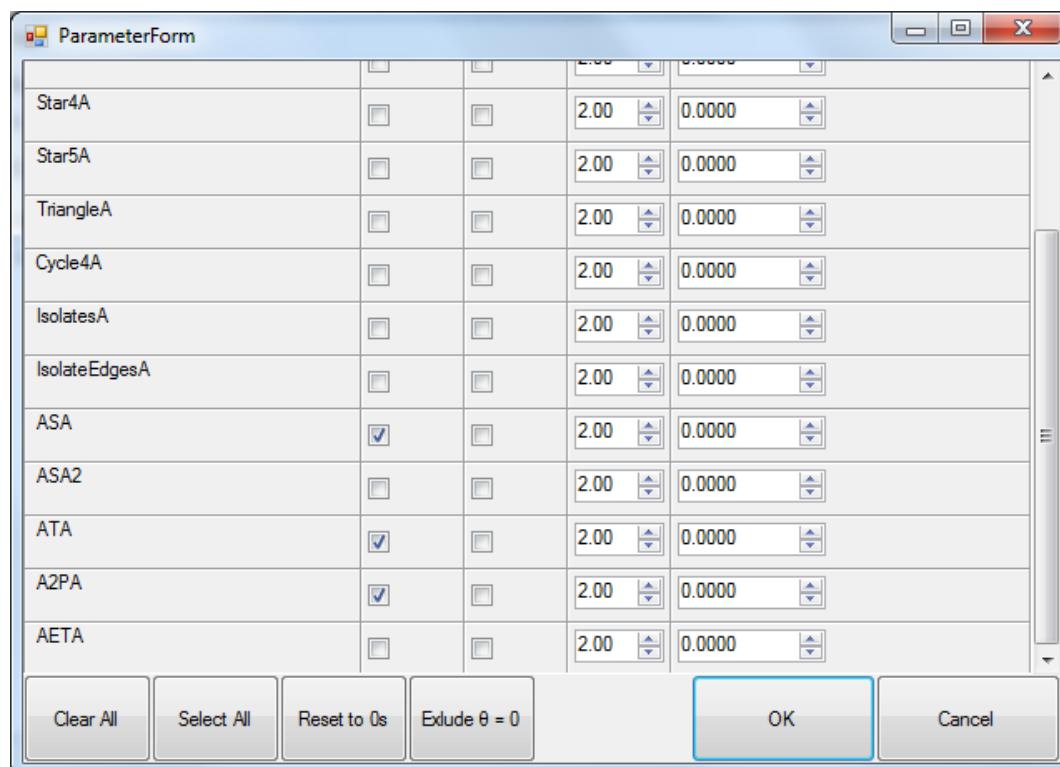# Exercise

Estimating social circuit models

Swinburne University of Technology, Australia ©

Exercise: Estimating social circuit models

Network Data: **FishermenTies.txt**

12.1 Social circuit models

Try estimating a social circuit model to the Fishermen data.

Choose 'Select parameters...'' and then > **Clear All**.
Under "Higher order parameters", select the following social circuit parameters:
>   Edge
>   ASA (alternating stars)
>   ATA (alternating triangles)
>   A2PA (alternating 2-paths).

The model should successfully converge with results like:

Effects	Lambda	Parameter	Stderr	t-ratio	SACF	
EdgeA	**2.0000**	**-3.3365**	**0.585**	**-0.092**	**0.634**	*
ASA	**2.0000**	**-0.2658**	**0.251**	**-0.089**	**0.650**	
ATA	**2.0000**	**0.8061**	**0.120**	**-0.073**	**0.596**	*
A2PA	**2.0000**	**0.0436**	**0.045**	**-0.064**	**0.655**	

Interpretation*:*
- The negative and significant edge effect indicates that we see few edges unless they are contained within other network structures. That is, there are relatively few isolated edges between dyads in this data – thus where edges occur they are within other network structures.

- The positive AT (alternating-triangle) parameter indicates there are many triangles in the network, given the other effects in the model

12.2 Social circuit models: Some simulations

In this exercise, simulations will illustrate some features of social circuit models. In order to make comparisons across different simulations, we will simulate distributions of 30 node (undirected) graphs with a fixed density of 0.10 (43 edges). **Because of the fixed density, the models will not have an edge parameter.**

In setting up the simulation, make sure that these settings are entered.

