EXPLORATORY GRAPH ANALYSIS WITH ESTIMATED CONSUMER UTILITIES: A PILOT STUDY FOR UNCOVERING THE NETWORK DIMENSIONALITY OF ICED TEA CLAIMS

A RESEARCH PAPER

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

BY

THOMAS J. NICOLETTI

W. HOLMES FINCH, PHD – ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

DECEMBER 2020

Exploratory Graph Analysis with Estimated Consumer Utilities:

A Pilot Study for Uncovering the Network Dimensionality of Iced Tea Claims

Thomas J. Nicoletti

W. Holmes Finch, PhD

Department of Educational Psychology, Ball State University

Research 697 – Research Paper for Educational Psychology

December 2020

Author Note

The author would like to express great appreciation to SKIM, a global market research insights agency, for providing the data used in this study.

Abstract

Using estimated consumer utilities for 21 different iced tea claims, this study aimed to examine how typical factor extraction approaches, as well as exploratory graph analysis, performed comparatively. Data was collected by an international market research agency on behalf of a consumer goods organization headquartered in a western-culture European country. The conceptual and statistical worth of exploratory graph analysis within the greater consumer beverages context was then evaluated. Exploratory graph analysis was able to reorganize claims into dimensions that were not only in agreement with most other factor extraction methods, but also those which were distinct and conceptually appropriate. Despite achieving such clear theoretical groundwork for future efforts on latent constructs in market research, much was still desired in terms of the statistical fit of this particular iced tea claims model. This provides applied market researchers, as well as social science academics, a unique opportunity to better understand the network dimensionality of products and services using consumer utilities. As such, continuation of this interdisciplinary, original research is highly encouraged.

Keywords: exploratory graph analysis, consumer utilities, iced tea claims, network psychometrics, dimensionality reduction, market research, consumer goods

Exploratory Graph Analysis with Estimated Consumer Utilities:

A Pilot Study for Uncovering the Network Dimensionality of Iced Tea Claims

In our modern, data-driven business world, consumer goods organizations often seek market insights by partnering with research agencies to better understand consumer behavior in a variety of contexts. Through this interdisciplinary, practical approach to answering key business questions, concepts from economics, market research, psychology, and sociology are weaved together into a strategic story for decision-making. Organizations are then able to better tailor their products and services towards the preferences of their consumers, ultimately yielding financial growth, as well as associated increases to organizational key performance indicators (Kim et al., 2003; Narayana & Markin, 1975). And as applied research in this area continues to evolve with technological and methodological advancements, opportunities arise that could further benefit the field of market research, namely the application of novel psychometric techniques (Carroll & Green, 1997; Wedel & Kannan, 2016).

Topics in psychometrics span great breadths and depths by marrying together research designs and conceptual frameworks critical to psychology with the rigorous application of classical and contemporary statistics. One such area of great impact and importance to psychometrics is dimensionality reduction, or a family of methods useful for discovering the latent constructs underlying some set of observable psychological variables. One major method within this family is exploratory factor analysis (EFA), which is similar to the well-known principal components analysis. Although EFA is a widely used technique in psychological research (Cattell, 1978; Golino & Epskamp, 2017), it also wields conceptual and statistical intricacies that can introduce unwanted researcher subjectivity (Henson & Roberts, 2006). One such concern revolves around the age-old question, “how many factors should I extract for this dimensionality analysis?”

Of course, a plethora of options exist to help inform researchers on the number of factors to extract for EFA. More specifically, such examples include the Kaiser-Guttman rule, parallel analysis, the very simple structure approach, Vellicer’s minimum average partial, and various maximum likelihood approaches (Golino & Epskamp, 2017; Morton & Altschul, 2019). But in most applied cases, it is quite difficult to objectively determine which of these methods yield more appropriate recommendations when discrepancies manifest. In a recent development by Golino and Epskamp (2017), a robust new method titled exploratory graph analysis (EGA) was added to the psychometrician’s toolbox, which leverages a more objective, network model approach to determining the optimal number of factors – now dimensions – to extract. In its seminal introduction to the greater scientific literature, it yielded promising results in its performance compared to the aforementioned alternatives (Golino & Epskamp, 2017). Due to its graphical, network-based output, more information can be provided to the researcher than ever before. For example, researchers can now quickly identify which variables are present within each estimated dimension.

