Social capital factors: Plausible theory about causal relationships

Discussion Paper Draft 2a

Paul Bullen
www.mapl.com.au
2007

Social capital factors

The study Measuring Social Capital in Five Communities in NSW identified 8 factors:

A. Community connections
B. Proactivity/ Social agency
C. Trust and safety
D. Neighbourhood connections
E. Family and friends
F. Tolerance of diversity
G. Value of life
H. Work connections

Each of these factors contributed to general social capital. Work connections only applied to people in the workforce. This paper focuses on factors A to G.

The seven factors A to G were measured using 31 survey questions.

Are there causal connections between these seven factors and if so what are these causal connections?

Plausible causal connections

Different theoretical positions suggest different approaches. For example, two possible causal paths could be:

- Family and friends --> Neighbourhood connections --> Trust and safety --> Community connections
- Trust and safety & Proactivity --> Community connections --> Neighbourhood connections --> Tolerance of diversity.

Given there are 7 factors and each factor could be connected to multiple other factors there are a very large number of possible causal models. There are also multiple theoretical perspectives also leading to multiple possible causal models.

What causal models are most plausible?

Structural equation modelling

Structural equation modelling can be used to test the plausibility of theories of causal relationships among factors.

Structural equation modelling in its most general form is a combination of path analysis and factor analysis. "In path analysis, the concern is with the predictive ordering of *measured variables*. For example X --> Y --> Z is a path analysis model in which X, Y and Z are measured variables and the arrows represent the hypothesised causal effects. In a full structural equation model, the concern is the predictive ordering of *factors.* A structural equation model in which the Fs are factors is F1 --> F2 --> F3, and just as in path analysis, the arrows represent hypothesized causal effects." (Laura Klem, Structural Equation Modeling in Reading and Understanding More Multievariate Statistics edited by L.G. Grimm and P.R. Yarnold, American Psychological Association, 2000).

This paper explores causal connections between social capital factors using structural equation modelling.

Gathering the data

Since undertaking the Five Communities study the "best 31 questions" for measuring social capital in communities that were identified in that study have been used in many other studies. Data from nine studies representing 9 communities and 4 groups (6,249 individuals) was gathered to explore the causal connections between social capital factors.

Confirming the factor structure

As noted above structural equation modelling in its most general form is a combination of path analysis and factor analysis. The path model identifies the causal connections between the factors. The factor model identifies the relationships between the measures (eg questionnaire responses) and the factors.

The study Measuring Social Capital in Five Communities in NSW identified 7 factors measured by 31 questions. A factor analysis of the data from the nine studies gathered here was undertaken. It replicated the factor structure in the original study. For 28 out of the 31 questions the largest factor loading was on the same factor as the original study. For three questions the largest loading was on a different factor (significant loadings were also on the original factor).

A hierarchical factor analysis was also undertaken. All 31 questions had loadings of between .28 to .48 on the second order factor (general social capital) - confirming the findings in the original study of the existence of the second order factor.

A confirmatory factor analysis was undertaken using structural equation modelling. See the statistical notes for details. It also confirmed the factor structure.

See the list of questions in *Attachment 2: Questions* which lists the questions most associated with each factor.

The exploratory process

The path model identifies the causal connections between the factors. What are plausible path models?

We started the analysis by exploring models suggested by different theoretical perspectives. For example elements of paths that were explored included:

- Family and friends --> Neighbourhood connections --> Community connections
- Trust and safety & Proactivity --> Community connections --> Neighbourhood connections --> Family and friends
- Family and friends --> Neighbourhood connections --> Trust and safety & Tolerance of diversity --> Community connections
- Trust and safety & Proactivity --> Community connections --> Neighbourhood connections --> Tolerance of diversity.

In exploring over 20 different models (each including all 7 factors) being suggested by theory and/or their opposites (ie, models going in the opposite causal direction being suggested by theory) we found that many of the proposed models worked 'moderately well enough' to 'not so well'. None of them worked 'well' or 'very well'.

Given there are 7 factors, there are 42 different paths (A-->B, A-->C, etc). These 42 different paths can be structured in many different ways (eg, A-->B or A-->B & C, or A-->B&C&D, etc). So it is not surprising that we didn't get a model that worked 'well' or 'very well' at the beginning of the explorations.

We did notice in these models that the path coefficients varied greatly from one path to another (eg the path coefficient A-->B was relatively high, and the path coefficient A --> F was relatively low).

Usually, structural equation modelling is theory driven, that is, it is used to test the plausibility of a theory. For example if there are two competing theories, two path models can be generated and structural equation modelling can identify which model best fits the data.

Because there were so many possible models that could be theoretically justified we started asking the question in a different way: What would be the distribution of the path coefficients in a large number of randomly generated structural models?

If some paths were more likely to be cause and effect linkages than others, on average those paths would have higher path coefficients than paths that did not have cause and effect linkages.

To explore this idea we generated 138 path models, chose those that were a moderate or better fit (136 models) and for those 136 models found the mean and standard deviation of the path coefficients for each of the 42 possible paths (A>B, A>C, etc).

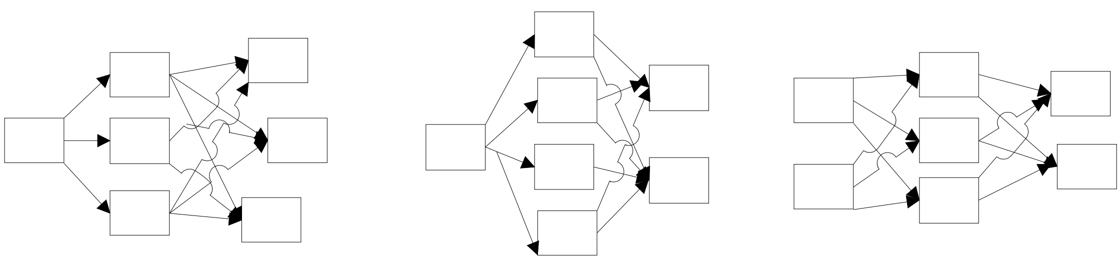
The means and standard deviations of the path coefficients were then used to develop two well fitting path models and a further 2 slightly more complex path models.

Generating the 138 SEM models

To generate models for the exploratory process we set some parameters:

> Each model would be a three step model(ie have a beginning, a middle and an end).
> Each model would include all seven factors.
> Each factor would connect to all the factors in the next step of the model.

There are fifteen different structures that meet the requirement of being a three step model and including 7 factors. Here are three examples of these 15 structures:



Structure A   Structure B   Structure C

For each of these fifteen structures the seven factors were randomly allocated within the structure. For example in structure A the first model included the factors C; E,G,B; F,B,A in the three steps. This process was repeated 7 times for each structure. In structure A the second model included the factors A; D,B,F; C,G,B.

This generated 105 structural equation models (7 for each of 15 structures).

These 105 models were reviewed and the structure what was producing the best fitting models (Structure C above) was used to randomly generate a further 33 models giving a total of 138 models.

