

Natural Language Use as Bipartite Networks in Psychology

Juan C. Correa

Fundación Universitaria Konrad Lorenz, Bogotá, Colombia

Natural language as a data source is quite common in different divisions of psychology. Among the several ways to analyze the information conveyed by natural language, psychologists rarely use bipartite networks despite the strong potential that this network perspective has for enriching psychology's research toolbox. This opinion article aims to provide a viewpoint on current advances and promising future research directions on modeling natural language as a bipartite network structure, using word-of-mouth as the basis for a tutorial exposition that paves the way for others to leverage the opportunities provided by network theory.

Keywords: Bipartite Networks, Natural Language Use, Text Mining, Affiliation Network, Textual Data Analysis

In psychology, neither natural language as a data source (Tausczik & Pennebaker, 2010) nor network analysis applications (Borsboom & Cramer, 2013) are entirely new, as they have their respective roots in Wundt's seminal works on psycholinguistics (Levelt, 2018) and Moreno's seminal works on sociometry and psychotherapy (Luke, 2015). Among the different ways natural language can be modeled or represented as a network structure, the concept of bipartite networks (BN) is a promising framework for extracting and analyzing the information conveyed in natural language. Although network scientists agree on the idea that BNs are convenient for understanding affiliation relationships (Kolaczyk, 2009; Newman, 2010; Wasserman & Faust, 1994), this framework is rarely applied by psychologists, being the study of Vallejo-Medina et al. (2020) a recent exception.


The omission of BNs in specialized social sciences textbooks (Dodsworth & Benton, 2019) and the massive number of research developed outside psychology's realm (Analytis et al., 2020; Corrêa Jr et al., 2018; Dhillon, 2001; Jesus et al., 2009; Kaya, 2020; Koskinen & Edling, 2012; Liu et al., 2018; Zha et al., 2001), might be reasonable factors that explain the scarce use of BNs that psychologists have currently shown. Above and beyond these reasons, an illustration of BNs and how to apply them for the analysis of psycho-linguistic phenomena deserves its own exposition. BNs have strong potential to enrich psychologists' research toolbox and provide a convenient framework for building and

testing psychological theories with techniques and concepts developed in the so-called complex systems sciences that are not well-known by psychologists either (Brusco et al., 2019; Correa, 2020). In what follows, the next section defines BNs, summarizes some examples, and provides a generalized pipeline that allows the analysis of natural language. To further illustrate the utility of BNs to the psychological community, the paper defines the concept of word-of-mouth as an example to develop a tutorial description. The goal of this perspective or opinion article is to provide a viewpoint on current applications of BNs and how they can stimulate future research in different branches of psychology.

Bipartite Networks: A brief review

A network represents any system, with two essential elements: nodes and edges (Wasserman & Faust, 1994). Nodes represent a particular entity (e.g., a neuron, a word, an individual, a city, or a country), while edges represent a specific relationship between these entities (e.g., a synapse, a collocation, a marriage, a highway, or an air route). A bipartite network, also known as an *affiliation*, *two-mode*, or *recommender* network (Newman, 2010), is a particular network with two types of nodes. Links or edges exist between different types of nodes but not between the same type of node. Thus, in a bipartite network, the interest lies in tracking node affiliations or memberships. While the formal representation of a unipartite network is through an $N \times N$ symmetric matrix, a bipartite network can be formally represented as an $N \times E$ *incidence matrix*, where N stands for the number of nodes, while E stands for the number of edges, with $N \leq E$. Let's look at the following matrices A and U .

Although A is an incidence matrix, its consideration as a bipartite matrix occurs when an explicit criterion splits the nodes into two categories. In matrix U the presence of a link between any pair of node $[N_i, N_j]$ is represented as 1 and

