

Subgrouping with Graphical VAR

Jonathan J. Park
The Pennsylvania State University
Department of Human Development and Family Studies

September 21, 2020



Opening Thoughts

“Ludwig Boltzmann, who spent much of his life studying Statistical Mechanics, died in 1906, by his own hand. Paul Ehrenfest, carrying on the work, died similarly in 1933. Now it is our turn to study Statistical Mechanics.”

“Perhaps it will be wise to approach the subject cautiously.”

in *States of Matter*, by David. L. Goodstein, 1975, Dover N.Y.



Overview

- VAR and gVAR
- 'Multilevel' Graphical VAR
- Subgrouping
- Demonstration

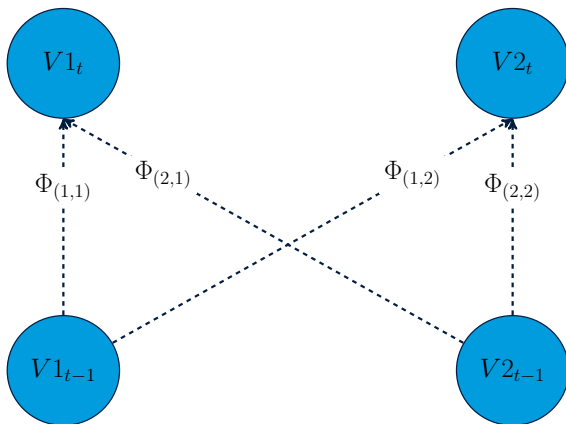


General Ideas & Goals

Extend the chain graphical VAR to handle heterogeneous samples



Standard VAR Model



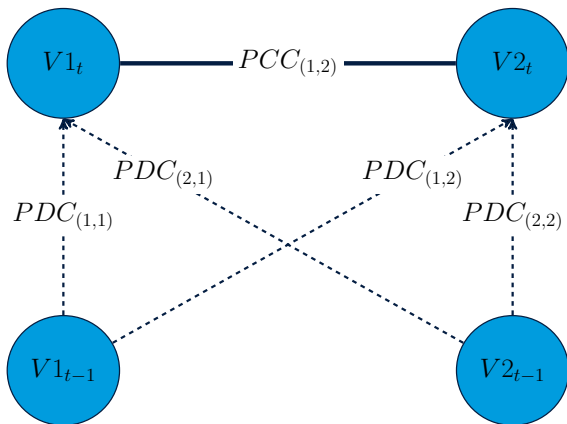
Standard VAR

$$\eta(t) = c + \Phi\eta(t-1) + \zeta(t)$$

- $\eta(t)$ = p variate vector of scores at time, t
- c = p variate vector of constants
- Φ = $p \times p$ dimensional matrix of lagged regression coefficients
- $\eta(t-1)$ = p variate vector of scores at a given lag
- $\zeta(t)$ = p variate vector of residuals

$$\zeta \sim N(0, \Psi)$$

Graphical VAR



Graphical VAR

- Graphical VAR builds upon the VAR and estimates a network of contemporaneous partial correlations using the inverse residual covariance matrix
 - i.e., $\mathbf{K} = \boldsymbol{\Psi}^{-1} = \text{cov}[\zeta_{(t)}, \zeta_{(t)}^t]^{-1}$
- Contemporaneous effects interpreted as $X \leftrightarrow Y$ conditioned upon all other pairwise associations

VAR \rightarrow gVAR - The Partial Contemporaneous Network

The inverse of the residual covariance matrix (i.e., $\mathbf{K} = \Psi^{-1}$) can be transformed into a network of partial contemporaneous correlations (PCCs) using the following:

$$PCC(X_{i(t)}, X_{j(t)}) = -\frac{K_{ij}}{\sqrt{K_{ii}K_{jj}}}$$

Where:

- K_{ij} = the element of \mathbf{K} at coordinates (i, j)
- K_{ii} = the diagonal element of \mathbf{K} associated with item, i
- K_{jj} = the diagonal element of \mathbf{K} associated with item, j

VAR → gVAR - The Partial Directed Network

The lagged relationships of the VAR can also be standardized using information from \mathbf{K} to form a network of partial directed correlations (PDCs) using the following formula:

$$PDC(X_{i(t)}, X_{j(t-1)}) = \frac{\Phi_{ij}}{\sqrt{\Psi_{ii}K_{jj} + \Phi_{ij}^2}}$$

Where:

- Φ_{ij} = the regression coefficient of i on j
- Ψ_{ii} = the residual variance of the outcome at time, t
- K_{jj} = the diagonal element of \mathbf{K} associated with item, j

