5.5 Summary

In this chapter we discussed three different types of BN which can be found from the data. The PSEM takes the SEM developed from theory by the researcher and converts it into a BN. This allows SEM practitioner to take a step forward by adding the capability to perform what-if analysis onto the network. On the other hand we have the EBN, which can be used without any prior knowledge regarding the data, as the process is purely data-driven. This can assist the practitioner in dynamically exploring the data. We can also generate a network that makes use of both the theory and data, called Semi-PSEM, where factors are defined according to the theory and the structural paths are constructed using a data-driven unsupervised approach. These BNs have been applied to the Facebook advertisement data which are displayed once again in figures 5.17, 5.18, 5.19 for convenience. They have all given results which correspond with a SEM performed with SPSS AMOS.

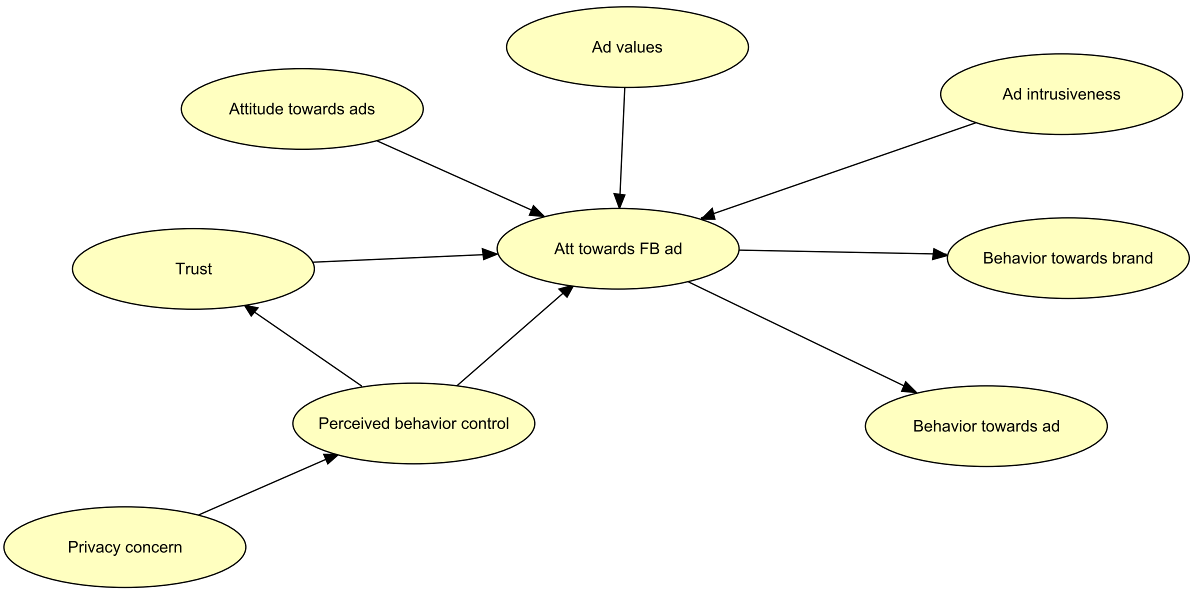


Figure 5.17: PSEM

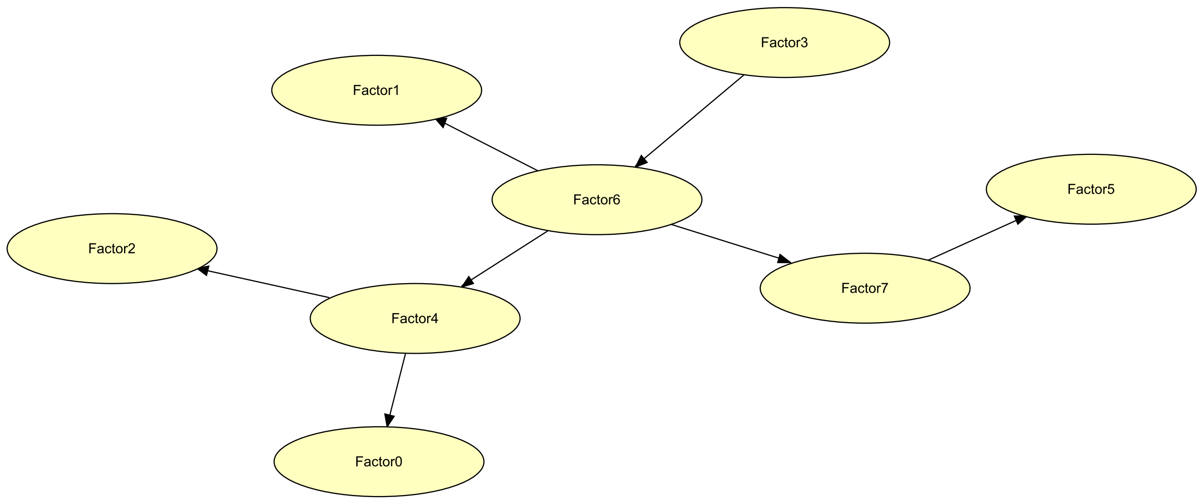


Figure 5.18: EBN

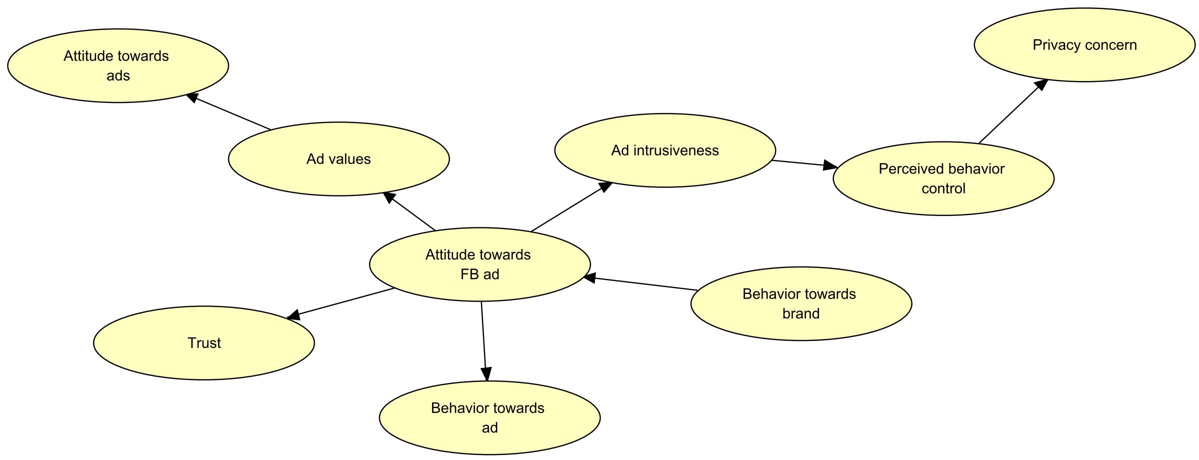


Figure 5.19: Semi-PSEM

Chapter 6

Conclusions

This chapter summarises the main results of the work done in this dissertation. Section 6.1 wraps up the key points of the dissertation and section 6.2 considers future work which can come from this dissertation.

6.1 Summary of Conclusions

Chapter 2 discussed the developments, both historic and current, as well as some interesting applications of SEM in the form of a literature review. The topic of SEM was followed in chapter 3, where the principles and processes behind SEM were covered. We then moved on to the field of graph theory in chapter 4. In chapter 4, we concluded that both SEM and BN are DAGs and it is possible for the structure of a SEM to be learned from data, using algorithms such as maximum weight spanning tree (MWST) and equivalence class (EQ) to find the initial network, cluster the variables according to mutual information, induce a factor for each cluster and learn the network among factors using Taboo algorithm to obtain an exploratory Bayesian Network (EBN).

This idea of EBN was applied along with other types of BNs in chapter 5. EBN is a BN derived from a data-driven perspective, parallel to the researcher's theory-based SEM. The researcher need not necessarily use the information from EBN but instead directly convert the SEM into a PSEM to conduct what-if analysis. It is also possible to specify the factors according to the theory and determine the structural path using the data. This results in a semi PSEM.