After each simulation, either delete the files in the session folder, or if you want to save them, move them to another folder. This will remove any problems with overwriting files.

For the most part in this exercise, we will simply use VPNet to visualise the final graph in the simulation, so as to get a qualitative understanding of the effects of the model parameterisation on the graph structure (you will need to create an 'attributes' file for VPNet – see earlier exercises for the correct format).

12.3 Triangulation

(a) <u>Markov models</u>: Run three simulations with the only parameter being a Markov Triangle parameter with values 0.5, 1.0 and 1.5, respectively. (Recall:

click Select Parameters ... to enter these values.) In each case, use Pajek to visualise the final graph in the simulation, "end_statistics_*session name*.txt". If you want, you can calculate the average number of triangles across your simulated sample from the output statistics file "simulation_*session name*.txt". If you have SPSS on your computer, the SPSS syntax file "sim_*session name*.sps" will do this automatically for you.

You should notice that there is some increase in the number of triangles in the first two models but then a massive jump in the third model to produce a clique-like structure. This illustrates the *phase transition* that often occurs in Markov models as parameters increase, one of the reasons why Markov models are often degenerate.

(b) <u>Social circuit models</u>: Now run three simulations with the only parameter being the alternating triangle, ATA. Use the same parameter values as before, 0.5, 1.0 and 1.5. To count the number of triangles in the simulated graphs, select the Markov triangle parameter, but <u>make sure it is set to 0</u>. You will see that the number of triangles does not increase to the maximum possible. Not that these graphs (especially for ATA=0.5) have the triangles dispersed across the network (i.e. there are not many higher order cliques, or a core to the network.)

(c) <u>The lambda value in ATA</u>: The lambda value controls the size of the denser regions of the networks. Choose a modest AT parameter value of 0.5. From the previous example, you know that this produces graphs without many higher order cliques. But this time change the lambda value to 10. This produces a more clique-like, core-periphery structure with more higher-order cliques.

12.4 Degrees

(a) <u>Positive and negative alternating stars:</u> Clear all the previous parameters. Run two simulations with only an alternating star (ASA) parameter. Use parameter values, -1.0 and +1.0, respectively. On the main MPNet window, select **Pick up sample degree distribution**. This command will produce a separate output file, "*session name*_degreeA.txt", with the degree distribution for every graph sampled from the simulation.

On the degree distribution output file, the variable *dn* indicates the number of nodes in the graph of degree *n*. The last two variables are the standard deviation and skew of the degree distribution. You can enter this file into a statistical package such as SPSS or Excel to look at the degree distribution across the simulation (e.g. using boxplots) or simply to examine the standard deviation and skew.

For a parameter value of -1.0, the average standard deviation and skew should be about 1.2 and 0.4 respectively. For a parameter of +1.0, the average standard deviation should have increased by a lot but the skew remains about the same. So, an increasingly positive ASA parameter leads to

increased dispersion in the degree distribution (that is, more high degree nodes).

A little further experimentation will show that increasing the ASA parameter by a lot (say, to 10) will produce a clique-like structure of a small number of nodes of equal high degrees and many other nodes with low degrees. On the other hand, leaving the ASA parameter value at 2 but increasing the lambda value leads to concentration of ties on one or two very high degree hubs.

(b) <u>Alternating and Markov-star parameters</u>: Sometimes it is helpful to include both an alternating star parameter and a Markov star parameter, usually a 2-star parameter. This is particularly useful for modelling highly centralized graphs with a handful of very high degree nodes. When fitting models to such data, we often find that the alternating star parameter is negative and the 2-star parameter is positive: the negative alternating star parameter handles a relatively equal distribution of degrees among low degree nodes, while the Markov parameter caters for the high degree nodes. Try a simulation with an alternating star parameter of – 1, and a Markov 2-star parameter of + 0.3.

12.5 Stars and triangles together

Finally, try a simulation with an alternating triangle parameter of 2.0 and an alternating star parameter of –1.0 (both with lambda of 2). Inspect the final graph using Pajek. You should see that strong triangulation, consistent with the positive triangle parameter is retained. Yet, instead of producing a core of high density, now it is distributed across a number of smaller, denser regions of the network, sometimes connected and sometimes in separate network components. Here the effect of the negative star parameter is to break down the core into smaller, clique-like regions, as the parameter attempts to limit the number of high degree nodes.

