

Graphs as Not Only Relational Databases for Behavior Science

Juan C. Correa^{1,2*}

^{1*}Departamento de Estudios Empresariales, Universidad Iberoamericana,
Mexico City, Mexico, 1219.

^{2*}Research and Development Unit, Critical Centrality Institute.

Corresponding author(s). E-mail(s): juan.correa@ibero.mx;

Abstract

Behavior scientists have multiple options for managing research data. While relational databases offer robust tools for storage and analysis, they impose structural constraints that limit the representation of complex behavioral phenomena. This article argues that graph databases provide a rigorous and conceptually richer alternative, enabling the modeling of human behavior as an interconnected system rather than a set of isolated variables. Beyond a technological shift, adopting graph-based approaches invites a paradigm change in behavior science—one that embraces complexity, dynamic relationships, and multi-level contingencies. Practical implications are illustrated through examples from clinical, consumer, and industrial/organizational psychology.

Keywords: key, dictionary, word

JEL Classification: D8 , H51

MSC Classification: 35A01 , 65L10

1 Introduction

In a recent paper, [Soto \(2025\)](#) introduced relational databases for behavior science and used real-world examples to illustrate how relational databases have been used by behavior scientists. Even though relational databases represent the dominant paradigm inside and outside academic research settings, other paradigms are gaining traction. The so-called “Not Only Relational Database” encompasses a series of database

management systems that use single data structures to hold information. There are several instances of single data structures such as lists, key-value pairs, wide columns, documents, matrices, or graphs.

In this article, I examine graphs as a different paradigm from the traditional paradigm of relational databases, where nodes and edges (instead of tables and joins) represent the basic elements of any behavior that can be represented as a network or complex system. Networks have a long history in mathematics as “*graph theory*” (Estrada, 2011). In sociology and social sciences, graph theory is known as “social network analysis” (Wasserman & Faust, 1994), and psychologists have leveraged this framework to analyze the structure of psychopathology (Borsboom & Cramer, 2013), estimate the correct number of dimensions in psychological and educational instruments (Golino & Epskamp, 2017), or understand the measurement of organizational climate (Menezes, Menezes, Moraes, & Pires, 2021).

2 Network as a collection nodes and edges

One of the easiest way to grasp the idea of a network is by looking its visual representation (see Figure 1). According to Estrada (2011), a network is a collection of points (called nodes) joined together in pairs by lines (called arcs or edges). Scientists can model different kinds of networks from physical networks (e.g., flights between airports) to biological networks (e.g., protein-protein interactions), and social networks (e.g., who follows whom in LinkedIn or X). In this context, the term “social network” should not be confused with online platforms such as Facebook or Instagram, as they are technological implementations that do not necessarily represent all aspects of social networks as an academic discipline.

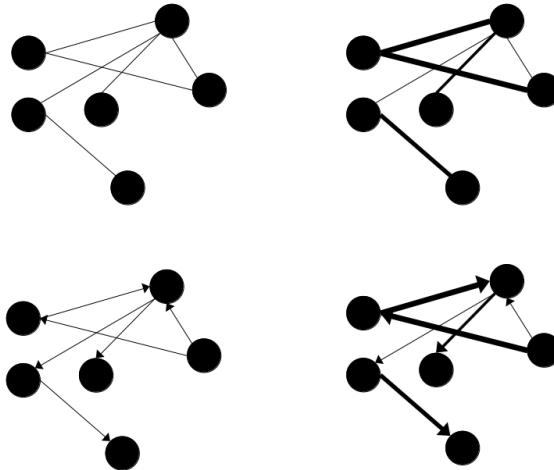


Fig. 1 A visual representation of four types of simple networks: non-directed unweighted network (upper left), non-directed weighted network (upper right), directed unweighted network (bottom left), and directed weighted network (bottom right).