These BNs can offer significant insight into the data, as the researcher can then explore the data by instantiating on different nodes of the network (also called 'what-if' analysis). Because the direction of the inference is not an issue, various scenarios can be simulated using the BN.

The augmentation of SEM with BN provides significant contributions to the field:

Firstly, structural learning can mine data for additional causal information which is not necessarily clear when hypothesising causality from theory. This is particularly useful when two opposing theories exist (for example, whether the brand or positive media coverage is more effective in improving a company's image) and the learned structure can confirm one theory above the other.

Secondly, the inference ability of the BN provides not only insight as mentioned before, but acts as an interactive tool as the 'what-if' analysis is dynamic. This has been found to be a powerful knowledge transfer platform, specifically in participatory research [8].

6.2 Future Work

Although using a tree structure can quickly find a network for the given variables, it cannot assign more than one parent node for a child node. This implies that data-driven methods such as MWST will not be able to suggest multiple causes for a single variable, unlike theoretical models suggested by the researcher. Therefore a possible research topic can be to find efficient algorithms which can offer network structures which are more complex than tree structures, possibly cyclic [25]

Another way in which this work can be extended is to apply it in other research fields as mentioned in section 2.2, such as finance, investment and economics. The theories which have governed in these fields can, with the help of techniques covered here, be confirmed or be given a new perspective [3].

Bibliography

[1] Barbara M Byrne. *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Routledge, 2010.

[2] Yingxia Cao, Haya Ajjan, and Paul Hong. Using social media applications for educational outcomes in college teaching: A structural equation analysis. *British Journal of Educational Technology*, 44(4):581–593, 2013.

[3] Chingfu Chang, Alice C Lee, and Cheng F Lee. Determinants of capital structure choice: A structural equation modeling approach. *The quarterly review of economics and finance*, 49(2):197–213, 2009.

[4] Stefan Conrady and Lionel Jouffe. Tutorial on driver analysis and product optimization with bayesialab, 2013.

[5] Stefan Conrady and Lionel Jouffe. *Bayesian Networks and BayesiaLab: A Practical Introduction for Researchers.* Bayesia USA, 2015.

[6] Adnan Darwiche. *Modeling and reasoning with Bayesian networks.* Cambridge University Press, 2009.

[7] A De Waal and T Ritchey. Combining morphological analysis and bayesian networks for strategic decision support. *ORiON*, 23(2):105–121, 2007.

[8] Meike Düspohl, Sina Frank, and Petra Döll. A review of bayesian networks as a participatory modeling approach in support of sustainable environmental management. *Journal of Sustainable Development*, 5(12):1, 2012.

[9] G David Garson. Path analysis. *from Statnotes: Topics in Multivariate Analysis. Retrieved*, 9(05):2009, 2008.

[10] Fred Glover. Tabu search: A tutorial. *Interfaces*, 20(4):74–94, 1990.

[11] Thomas F Golob. Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, 37(1):1–25, 2003.

[12] Thomas F Golob, Seyoung Kim, and Weiping Ren. How households use different types of vehicles: A structural driver allocation and usage model. *Transportation Research Part A: Policy and Practice*, 30(2):103–118, 1996.

[13] Peter Grünwald. Introducing the minimum description length principle. *Advances in minimum description length: Theory and applications*, page 3, 2005.

[14] Michael Haenlein and Andreas M Kaplan. A beginner's guide to partial least squares analysis. *Understanding statistics*, 3(4):283–297, 2004.

[15] Robert M Hauser and Arthur S Goldberger. The treatment of unobservable variables in path analysis. *Sociological methodology*, 3:81–117, 1971.

[16] Rick H Hoyle. *Handbook of structural equation modeling.* Guilford Press, 2012.

[17] Karl G Jöreskog. Analysis of covariance structures. *Multivariate analysis*, 3:263–85, 1973.

[18] Lionel Jouffe and Stefan Conrady. Probabilistic latent factor induction with bayesialab. April 2014.

[19] Sungduk Kim, Sonali Das, Ming-Hui Chen, and Nicholas Warren. Bayesian structural equations modeling for ordinal response data with missing responses and missing covariates. *Communications in StatisticsTheory and Methods*, 38(16-17):2748–2768, 2009.