The 'Multilevel' Graphical VAR

A.K.A., Chain Graphical VAR; Epskamp et al., (2018)

$$\eta_{(t,q)} = \mu_q + \Phi_p(\eta_{(t-1,q)} - \mu_q) + \zeta_{(t,q)}$$

where

- $\eta_{(t,q)}$ = p variate vector of scores at time, t , for subject q
- μ_p = the person-specific mean vector for subject, q
- Φ_p = $p \times p$ dimensional matrix of lagged regression coefficients for subject q
- $\zeta_{(t,q)}$ = p variate residual vector at time, t for subject q

$$\begin{aligned}\zeta_{(T,q)} &\sim N(0, \Psi_q) \\ \Psi_q^{-1} &= \mathbf{K}_q^{(\Psi)}\end{aligned}$$

Assumptions

Assuming a subject picked at random, and data are grand-mean centered, we expect the following (Epskamp et al., 2018):

$$\begin{aligned}\mathbb{E}(\mu_Q) &= 0 \\ \mathbb{E}(\mathbf{B}_Q) &= \mathbf{B}_* \\ \mathbb{E}(\mathbf{K}_Q^{(\Psi)}) &= \mathbf{K}_*^{(\Psi)}\end{aligned}$$

where

- \mathbf{B}_* = average $p \times p$ dimensional lagged effects matrix
- $\mathbf{K}_*^{(\Psi)}$ = average $p \times p$ dimensional precision matrix

e.g., $(\mathbf{B}_p - \mathbf{B}_*)$ would be the 'random effects'

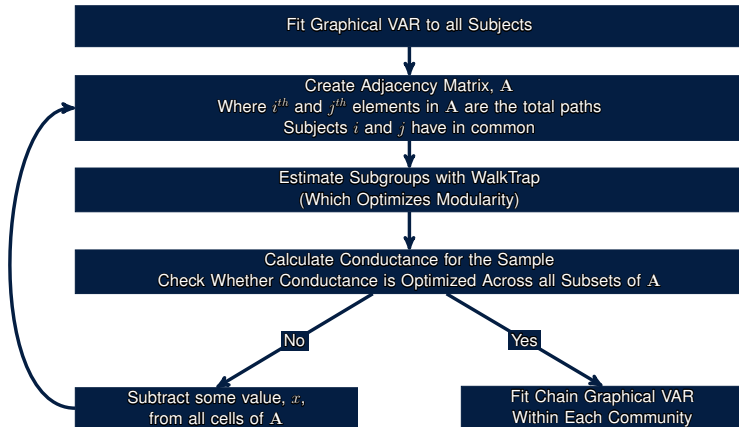
Recap

- A graphical VAR model can be fit to the chained time-series of multiple subjects
 - The resulting lagged and contemporaneous 'fixed effects' networks represent the average effects in a sample
- Given severe heterogeneity within a sample these assumptions could be problematic

Why Subgrouping

- A chain gVAR fit to data comprised of multiple, distinct subgroups could produce 'fixed effects' networks that are unrepresentative
- Subgrouping individuals prior to fitting a chain gVAR to the data can help with this

Our Approach



WalkTrap

Gates et al., 2017; Pons & Latapy, 2006

- Adjacency matrix, \mathbf{A} , is recast into a transition matrix, \mathbf{P}_{ij}^k
 - \mathbf{P}_{ij}^k depicts the probability of going from edge i to j in k -steps
- Distances are then computed between all subjects:

$$z_{ij} = \left\| D^{-\frac{1}{2}} P_i^k - D^{-\frac{1}{2}} P_j^k \right\|$$

where

- D = the average weight in the adjacency matrix
- P = the transition probability vector for the i^{th} or j^{th} subject
- $\|\cdot\|$ = Euclidean norm of R^n

WalkTrap Continued

Gates et al., 2017; Pons & Latapy, 2006

- Once the distance matrix is constructed, Wards algorithm is used to begin merging subjects then clusters beginning with each subject as their own cluster:

$$\sigma_q = \frac{1}{2} \sum_{C \in P_q} \sum_{i \in C_q} z_{iC}^2$$

where

- q -steps are repeated iteratively by selecting partitions, P_q which minimize the average squared distances between each subject with its community, σ_q
- Modularity is used to select the best partition

Modularity

From within WalkTrap; Newman (2004)

$$Q(G) = \sum_{C \in G} (e_C - a_C^2)$$

where

- Q = Modularity for graph, G
- C = Cluster within graph G
- e = probability of an edge within cluster, C
- a = probability of an edge bounded to cluster, C