In greater depth, EGA is categorized as a network psychometrics technique that provides a graphical model for visual interpretation. It abides by the fundamental rule that clusters in networks are in fact latent constructs (Golino & Epskamp, 2017). On the mathematical end, EGA begins by estimating the correlation matrix of all observable variables. Following this, graphical least absolute shrinkage and selection operator (lasso) estimation is used to acquire the sparse inverse covariance matrix. And finally, a walktrap algorithm is employed to determine the number of clusters present (Golino & Epskamp, 2017). Golino and Epskamp (2017) explained that EGA is likely to "present a high accuracy in estimating the number of dimensions in psychology-like datasets due to the use of the [lasso] technique". In other words, because partial correlations are often used in estimating network models, the usage of the L1 regularization technique is important in minimizing spurious relationships between variables (Golino and Epskamp, 2017).

Dimensionality reduction has proven its worth for the field of psychometrics across the decades (Cattell, 1978), and it can likely produce similar benefits for tangential fields like market research. When it comes to developing new products and innovations, reimagining pack designs and messaging, or overcoming challenging disruptions to habitual shopping behaviors, consumer goods organizations need to make informed decisions to better their offerings to the market. EGA could prove useful in simplifying the complex nature of human decision-making by elucidating latent constructs that guide people in their consumption choices. For example, consumers might not see hundreds of different beverage claims, but rather they associate each claim with a larger, encompassing dimension with which they hold great preference towards.

Using estimated consumer utilities for iced tea claims, we will first examine how typical extraction approaches, as well as EGA, perform comparatively when determining the number of factors to extract. Following this, we will evaluate the conceptual worth of EGA within the iced tea market research context. And finally, confirmatory factor analysis (CFA) will be conducted based on the EGA-informed solution. Taken together, this pilot study aims to introduce market researchers to an innovative psychometric technique, explore both its strengths and drawbacks, and ultimately spark new interest within the field of consumer goods marketing.

Methods

Research Design

Data was collected by an international market research agency on behalf of a consumer goods organization headquartered in a western-culture European country. Twenty-one unbranded iced tea claims (see Table A1) were drafted by the consumer goods organization for testing using an online MaxDiff, or best-worst scaling, design. Participants were asked to choose their most and least preferred claims across multiple iterations, which varied between exercises and across participants. To yield the interval-level estimated consumer utilities, hierarchical Bayes estimation was used on the raw choice data. For more information on this process, please see Sawtooth Software (n.d.).

Participants

Data collected by the international market research agency was bound by contractual law with their client, as well as global market research ethics organizations (ESOMAR, n.d.). As such, information specifically about this non-probability panel sample (N = 203) was not provided for this report beyond the anonymous estimated consumer utilities. However, as in typical fashion for this agency, it is assumed that the gender distribution was inclusive and uniform between females and males, and that the age distribution was approximately normally distributed between the ages of 18 and 65 years old. Further extrapolation of sample characteristics will not be assumed, though such missing information need be brought to the attention of those interested in this research.

Analysis Plan

To begin, factor extraction insights will be generated using A) the Kaiser-Guttman rule, B) parallel analysis, C) the optimal coordinates approach, D) the acceleration factor, E) Zoski and Jurs regression, F) the very simple structure approach, G) Vellicer's minimum average partial, H) the maximum likelihood estimation chi-square test, and I) EGA. In far greater depth, EGA results will be evaluated for conceptual relevance given the context of iced tea claims in a market research environment. Following this step, a CFA will be conducted to statistically validate the EGA-informed solution.

Statistical Programming

Resources are widely available for researchers interested in replicating this study or investigating the exciting applications of EGA for future studies. Such resources include the open programming language R (The R Foundation, n.d.), as well as its associated EGA package, *EGAnet* (Golino et al., 2018). During the time this study took place, no alternative method for computing EGA existed outside of the R environment. As such, reproducible R code is also provided below in Appendix B for both reference and accountability.

Results

Factor Extraction Methods

Out of the eight non-EGA factor extraction methods calculated for this analysis, a majority of three agreed that a six-factor solution would be optimal for the data. More specifically, these three methods were the Kaiser-Guttman rule, parallel analysis, and the optimal coordinates approach (see Figure A1). The acceleration factor determined the optimal solution to be two factors and disagreed with all other methods (see Figure A1). Zoski and Jurs regression (b = 4) and the very simple structure approach agreed that a four-factor solution was ideal (see Figure A2). And finally, Vellicer's minimum average partial (MAP = 0.07) and the maximum likelihood estimation chi-square test ($\chi^2 = 4{,}734.22$, $p < 0.0001$, RMSR = 0.02, RMSEA = 0.57) found an optimal, yet mediocre solution of eight factors for the data. It should be noted that the latter method was compared against – and outperformed – the two-, four-, and six-factor solutions as well. Given these findings, the Kaiser-Meyer-Olkin test for sampling adequacy was computed post-hoc and found poor reliability for the factorability of the data (MSA = 0.12).

