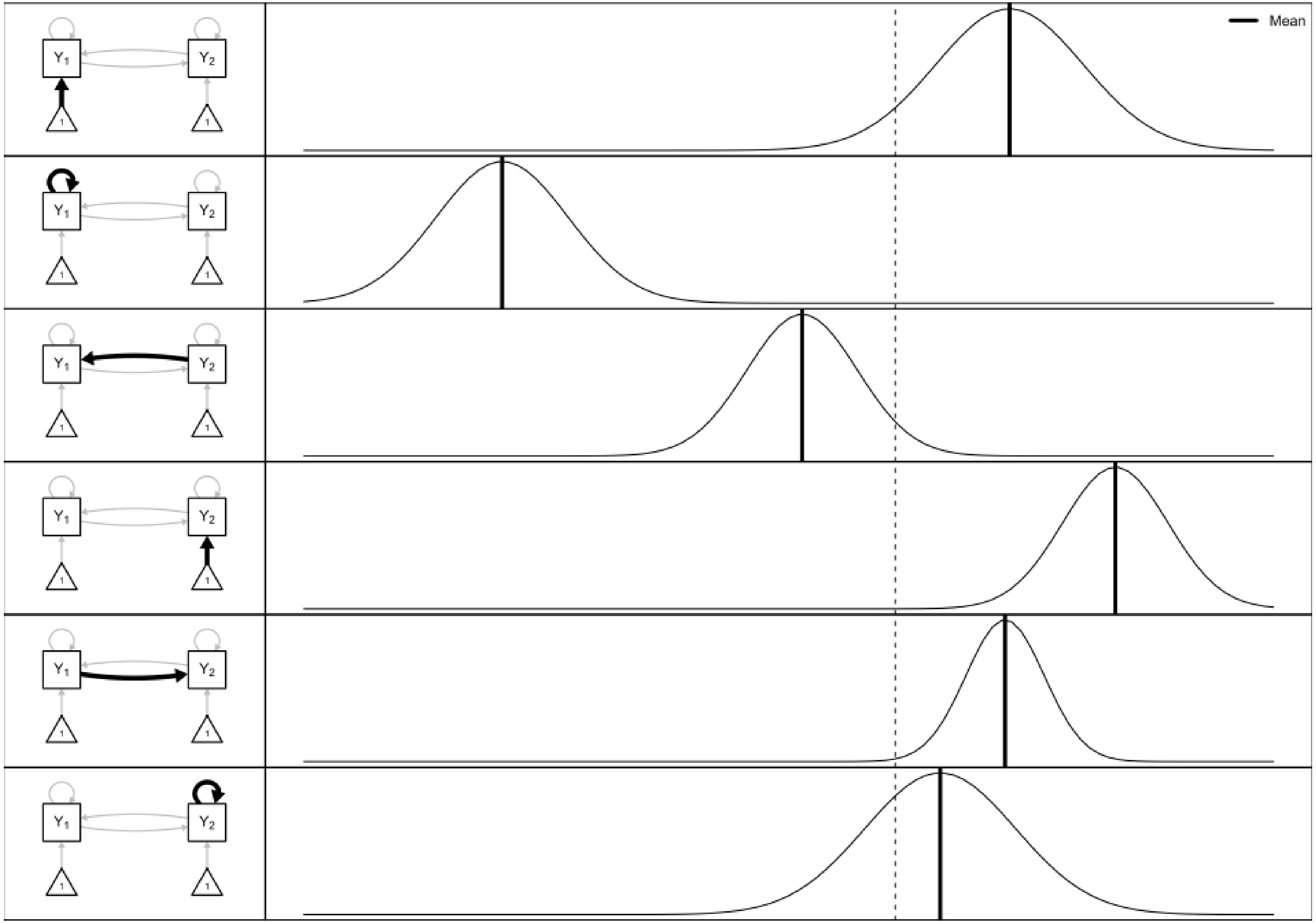
Individual Networks



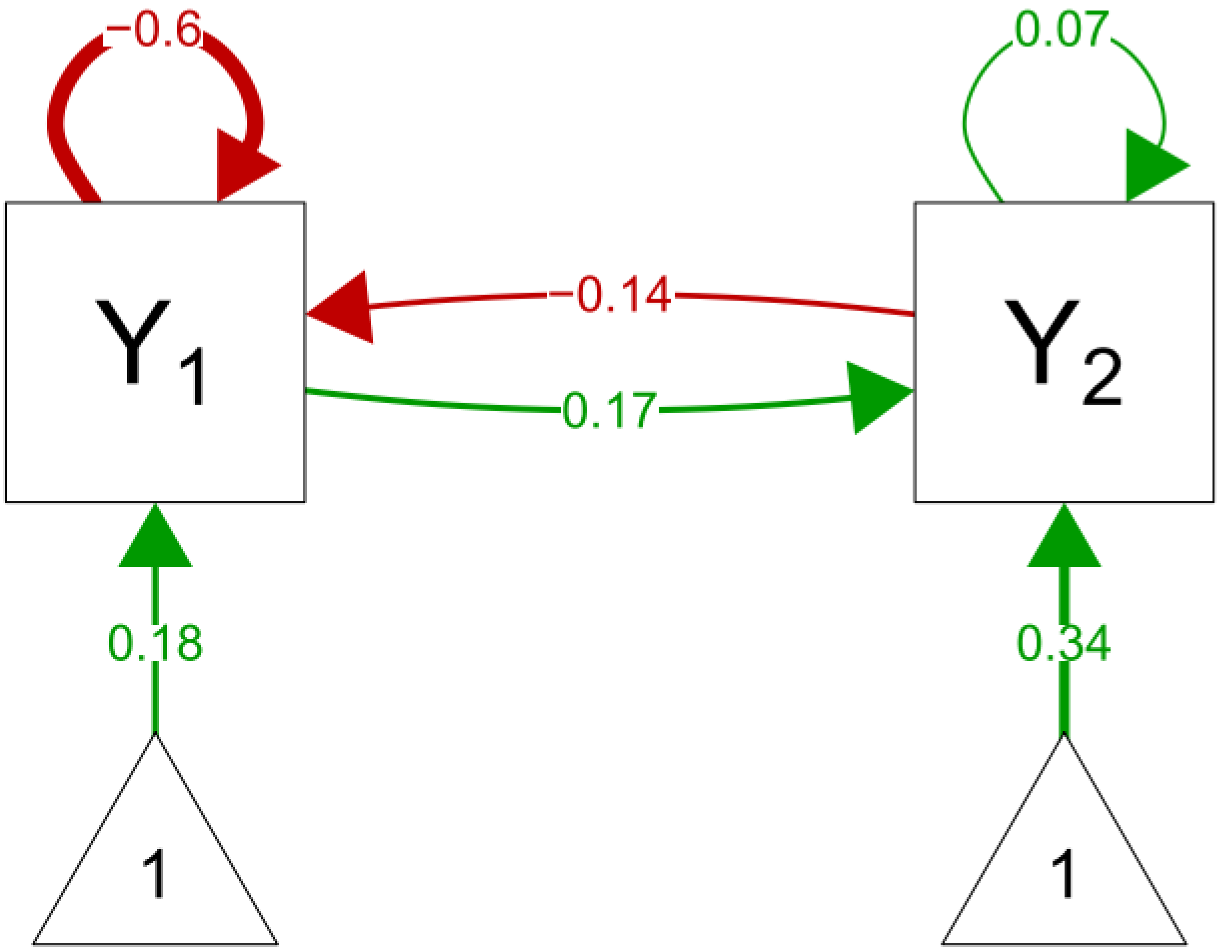
Random Effects



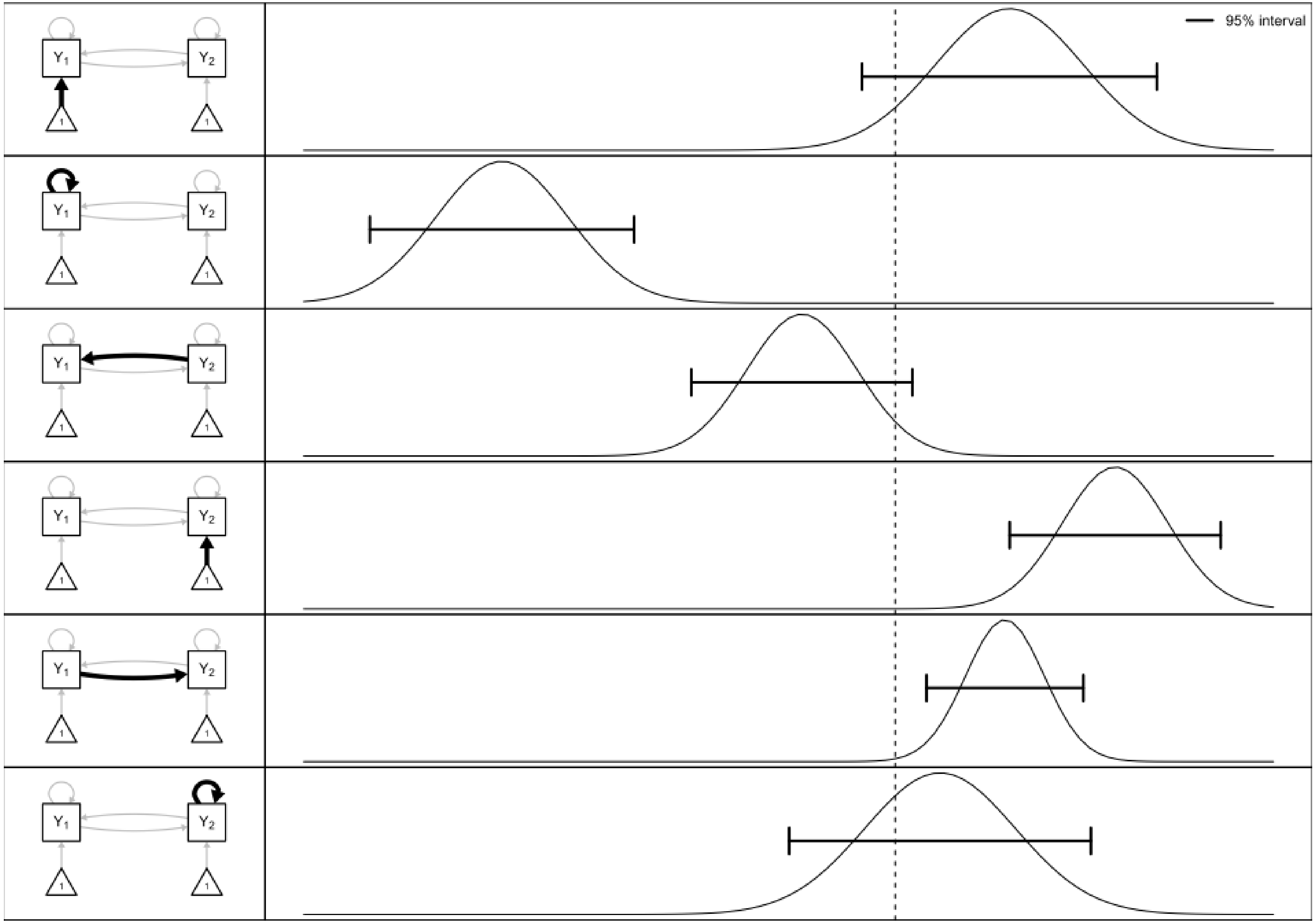
Fixed Effects

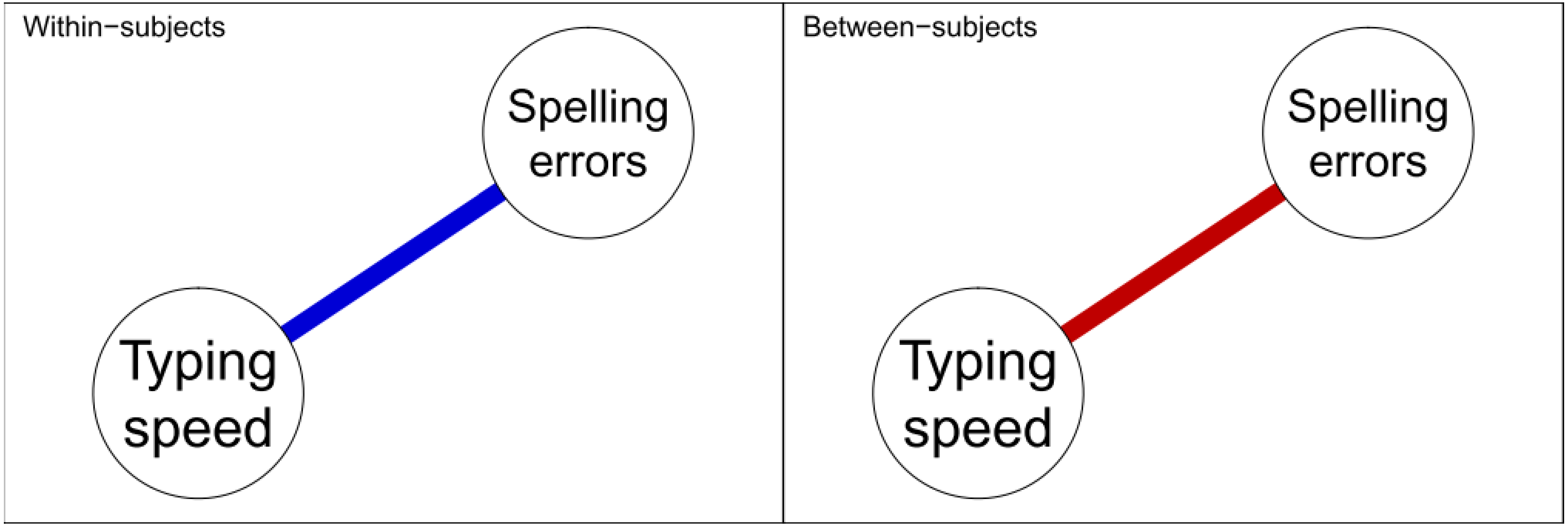


Fixed Effects

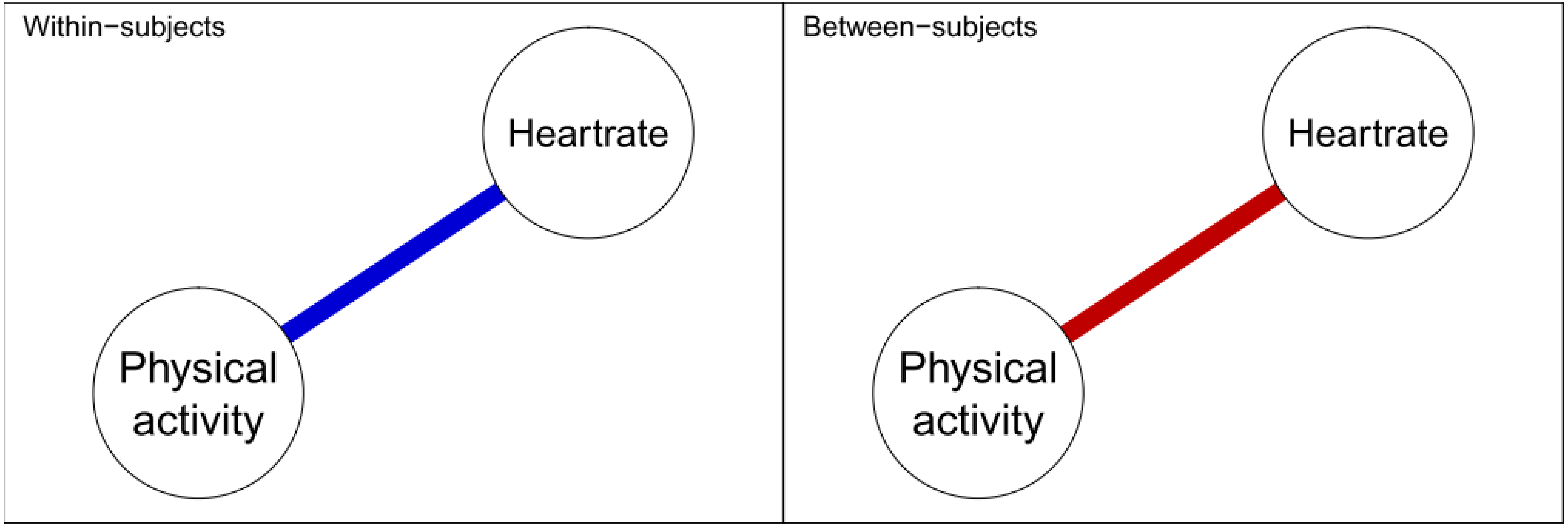


Individual Differences





Example based on Hamaker, E. L. (2012). Why Researchers Should Think 'Within-Person': A Paradigmatic Rationale. *Handbook of Research Methods for Studying Daily Life*. The Guilford Press New York, NY, 43–61.



Example kindly provided by Ellen Hamaker and based on Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York, NY, USA: Routledge.

Estimation

Both methods use within-person centring using the sample mean per person. This will highly bias results with too few measurements per person (< 20)!

Two-step Multilevel VAR

Step 1:

Node-wise multilevel regressions of variables on within-person centred lagged predictors (temporal effects) and person-wise means (between-subject effects)

Step 2:

Node-wise multilevel regressions using residuals from step 1 (contemporaneous effects)

