Science is about theory not method

Lourens Waldorp **Maarten Marsman**

waldorp@uva.nl m.marsman@uva.nl

University of Amsterdam, Psychological Methods

Recently, Forbes, Wright, Markon, and Krueger [3] expressed their concerns with the network approach. We believe such concerns are justified, but for completely different reasons than provided.

The arguments put forth by Forbes et al. [3] are mostly about the quality of the data and general statistics and analysis, and are true but which apply to any analysis. However, there are two arguments that seem to identify network analysis as particularly bad for scientific knowledge, in particular about psychopathology. The first is that networks are in general unreliable and second, or maybe more as one of the reasons why it is unreliable, network analysis uses partial correlations. We will detail why these arguments are wrong.

First, networks are basically a series of regressions [2;4]. Because they can be considered as such, arguing that networks are unreliable generally, implies that regression is unreliable. But for the social sciences regression is the foremost important tool of modelling. Indeed, in many sciences it is the dominant tool. For instance, in neuroscience and econometrics, regression is used to model anything from communication between brain areas [7] to the effect of government intervention on society [5]. This does not imply that regression is necessarily a good choice, of course. This brings us to the second argument [3], partial correlations. The argument is that by using partial correlations the baby (shared variance) is tossed out with the bathwater (overlap between variables). If it were true that in regression the shared variance is 'thrown out', then this implies that adding correlated predictors to any regression will decrease the R-squared (the measure used to determine the explained variance). But this is obviously not true, as is well known [1;6].

This leaves us with what we believe is the most relevant issue, not mentioned by Forbes et al. [3], but certainly present between the lines: The network model is wrong. This may well be true and is worth discussing and certainly worth investigating scientifically. We believe that one of the most important ways to approach such a debate is by considering what predictions a model makes and how this can be verified or falsified empirically. We will give an example and discuss both the latent variable and network model.

As an example we consider depression. Suppose we are in a rather luxurious position and we have obtained data from a single subject with 500 repeated measures on all 9 symptoms. Then we fit both a unidimensional latent variable model and a network (partial correlations). We now want to make predictions from these models. In the latent variabel model we predict that if the latent variable increases then the symptoms increase (where we assumed that all symptoms are positively correlated and so all regression coefficients between the latent variable and the observed symptoms are positive). In the network model we predict that if we increase one of the symptoms, then only the connected set of symptoms will immediately increase (others may follow later if they are indirectly connected). To determine which of the models is correct, we could set up an experiment with this particular subject. Suppose that we intervene by giving a sleeping pill to the subject to decrease the 'lack of sleep' symptom. In the latent variable model, this should have no effect on all other symptoms, because only the latent variable can affect the symptoms. But in the network model we should see a decrease in the connected symptoms.

Obviously we are not in such a luxurious position, but we could approximate the situation with cross-sectional data. Of course, we need to assume a form of weak ergodicity, that suggests that the sample we have is to some extent homogeneous. And in that case we could perform the above experiment and see what happens. Will really nothing happen when we provide a sleeping pill to increase sleep?

References

[1] Bilodeau, M., Brenner, D., 1999. Theory of multivariate statistics. New York: Springer-Verlag.

[2] Bühlmann, P., van de Geer, S., 2011. Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer.

[3] Forbes, M. K., Wright, A. G., Markon, K. E., Krueger, R. F., 2019. The network approach to psychopathology: promise versus reality. World Psychiatry 18 (3), 272.

[4] Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the lasso. The Annals of Statistics 34 (3), 1436–1462.

[5] Morgan, S., Winship, C., 2007. Counterfactuals and causal inference. Cambridge University Press, Cambridge.

[6] Seber, G. A., Lee, A. J., 2012. Linear regression analysis. Vol. 936. John Wiley & Sons.

[7] Sporns, O., 2011. Networks of the brain. MIT Press, Cambridge, MA.