Juan C. Correa  <https://orcid.org/0000-0002-0301-5641>

Correspondence concerning this opinion article should be addressed to Juan C. Correa (<https://correajc.com/>), Fundación Universitaria Konrad Lorenz, Escuela de Posgrados, Bogotá, Colombia, Carrera 10 #64-61. E-mail: juanc.correan@konradlorenz.edu.co

$$\begin{array}{rcc}
 & & \begin{array}{cccc} N_1 & N_2 & N_3 & N_4 \end{array} \\
 U = & \begin{array}{c} N_1 \\ N_2 \\ N_3 \\ N_4 \end{array} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \\
 \\
 & & \begin{array}{ccccc} E_1 & E_2 & E_3 & E_4 & E_5 \end{array} \\
 A = & \begin{array}{c} N_1 \\ N_2 \\ N_3 \\ N_4 \end{array} & \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}
 \end{array}$$

0 otherwise. Elements above and below the main diagonal of matrix U are the same, as there is no interest in distinguishing the direction of a link (as in the case of a directed network, where it is important to differentiate which node goes in the first position, and which one goes in the second position). Thus, the five non-ordered connections in matrix U ($N_1 - N_2, N_1 - N_3, N_1 - N_4, N_2 - N_4, N_3 - N_4$), can be re-expressed as matrix A . Although A is an incidence matrix, it is still considered a unipartite matrix unless its nodes are differentiated into two categories. Here, psychologists have a virgin ground on which substantive theory can be developed, as we will show below. Before turning the attention to particular ways in which node categorization might be framed in psychological terms, some examples from other disciplines are worth mentioning.

Park et al. (2021) used BNs to analyze the relationship between the ingredients of different foods and compounds chemicals of its flavor for generating food representations and recommending food pairings, which has the potential to introduce innovations in the food industry. Their work relied on the paper of Ahn et al. (2011) who coined the concept of food pairing as a mechanism that describes cultural gastronomical differences between Western and Eastern cuisines. In economics, Hidalgo and Hausmann (2009) used BNs to understand international trade data by representing countries as a particular type of node and products as a second particular type of node, and develop the “economic complexity index” as a robust predictor of countries’ future GDP per capita growth. In health psychology, BNs were useful for knowing if the readability of the written materials that accompany a health promotion program corresponds to the target population it was designed for, representing the number of sessions as a particular type of node and the most common words in each session, as the other type of node (Vallejo-Medina et al., 2020). While these applications are relevant for those trying to apply quantitative analyses to qualitative information (Fakis et al., 2014), BNs nicely fit into psychology’s research toolbox. In consumer psychology, for example, BNs can be used to analyze customers’ word-of-mouth data, as this particular type of data is a robust input for predicting and under-

standing consumers’ behavior (Berger, 2014; Berger et al., 2020). Extracting useful information from natural language requires the application of the network mapping process (Kolaczyk, 2009) illustrated in Figure 1. Here, the input is a single text or a collection of texts, and the output is a BN, whose structure becomes the input for subsequent quantitative analyses. Nonetheless, two preliminary considerations are mandatory. Firstly, the textual unit represented as nodes of a network should be explicitly defined (e.g., a word, a sentence, a paragraph, a comment). The criteria defining nodes affiliation should also be clarified (e.g., a semantic relationship, a contentwise link, or a psychologically-grounded affiliation). While theoretical explanatory power should guide the pursuit of nodes-bonding, the process of network mapping to textual data becomes vital, as “[This] is usually a non-trivial task of significant interest in itself” (Kolaczyk, 2009, p. 9).

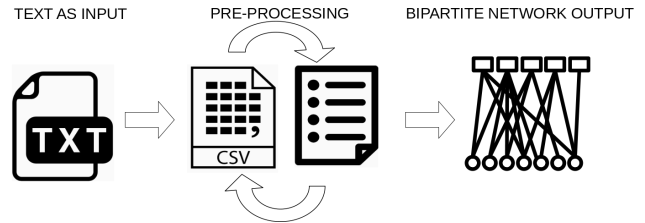


Figure 1

A generalized pipeline to represent natural language as bipartite networks

Recognizing natural language as a system from which useful information can be extracted offers exciting opportunities for developing theories based on psychological contents with the potential of being an interdisciplinary object of study (Correa, 2020). Such an orientation is fairly welcomed in consumer psychology, where traditional theoretical focus prioritizes psychological processes above psychological contents (Pham, 2013). Put it differently, BNs are useful when the unit of analysis is the collective trend of what people say. This analytical unit is fairly different from previous attempts using referral networks with edges between persons (Reingen & Kernan, 1986), instead of representing words clustered by topics for example (Teichert et al., 2020). Given the potential of psychological research focused on the analysis of natural language as a BN structure, in my viewpoint, a tutorial exposition is well-deserved as it helps newcomers to leverage this framework for both theoretical or applied purposes.