Implemented in `mlVAR`

Pooled and individual LASSO estimation

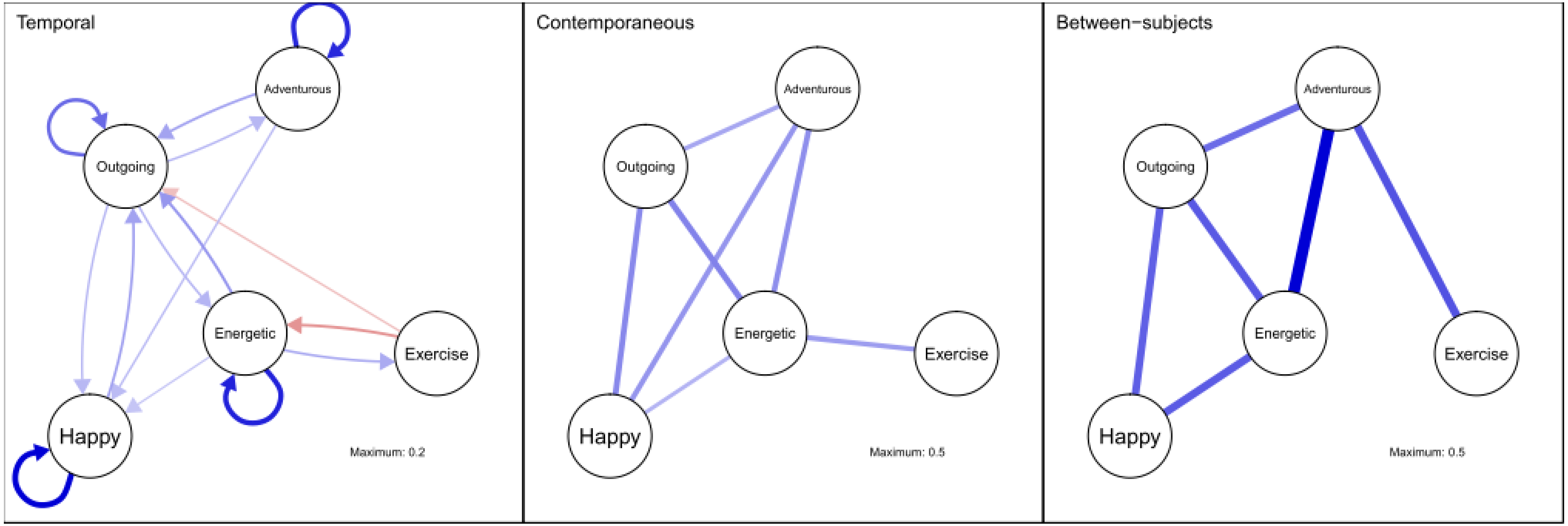
Pool all within-person centred datapoints and estimate graphical VAR model to obtain fixed effects.

Estimate GGM on means of subjects for between-subject effects.

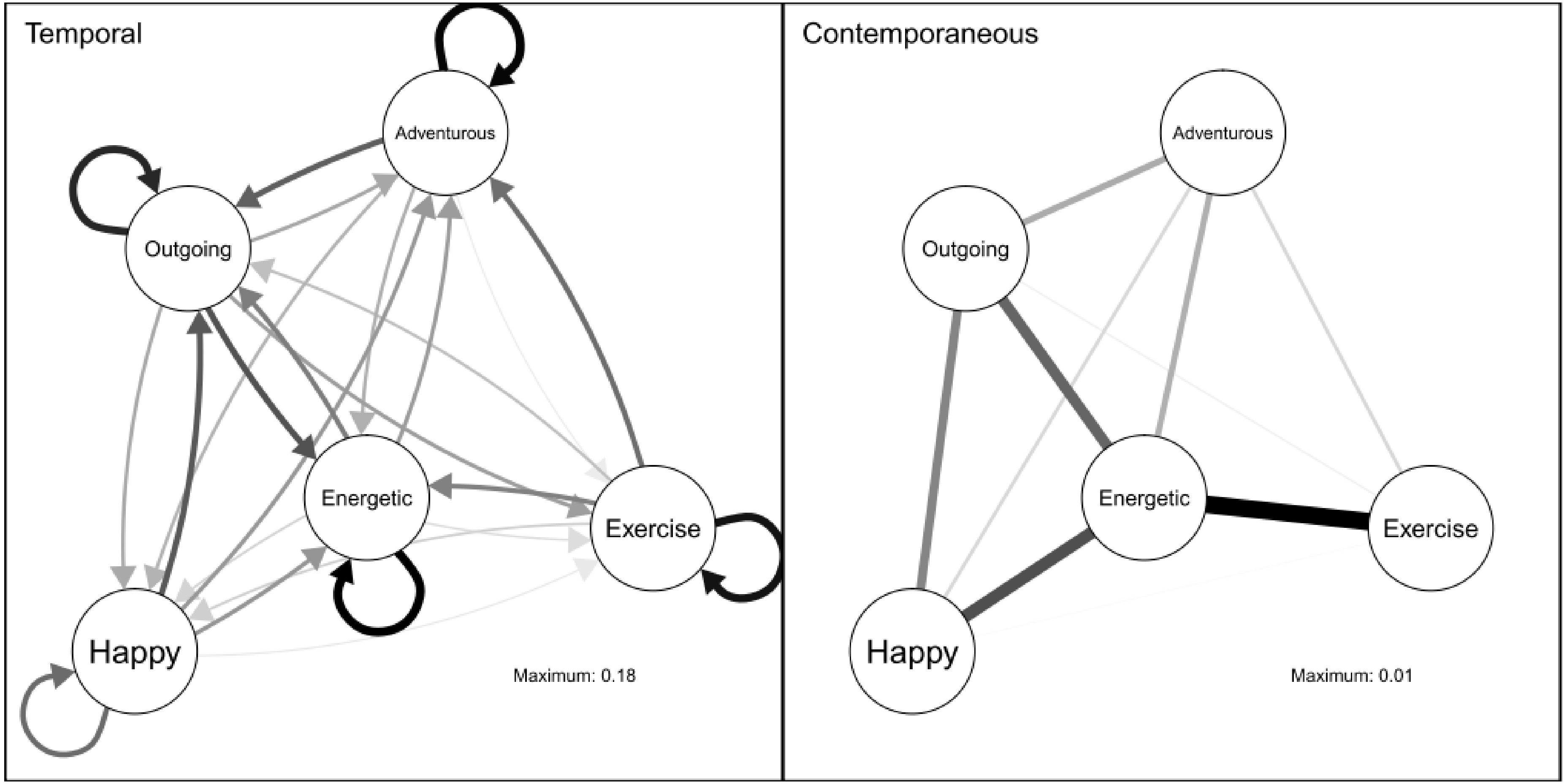
Estimate N = 1 networks for every subject for individual effects.

