

Supplemental Material

CBE—Life Sciences Education

Esparza *et al.*

Appendix 1. Primer for constructing and interpreting partial correlation networks (PCNs).

In the PCN models, partial correlations employing regularization techniques to correct for over-parameterization of the fitted models (Hastie *et al.*, 2015) are used to deal effectively with sparsity and the resulting low-rank covariance matrices. This prevents the models from becoming too noisy and decreases the number of spurious non-zero partial correlations, which will have the effect of simplifying and elucidating interpretations. Relatedly, a major challenge associated with these models is demonstrating stability of the observed association structures revealed by the models. Thus, we employed post-hoc analysis of the models to demonstrate stability and strengthen research conclusions.

As described in our article, the Fruchterman-Reingold algorithm was used to construct all PCNs. This algorithm plots the nodes so that nodes with greater levels of correlation are grouped closer together and, consequently, have shorter edge lengths. Conversely, if two edges are very far apart, then the level of correlation is weak relative to the other nodes. When interpreting these network plots, it is imperative to not assume directionality. This, in effect, produces visible clusters of nodes. In particular, if two nodes exhibit strong association, this does not convey a causal relationship or even a sequence of activities (though, in fact, this may be occurring). Instead, the associations should be strictly interpreted as the association of two behaviors in a cross-sectional snapshot. When interpreting PCNs, it is likewise critical to note that when two variables (nodes) are unconnected, they are independent when conditioning on all other variables. Otherwise, a non-zero partial correlation indicates that a relationship exists between two variables (with latent variables directing the observed variables). This implies that one variable causes the other, there is some mutually-reciprocal relation, or both variables cause some other variable(s) in the data. In these cases, red edges indicate that the partial correlations are non-zero and negative while green edges indicate that the non-zero partial correlations are positive. The weight of the edge indicates the magnitude of the partial correlation. Thus, heavier and wider edges indicate that the partial correlation is closer to 1 or -1, and thinner and sparser edges indicate that the partial correlations are non-zero but of lesser magnitude. In our analyses, the hyperparameter controlling the appearance of spurious edges was set to three values (0, 0.2, and 0.4). The choice threshold value of 0.2 was determined by examining the closeness, betweenness, and degree centrality of networks for the three values of hyperparameters considered. The other values (0 and 0.4) resulted in too dense or too sparse networks, respectively. The tuning parameter was estimated for each PCN to optimize fit using a shrinkage estimation procedure. These values ranged from 0.028 to 0.064 but averaged close to 0.050. Furthermore, multiplicity was controlled using the false discovery rate (FDR) correction at 10% (Benjamini & Hochberg, 1995). This ensures that only 10% of the drawn edges occur by chance with the remainder representing true positives. Network accuracy was assessed using bootstrapping and analyzing the bootstrap confidence intervals and correlation stability estimators that resulted. Such an approach provides information about how stable the model is from sample to sample and lends credibility to the presented results.

In the context of our own research, nonparametric and single case deletion bootstrapping methods indicated that, for each PCN, the estimated parameters demonstrated reasonable levels of closeness, betweenness, and edge strength. The bootstrap confidence intervals for the edge weights confirm the edges showing up in the PCNs and do not raise any questions about the stability of the models. The correlational stability (CS) was assessed for all models reported in the accompanying article. The CS coefficient exceeded 0.9 for all models. Usually, this value should exceed 0.5; thus, our models possess stronger-than-average correlational stability.

Additionally, there are some general patterns to look for in PCNs, including those presented in the accompanying article. For example, if multiple nodes are connected to a single variable, this indicates the presence of a multiple regression relation among this set of variables. PCNs can also be used to identify mediating variables by finding relationships such as: A is linked to B is linked to C. In this case, A and C may or may not be dependent, but the PCN identifies a possibly indirect relationship. Latent variables can also be identified in a PCN by looking for independent clusters among the nodes. We investigated all of these patterns in the PCNs generated as a result of our research in order to gain insight into how the data are related and co-vary. All analyses were performed in *R* (R Core Team, 2018) using the following packages: *qgraph* (Epskamp *et al.*, 2012) and *bootnet* (Epskamp *et al.*, 2018).

Appendix 1. Primer for constructing and interpreting partial correlation networks (PCNs).

References

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300.
- Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(4), 1-18.
- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195-212.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations*. Boca Raton, FL: CRC Press.

SupplementalTable S1. LOPUS codes and descriptions.