[20] Rex B Kline. *Principles and practice of structural equation modeling.* Guilford publications, 2015.

[21] Jonathan Kohn and Sarah K Bryant. Factors leading to the us housing bubble: A structural equation modeling approach. *Research in Business and Economics Journal*, 3:D1, 2011.

[22] Kevin B Korb and Ann E Nicholson. *Bayesian artificial intelligence.* CRC press, 2010.

[23] Reinhold Kosfeld and Jørgen Lauridsen. Factor analysis regression. *Statistical Papers*, 49(4):653–667, 2008.

[24] Timo Koski and John Noble. *Bayesian networks: an introduction*, volume 924. John Wiley & Sons, 2011.

[25] Pedro Larrañaga, Mikel Poza, Yosu Yurramendi, Roberto H. Murga, and Cindy M. H. Kuijpers. Structure learning of bayesian networks by genetic algorithms: A performance analysis of control parameters. *IEEE transactions on pattern analysis and machine intelligence*, 18(9):912–926, 1996.

[26] Ross L Matsueda and Guilford Press. Key advances in the history of structural equation modeling. *Handbook of structural equation modeling*, pages 17–42, 2012.

[27] Paul Munteanu and Mohamed Bendou. The eq framework for learning equivalence classes of bayesian networks. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 417–424. IEEE, 2001.

[28] Kevin P Murphy. *Machine learning: a probabilistic perspective.* MIT press, 2012.

[29] Richard E Neapolitan et al. *Learning bayesian networks*, volume 38. Pearson Prentice Hall Upper Saddle River, NJ, 2004.

[30] Norm O'Rourke and Larry Hatcher. *A step-by-step approach to using SAS for factor analysis and structural equation modeling.* Sas Institute, 2013.

[31] Randall E Schumacker and Richard G Lomax. *A beginner's guide to structural equation modeling.* Psychology Press, 2004.

[32] Charles Spearman. " general intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292, 1904.

[33] Sheridan Titman and Roberto Wessels. The determinants of capital structure choice. *The Journal of finance*, 43(1):1–19, 1988.

[34] Xiao-fu Xu, Jian Sun, Hong-tao Nie, De-kui Yuan, and Jian-hua Tao. Linking structural equation modeling with bayesian network and its application to coastal phytoplankton dynamics in the bohai bay. *China Ocean Engineering*, 30(5):733–748, 2016.