Graphs offer fundamental concepts for understanding how entities (nodes) and their relationships (edges) form interconnected structures. These concepts, thoroughly covered in introductory texts (Newman, 2010), will not be revisited here. Instead, this article focuses on their implications for behavior science. Network structures underpin a wide range of behavioral phenomena—from disease transmission and social clustering to the spread of information and misinformation. As we have mentioned above, the recognition of these patterns requires tools that go beyond the rigid tabular constraints of relational databases. My goal is to illustrate how graph-based databases can enrich the methodological toolbox of behavior scientists, enabling analyses that embrace complexity, dynamic relationships, and multi-level contingencies.

3 Graph databases: A gentle introduction

A configuration is a specific pattern of connections between nodes. The most elementary configuration is a dyad (two connected nodes). Other common configurations include triangles (three nodes, all connected to each other) and k-stars (a central node connected to k other nodes). These configurations, along with others, make up the larger connected subgraphs of a network, known as components. A component is a set of nodes where a path exists between every pair of nodes. A component is a maximal connected subgraph, where all nodes within it are reachable from each other, and it is not connected to any other nodes. A component can be as simple as a long line of connected nodes with no triangles or k-stars. A path is a fundamental concept that describes a sequence of connections between nodes. It is the route one can follow to travel from one node to another. Paths are essential for understanding properties like the shortest distance between any pair of nodes. A special type of path is a cycle, which is a closed path with no repeated nodes or edges other than the start and end nodes. A cycle has a minimum of three links (e.g., a triangle). From the network science perspective, geometric structures like squares, pentagons, or decagons can all be seen as cycles. The four types of configurations in a network are depicted in Figure 2.

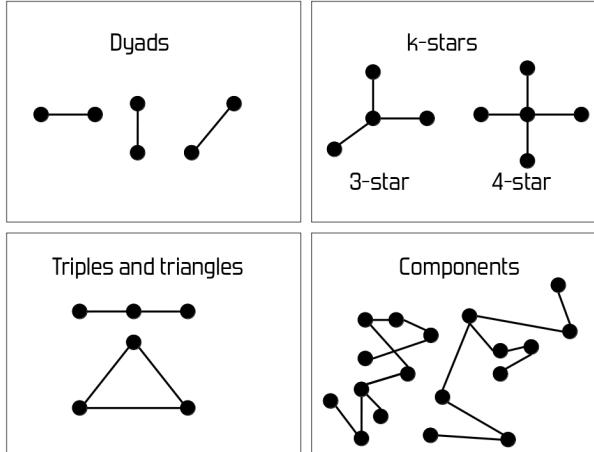


Fig. 2 Basic configurations in a network: dyads (top left), k-stars (top right), triples and triangles (bottom left), and components (bottom right).

In the analysis of social networks, it is very important to distinguish between connected and disconnected networks (Wasserman & Faust, 1994). When a network is connected there is a path between every pair of nodes; that is, you can reach any node from any other node. In a disconnected network some paths do not exist (i.e., some nodes are isolated and not reachable). The nodes in a disconnected network can be partitioned into two or more subsets in which there are no paths between the nodes in different subsets. The connected sub-networks in a network are called components. A component is a sub-network in which there is a path between all pairs of nodes and there is no path between a node in a component and any node other node in the other component.

3.1 Nodal Centrality

Nodal centrality relates to the amount of connections a given node has, and can be computed in different ways (Oldham et al., 2019). Nodal degree is the most basic way of estimating a node centrality. It requires counting the number of lines incident with it. For example, the nodal degree of the central node in the 3-star network depicted in the top right of Figure 2 is three.

Note that nodal centrality is a property of every single node in a network, and when the analysis of a social network focuses on nodes centrality, this estimation is done to every single node in the network. The analysis of nodes centrality facilitates enables the differentiation of most important nodes from most peripheral nodes and organize nodes in order of importance or number of connections in the network.