Exploratory Graph Analysis

A parametric bootstrapped EGA was conducted with 500 iterations using the graphical lasso technique. Results from this network model yielded an optimal solution of six dimensions, which agreed with the majority of typical factor extraction methods tested previously (see Figure A3). On this network graph, nodes signify each iced tea claim (see Table A1), node colors represent different clusters (i.e., latent constructs), and edges depict positive and negative partial correlations (i.e., blue and red edges, respectively) based on the median network structure across the bootstrap.

Confirmatory Factor Analysis. For the EGA-informed six-dimension solution, a CFA was calculated and found suboptimal fit for this model ($\chi^2 = 7{,}003.32$, $p < 0.0001$, SRMR = 0.17, RMSEA = 0.44, CFI = 0.27, TLI = 0.11). This result was in alignment with the previous finding via the Kaiser-Meyer-Olkin test for sampling adequacy. It should be noted that Mardia’s test for multivariate normality was also conducted and found a violation to this assumption due to non-normal kurtosis ($K = -5.16$, $p < 0.0001$; see Figure A4).

Discussion

Exploratory Graph Analysis: Conceptual Value

In terms of results yielded from the array of tested factor extraction methods, EGA was a majority member agreeing with the optimal six-dimension solution. From these objectively estimated dimensions, it appears that clear, separate latent constructs exist for iced tea claims. More specifically, the following categories manifested through this technique: 1) Sugar Alternatives containing claims E, F, M, N, and O; 2) Low and Real Sugar containing claims H, I, K, and L; 3) Competitive Value containing claims C, D, J, and P; 4) Calorie Counts containing claims G, T, and U; 5) Precise Measurements containing claims R and S; and 6) Ideal Sweetness containing claims A, B, and Q (see Table A1; see Figure A3). Through this dimensionality reduction estimation, it is reasonable to then conclude that consumers do in fact hold preferences towards some hierarchical latent construct that weaves together similar, yet descriptively different marketing messages.

In taking a closer look at these dimensions, such conclusions about the psychological underpinnings of consumer behavior begin to emerge more clearly. For example, the first dimension Sugar Alternatives contains claims revolving around Stevia (i.e., an alternative sweetener), Agave syrup, and honey in addition to or in replacement of sugar. The second dimension Low and Real Sugar contains claims that home in on real sugar and the use of it in controlled quantities. The third dimension Competitive

Value contains claims that reference soft drinks, iced teas, and fruit juice in conjunction with and comparison to the tested product. The fourth dimension Calorie Counts is unanimously dedicated to the number of calories within the tested product. Similarly, the fifth dimension Precise Measurement is entertained by claims that provide specific facts about the use of sugar in grams per milliliters. And finally, the sixth dimension The Right Sweetness focuses on high-level characteristics about the saccharinity of the tested product. Taken together, EGA was able to reorganize claims into clusters that were not only in agreement with most other factor extraction methods, but also distinct and conceptually appropriate within the greater market research context. In addition, this graphical solution acts as a very useful tool for both storytelling and decision-making purposes.

Exploratory Graph Analysis: Statistical Value

In the applied business world, conceptual worth is often regarded in equal or greater value in comparison with the statistical properties of a given model in relatively low-risk environments. However, striking a healthy medium between the two is often ideal. Despite achieving such clear theoretical groundwork for future research on latent constructs in market research, much is still desired in terms of the statistical fit of this particular iced tea claims model. More specifically, between the suboptimal results yielded from the Kaiser-Meyer-Olkin test for sampling adequacy and the CFA, this iced tea claims data is simply not suited for proper factor analytic procedures; this makes it quite difficult to evaluate the accuracy of EGA when compared to other factor extraction methods. The data has poor reliability in terms of factorability and unfortunately violated the assumption of multivariate normality, ultimately dampening model interpretability.