Word-of-mouth as bipartite network: A tutorial

As an oral or written statement about the pros and cons of products or services, word-of-mouth is a well-documented topic in consumer psychology (Teeny et al., 2020; Weaver

& Hamby, 2019), and recent studies have shown how to collect word-of-mouth data through web mining techniques (Correa et al., 2019). By using the web scraping procedure described by Teichert et al., 2020 the following tutorial shows a hands-on approach with an ad-hoc data set with the following four columns: i) Restaurant, ii) Ranking, iii) text, and iv) the total number of comments that each restaurant received at the moment of data collection. For interested readers, the following syntaxes are available here (https://github.com/jcorrean/Text_as_Bipartite_Network/blob/main/Example.R). The process begins with the data loading, as follows

```
library(quanteda)
load("/home/juan/Comments.RData")
my_corpus <- corpus(Comments$text)
mycorpus <- data.frame(summary(
  my_corpus, n = nrow(Comments)))
summary(mycorpus)
```

The column `text` is the input data for creating a standard corpus with the “`quanteda`” package (Benoit et al., 2018). Further quantitative characterization of customers’ comments proceeds by creating a document-term matrix from customers’ word-of-mouth (i.e., each comment is considered a document). In creating this document-term matrix, numeric characters and stopwords (i.e., words without semantic meaning like articles, prepositions, etc.) should be removed.

```
spanishstopwords <- c("q", stopwords("spanish"))
Restaurant <- dfm(corpus(
  my_corpus),
  remove_numbers = TRUE,
  remove = spanishstopwords,
  stem = TRUE, remove_punct = TRUE)
restaurant <- textstat_simil(Restaurant,
  margin = "documents",
  method = "jaccard")
restaurantdf <- data.frame(as.matrix(
  restaurant))
restaurantdf[is.na(restaurantdf)] = 0
restaurant <- data.frame(
  jaccard = restaurant[lower.tri(restaurant,
  diag = FALSE)])
```

As the document-term matrix has the form of an $N \times E$ matrix, it could be regarded as an incidence matrix that facilitates the affiliation represented by each customer and his/her words. Although this representation is compatible with qualitative interpretations of network analysis (Decuypere, 2020), psychologists can proceed differently for theoretical development purposes. Here, clustering techniques are fundamental to explore the structure of customers’ experiences

from a particular sample (Teichert et al., 2020) or test if such a structure preserves regardless of the sample it belonged to. The theoretical orientation of these efforts is recently recognized by scholars working on blockmodeling techniques (Brusco et al., 2019) with foundations going back to the structural balance theory (Cartwright & Harary, 1956). Given the variety of clustering techniques to analyze textual data represented as BN, it is impossible to cover them all properly. Nonetheless, the following treatment should suffice to inspire other applications. The document-term matrix can be re-expressed as a similarity matrix whose entries include the Jaccard distance index quantifying the similarity between every pair of comments, A and B , calculated as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

A reasonable step is evaluating the matrix structure by applying the Hopkins statistic from the “`clustertend`” R package (YiLan & RuTong, 2015).

```
library(clustertend)
set.seed(123)
hopkins(restaurantdf, n = nrow(restaurantdf)/10)
## $H
## [1] 0.06583476
```