4 Applications of Network Modeling

Network modeling has been used by psychologists and behavioral scientists in several ways that fit into the professional goals of well-established APA's divisions such as

quantitative and qualitative methods (division 5), clinical psychology (division 12), and industrial and organizational psychology (division 14). The following subsections summarize how network modeling has been applied in these subdisciplines.

4.1 Network modeling as a quantitative method

Golino and Epskamp (2017) illustrated the principles by which network modeling can be used as a framework for estimating the correct number of dimensions in psychological and educational instruments. While traditional techniques such as “parallel analysis” and “minimum average partial” are widely utilized, they often underestimate the number of factors in scenarios involving high correlations between latent variables, small sample sizes, or few indicators per factor. Conversely, the widely used Kaiser-Guttman rule is frequently criticized for overestimating the number of factors, particularly as sample sizes and the number of items increase.

In response to these limitations, “exploratory graph analysis” has been developed as a new approach derived from the field of network psychometrics (Epskamp, Maris, Waldorp, & Borsboom, 2018). This framework utilizes Markov random fields to model the interaction between random variables as a network of nodes (variables) and edges (direct relationships). Specifically, EGA employs the **Gaussian Graphical Model (GGM)**, which models the multivariate distribution through the inverse covariance matrix. When these elements are standardized, they represent **partial correlation coefficients**. A fundamental tenet of this approach is the “Clusters in network = latent variables” rule: if a latent variable model represents the true causal structure, its indicators will manifest as strongly connected clusters (or cliques) within the network because they cannot become independent after conditioning on other observed variables.

The EGA Procedure The EGA framework follows a rigorous three-step statistical sequence to identify the underlying dimensionality of a dataset:

1. **Estimation:** The correlation matrix of observable variables is estimated (e.g., tetrachoric or polychoric correlations for categorical data).

2. **Regularization:** To prevent overfitting and the inclusion of spurious correlations common in typical psychological sample sizes, the **graphical LASSO (least absolute shrinkage and selection operator)** is applied. The sparsity of the resulting network is optimized by selecting a regularization parameter that minimizes the **Extended Bayesian Information Criterion (EBIC)**.

3. **Community Detection:** The **walktrap algorithm**—a random walk-based procedure—is used to identify dense subgraphs or communities within the regularized partial correlation matrix. The number of detected communities corresponds to the estimated number of latent dimensions.

Empirical Performance and Accuracy Extensive simulation research involving **32,000 datasets** across 64 conditions has demonstrated that EGA is highly robust. While it performs comparably to traditional methods in two-factor structures, EGA significantly outperforms them in **four-factor structures with high correlations (.70)** between dimensions. In these complex scenarios, EGA was often the only technique capable of maintaining high accuracy, reaching 100% accuracy with a sample

size of 5,000. Furthermore, analysis of variance (ANOVA) indicates that EGA's accuracy is the **least affected** by varying experimental conditions, such as sample size and factor correlation, compared to traditional techniques.

Additional Methodological Benefits Beyond estimating the total number of dimensions, EGA provides a distinct advantage by automatically identifying **which specific items indicate each retrieved dimension**. This dual functionality—dimensionality estimation and item membership identification—provides a more comprehensive output for construct validation than traditional factor-retention rules. Empirical application to the **Inductive Reasoning Developmental Test (IRDT)** further validates the framework, as EGA correctly identified seven dimensions that matched the instrument's theoretical developmental stages, while other methods like Parallel Analysis suggested only four.

4.2 This is an example for second level head—subsection head

4.2.1 This is an example for third level head—subsubsection head

Sample body text. Sample body text.

5 Equations

Equations in L^AT_EX can either be inline or on-a-line by itself (“display equations”). For inline equations use the $\$...$$ commands. E.g.: The equation $H\psi = E\psi$ is written via the command $\$H \backslash\psi = E \backslash\psi\$$.