However, this finding lends itself to the core research question of whether such market research data can be reduced to uncover latent psychological constructs. The insights provided in this paper indicate that conceptual support is both evident and interpretable, but that further research on the feasibility of factor analysis within this space is required. More specifically, avenues for improvement and optimization include running these procedures using consumer utilities data that has been properly factor analyzed. This would provide greater understanding into how EGA performs statistically by first comparing it to those previous exploratory results, and then assessing the related confirmatory model(s). Another option includes treating the iced tea claims tested in this study as traditional factor analysis items in psychological research, with participants rating each using carefully designed Likert scales. This would be especially beneficial if the use of consumer utilities is the driving force muddying the statistical waters. And finally, it is worth considering how principal components analysis compares to EGA as a dimensionality reduction tool, rather than an opportunity to uncover latent constructs. As this area of research continues to evolve, the knowledge gained from each consecutive study will be able to better inform this process and allow researchers to draw more appropriate statistical conclusions.

Limitations

As a pilot study taking a novel psychometric technique into unknown market research territories, exploratory knowledge was sought over perfect scientific findings, with drawbacks as foci for future iterations and replications. As such, the following items consist of the most predominant limitations in this study: A) Sample size and participant quality, B) the estimation process of consumer utilities, and C) data factorability and appropriate model validation procedures. For follow-up studies, it is therefore suggested that sample size be increased to at least 500 participants minimum and that they be randomly sampled from the population of interest. Additionally, researchers should begin assessing the impact of hierarchical Bayes estimated consumer utilities on factorability within different market research contexts (i.e., brand or category), as well as varying research designs (e.g., conjoint analysis). And finally, once all other issues are remedied, it is recommended that proper model validation procedures be conducted to verify the statistical properties of EGA within this exciting research space.

Conclusion

By leveraging the innovative advancements and time-tested tools of psychometrics in fields like market research, consumers and organizations have the potential to experience great benefits. And by furthering research on the latent psychological constructs that guide shopping behavior, new product development and consumer decision-making will become streamlined and supported by both conceptual thought and empirical results. EGA provides applied market researchers, as well as social science academics, a unique opportunity to achieve these related goals by understanding the network dimensionality of products and services using consumer utilities. As such, continuation of this interdisciplinary, original research is highly encouraged.

References

Carroll, J. D., & Green, P. E. (1997). Psychometric methods in marketing research: Part II, multidimensional scaling. *Journal of Marketing Research, 34*(2), 193–204. **https://doi.org/10.1177%2F002224379703400201**

Cattell, R. (1978). *The scientific use of factor analysis in behavioral and life sciences*. Springer. **https://doi.org/10.1007/978-1-4684-2262-7**

ESOMAR (n.d.). *ESOMAR is the only truly global voice of the data, research, analytics and insights community.* ESOMAR. **https://www.esomar.org/**

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). **https://doi.org/10.1371/journal.pone.0174035**

Golino, H. F., Shi, D., Garrido, L. E., Christensen, A. P., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2018). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *PsyArXiv*. **https://doi.org/10.31234/osf.io/gzcre**

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393–416. **https://doi.org/10.1177%2F0013164405282485**

Kim, H., Gon Kim, W., & An, J. A. (2003). The effect of consumer-based brand equity on firms' financial performance. *Journal of Consumer Marketing, 20*(4), 335–351. **https://doi.org/10.1108/07363760310483694**

Morton, F. B., & Altschul, D. (2019). Data reduction analyses of animal behaviour: Avoiding Kaiser's criterion and adopting more robust automated methods. *Animal Behaviour, 149*, 89–95. **https://doi.org/10.1016/j.anbehav.2019.01.003**

Narayana, C. L., & Markin, R. J. (1975). Consumer behavior and product performance: An alternative conceptualization. *Journal of Marketing, 39*(4). **https://www.jstor.org/stable/i253357**

Sawtooth Software (n.d.). *Hierarchical Bayes estimation.* Sawtooth Software. **https://sawtoothsoftware.com/resources/technical-papers/categories/hierarchical-bayes-estimation**

The R Foundation (n.d.). *What is R?* R Project. **https://www.r-project.org/about.html**

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing, 80*(6), 97–121. **https://doi.org/10.1509%2Fjm.15.0413**

Appendix A

Table A1

List of Iced Tea Claims

A	Slightly sweet
B	Just a little sweet
C	Sweetened from natural sources
D	30% less sugar than soft drinks
E	With Stevia plant sweetener
F	Low in sugar
G	Low in calories
H	Low in sugar and calories
I	Just a touch of real sugar
J	Half the calories of most other iced teas
K	Naturally low in sugar and calories
L	Made with real sugar
M	Made with Agave syrup
N	Made with honey
O	With a touch of sugar and Stevia plant sweetener
P	Made with real fruit juice
Q	Not too sweet
R	Only 2.5g of sugar/100ml
S	Only 3.0g of sugar/100ml
T	Only 50 cal/bottle
U	Only 25 cal/serving