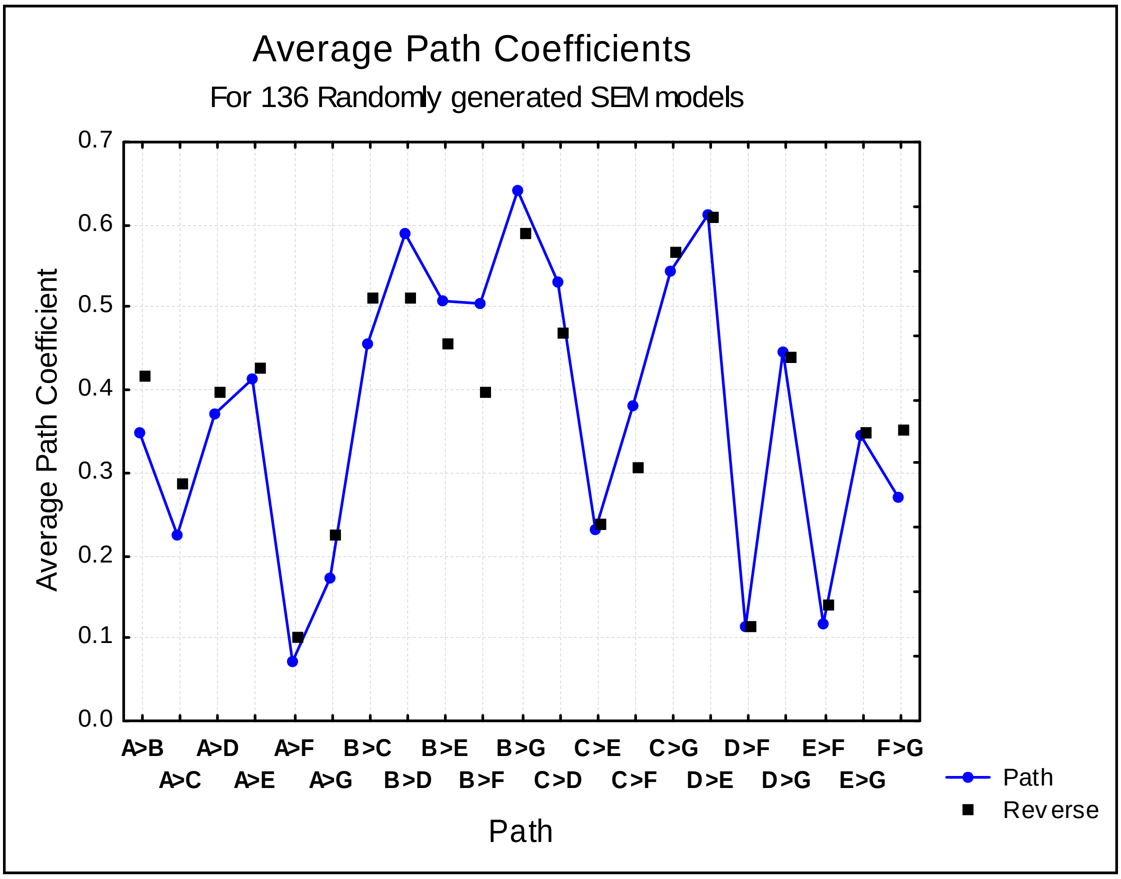
All the models where the RMSEA was .1 or less were taken to be a good enough fit for further analysis - 136 models.

Path coefficients

There are 42 possible paths (A>B, A>C etc). The paths fall into two groups of 21. The second 21 paths are the reverse of the first 21 paths (A>B and B>A).

The 1370 path coefficients in these 136 models were grouped into their respective paths (A-->B, A-->C etc) and the mean and standard deviation calculated for each path.

The following chart shows the means for the path coefficients. There are 21 possible paths: A-->B, A-->C, A-->D etc. There are a further 21 reversed paths B-->A, C-->A, D-->A, etc



This data suggested:

- Some paths are more likely to work in one direction than the other. For example, on average in the 136 models the path B-->D has higher path loadings (0.6) than the path D-->B (0.5). Most paths beginning with B or C have higher average path coefficients than their reverse (ending with B or C).
- Some paths have very low path loadings in both directions, eg, A-->F (F-->A), D-->F, (F-->D)
- Some paths have moderate to high loadings in both directions, eg, A-->E, E-->A.

In the light of these distributions of path coefficients a structural equation model was developed that:

- Began the model with B and C to reflect the facts that most paths beginning with B or C have higher average path coefficients than their reverse (ending with B or C) and that the biggest differences in path directions were in paths including B or C
- Included the paths that were moderate to high and had higher loadings than their reverse, eg. B-->D, B-->F
- Included the paths that were moderate to high loadings
- Excluded all paths with negligible or very low loadings, eg, A-->F, F-->A, C-->E, E-->C, E-->C, D-->F, F-->D

In developing the model there were some constraints from the nature of structural equation modelling:

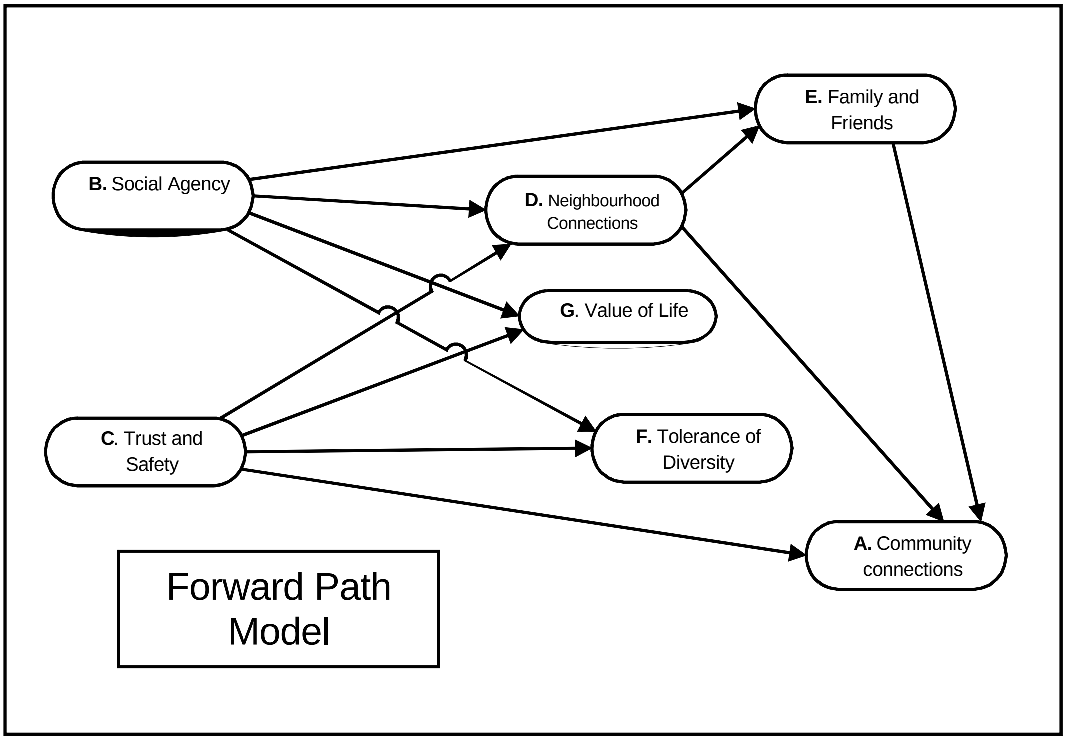
Paths could not be circular. There must be a beginning and an end.

The same factors could not be in two different places in a causal sequence.

In the real world these two assumptions are probably false and highlight the oversimplification taking place in this analysis.