Appendix A

Maximum weight spanning tree

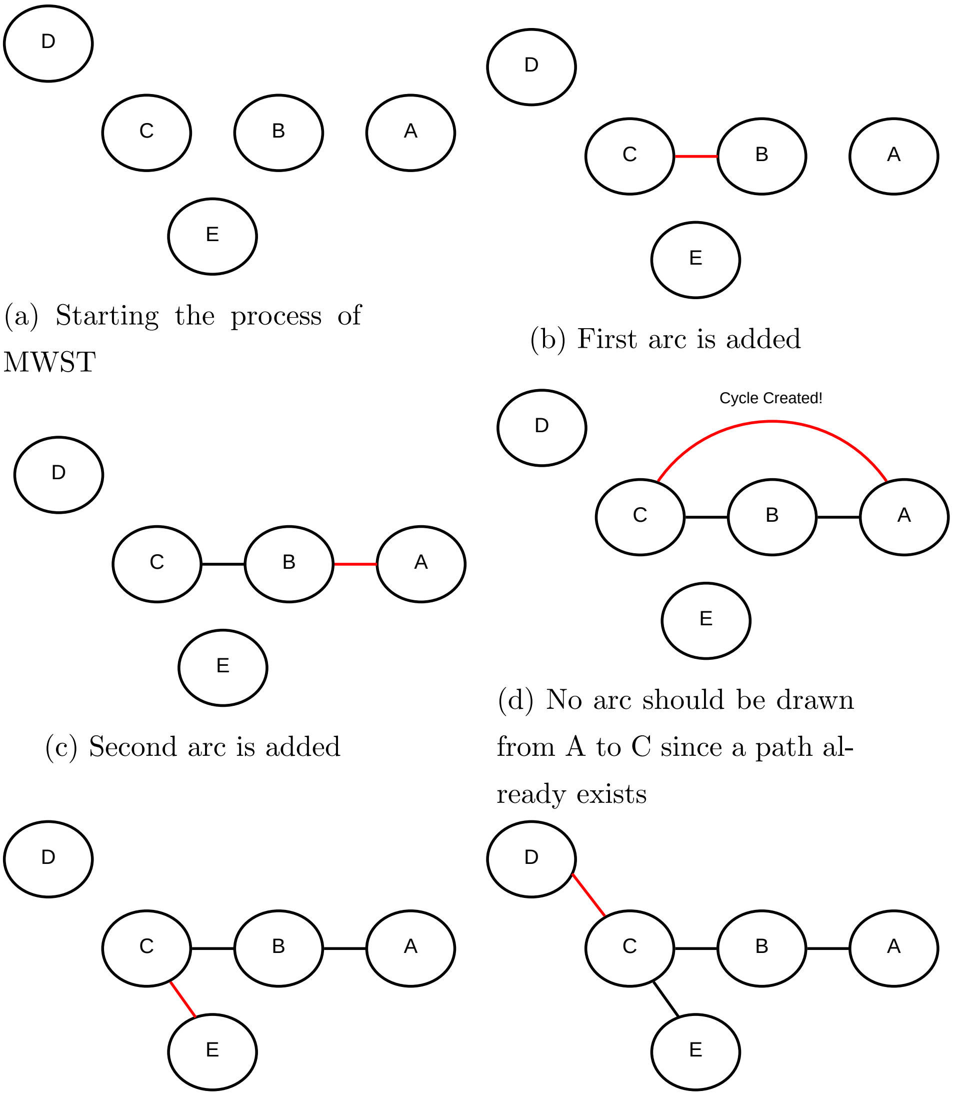
A.1 MWST example

Let us suppose there are 5 variables with which we want to create a MWST. The first step we need to take is to calculate mutual information as given by equation 4.8 for all possible pairs of variables.

Table A.1: Pairwise Mutual information, sorted descending

Var1,Var2	MI
B,C	0.83
A,B	0.71
A,C	0.63
C,E	0.58
B,E	0.22
A,E	0.19
C,D	0.15
A,D	0.13
B,D	0.11
D,E	0.08

Table A.1 shows a fictitious set of values for the mutual information of 5 variables, sorted descending according to their MI.



(a) Starting the process of MWST

(b) First arc is added

(c) Second arc is added

(d) No arc should be drawn from A to C since a path already exists

(e) Third arc is added

(f) Fourth arc is added

Figure A.1: Directed and undirected trees.

Initially we start off with a fully unconnected network as shown in figure A.1a. Next we start connecting the pair of variables as we move down the rows of table A.1. The highest value of MI is between variables B and C and so we draw an arc between those two nodes, illustrated in figure A.1b. The next highest MI is present between A and B so an arc is drawn to connect those two (figure A.1c).

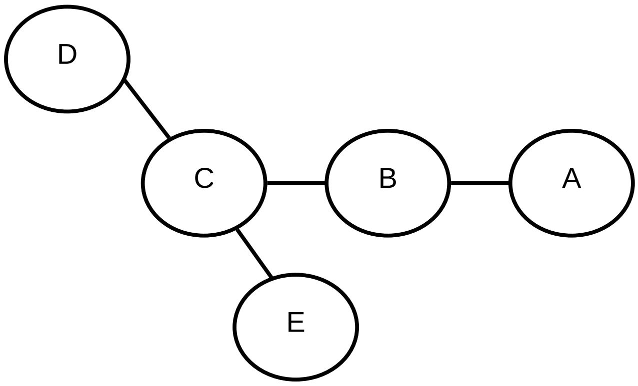


Figure A.2: Final Tree

The next highest MI is between variables A and C but we should not connect those two, as there is already a path between those two variables (through$B$). Put differently, drawing an arc between A and C creates a cycle with variables A, B and C as shown in figure A.1d. Hence we do not change the network and proceed to the next highest MI. The next arc added is between C and E. This means the next two paths (B, E & A, E) will not be drawn (figure A.1e). Finally, an arc is drawn between C, D and this causes all subsequent MIs to be forbidden (figure A.1f). The final tree is given by figure A.2.