For display equations (with auto generated equation numbers) one can use the equation or align environments:

$$\|\tilde{X}(k)\|^2 \leq \frac{\sum_{i=1}^p \|\tilde{Y}_i(k)\|^2 + \sum_{j=1}^q \|\tilde{Z}_j(k)\|^2}{p+q}. \quad (1)$$

where,

$$\begin{aligned} D_\mu &= \partial_\mu - ig \frac{\lambda^a}{2} A_\mu^a \\ F_{\mu\nu}^a &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc} A_\mu^b A_\nu^c \end{aligned} \quad (2)$$

Notice the use of `\nonumber` in the align environment at the end of each line, except the last, so as not to produce equation numbers on lines where no equation numbers are required. The `\label{}` command should only be used at the last line of an align environment where `\nonumber` is not used.

$$Y_\infty = \left(\frac{m}{\text{GeV}} \right)^{-3} \left[1 + \frac{3 \ln(m/\text{GeV})}{15} + \frac{\ln(c_2/5)}{15} \right] \quad (3)$$

The class file also supports the use of `\mathbb{R}`, `\mathscr{R}` and `\mathcal{R}` commands. As such `\mathbb{R}`, `\mathscr{R}` and `\mathcal{R}` produces \mathbb{R} , \mathscr{R} and \mathcal{R} respectively (refer Subsubsection 4.2.1).

Table 1 Caption text

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700

6 Tables

Tables can be inserted via the normal `knitr::kable()` function or other table-generating packages.

Tables can also be inserted via the normal `table` and `tabular` environment. To put footnotes inside tables you should use `\footnotetext[]{...}` tag. The footnote appears just below the table itself (refer Tables~\ref{tab:example} and \ref{tab:example2}). For the corresponding footnotemark use `\footnotemark[...]`

Table 2 Caption text

Column 1	Column 2	Column 3	Column 4
row 1	data 1	data 2	data 3
row 2	data 4	data 5 ¹	data 6
row 3	data 7	data 8	data 9 ²

Source: This is an example of table footnote. This is an example of table footnote.

¹Example for a first table footnote. This is an example of table footnote.

²Example for a second table footnote. This is an example of table footnote.

The input format for the above table is as follows:

```
\begin{table}[<placement-specifier>]
\caption{<table-caption>} \label{<table-label>}%
\begin{tabular}{@{}llll@{}}
\toprule
Column 1 & Column 2 & Column 3 & Column 4 \\
\midrule
row 1 & data 1 & data 2 & data 3 \\
row 2 & data 4 & data 5\footnotemark[1] & data 6 \\
row 3 & data 7 & data 8 & data 9\footnotemark[2]\\
\botrule
\end{tabular}
\footnotetext{Source: This is an example of table footnote.\\
This is an example of table footnote.}
```

```

\footnotetext[1]{Example for a first table footnote.  
This is an example of table footnote.}
\footnotetext[2]{Example for a second table footnote.  
This is an example of table footnote.}
\end{table}

```

Table 3 Example of a lengthy table which is set to full textwidth

Project	Element 1 ¹			Element 2 ²		
	Energy	σ_{calc}	σ_{expt}	Energy	σ_{calc}	σ_{expt}
Element 3	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 4	500 A	961	922 ± 10	900 A	1268	1092 ± 40

Note: This is an example of table footnote. This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote.

¹Example for a first table footnote.

²Example for a second table footnote.