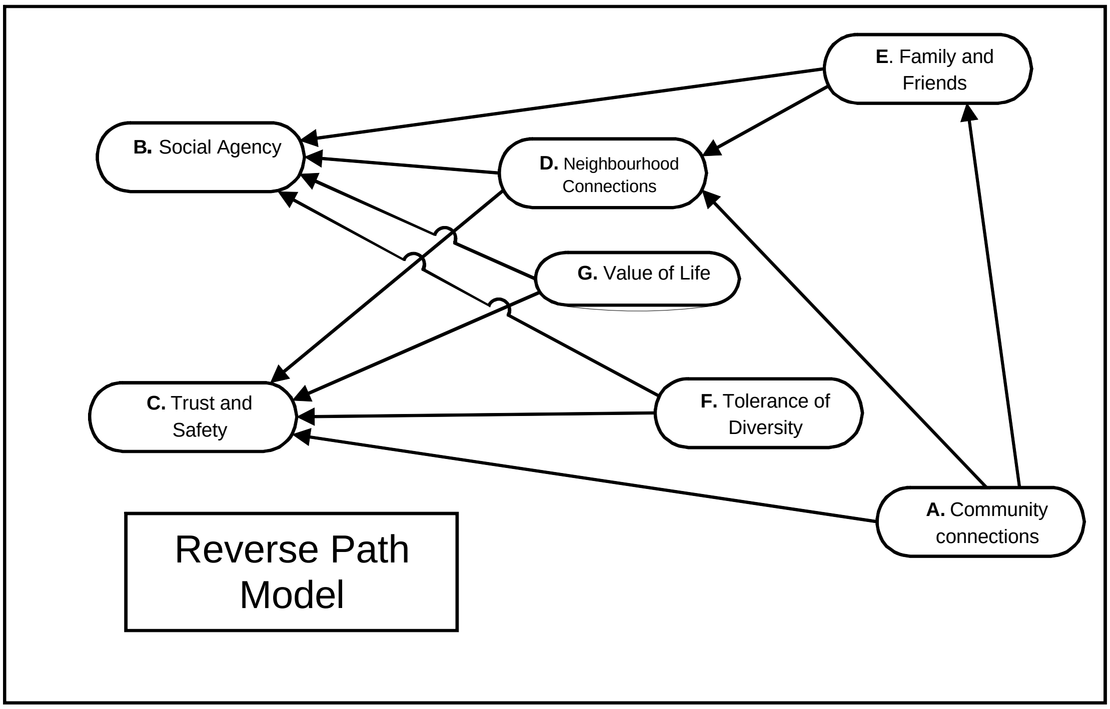
The Path models

The process outlined above generated the following path model. In the first versions of the model D>G and G >F were included. However, in practice they did not work (either having low path coefficients or causing statistical anomalies), so these two paths were left out in the final version.



This model was a better fit than any of the randomly generated models and any of the initial theory generated models.

While the Forward model above was a good fit, an obvious next step was to test the reverse model:



This model is a good fit but not as good as the Forward model. All but one of the path loadings in this model are lower than the forward model. Nonetheless it is a plausible model.

The structural equation models

These path models were used in combination with the factor analysis of the 31 questions to develop complete structural equation models. The Forward and Reverse models above are shown below in detail

The Model A (forward and reverse) shows:

- The causal paths for the factors
- The path coefficients for these causal paths
- For each factor the loadings on the questions
- For each question the unexplained variance.
- For each dependent factor the unexplained variance

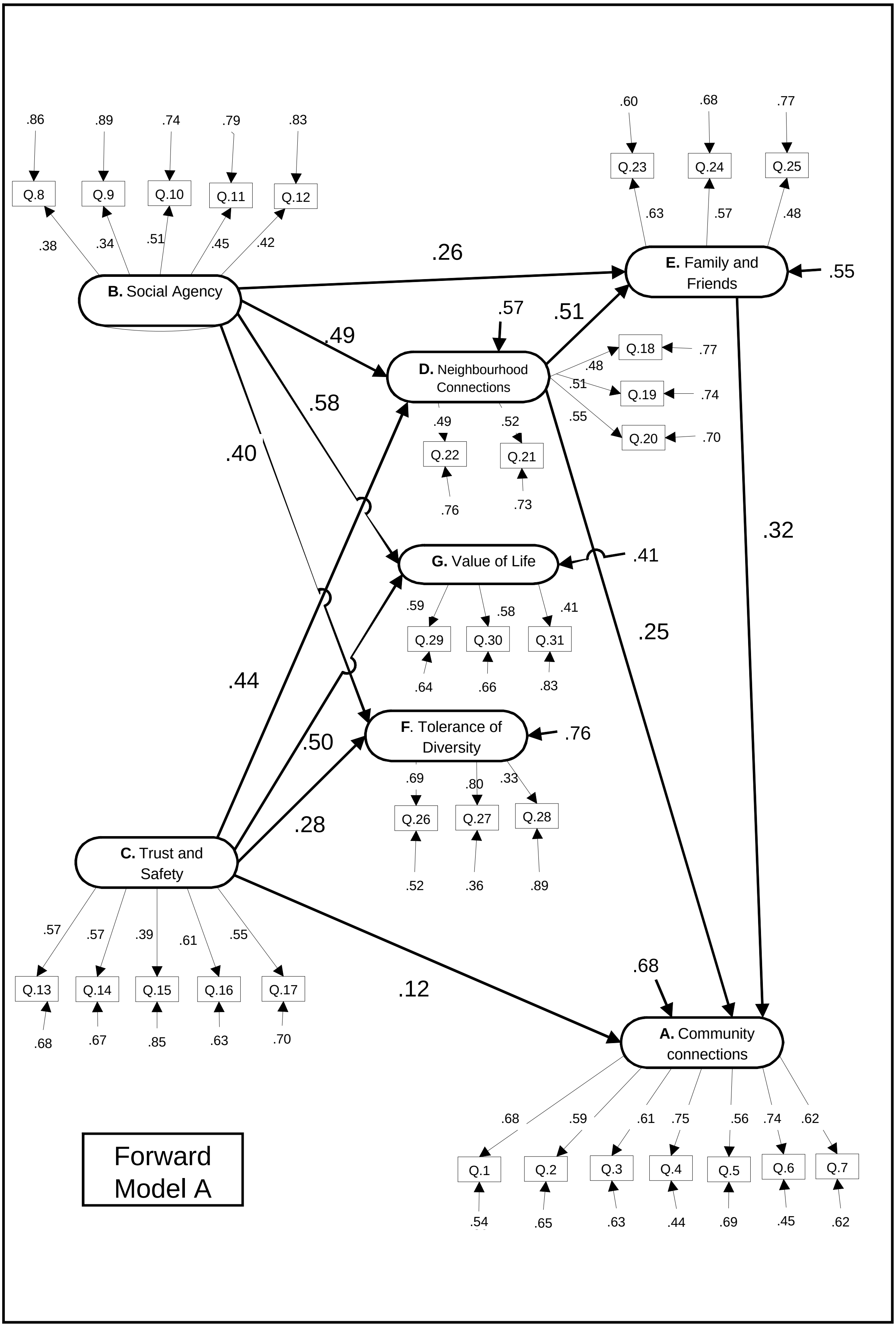
The model is standardised and all variables have a variance of 1. The questions are in *Attachment 2: Questions*.