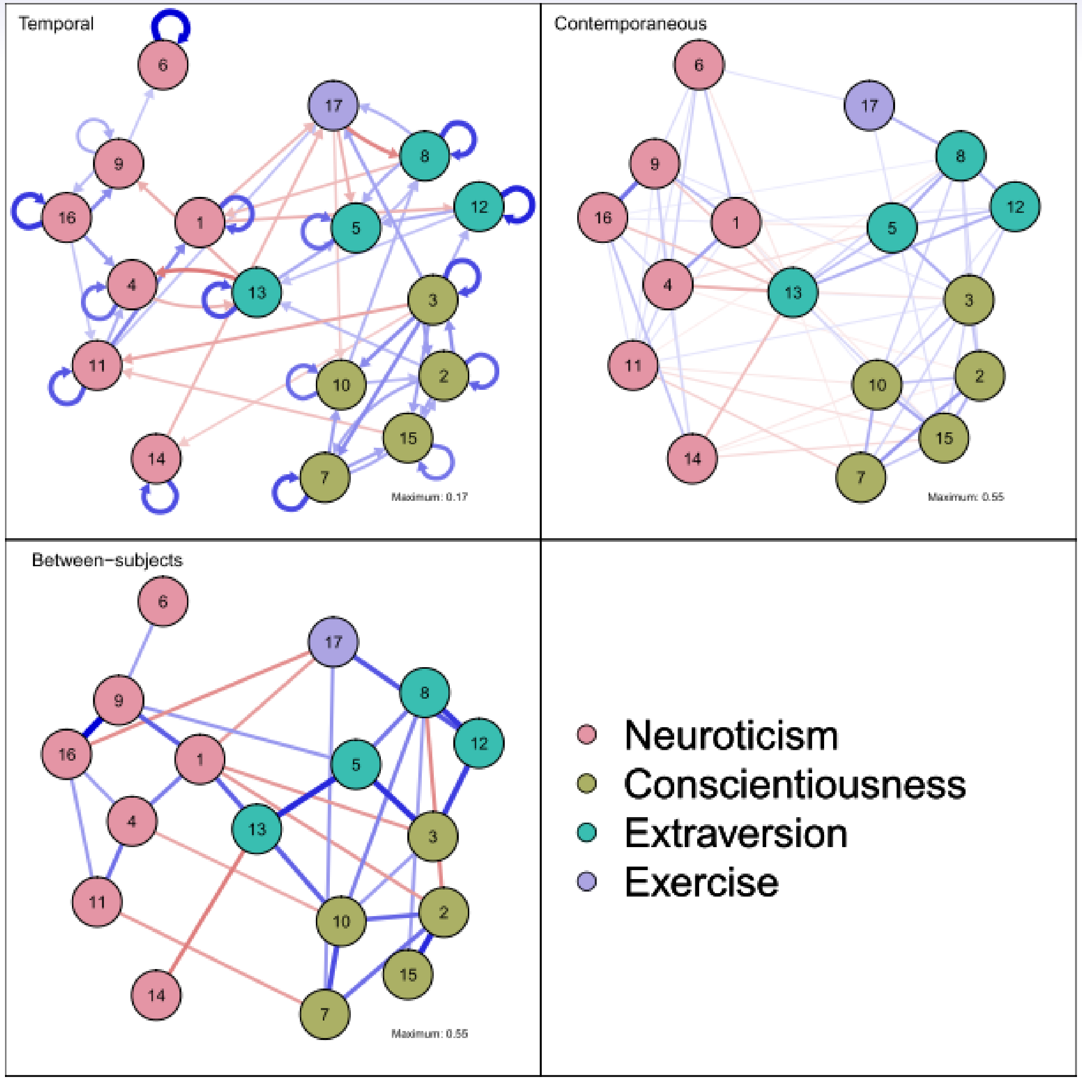
Implemented in `graphicalVAR` (`mlGraphicalVAR`)

