Intelligence xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Intelligence

Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis

Hudson F. Golino [a], Andreas Demetriou [b,*]

[a] Universidade Salgado de Oliveira, Brazil
[b] University of Nicosia, Cyprus

ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form 6 January 2017
Accepted 20 February 2017
Available online xxxx

ABSTRACT

This study compared various exploratory and confirmatory factor methods for recovering factors of cognitive test-like data. We first note the problems encountered by several widely used methods, such as parallel analysis, minimum average partial procedure, and confirmatory factor analysis, in estimating the number of dimensions underlying performance on test batteries. We then argue that a new method, Exploratory Graph Analysis (EGA), can more accurately uncover underlying dimensions or factors and demonstrate how this method outperforms the other methods. We use several published data sets to demonstrate the advantages of EGA. We conclude that a combination of EGA and confirmatory factor analysis or structural equation modeling may be the ideal in precisely specifying latent factors and their relations.

1. Introduction

Discovering dimensions (or factors) underlying human behavior and cognitive ability is central in psychology and the cognitive sciences. Factor analysis was developed to uncover the dimensions underlying a large number of measures of the behaviors or abilities of interest (Spearman, 1904; Carroll, 1993; Jensen, 1998). However, there is no agreement yet about the method of choice for identifying the best number of dimensions and their relations under various conditions of measurement, sampling of persons, and between dimensions' relations. Principal component analysis, factor analysis of various types and rotations, and, recently confirmatory factor analysis and structural equation modeling were advanced to cope with these problems.

Recently, Keith, Caemmerer and Reynolds (2016) showed that both parallel analysis (PA) and minimum average partial procedure (MAP) underestimate the number of dimensions in many realistic data condition, especially when the correlation between factors are high (.70) and the number of indicators per factor are low. Their results align with earlier research showing that both PA and MAP work well when there is a low or moderate correlation between factors, when the sample size is equal to or >500 and when the factor loadings are from moderate to high (Buja & Eyuboglu, 1992; Crawford et al., 2010; Garrido, Abad & Ponsoda, 2011; Green, Redell, Thompson & Levy, 2016; Timmerman & Lorenzo-Seva, 2011; Velicer, Eaton, & Fava, 2000, Velicer, 1976; Zwick & Velicer, 1986).

Simulation studies point to a relevant issue: PA and MAP fail to uncover the correct number of factors in situations approaching real intelligence datasets. Keith et al. (2016) suggested that researchers must use confirmatory factor analysis (CFA) guided by a relevant theory, because CFA was more accurate than other methods in recovering the correct number of dimensions in their simulation study. To deal with the fact that exploratory techniques did not correctly recover the number of factors in realistic data conditions, these authors strongly suggested using theory to guide the analysis.

Although useful when available, theory, in principle, may suggest but, cannot specify either the true number of dimensions in an instrument or how items in the instrument may relate to these dimensions. Obviously, there may be factors in the battery that are ignored or overlooked by the theory. Thus, CFA may be used to test if the theoretically expected dimensions are present in the data. However, when a CFA theory-based model fails to fit empirical data, other tools are needed to explore the structure of the instrument as precisely as possible.

We argued above that PA, MAP and other traditional techniques are not robust enough to estimate the number of factors underlying a given instrument, when the correlation between factors is high and the number of variables per factor is low. Thus, new robust techniques are needed. This paper presents a new method, Exploratory Graph Analysis (EGA; Golino & Epskamp, 2016), that is more powerful than earlier methods to estimate the number of dimensions in intelligence-like data. EGA was shown to outperform PA and MAP in conditions where these methods are not accurate: That is, when correlations between factors are high and the number of items per factor is low (Golino & Epskamp, 2016). Specifically, EGA is better in estimating the number of factors in situations which (1) are very close to what we find in real

* Corresponding author.
E-mail addresses: hfgolino@gmail.com (H.F. Golino), ademetriou@ucy.ac.cy (A. Demetriou).

http://dx.doi.org/10.1016/j.intell.2017.02.007
0160-2896/© 2017 Published by Elsevier Inc.