In reviewing the randomly generated models the better fitting models tended to be the more complex models. The fit of the two models (A) here can be improved by adding three paths to the forward model (C-->B, B-->A and E-->G) and four paths to the Reverse model (F-->G, G-->D, G-->E and A-->B).

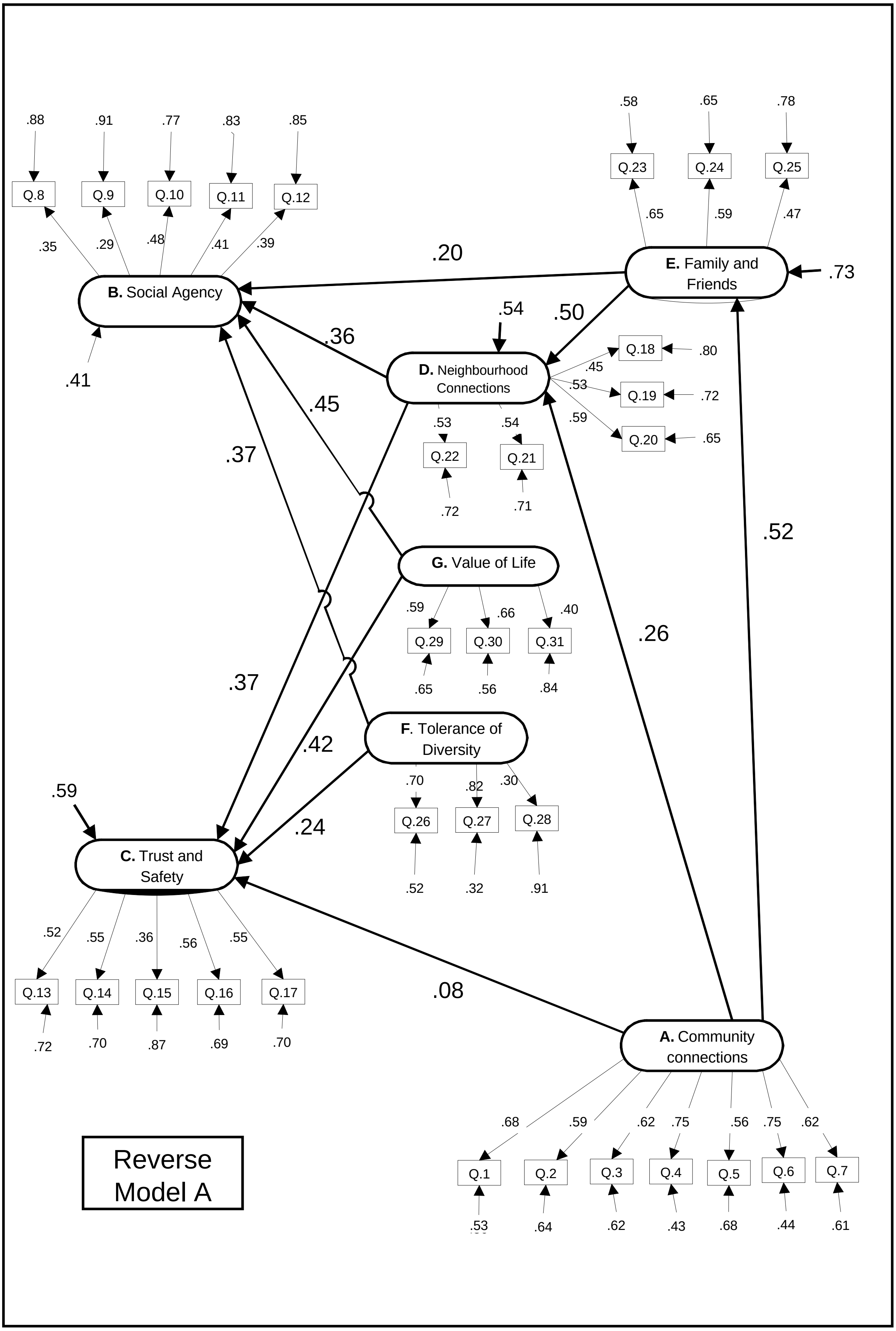
These two models are shown below as Model B Forward and Model B Reverse. In these models the same paths are shown for the Forward and Reverse models, however, note that in the forward model D–>G and G–>F are effectively zero and in the Reverse model B–>C is zero.

The increased complexity requires larger sample sizes and so in the analysis below comparing communities, groups and gender the less complex model was used.

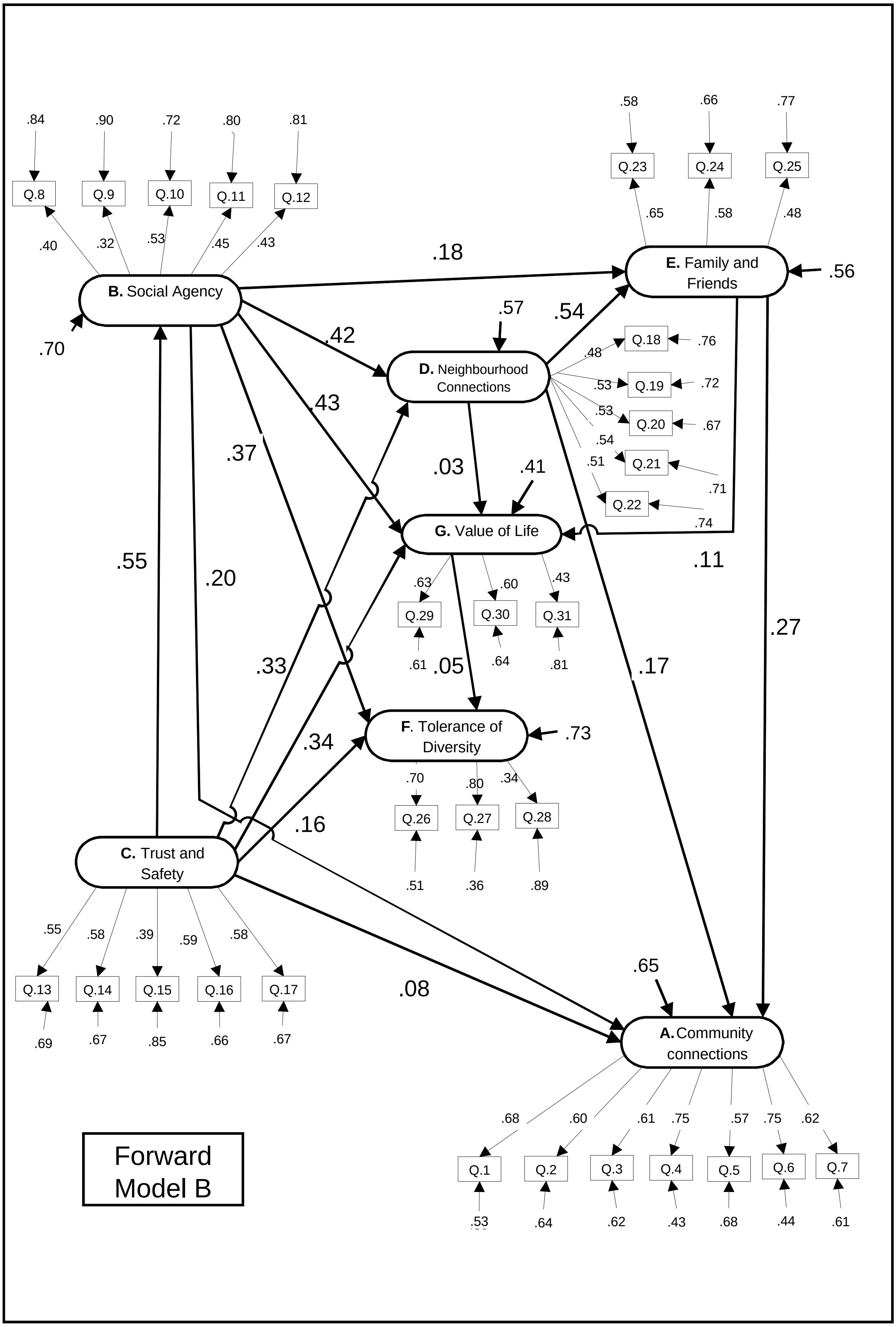
The Forward model (A)