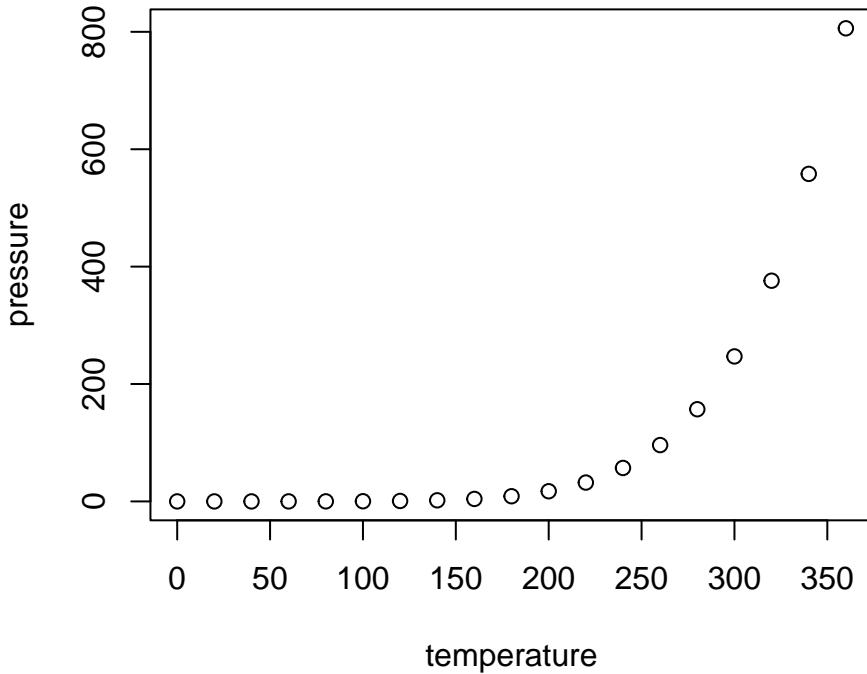


Fig. 3 This is an example of a caption

In case of double column layout, tables which do not fit in single column width should be set to full text width. For this, you need to use `\begin{table*} ... \end{table*}` instead of `\begin{table} ... \end{table}` environment. Lengthy tables which do not fit in textwidth should be set as rotated table. For this, you need to use `\begin{sidewaystable} ... \end{sidewaystable}` instead of `\begin{table*} ... \end{table*}` environment. This environment puts tables rotated to single column width. For tables rotated to double column width, use `\begin{sidewaystable*} ... \end{sidewaystable*}`.

7 Figures

As per the \LaTeX standards you need to use eps images for \LaTeX compilation and pdf/jpg/png images for PDF \LaTeX compilation. Use the `dev knitr` option to use the appropriate format. This is one of the major difference between \LaTeX and PDF \LaTeX . Each image should be from a single input .eps/vector image file. Avoid using subfigures. The command for inserting images for \LaTeX and PDF \LaTeX can be generalized. The package used to insert images in $\text{\LaTeX}/\text{PDF}\text{\LaTeX}$ is the `graphicx` package. Figures can be inserted via the normal figure environment as shown in the below example:

Table 4 Tables which are too long to fit, should be written using the "sidewaystable" environment as shown here

Projectile	Element 1 ¹			Element 2 ²		
	Energy	σ_{calc}	σ_{expt}	Energy	σ_{calc}	σ_{expt}
Element 3	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 4	500 A	961	922 ± 10	900 A	1268	1092 ± 40
Element 5	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 6	500 A	961	922 ± 10	900 A	1268	1092 ± 40

Note: This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote.

¹This is an example of table footnote.

²This is an example of table footnote.

8 Algorithms, Program codes and Listings

Packages `algorithm`, `algorithmicx` and `algpseudocode` are used for setting algorithms in L^AT_EX using the format:

```
\begin{algorithm}
\caption{<alg-caption>}\label{<alg-label>}
\begin{algorithmic}[1]
...
\end{algorithmic}
\end{algorithm}
```

You may refer above listed package documentations for more details before setting `algorithm` environment. For program codes, the “program” package is required and the command to be used is `\begin{program} ... \end{program}`. A fast exponentiation procedure:

Similarly, for `listings`, use the `listings` package. `\begin{lstlisting} ... \end{lstlisting}` is used to set environments similar to `verbatim` environment. Refer to the `lstlisting` package documentation for more details.