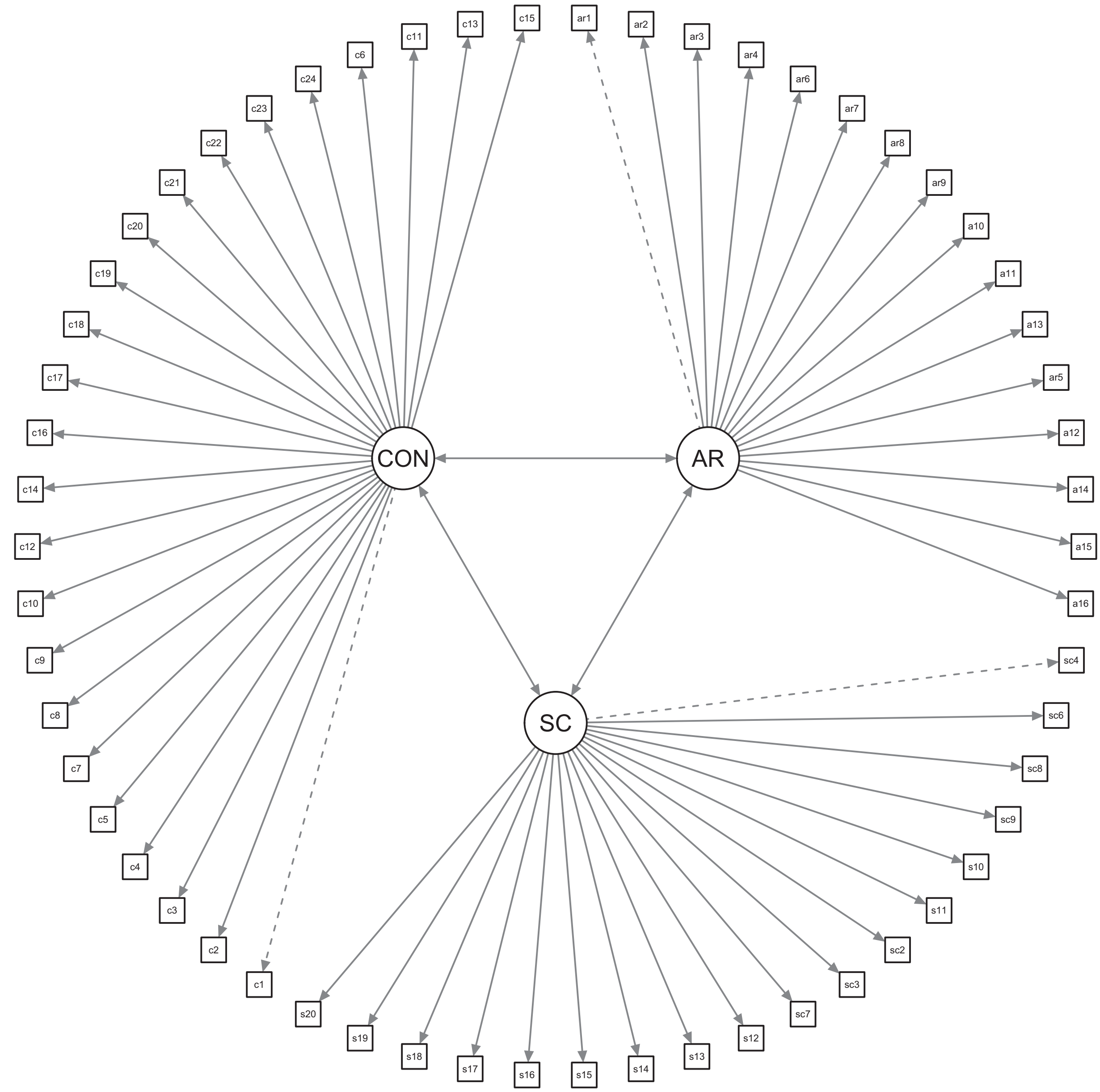

Fig. 1. The theoretical structure of the NIT subtests used in the current study. AR = arithmetical reasoning; SC = sentence completion; CON = concepts.

intelligence datasets but (2) traditional techniques underestimate the dimensions involved.

This paper is organized as follows. In the first section, EGA is briefly presented as part of a new area called *network psychometrics* (Epskamp, Maris, Waldorp & Borsboom, 2017). In the second section, EGA is applied on the datasets simulated by Keith et al., where PA and MAP presented low accuracy (i.e., high factor correlations, low factor loadings, and $N = 500$). Finally, in the third section, EGA is applied in three previously published datasets (Must & Must, 2013, 2014; Demetriou & Kazi, 2006; Žebec, Demetriou, & Kotrla-Topić, 2015) to show how this new technique can guide researchers in their search for the underlying dimensionality of intelligence like data.

1.1. Exploratory graph analysis: a brief overview

Exploratory Graph Analysis is part of a new area called network psychometrics (see Epskamp et al., 2017), which focuses on the estimation of undirected network models (i.e. Lauritzen, 1996a, b) to psychological datasets. This area has been applied in different areas of psychology, including psychopathology (e.g., Borsboom et al., 2011; Borsboom & Cramer, 2013; Fried et al., 2015) and developmental psychology (Kossakowski et al., 2015; van der Maas et al., 2006). In network psychometrics, the nodes represent psychological variables (e.g., test and/or questionnaire items, psychopathological symptoms, etc.) and the connection between nodes (i.e., edges) represents statistical relationships to be estimated (Epskamp & Fried, 2016). Thus, there is a fundamental distinction between network psychometrics and other types of network models, in which the links between nodes do not need to be estimated, such as social networks analysis (Epskamp & Fried, 2016). When analyzing data generated by psychological instruments, one may want to know if nodes are connected with each other, forming clusters standing for underlying latent variables. If a latent variable model is the true underlying causal model, we would expect indicators in a network model to form strongly connected clusters for each latent variable. Network models may be shown to be mathematically equivalent under certain conditions to latent variable models in both binary (Epskamp et al., 2017) and Gaussian datasets (Chandrasekaran, Parrilo & Willsky, 2010).

Please cite this article as: Golino, H.F., & Demetriou, A., Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis, *Intelligence* (2017), http://dx.doi.org/10.1016/j.intell.2017.02.007