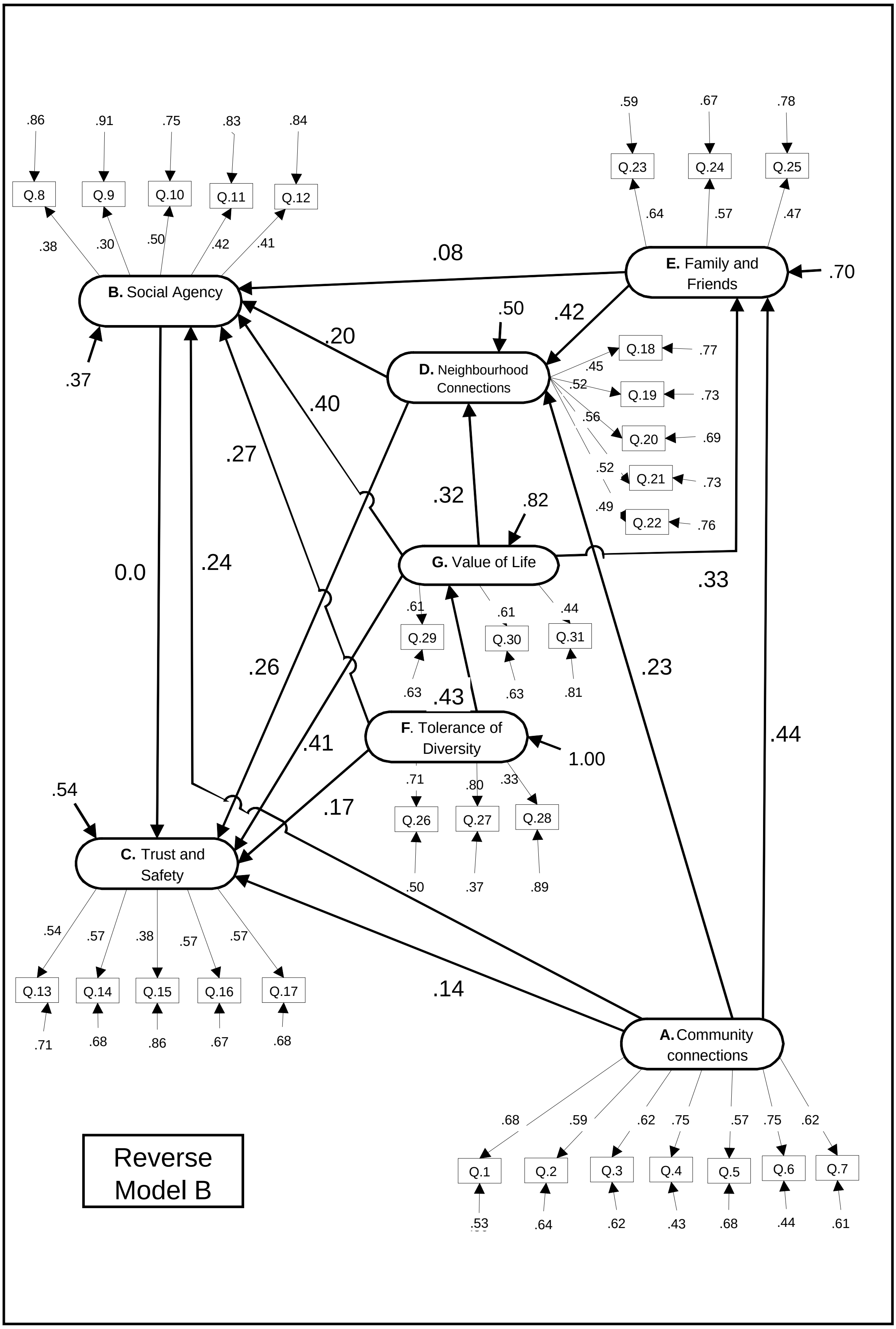
The Reverse model (A)



Forward model (B) - more complex



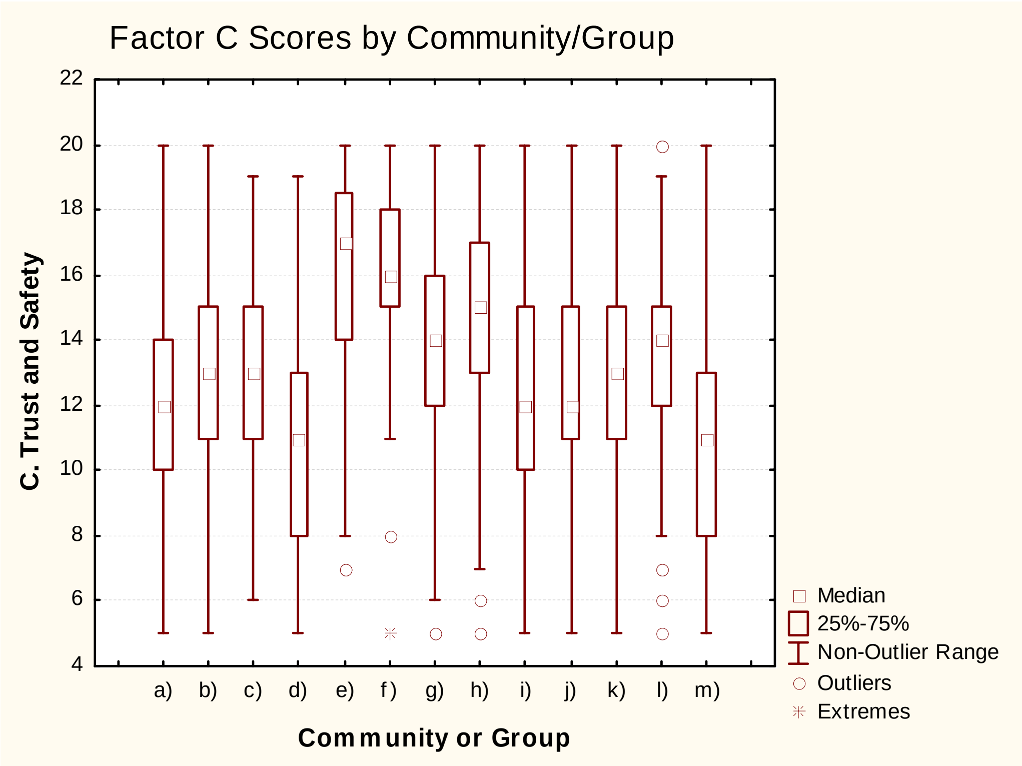
Reverse model (B) - more complex



Comparing communities and groups

The Forward and Reverse structural equation models (A) above can be used to explore the question: Are the causal connections between social capital factors the same across communities, or different from one community to another.