The Hopkins statistic should be below the threshold of 0.5 to consider that the matrix is clusterable (Kassambara, 2017); otherwise, the idea of a bipartite network would be unfounded. The preliminary result (Hopkins statistics = 0.0658) indicates that the matrix is clusterable, and further pre-processing tasks might follow. The “`mclust`” R package (Scrucca et al., 2016) is now required to employ a model-based clustering, which considers the data as coming from a distribution that is a mixture of two or more clusters. According to Kassambara, 2017, the model-based clustering uses a soft assignment, where each data point (i.e., each written comment) has a probability of belonging to each topic. The model-based clustering is a technique that relies on parametric finite Gaussian mixture models that employ the so-called “expectation maximization” (EM) algorithm initialized by hierarchical model-based agglomerative clustering. With the Bayesian Information Criterion, this technique results in an optimal model that suggests the number of clusters in a data set.

```
library(mclust)
fit <- Mclust(restaurantdf)
summary(fit)
-----
Gaussian finite mixture model fitted by EM algorithm
-----
Mclust EII (spherical, equal volume) model with
9 components:
```

```
log-likelihood    n    df      BIC      ICL
      5089368 1830 16479 10054944 10054944
```

Clustering table:

```
  1  2  3  4  5  6  7  8  9
148 792 303 176 139 90 79 78 25
```

The outcome from this analysis suggests that customers' comments can be summarized in terms of nine different clusters. A total of 148 statements belong to cluster one, 792 belongs to cluster two, 303 belongs to cluster three, and so on. The "topicmodels" R package (Grün & Hornik, 2011) helps identify the membership of unique words to each of these clusters with the Latent Dirichlet Allocation (LDA) algorithm. Because this algorithm returns the so-called hyperparameter β ("beta") that refers to the probability of each word belonging to each topic, it is convenient the creation of a single data frame that contained the associated beta for the set of most common words (also called essential words) employed by customers to express their experiences with the online food delivery platform.

```
library(topicmodels)
restaurantC <- convert(Restaurant,
  to = "topicmodels",
  docvars = NULL)
pave <- LDA(restaurantC, k = 9)
library(tidytext)
restaurant_topics <- tidy(pave, matrix = "beta")
restaurant_topics$Category <- "restaurant"
```

These essential words are manipulated in a relational data set with the aid of the "dplyr" R package (Wickham & Grolemund, 2017). Then, the process of extracting bipartite network data is pretty straightforward with the "igraph" R package (Kolaczyk & Csárdi, 2014).

```
library(broom)
EssentialWordsrestaurant <- tidy(pave)
library(dplyr)
topics <- EssentialWordsrestaurant %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)
```

```
library(igraph)
g_rest <- graph.data.frame(topics,
  directed = FALSE)
bipartite.mapping(g_rest)
V(g_rest)$type <- bipartite_mapping(g_rest)$type
V(g_rest)$color <- ifelse(V(g_rest)$type,
  "lightblue", "red")
```

```
V(g_rest)$shape <- ifelse(V(g_rest)$type,
  "circle", "square")
E(g_rest)$color <- "black"
V(g_rest)$size <- eccentricity(g_rest) * 2.5
V(g_rest)$label.cex <- degree(g_rest) * 2.5
plot(g_rest,
  vertex.label.cex = 0.8,
  vertex.label.color = "black",
  layout = layout_with_dh)
```

After running previous syntaxes, the resulting network visualization should be similar as Figure 2. Any variation in the way essential words are identified, should result in a different bipartite network structure. For example, the number of essential words in this example was arbitrarily set to 10 (i.e., each cluster is connected to 10 words, but words can be linked to more than one cluster). Nonetheless, this number could be defined differently, affecting the size of the network accordingly. An emerging feature relates to some words, represented with light blue circles, that share ties with more than one topic cluster, represented by red boxes. A remark here is that the proximity between clusters in this visualization has nothing to do with their semantic similarity. Instead, the mere counting of shared words by topic clusters could help identify subtle semantic differences among topics. In this context, Correa, 2020 has asserted that such subtleties might be observed in recommendations (e.g., a strong recommendation without hesitation on any aspect of the service versus a recommendation accompanied by a warning regarding food variety). Other alternative analyses might focus on evaluating customers' word-of-mouth network differences as a function of commercial brands, restaurant categories, or any other marketing-oriented variable. Here, consumer psychologists might leverage the combined use of natural language processing and bipartite network analysis for similar purposes. Such an orientation, in turn, is compatible with previous arguments that point out the relevance of paying much more attention to the psychological contents of consumer behavior, as opposed to its psychological processes, and expand the epistemology beyond the traditional theory-driven path, and adopt more phenomenon-based research (Pham, 2013).