Figure A1

Scree Plot for the Kaiser-Guttman Rule, Parallel Analysis, the Optimal Coordinates Approach, and the Acceleration Factor

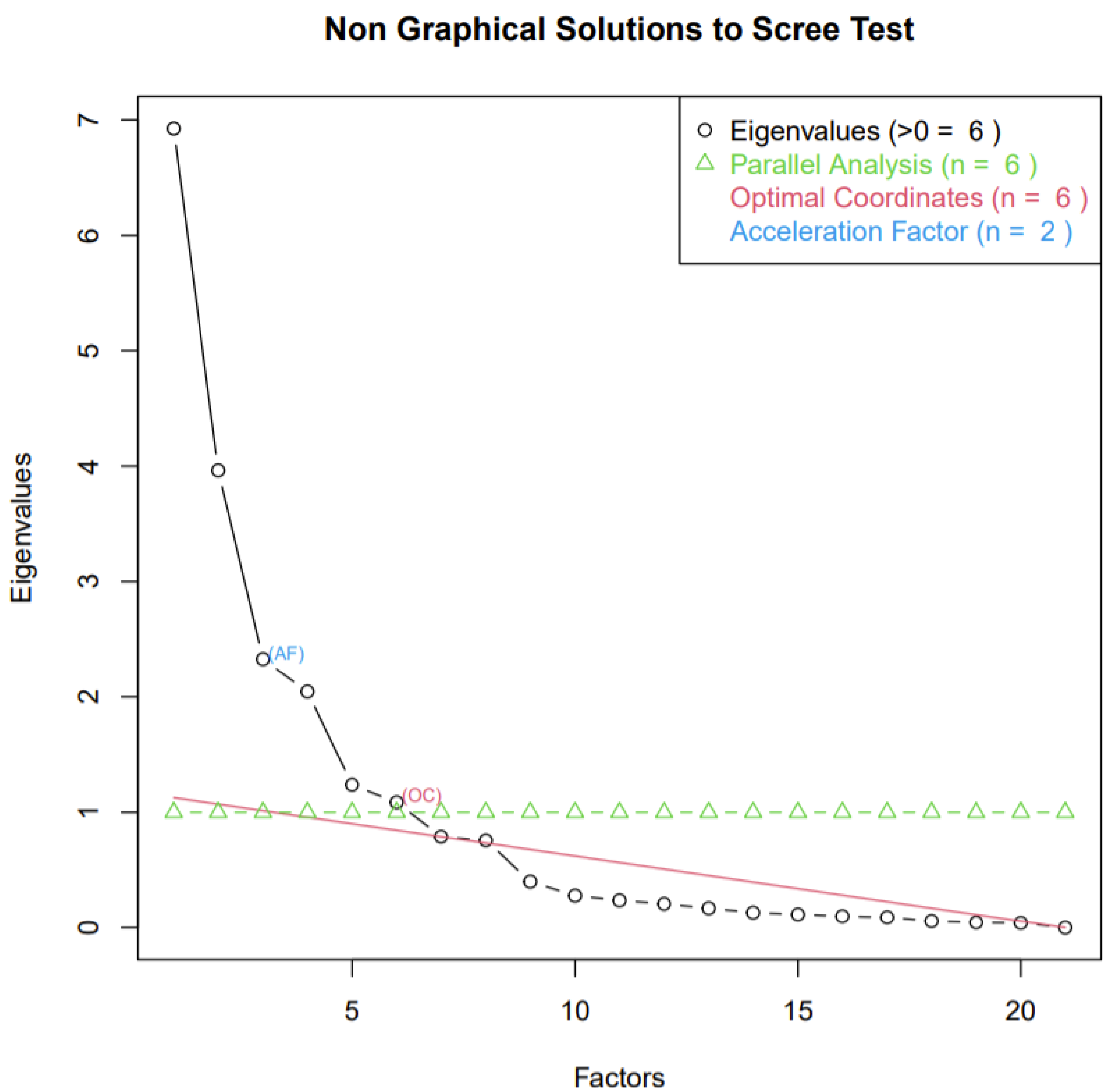


Figure A2

Line Plot for the Very Simple Structure Approach

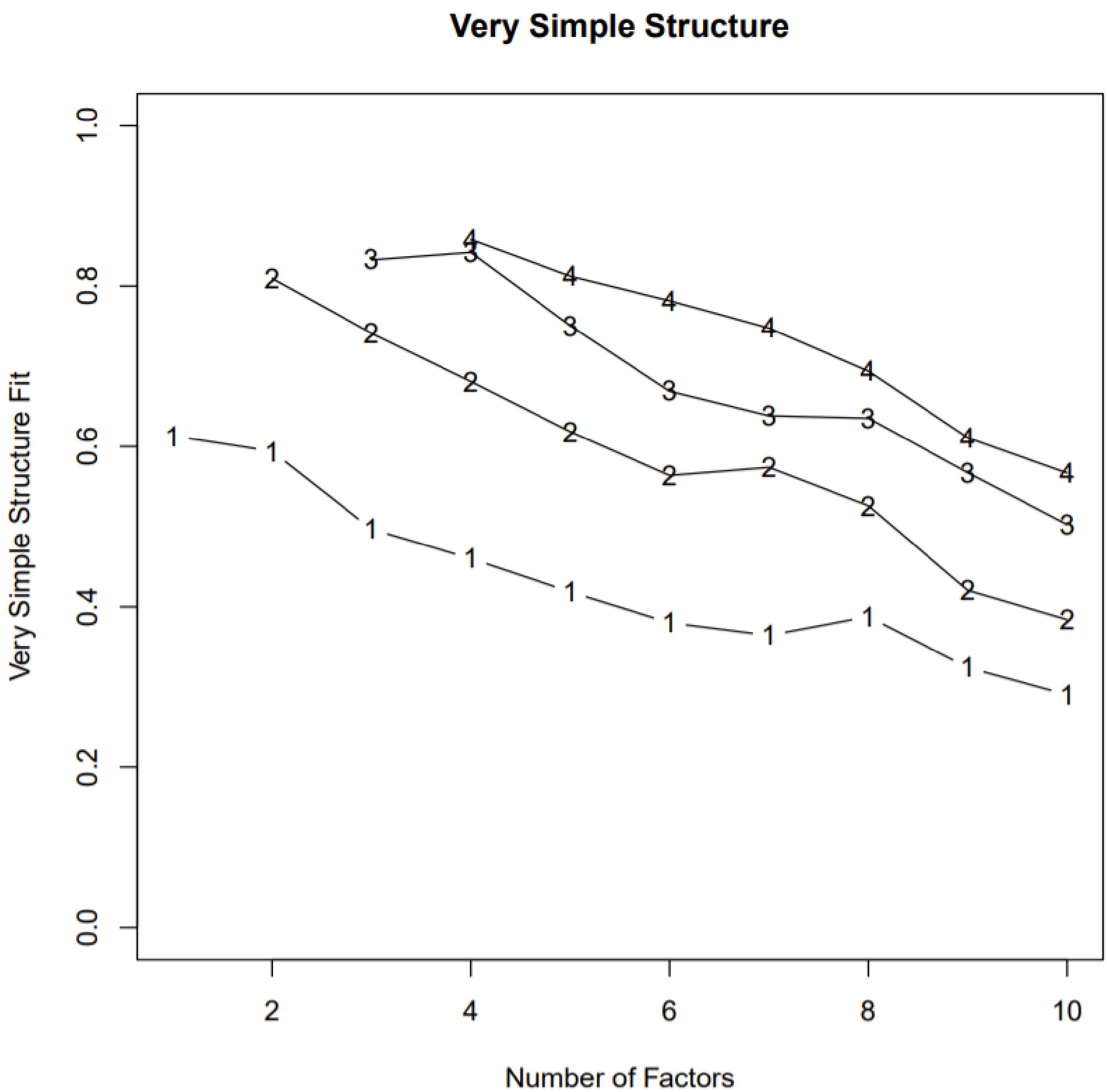


Figure A3