WalkTrap

Some thoughts

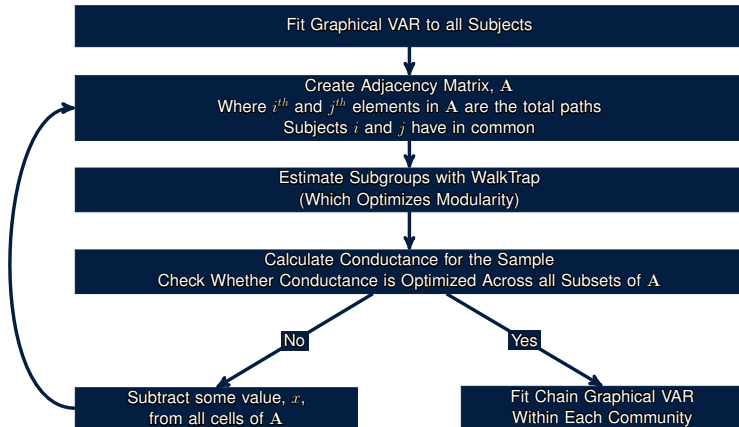
- Modularity compares observed edges within a community to the expected number of edges in a random graph
- Generally prefers communities of a given size regardless of network size
- Is not very adaptive to varying networks types and gives consistent results

Optimization of **A**

Why?

- Communities should be more densely connected to same-community members than they are to members of other communities
- For example, if subjects i and j are in the same community, they should share more connections with one another than they would if they belonged to separate communities
- S-GIMME, by default, subtracts the minimum value from all cells to induce sparsity
- This makes sense as we would expect the minimum value to exist in the space between communities

Recall



Conductance

- Conductance attempts to minimize the number of edges with anchors outside of the community relative to the total number of edges associated with that community
- Community structures are dynamic based on network type and size
- Maps on more to 'real-world' networks

Conductance

What we are maximizing

$$\phi(S_k) = \frac{\sum_{i \in S_k, j \notin S_k} \mathbf{A}_{ij}}{\min[A(S_k), A(\bar{S}_k)]}$$

where

- $\phi(S_k)$ = the conductance for the k^{th} subgroup
- \mathbf{A} = the adjacency matrix
- $A(S_k)$ = number of edges which terminate in subgroup k
- $A(\bar{S}_k)$ = number of edges which do not terminate in subgroup k

Conductance

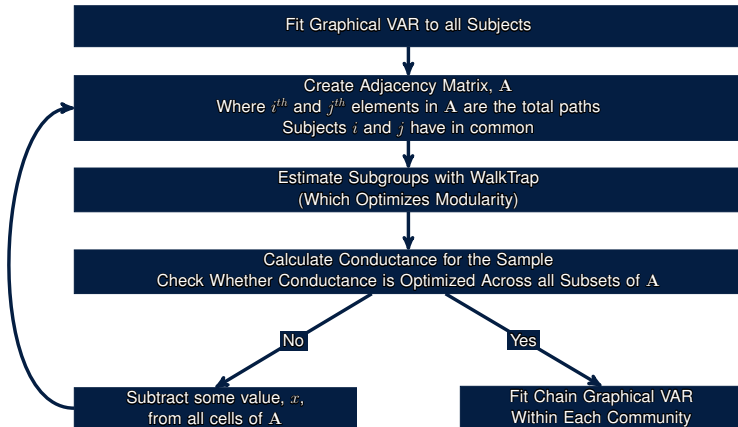
What we are maximizing

$$\phi(G) = 1 - \frac{1}{k} \sum_k \phi(S_k)$$

where

- $\phi(G)$ = conductance for graph, G
- This is just the mean of the conductance for all partitions

A Final Review



Simulation Example

A demonstration of subgrouping with Chain Graphical VAR

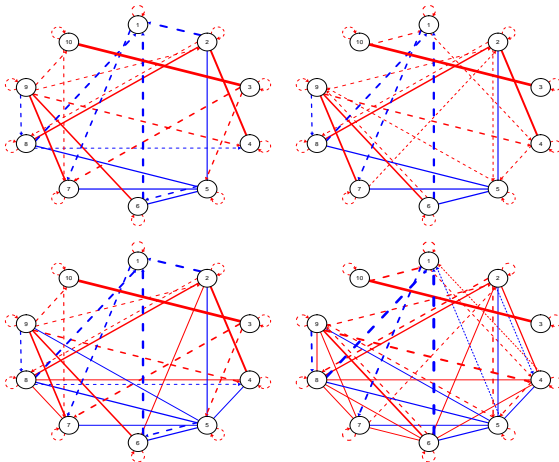
- Simulate a VAR (e.g., Φ , \mathbf{K})
- Test subgroup recovery via the Adjusted Rand Index

Preliminary Results (100-replications)

- $M_{ARI} = 0.97$; $SD_{ARI} = 0.15$
- Fewer than 4-replications with $ARI < 0.80$
 - 3 of which Subgroup B was perfect subset of Subgroup A

Demonstration

True Subgroup Models (Top Row) and Estimated Subgroups (Bottom Row)



Thank you!

- E-mail: JPark@psu.edu
- Web: IAmJonathanPark.com