Communities vary greatly in the levels and mix of social capital. Many studies have shown this. For example the following chart shows the factor scores for Factor C. Trust and Safety for the 9 communities and 4 groups in this analysis.



West Wyalong (e) has the highest levels of trust and safety. Greenacre (d) and Family Support clients (m) the lowest.

These 8 communities and 4 groups are very different on each of the social capital factors. These differences tell us about the amount of social capital not the causal connections between factors.

Two approaches to exploring the causal connections between factors are:

a) using the Forward and Reverse structural models (A) that have been identified above see if there are significant differences between the path coefficients in each community and/or group.

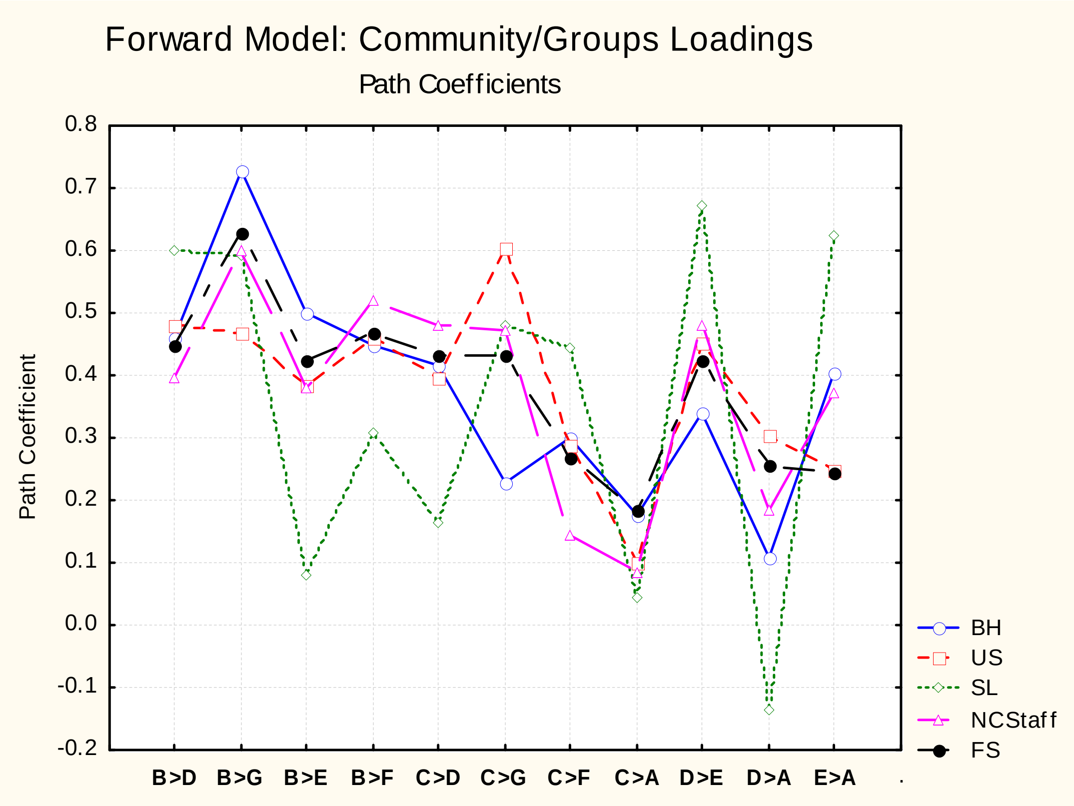
b) identify where there are different models to the Forward and Reverse models (A) above that better suit particular communities.

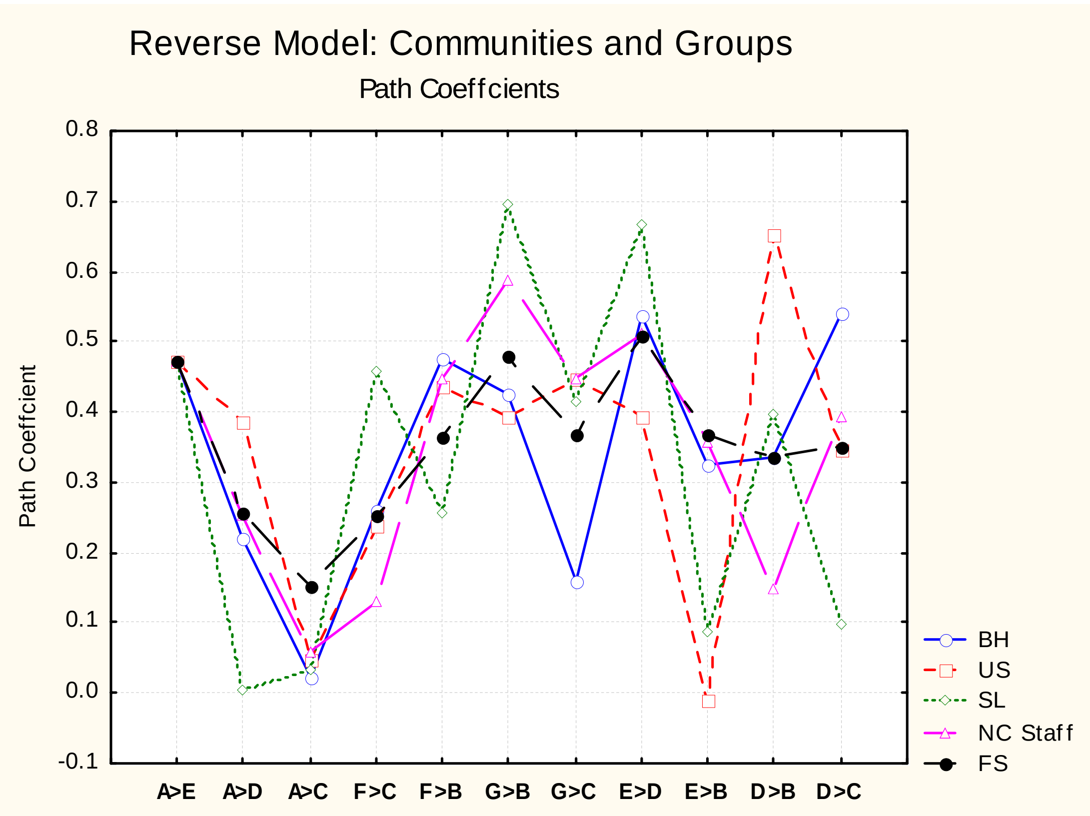
Only the first approach is taken here.

The path coefficients are explored for three communities and two groups separately. The three communities that are included are the three with the biggest sample sizes. The two groups that are included also have large sample sizes but had been included because of who they are - Neighbourhood Centre staff and Family support clients. They are very different from each other on the amount of social capital and mix of factors.

The analysis was undertaken in such a way that there three communities and two groups were combined into one analysis which was used to generate a measurement model within which each of the three communities and two groups could have path coefficients unique to the community or group respectively.

The following two graphs show the path coefficients for each community and group for the Forward and Reverse path models.