Appendix B

Facebook advertisement data questionnaire

Table B.1: Questionnaire for Facebook advertisement

Q8Q1	Facebook is a trustworthy social network
Q8Q2	Facebook can be relied on to keep its promises
Q8Q3	Even if not mentioned, I would trust Facebook to do the job right
Q8Q4	I believe that Facebook would use my data only for purposes that I have approved
Q8Q5	I can count on Facebook to protect my privacy
Q8Q6	I can count on Facebook to protect my personal information from unauthorized use
Q9Q1	I consider advertising a good thing
Q9Q2	In general, I like advertising
Q9Q3	I consider advertising essential
Q9Q4	Having advertisements are important to me
Q9Q5	Advertisements in general are interesting to me
Q9Q6	I would describe my overall attitude towards advertising as favourable
Q11Q1	I consider ads on my Facebook page a good thing
Q11Q2	I like ads on my Facebook page
Q11Q3	I consider ads on my Facebook page essential
Q11Q4	Having ads on my Facebook page are important to me
Q11Q5	Ads on my Facebook page are interesting to me
Q11Q6	I would describe my overall attitude towards ads on my Facebook page as favourable

Table B.2: Questionnaire continued

	When I see an advertisement on my Facebook page, I generally
Q12Q2	click on the ad to find more information
Q12Q3	'like' or 'comment' on the ad
Q12Q4	'share' or 'repost' the ad to my friends
Q12Q11	become a fan of the company/brand
Q12Q12	visit the company/brands website
Q12Q13	purchase the advertised product/service
	It is important to me that I can
Q14Q1	only receive ads on my Facebook page if I have previously provided permission
Q14Q2	control the permission to receive ads
Q14Q3	refuse to receive advertising on my Facebook page
Q14Q4	filter advertising on my Facebook page to match my needs
	I find advertisements on my Facebook page
Q16Q1	distracting
Q16Q2	intruding on my privacy
Q16Q3	interfering
Q16Q4	invading my privacy
Q16Q5	deceptive
Q16Q6	confusing
Q16Q7	annoying
Q16Q8	irritating
Q16Q9	compromising my privacy
	Facebook advertising is
Q17Q1	useful
Q17Q2	valuable
Q17Q3	important
Q20Q1	All things considered, the Internet causes serious privacy problems
Q20Q2	Compared to others, I am more sensitive about the way online companies handle my personal information
Q20Q3	To me, it is very important to keep my privacy intact/unharmed from online companies
Q20Q4	I believe other people are not concerned enough with online privacy issues
Q20Q5	Compared to other subjects on my mind, personal privacy is very important
Q20Q6	I am concerned about the threat to my personal privacy today

Appendix C

Facebook data SEM using AMOS

C.1 Summary

Figure C.1 shows how the SEM for the Facebook data was constructed in SPSS AMOS. All path coefficients as well as error variance were significant. Figure C.2 shows the diagram with estimated standardised path coefficients and squared multiple correlation for endogenous variables.

The value of -.13 between $Advertising\,intrusiveness$ and $Attitudes\,towards\,FB\,ad$ indicates that there is an inverse relationship, albeit relatively weak, between the two variables, where an increase of 1 standard deviation in $Advertising\,intrusiveness$ will lead to a decrease of 0.13 standard deviations in $Attitudes\,towards\,FB\,ad$. A strong positive relationship exists between $Attitudes\,towards\,FB\,ad$ and $Behaviour\,towards\,ad$, indicated the coefficient value of .79. The value of .63 for squared multiple correlation of $Behaviour\,towards\,ad$ shows that $Attitudes\,towards\,FB\,ad$ explains 63% of the variance in $Behaviour\,towards\,ad$. Other values in the diagram can be interpreted in the same way.

Furthermore, table C.1 shows values for the model goodness of fit. CFI is larger than 0.9 while $RMSEA$ is less than 0.055, which are both indicative of an adequate overall model fit.