Median Network Structure for the Bootstrapped Exploratory Graph Analysis

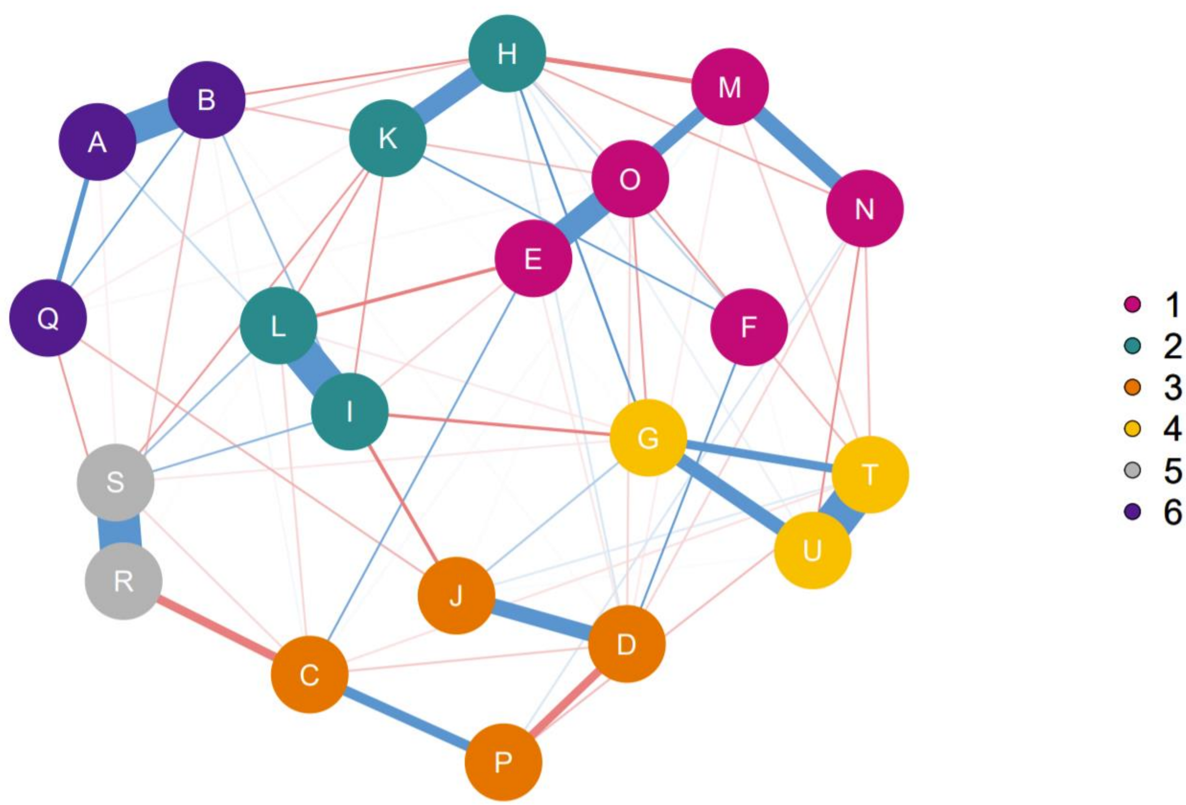
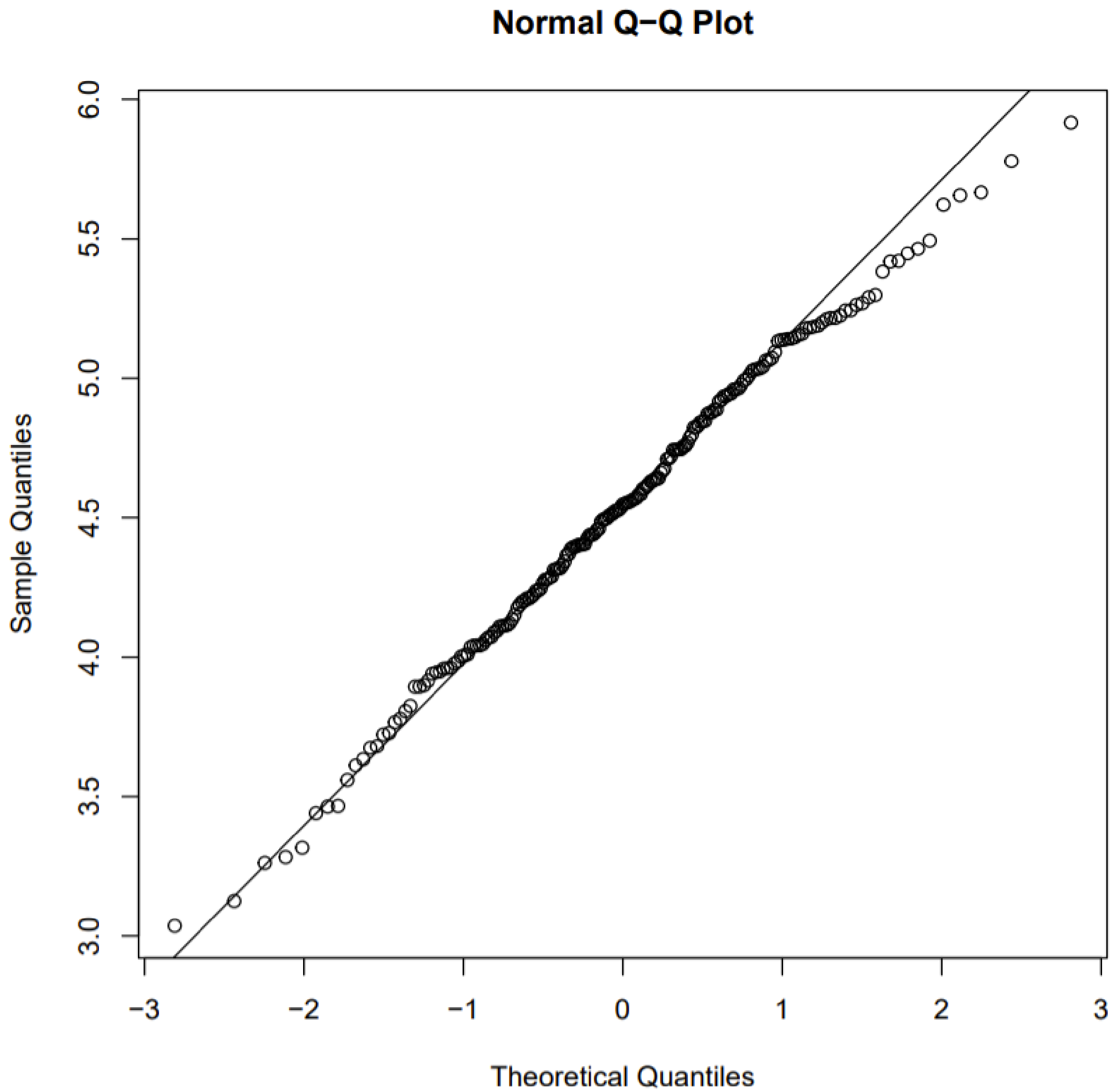