A fast exponentiation procedure:

```
begin
  for i:=1 to 10 step 1 do
    expt(2,i);
    newline() od
  where
  proc expt(x,n) ≡
    z := 1;
    do if n = 0 then exit fi;
    do if odd(n) then exit fi;
      comment: This is a comment statement;
      n := n/2; x := x * x od;
    { n > 0 };
    n := n - 1; z := z * x od;
  print(z).
end
```

```
for i:=maxint to 0 do begin \{ do nothing \} end; Write('Case
insensitive'); Write('Pascal_keywords.');
```

9 Cross referencing

Figures and tables are labeled with a prefix (fig or tab, respectively) plus the chunk label. Other environments such as equation and align can be labelled via the `\label{#label}` command inside or just below the `\caption{}` command. You can then use the label for cross-reference. As an example, consider the chunk label declared for Figure 3 which is fig1. To cross-reference it, use the command `Figure \ref{fig:fig1}`, for which it comes up as “Figure 3”.

Algorithm 1 Calculate $y = x^n$

Require: $n \geq 0 \vee x \neq 0$
Ensure: $y = x^n$

```
1:  $y \Leftarrow 1$ 
2: if  $n < 0$  then
3:    $X \Leftarrow 1/x$ 
4:    $N \Leftarrow -n$ 
5: else
6:    $X \Leftarrow x$ 
7:    $N \Leftarrow n$ 
8: end if
9: while  $N \neq 0$  do
10:  if  $N$  is even then
11:     $X \Leftarrow X \times X$ 
12:     $N \Leftarrow N/2$ 
13:  else[ $N$  is odd]
14:     $y \Leftarrow y \times X$ 
15:     $N \Leftarrow N - 1$ 
16:  end if
17: end while
```

To reference line numbers in an algorithm, consider the label declared for the line number 2 of Algorithm 1 is `\label{algln2}`. To cross-reference it, use the command `\ref{algln2}` for which it comes up as line 2 of Algorithm 1.

9.1 Details on reference citations

For citations of references, use `? or (?)`.

10 Examples for theorem like environments

The documentclass for springer `sn-jnl.cls` contains 3 styling that you can use to set new default for theorems and proofs type

`thmstyleone` Numbered, theorem head in bold font and theorem text in italic style
`thmstyletwo` Numbered, theorem head in roman font and theorem text in italic style
`thmstylethree` Numbered, theorem head in bold font and theorem text in roman style

For mathematics journals, theorem styles can be included as shown in the following examples.

Theorem 1. *Example theorem text. Example theorem text.*

To add labels and subheadings, use LaTeX notation

Theorem 2 (Theorem subhead). *Example theorem text. Example theorem text.*

Other environments are proposition, example, remark, definition, proof and quote

Sample body text. Sample body text.

Proposition 3. *Example proposition text. Example proposition text.*

Sample body text. Sample body text.

Example 1. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem.

Sample body text. Sample body text.

Remark 1. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem.

Sample body text. Sample body text.

Definition 1 (Definition sub head). *Example definition text. Example definition text.*

Additionally a predefined “proof” environment is available. This prints a “Proof” head in italic font style and the “body text” in roman font style with an open square at the end of each proof environment.

Proof. Example for proof text. Example for proof text.

□

Sample body text. Sample body text.

11 Methods

Topical subheadings are allowed. Authors must ensure that their Methods section includes adequate experimental and characterization data necessary for others in the field to reproduce their work. Authors are encouraged to include RIIDs where appropriate.