There are significant differences between communities and groups in the Forward model. For example:

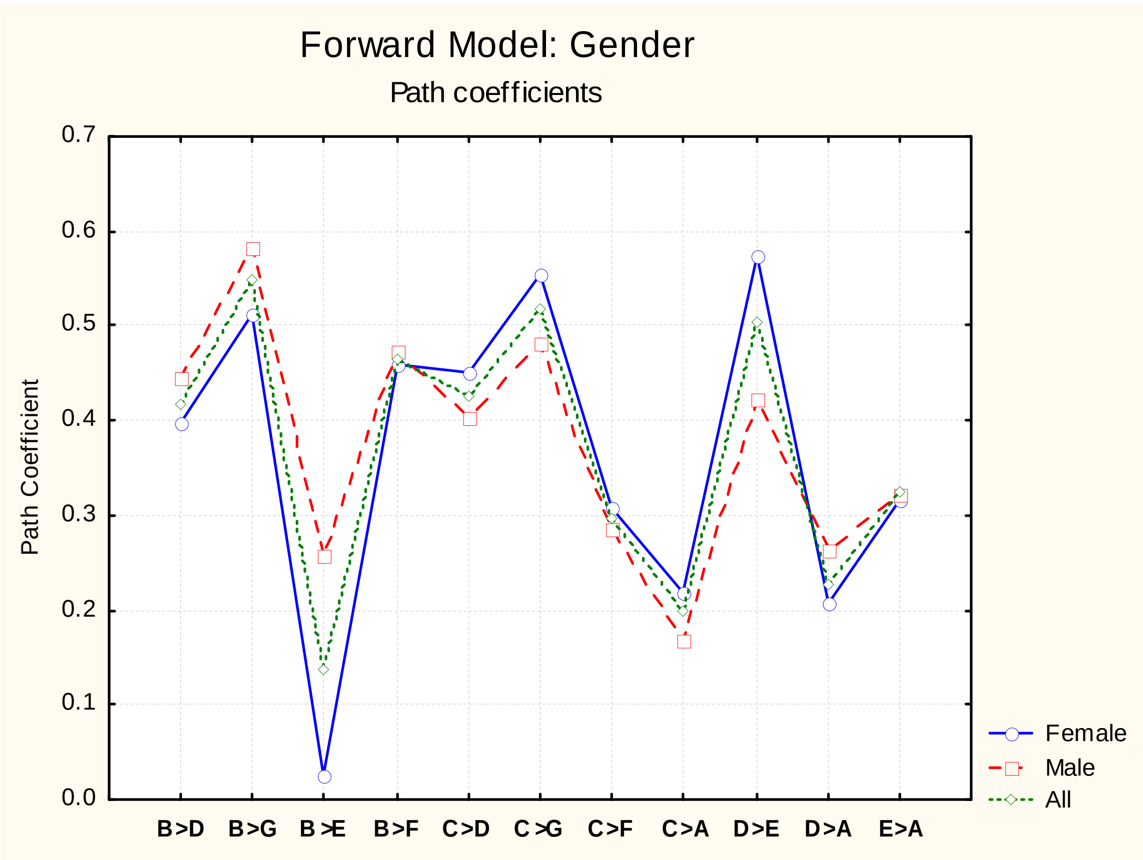
- Broken Hill has a comparatively high B-->G path coefficient and comparatively low C-->G path coefficient
- South Lake has comparatively very low B-->E, C-->D, and D-->A path coefficients
- The US Mid West has a comparatively high C-->G path coefficient

There are also significant differences between communities and groups in the Reverse model. For example:

- Broken Hill has a comparatively high D-->C path coefficient and comparatively low G-->C path coefficient
- South Lake has comparatively high E-->D and G-->B path coefficients
- The US Mid West has a comparatively low E-->B path coefficient

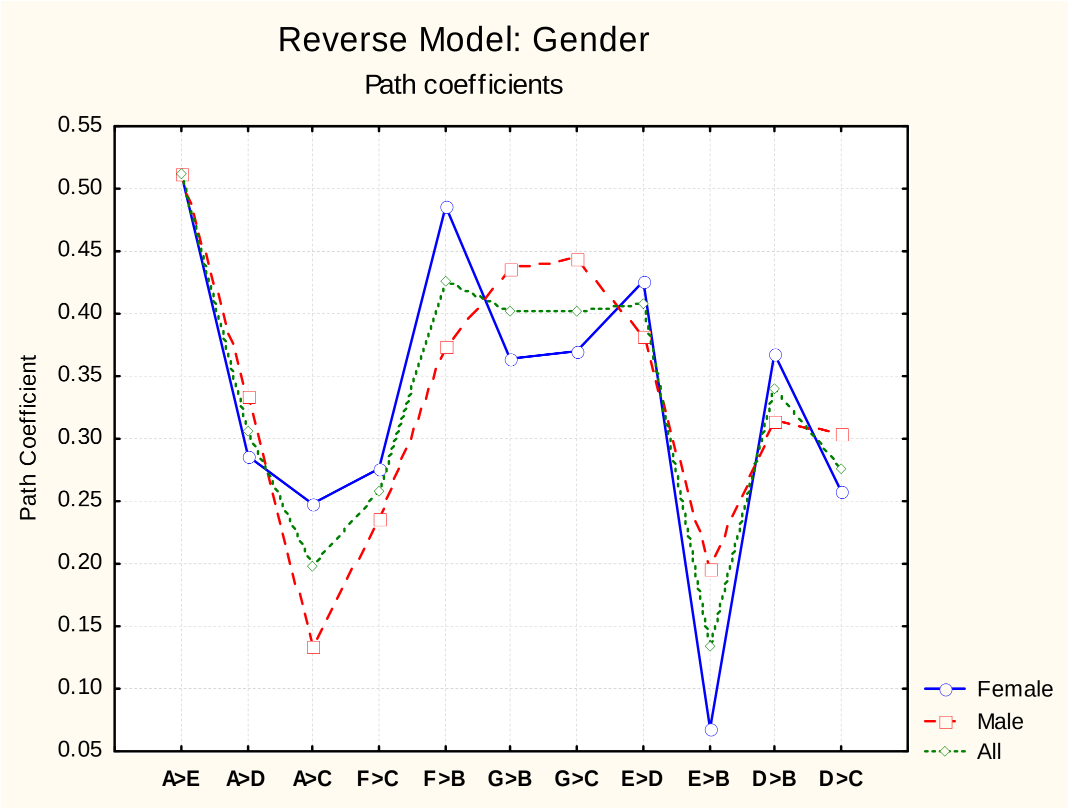
Gender differences

Similar questions can be asked about gender differences. The path loadings for the Forward and Reverse models showing gender differences are in the following two charts.



There are a few significant differences between genders.

- Females have comparatively lower path coefficients for B-->E and E-->B and comparatively higher path coefficients for D-->E
- Males have comparatively lower path coefficients for A to C and comparatively higher path coefficients for E-->B



Conclusions

- The factor structure is very stable across communities and groups.
- Many path models are possible, many are moderate to poor fits.
- Some path models are much more plausible than others (eg the Forward Model and Reverse Models described here)
- Path coefficient loadings in the Forward and Reverse models vary from one community to another suggesting that the mechanisms for social capital development vary from one community to another.
- Further exploration may identify structural models that suit one community better than another.
- A few path coefficient loadings vary for males and females suggesting that some of the mechanisms for social capital development vary between males and females.
- There are significant implications for community building strategies (one size does not fit all).

Attachment 1: Statistical analysis notes

Sample sizes N1 and N2

There are 6249 cases in the study (N1).