Fixed effects



Individual Differences

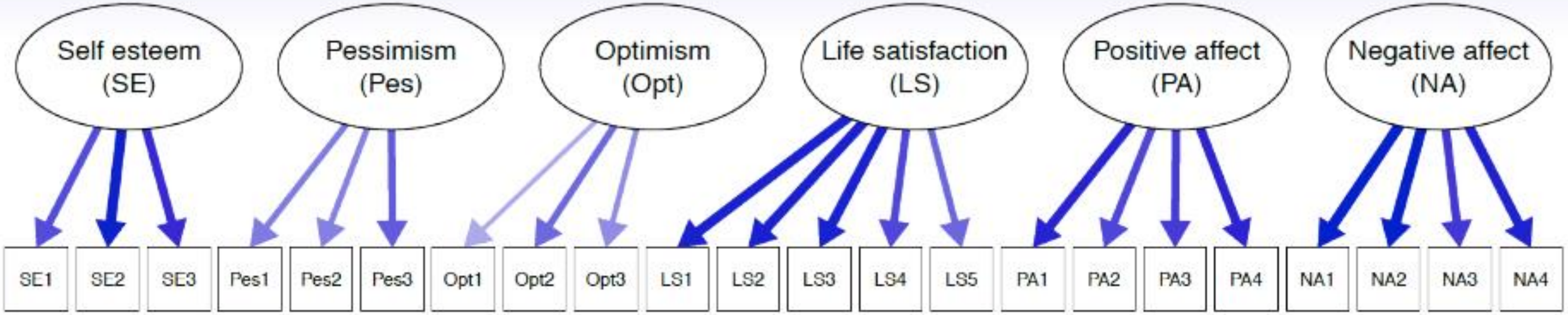




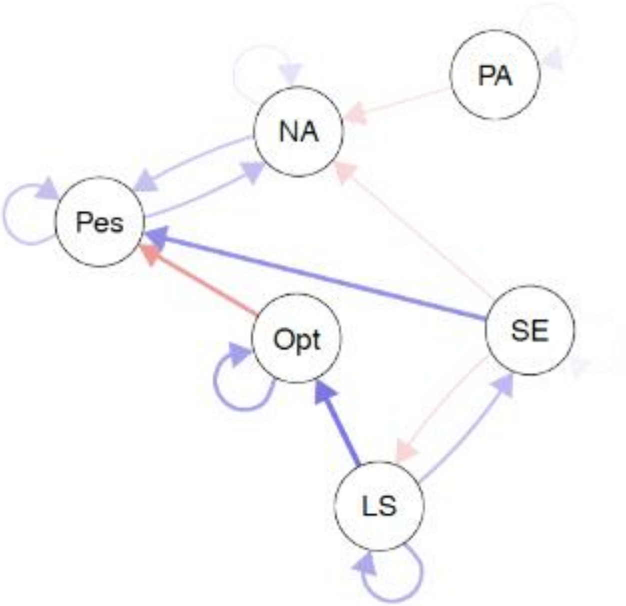
1 = "Worried"; 2 = "Organized"; 3 = "Ambitious"; 4 = "Depressed"; 5 = "Outgoing"; 6 = "Self-Conscious"; 7 = "Self-Disciplined"; 8 = "Energetic"; 9 = "Frustrated"; 10 = "Focused"; 11 = "Guilty"; 12 = "Adventurous"; 13 = "Happy"; 14 = "Control"; 15 = "Achieved"; 16 = "Angry"; 17 = "Exercise."

Panel data

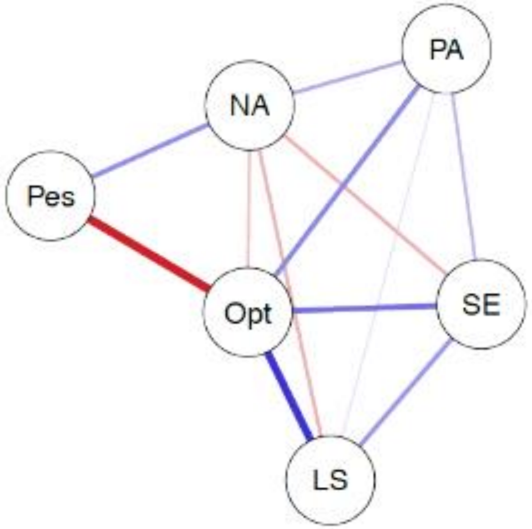
- An often occurring datatype between cross-sectional and time-series data is **panel data**
- **Many people** measured on a **few repeated occasions**
- When the measurement occasions are the same for each subject (e.g., baseline, one month after baseline, etcetera), structural equation modeling (SEM) offers many possible analyses methods closely related to network modeling:
  - e.g., latent change-score models, latent growth curve models, random-intercept cross-lagged panel models.
- Like `mlVAR`, this assumes **stationarity**