Final remarks

Even though the natural language is frequently used as a data source in psychological research, psychologists rarely analyze natural language as a bipartite network structure. Based on the idea that network visualization is not a trivial task for the reasons well-discussed in other sources (Kolaczyk & Csárdi, 2014; Luke, 2015), this paper aimed at providing a tutorial or hands-on approach for psychologists who use natural language as a data source from which information extraction becomes an essential task, either for theoretical or practical purposes. Because visual imagery proves essential

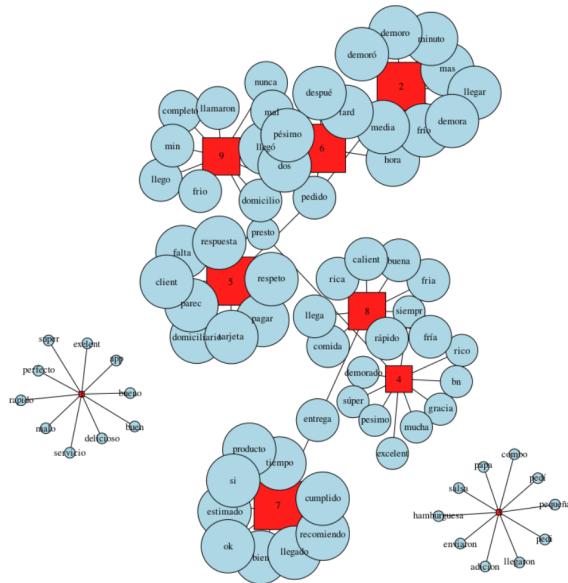


Figure 2
A bipartite network visualization of essential words associated with emerging topics from customers’ word-of-mouth.

in network data analysis (Kolaczyk, 2009), future research diving deep into those matters might be beneficial not only for psychologists but also for network scientists with a special interest in analyzing psychological data.

In my opinion, the more promising opportunities relate to the development of psychological theories that provide substantial arguments by which the information conveyed in natural language can be framed as bipartite network structures. Such endeavors would be helpful by defining criteria for nodes categorization purposes and nodes connections descriptions, just as previous contributions in economics (Hidalgo & Hausmann, 2009), food sciences (Ahn et al., 2011; Park et al., 2021), or health psychology (Vallejo-Medina et al., 2020). Inspired by these advances, health psychologists interested in nutrition could provide novel ways of understanding eating behavior by describing the relationship between ingredients and typical and atypical ways of preparing them through a bipartite network where nodes are categorized as either ingredients or the cooking actions (verbs) required to prepare a recipe.

Another discipline that benefits from modeling natural language as bipartite networks is social psychology in general and marital processes and romantic relationships in particular. Although research on these topics has always relied on observing verbal and non-verbal interactions between spouses (J. Gottman & Krokoff, 1989), more recent developments on these interactions (J. M. Gottman, 2015) have missed the potential benefits of modeling verbal interactions as suggested here. Educational psychology is another branch

where BNs can be used for analyzing and visualizing students’ learning based on multi-topic chat texts (Zhang et al., 2018). A third branch is industrial and organizational psychology. Here, several scholars agree on the idea of limited empirical research tackling how occupations and careers are changing (Spreitzer et al., 2017). This topic is pertinent to the potential impact that artificial intelligence and automation will have on the future of jobs (Rampersad, 2020). Nonetheless, the concept of bipartite networks remain unnoticed by organizational psychologists who highlighted the promises of computer-assisted text analysis in recent reviews (Campion & Campion, 2020). Likewise, in personnel psychology, the combination of natural language use with bipartite network analysis might help structure individuals’ LinkedIn profile, and according to Roulin and Levashina, 2019, this use is an example of research lagging far behind the practice. If personality data can be extracted from LinkedIn profiles (Roulin & Levashina, 2019), then network-based analyses, as illustrated by Costantini et al., 2015, might become as another innovative practice. Apart from substantive areas, recent reviews on network analysis for qualitative research (Decuyper, 2020; Luxton & Sbicca, 2020) do not even mention the term bipartite networks. Thus, qualitative researchers may leverage them as well. Arguably, the so-called network statistics (e.g., diameter, density, degree centrality, Erdos number) and their potential applications (not covered here) might be of great help for psychologists interested in the quantitative characterization of textual data sources.