There are 4618 cases in which there is no missing data on the social capital questions (N2). N2 is used in all structural equation and factor analyses.

Samples S1 and S2

The 4618 cases in N2 were randomly divided into two samples S1 and S2.

Most of the structural equation modelling analyses were undertaken separately with these two samples - to see if the results in the first sample could be replicated with the second sample.

Where results could not be replicated in the second sample they have not been relied on. For example Narellan, Greenacre and the other communities with smaller sample sizes were not able to replicate similar path coefficients for both S1 and S2 and so were not included in the analysis of differences between communities.

Factors and questions

The questions were grouped into factors on the basis of the original factors and question groupings. In addition the two 'other' questions which are not usually added together in the individual factor scores were included in the F and G factors (as they loaded on these questions (and others)).

Gender and communities/groups path coefficients

Path coefficients for *All*, *Male* and *Female* models. These analyses were replicated for S1 and S2. In generating the loadings for Samples 1 and 2 the measurement model is based on the 4618 cases and the path loadings are allowed to vary with each sample. The loadings in the *All* model are arrived at by constraining loadings in Samples 1 and 2 to be the same.

A similar approach is taken to the gender analyses and communities/groups analyses - the measurement model is based on all males and females in Sample 1, Sample 2 and the Total respectively - the path loadings are then allowed to vary by gender. In the communities/groups analysis the measurement model is based on all cases in Sample 1, Sample 2 and the Total respectively - the path loadings are then allowed to vary by community/group.

Statistical software

STATISTICA 7 used for all statistical analysis.

Structural equation solution

The structural equation models have been developed using a standardised solution via constrained estimation. this approach produces a solution where all latent and manifest variables variances of 1. This method, described by Browne and DuToit (1987), and Mels (1989), is a constrained Fisher Scoring algorithm. This is not the same as the problematic "standardised" solutions generated in LISREL VI and CALIS.

Model Fit

The Confirmatory Factor Analysis and the Forward and Reverse Models, are based on the combined data from all nine studies (nine communities and four groups). The fit statistics are:

	Forward	**Reverse**	**CFA**
Steiger-Lind RMSEA Index	.057	.060	.089
RMS Standardized Residual	.068	.089	.147
Adjusted Population Gamma Index	.906	.897	.801
Joreskog AGFI	.900	.891	.797
Independence Model Chi-Square	32979	32979	32979
Independence Model df	465	465	465
Bentler-Bonett Non-Normed Fit Index	.806	.803	.639
Bentler Comparative fit Index	.823	.803	.663

The Forward Model is the best fit.

The Reverse model is a good fit, but not as good as the Forward model.

The Confirmatory Factory Analysis is a good enough fit given the nature of social capital and the measurement tool.

Samples

The nine studies include the following samples

Communities	**N1**	**N2**
a) Ultimo/Pyrmont, Inner City Sydney, NSW	247	199
b) Deniliquin, Rural, NSW	266	230
c) Narellan, Outer Sydney, NSW	233	183
d) Green Acre, Wester Sydney, NSW	256	186
e) West Wyalong, Rural NSW	209	192
f) Melany, Rural QLD	137	113
g) Broken Hill, Western NSW	635	494
h) US Mid West	496	379
i) South Lake, Urban, WA	976	586
Neighbourhood Centres (NC) and Family Support services		
j) NC Group participants (NSW)	944	649
k) NC volunteers (NSW)	378	246
l) NC staff (NSW)	796	644
m) Family support clients (NSW)	676	517

Sources of data

The Ultimo/Pyrmont, Deniliquin, Narellan, Green Acre and West Wyalong data is from the study "*Measuring Social Capital in Five Communities in NSW*", Jenny Onyx and Paul Bullen

The Melany data is from *“Maleny: Social Capital and the Development Paradox”*, December 2005, Melissa Edwards, Jenny Onyx, Ann Dale CACOM Working Paper No 70

The South Lake Ottey and Family and Neighbourhood Centre provided the South Lake data from the study *"The Community of South Lake Measuring Social Capital and Community Pride*" 2002 (Margaret Auld and Margaret O'Neil)

The Broken Hill data is from *Citizens Views of Broken Hill*, June 2005, Jenny Onyx, Lynelle Osburn, Paul Bullen CACOM Working Paper No 67 and *Social Capital: A Rural Youth Perspective*, June 2005 Jenny Onyx, Craig Wood, Paul Bullen, Lynelle Osburn, CACOM Working Paper No 68.

Megan O'Brien provided the data form the US Mid West (a community of approximately 350,000 residents) from the study "*Further development of an Australia-based measure of social capital in a US sample*" 2004

The Neighbourhood and Community Centre data is from the study "*Social Capital and Family Support Services and Neighbourhood and community centres in NSW*" 1999, 2005 (Paul Bullen and Jenny Onyx).

Attachment 2: Questions

The following are the 31 questions used in the studies. The numbers in brackets are the number of the questions in the original five communities study. The numbers in the left hand column are the question numbers used in the structural equation models above.

Participation in the Local Community

1. Do you help out a local group as a volunteer? (16)
2. Have you attended a local community event in the past 6 months (eg, church fete, school concert, craft exhibition)? (29)
3. Are you an active member of a local organisation or club (eg, sport, craft, social club)? (31)
4. Are you on a management committee or organising committee for any local group of or organisation? (44)
5. In the past 3 years, have you ever joined a local community action to deal with an emergency? (46)
6. In the past 3 years have you ever taken part in a local community project or working bee? (48)
7. Have you ever been part of a project to organise a new service in your area (eg, youth club, scout hall, child care, recreation for disabled)? (50)

Proactivity in a social context

8. Have you ever picked up other people's rubbish in a public place? (14)
9. Do you go outside your local community to visit your family? (37)
10. If you need information to make a life decision, do you know where to find that information? (41)
11. If you disagree with what everyone else agreed on, would you feel free to speak out? (54)
12. If you have a dispute with your neighbours (eg, over fences or dogs) are you willing to seek mediation? (56)

Feelings of Trust and Safety

13. Do you feel safe walking down your street after dark? (17)
14. Do you agree that most people can be trusted? (18)
15. If someone's car breaks down outside your house, do you invite them into your home to use the phone? (19)
16. Does your area have a reputation for being a safe place? (24)
17. Does your local community feel like home? (33)

Neighbourhood Connections

18. Can you get help from friends when you need it? (21)
19. If you were caring for a child and needed to go out for a while, would you ask a neighbour for help? (26)
20. Have you visited a neighbour in the past week? (28)

21. When you go shopping in your local area are you likely to run into friends and acquaintances? (39)
22. In the past 6 months, have you done a favour for a sick neighbour? (45)

Family and Friends Connections

23. In the past week, how many phone conversations have you had with friends? (34)
24. How many people did you talk to yesterday? (35)
25. Over the weekend do you have lunch/dinner with other people outside your household? (36)

Tolerance of Diversity

26. Do you think that multiculturalism makes life in your area better? (57)
27. Do you enjoy living among people of different life styles? (59)

Other

28. If a stranger, someone different, moves into your street, would they be accepted by the neighbours? (60)

Value of Life

29. Do you feel valued by society? (1)
30. If you were to die tomorrow, would you be satisfied with what your life has meant? (3)

Other

31. Some say that by helping others, you help yourself in the long run. Do you agree? (15)