Student Behavior Codes	Description
<i>Typical Behaviors</i>	
L	Listening to the instructor
Lab	Engaging in laboratory activities
TQ	Completing a test or quiz
<i>Interactive Behaviors</i>	
SQ	An individual student or group of students asking the instructor a question with the entire class listening
X1	An individual student or group of students privately asking the instructor a question
WC	Engaging in whole-class discussion; such discussion may or may not be facilitated by the instructor
Prd	Making predictions about the outcome of an experiment or demonstration
SP	Giving a presentation
SI	Initiating one-on-one interaction with the instructor
WG	Students interacting with other students in their laboratory group (i.e., within-group interaction)
BG	Students interacting with peers in another laboratory group (i.e., between-group interaction)
<i>Noninstructive Behaviors</i>	
W	Waiting
Instructor Behavior Codes	Description
<i>Typical Behaviors</i>	
Lec	Lecturing to the entire class
RtW	Real-time writing on the board
FUp	Providing follow-up/feedback on one or more activities that the students have completed
DV	Conducting a demonstration or showing a video
M	Monitoring the class or individual groups of students
<i>Interactive Behaviors</i>	
PQ	Posing a non-rhetorical question related to laboratory activities
Tlk	Talking to an individual student or group of students one-on-one
TPQ	Posing a question to an individual student or group of students
TI	Initiating one-on-one interaction with individual students or a group of students
<i>Noninstructive Behaviors</i>	
Adm	Performing administrative tasks
W/W.1	Waiting

SupplementalTable S2. Comparison of instructor and student behaviors in CURE and non-CURE contexts stratified by courseinstructor.

Category	Non-CURE Instructors				CURE Instructors			
	Joe	Jen	Bob	Ella	Sam	Nat	Nia	Jon
Student Behaviors^a								
Typical Behaviors								
Listening to Instructor (L) ⁺	12%	20%	8%	41%	11%	34%	50%	56%
Performing Laboratory Exercise (Lab) ⁺	77%	65%	84%	58%	84%	74%	62%	29%
Test/Quiz (TQ)	7%	7%	5%	6%	0%	0%	0%	0%
Interactive Behaviors								
Individual Student/Group Asking Instructor a Question with Class Listening (SQ) ⁺	0%	0%	0%	0%	1%	10%	3%	2%
Individual Student/Group Asking Instructor a Question (X1) [*]	17%	7%	17%	2%	22%	30%	10%	16%
Whole-Class Discussion (WC) [*]	2%	0%	3%	7%	2%	10%	27%	1%
Making Predictions (Prd)	0%	0%	0%	0%	4%	3%	2%	0%
Giving a Presentation (SP)	0%	0%	0%	0%	0%	0%	0%	0%
Initiating One-on-One Interaction with the Instructor (SI) [*]	14%	7%	19%	2%	34%	33%	8%	16%
Students Interacting with Other Students in Their Group (WG) [*]	88%	74%	88%	69%	95%	88%	61%	38%
Students Interacting with Peers in Another Group (BG) [*]	10%	13%	13%	10%	49%	37%	28%	0%
Noninstructive Behaviors								
Waiting (W) ⁺	0%	1%	0%	1%	3%	1%	3%	14%
Instructor Behaviors^a								
Typical Behaviors								
Lecturing to Class (Lec) [*]	8%	10%	7%	26%	8%	29%	40%	37%
Real-time Writing on the Board (RtW) ⁺	0%	0%	0%	0%	1%	15%	25%	0%
Providing Follow-up/Feedback (FUp) ⁺	0%	3%	0%	0%	2%	23%	0%	0%
Demonstration or Video (DV)	8%	11%	3%	10%	19%	6%	5%	2%
Monitoring the Class/Groups (M) ⁺	44%	41%	62%	23%	22%	35%	4%	17%
Interactive Behaviors								
Posing a Non-Rhetorical Question (PQ) [*]	2%	1%	2%	7%	3%	16%	27%	2%
Talking to Students/Groups One-on-One (Tlk) [*]	39%	52%	64%	7%	85%	60%	47%	31%
Posing a Question to Individual Student or Group of Students (TPQ)	4%	9%	6%	3%	20%	22%	20%	5%
Initiating One-on-One Interaction with Individual Student or Group of Students (TI) [*]	21%	42%	53%	6%	75%	41%	42%	27%
Noninstructive Behaviors								
Performing Administrative Tasks (Adm) [*]	20%	41%	26%	54%	25%	31%	3%	17%
Waiting (W/W.1) [*]	39%	7%	10%	14%	0%	11%	0%	1%

^aValues represent the median of the percentage of time spent on various activities over six laboratory sessions. They do not add up to 100% because multiple behaviors can be observed during any 2-minute interval.

⁺ $p < 0.05$ for indicated comparisons, as determined by a series of Kruskal-Wallis tests comparing behaviors evidenced in CURE sections or non-CURE sections.

^{*} $p < 0.005$ for indicated comparisons, as determined by a series of Kruskal-Wallis tests comparing behaviors evidenced in CURE sections or non-CURE sections.