References

Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., & Barabasi, A.-L. (2011). Flavor network and the principles of food pairing. *Scientific Reports, 1*:196, 1–7. <https://doi.org/10.1038/srep00196>

Analytis, P. P., Barkoczi, D., Lorenz-Spreen, P., & Herzog, S. (2020). The structure of social influence in recommender networks. *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020, 2655–2661*. <https://doi.org/10.1145/3366423.3380020>

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30), 774.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology, 24*(4), 586–607.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of*

- Marketing*, 84(1), 1–25. <https://doi.org/10.1177/0022242919873106>
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121.
- Brusco, M., Doreian, P., & Steinley, D. (2019). Deterministic blockmodelling of signed and two-mode networks: A tutorial with software and psychological examples. *British Journal of Mathematical and Statistical Psychology*. <https://doi.org/10.1111/bmsp.12192>
- Campion, E., & Campion, M. (2020). Using computer-assisted text analysis (cata) to inform employment decisions: Approaches, software, and findings. *Research in Personnel and Human Resources Management*, 38, 285–325. <https://doi.org/10.1108/S0742-73012020000038010>
- Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of heider's theory. *Psychological Review*, 63(5), 277–293.
- Correa, J. C. (2020). Metrics of emergence, self-organization, and complexity for ewom research. *Frontiers in Physics*, 8, 35. <https://doi.org/10.3389/fphy.2020.00035>
- Correa, J. C., Garzón, W., Brooker, P., Sakarkar, G., Caranza, S. A., Yunado, L., & Rincón, A. (2019). Evaluation of collaborative consumption of food delivery services through web mining techniques. *Journal of Retailing and Consumer Services*, 46, 45–50. <https://doi.org/10.1016/j.jretconser.2018.05.002>
- Corrêa Jr, E. A., Lopes, A. A., & Amancio, D. R. (2018). Word sense disambiguation: A complex network approach. *Information Sciences*, 442, 103–113.
- Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, 54, 13–29. <https://doi.org/10.1016/j.jrp.2014.07.003>
- Decuyper, M. (2020). Visual network analysis: A qualitative method for researching sociomaterial practice. *Qualitative Research*, 20(1), 73–90. <https://doi.org/10.1177/1468794118816613>
- Dhillon, I. S. (2001). Co-clustering documents and words using bipartite spectral graph partitioning. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 269–274.
- Dodsworth, R., & Benton, R. (2019). *Language variation and change in social networks: A bipartite approach*. Routledge.
- Fakis, A., Hilliam, R., Stoneley, H., & Townend, M. (2014). Quantitative analysis of qualitative information from interviews: A systematic literature review. *Journal of Mixed Methods Research*, 8(2), 139–161.
- Gottman, J. M. (2015). *Principia amoris: The new science of love*. Routledge.
- Gottman, J., & Krokoff, L. (1989). Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology*, 57(1), 47–52. <https://doi.org/10.1037//0022-006x.57.1.47>
- Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30. <https://doi.org/10.18637/jss.v040.i13>
- Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26), 10570–10575.
- Jesus, R., Schwartz, M., & Lehmann, S. (2009). Bipartite networks of wikipedia's articles and authors: A meso-level approach. *Proceedings of the 5th international symposium on Wikis and open collaboration*. <https://doi.org/10.1145/1641309.1641318>
- Kassambara, A. (2017). *Practical guide to cluster analysis in r: Unsupervised machine learning* (Vol. 1). STDA.
- Kaya, B. (2020). Hotel recommendation system by bipartite networks and link prediction. *Journal of Information Science*, 46(1), 53–63. <https://doi.org/10.1177/0165551518824577>
- Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. Springer-Verlag New York.
- Kolaczyk, E. D., & Csárdi, G. (2014). *Statistical analysis of network data with r* (1st ed.). Springer-Verlag New York.
- Koskinen, J., & Edling, C. (2012). Modelling the evolution of a bipartite network-peer referral in interlocking directorates. *Social Networks*, 34(3), 309–322. <https://doi.org/10.1016/j.socnet.2010.03.001>
- Levelt, W. J. (2018). Is language natural to man? some historical considerations. *Current opinion in behavioral sciences*, 21, 127–131.
- Liu, C., Tang, L., & Shan, W. (2018). An extended hits algorithm on bipartite network for features extraction of online customer reviews. *Sustainability (Switzerland)*, 10(5). <https://doi.org/10.3390/su10051425>
- Luke, D. A. (2015). *A User's Guide to Network Analysis in R*. Springer.
- Luxton, I., & Sbicca, J. (2020). Mapping movements: A call for qualitative social network analysis. *Qualitative Research*. <https://doi.org/10.1177/1468794120927678>
- Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.