Ethical approval declarations (only required where applicable) Any article reporting experiment/s carried out on (i)~live vertebrate (or higher invertebrates), (ii)~humans or (iii)~human samples must include an unambiguous statement within the methods section that meets the following requirements:

1. Approval: a statement which confirms that all experimental protocols were approved by a named institutional and/or licensing committee. Please identify the approving body in the methods section
2. Accordance: a statement explicitly saying that the methods were carried out in accordance with the relevant guidelines and regulations
3. Informed consent (for experiments involving humans or human tissue samples): include a statement confirming that informed consent was obtained from all participants and/or their legal guardian/s

If your manuscript includes potentially identifying patient/participant information, or if it describes human transplantation research, or if it reports results of a clinical trial then additional information will be required. Please visit (<https://www.nature.com/nature-research/editorial-policies>) for Nature Portfolio journals, (<https://www.springer.com/gp/authors-editors/journal-author/journal-author-helpdesk/publishing-ethics/14214>) for Springer Nature journals, or (<https://www.biomedcentral.com/getpublished/editorial-policies/#ethics+and+consent>) for BMC.

12 Discussion

Discussions should be brief and focused. In some disciplines use of Discussion or ‘Conclusion’ is interchangeable. It is not mandatory to use both. Some journals prefer a section ‘Results and Discussion’ followed by a section ‘Conclusion’. Please refer to Journal-level guidance for any specific requirements.

13 Conclusion

Conclusions may be used to restate your hypothesis or research question, restate your major findings, explain the relevance and the added value of your work, highlight any limitations of your study, describe future directions for research and recommendations.

In some disciplines use of Discussion or ‘Conclusion’ is interchangeable. It is not mandatory to use both. Please refer to Journal-level guidance for any specific requirements.

Supplementary information. If your article has accompanying supplementary file/s please state so here.

Authors reporting data from electrophoretic gels and blots should supply the full unprocessed scans for key as part of their Supplementary information. This may be requested by the editorial team/s if it is missing.

Please refer to Journal-level guidance for any specific requirements.

Acknowledgments. Acknowledgments are not compulsory. Where included they should be brief. Grant or contribution numbers may be acknowledged.

Please refer to Journal-level guidance for any specific requirements.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading ‘Declarations’:

- Funding
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
- Ethics approval
- Consent to participate
- Consent for publication
- Availability of data and materials
- Code availability
- Authors’ contributions

If any of the sections are not relevant to your manuscript, please include the heading and write ‘Not applicable’ for that section.

Editorial Policies for:

Springer journals and proceedings: <https://www.springer.com/gp/editorial-policies>

Nature Portfolio journals: <https://www.nature.com/nature-research/editorial-policies>

Scientific Reports: <https://www.nature.com/srep/journal-policies/editorial-policies>

BMC journals: <https://www.biomedcentral.com/getpublished/editorial-policies>

Appendix A Section title of first appendix

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.

For submissions to Nature Portfolio Journals please use the heading “Extended Data”.

References

Borsboom, D., & Cramer, A. (2013, 03). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9(Volume 9, 2013), 91–121, <https://doi.org/10.1146/annurev-clinpsy-050212-185608>

Epskamp, S., Maris, G., Waldorp, L.J., Borsboom, D. (2018). Network psychometrics. In (pp. 953–986). John Wiley & Sons, Ltd.

- Estrada, E. (2011). *The structure of complex networks: Theory and applications*. Oxford University Press.
- Golino, H.F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, 12(6), e0174035, <https://doi.org/10.1371/journal.pone.0174035>
- Menezes, I., Menezes, A.C., Moraes, E., Pires, P.P. (2021). Measuring organizational climate via psychological networks analysis. *International Journal of Organization Theory & Behavior*, 24(3), 229–250, <https://doi.org/10.1108/IJOTB-08-2020-0142>
- Newman, M.E.J. (2010). *Networks: An introduction*. Oxford University Press.
- Oldham, S., Fulcher, B., Parkes, L., Arnatkevičiūtė, A., Suo, C., Fornito, A. (2019, 07). Consistency and differences between centrality measures across distinct classes of networks. *PLOS ONE*, 14(7), e0220061, <https://doi.org/10.1371/journal.pone.0220061>
- Soto, P.L. (2025, 12). Relational databases for behavior science. *Perspectives on Behavior Science*, , <https://doi.org/10.1007/s40614-025-00486-w>
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.