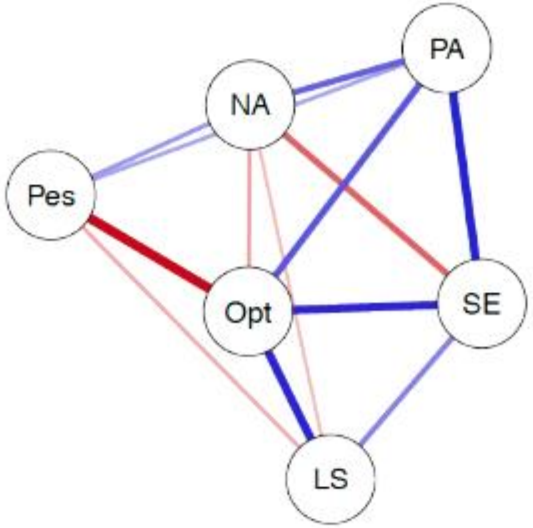
(a) Estimated factor loadings



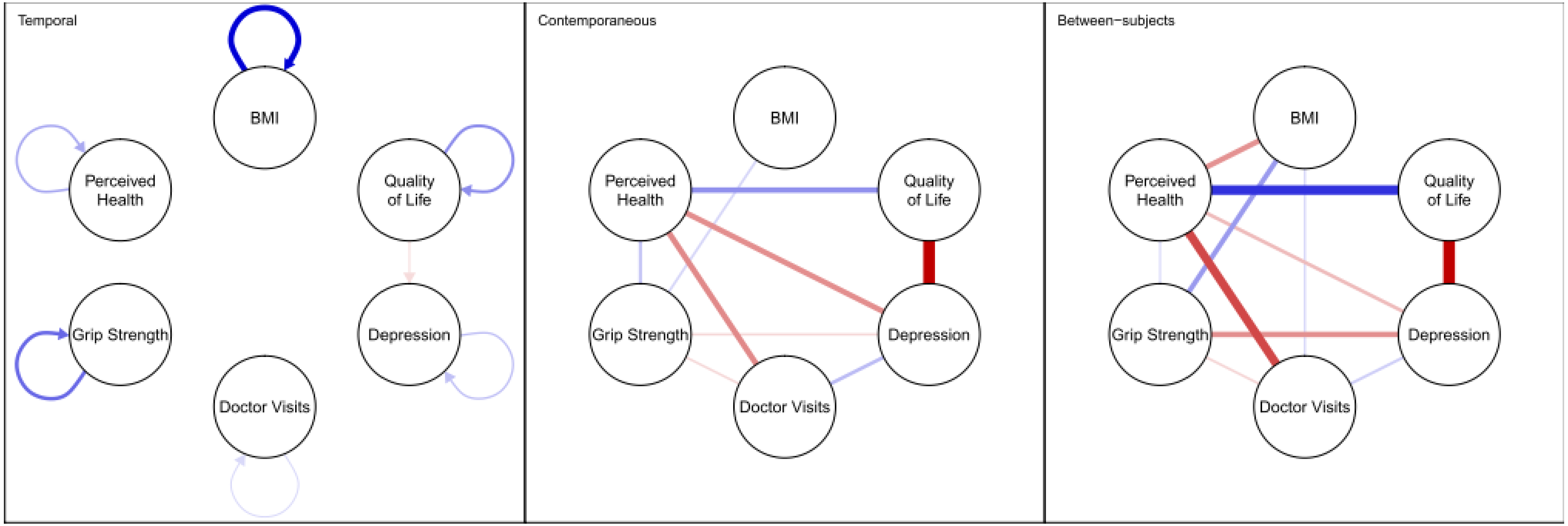
(b) Estimated temporal network, standardized to partial directed correlations.



(c) Estimated contemporaneous partial correlation network.



(d) Estimated between-subjects partial correlation network.


Temporal
BMI
Perceived Health
Quality of Life
Grip Strength
Depression
Doctor Visits
Contemporaneous
BMI
Perceived Health
Quality of Life
Grip Strength
Depression
Doctor Visits
Between−subjects
BMI
Perceived Health
Quality of Life
Grip Strength
Depression
Doctor Visits

Thank you for your attention and good luck with the assignment!