- Park, D., Kim, K., Kim, S., Spranger, M., & Kang, J. (2021). Flavorgraph: A large-scale food-chemical graph for generating food representations and recommending food pairings. *Scientific reports*, *11*(1), 1–13. <https://doi.org/10.1038/s41598-020-79422-8>
- Pham, M. T. (2013). The seven sins of consumer psychology. *Journal of Consumer Psychology*, *23*(4), 411–423.
- Rampersad, G. (2020). Robot will take your job: Innovation for an era of artificial intelligence. *Journal of Business Research*, *116*, 68–74. <https://doi.org/10.1016/j.jbusres.2020.05.019>
- Reingen, P. H., & Kernan, J. B. (1986). Analysis of referral networks in marketing: Methods and illustration. *Journal of Marketing Research*, *23*(4), 370–378.
- Roulin, N., & Levashina, J. (2019). LinkedIn as a new selection method: Psychometric properties and assessment approach. *Personnel Psychology*, *72*(2), 187–211. <https://doi.org/10.1111/peps.12296>
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 289–317. <https://doi.org/10.32614/RJ-2016-021>
- Spreitzer, G., Cameron, L., & Garrett, L. (2017). Alternative work arrangements: Two images of the new world of work. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 473–499. <https://doi.org/10.1146/annurev-orgpsych-032516-113332>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.
- Teeny, J., Deng, X., & Unnava, H. (2020). The “buzz” behind the buzz matters: Energetic and tense arousal as separate motivations for word of mouth. *Journal of Consumer Psychology*, *30*(3), 429–446. <https://doi.org/10.1002/jcpy.1151>
- Teichert, T., Rezaei, S., & Correa, J. C. (2020). Customers’ experiences of fast food delivery services: Uncovering the semantic core benefits, actual and augmented product by text mining. *British Food Journal*, *122*(11), 3513–3528. <https://doi.org/10.1108/BFJ-12-2019-0909>
- Vallejo-Medina, P., Correa, J. C., Gómez-Lugo, M., Saavedra-Roa, D. A., García-Montaño, E., Pérez-Pedraza, D., Niebles-Charris, J., García-Roncallo, P., Abello-Luque, D., Espada, J. P., & Morales, A. (2020). A text mining approach for adapting a school-based sexual health promotion program in Colombia. *Preventive Medicine Reports*, *18*, 101090. <https://doi.org/10.1016/j.pmedr.2020.101090>
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Weaver, K., & Hamby, A. (2019). The sounds of silence: Inferences from the absence of word-of-mouth. *Journal of Consumer Psychology*, *29*(1), 3–21. <https://doi.org/10.1002/jcpy.1067>
- Wickham, H., & Grolemund, G. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. O’Reilly.
- YiLan, L., & RuTong, Z. (2015). *Clustertend: Check the clustering tendency* [R package version 1.4]. <https://CRAN.R-project.org/package=clustertend>
- Zha, H., He, X., Ding, C., Simon, H., & Gu, M. (2001). Bipartite graph partitioning and data clustering. *Proceedings of the tenth international conference on Information and knowledge management*, 25–32.
- Zhang, H., Pan, F., Wu, Z., & Ji, Y. (2018). Analysis and visualization of students’ learning based on multi-topic chat text. *Proceedings of the 6th IEEE International Conference on MOOCs Innovation and Technology In Education, MITE 2018*, 90–97. <https://doi.org/10.1109/MITE.2018.8747084>