Figure A4

Q-Q Plot for Mardia's Test for Multivariate Normality



Appendix B

Reproducible R Code

```
library(EGAnet)
library(GPArotation)
library(lavaan)
library(nFactors)
library(psych)
utilities <- read.csv("utilities.csv", header = TRUE)

# Kaiser-Guttman Rule, Parallel Analysis, Optimal Coordinates, Acceleration Factor
eigenvalues <- eigen(cor(utilities))
pdf("Scree Plot.pdf")
plotnScree(nScree(x = eigenvalues$values, model = "factors"))
dev.off()

# Zoski and Jurs Regression
nMreg(utilities, model = "factors")

# Very Simple Structure, Vellicer's Minimum Average Partial
VSS <- vss(utilities, n = 10, rotate = "oblimin", diagonal = FALSE, fm = "pa",
        SMC = FALSE, plot = FALSE)
VSS
pdf("Very Simple Structure.pdf")
plot(VSS)
dev.off()

# Exploratory Graph Analysis
EGA <- bootEGA(utilities, 500, model = "glasso", type = "parametric",
        plot.typicalStructure = FALSE)
```

```
colors <- c("deeppink3", "darkcyan", "darkorange2", "goldenrod1", "gray70", "purple4",
        "darkblue", "gray30", "burlywood2", "cadetblue2")
pdf("Exploratory Graph Analysis.pdf")
plot(EGA, color = colors, borders = FALSE, vsize = 5, label.color = "white",
        posCol = "steelblue3", negCol = "lightcoral", legend = TRUE)
dev.off()

# Exploratory Factor Analysis
for (i in 2:8) {
  if (i %% 2 == 0) {
    FA <- fa(utilities, nfactors = i, residuals = TRUE, rotate = "oblimin", SMC = TRUE,
            fm = "ml")
    print(FA)
  }
}
KMO(cor(utilities))

# Confirmatory Factor Analysis
pdf("Multivariate Normality.pdf")
mardia(utilities)
dev.off()
model <- 'first=~E+F+M+N+O
        second=~H+I+K+L
        third=~C+D+J+P
        fourth=~G+T+U
        fifth=~R+S
        sixth=~A+B+Q'
CFA <- cfa(model, data = utilities)
summary(CFA, fit.measures = TRUE, standardized = TRUE, rsq = TRUE)
```