Table C.1: Goodness-of-fit Facebook data SEM

Model fit index	Index value
CFI	0.926
TLI	0.922
RMSEA	0.051
RMSEA upr90	0.054
RMSEA lwr90	0.049

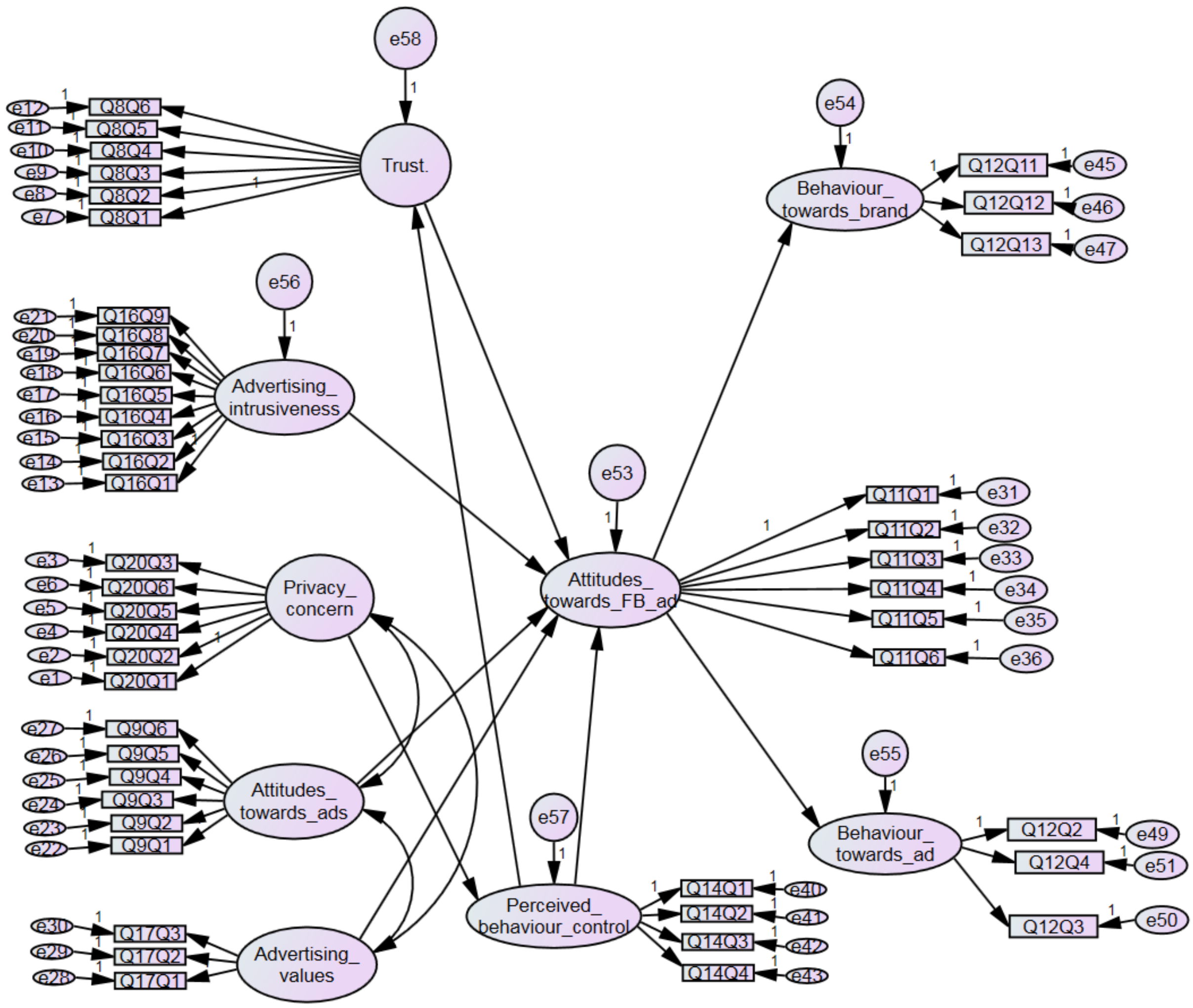


Figure C.1: Theoretical SEM for Facebook data, as drawn in SPSS AMOS

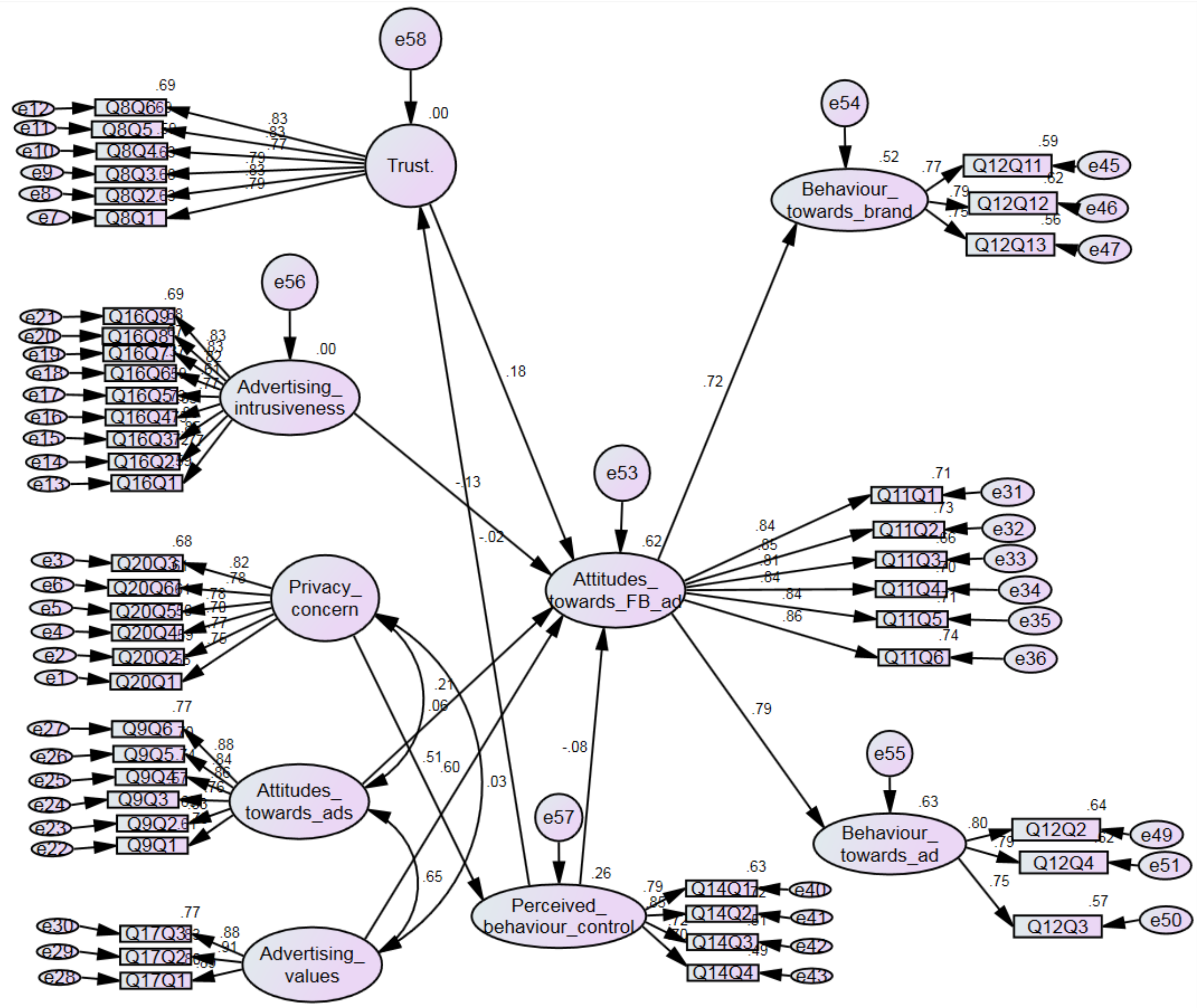


Figure C.2: Estimated coefficients for Facebook data SEM