Exercise: Goodness of Fit

Network Data: **FishermenTies.txt**
                  comm_undirected.txt

13.1 Social circuit models

Begin by estimating the following models for the both:
- "**Fishermen**" data
- "**comm_undirected.txt**" data

For both data sets, estimate:
1. A Bernoulli model
2. A social circuit model

As a reminder

Select parameters... > Clear All.
Then, select the following social circuit parameters:

   EdgeA
   ASA (alternating stars)
   ATA (alternating triangles)
   A2PA (alternating 2-paths).

The social circuit model for the **Fishermen** should successfully converge with results like this :
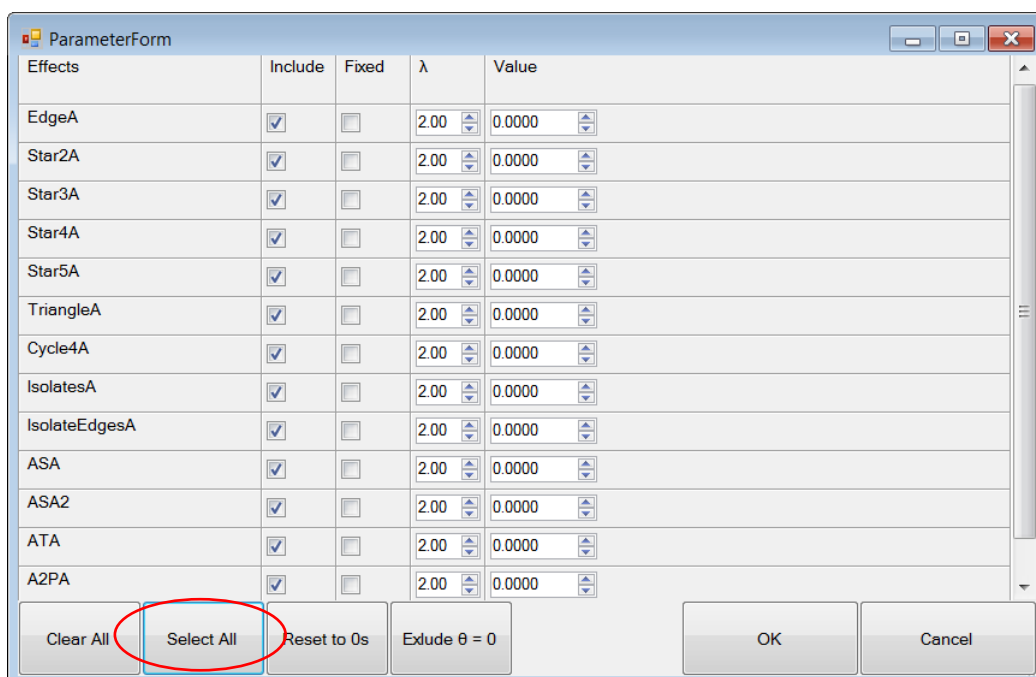
Effects	Lambda	Parameter	Stderr	t-ratio	SACF	
EdgeA	2.0000	-3.3626	0.570	-0.031	0.450	*
ASA	2.0000	-0.2502	0.242	-0.021	0.469	
ATA	2.0000	0.7906	0.126	0.046	0.449	*
A2PA	2.0000	0.0429	0.039	0.005	0.500	

13.2 Goodness of fit (GOF)

Select the **GOF** button. The Goodness of Fit (GOF) procedure takes a user defined set of estimates and simulates a distribution of graphs from that model. It then compares various statistics and graph features from the distribution of graphs against the data. In particular, when the data is extreme compared with the simulated distribution of graphs, the output will report a t-statistic that is greater than 2. In that case, the model does not fit that aspect of the data well.

- (**NB**: for graph statistics that are in the model, the t-statistic in the GOF output should be suitably small – e.g. less than 0.1 in absolute value – else this is evidence that the original estimation did not converge well, irrespective of the convergence statistics provided by the estimation procedure.)

To investigate particular graph features, select **Structural Parameters** and then click **Select Parameters...**. Make sure that you select all the parameters in the estimated model (in this case, EdgeA, ASA, ATA, A2PA). It is also good to select the Markov statistics, especially all the stars, for details of how well the model reproduces the degree distribution and triangle count, and the clique statistics, to investigate how well the higher order closure (no of cliques) is reproduced. You can choose others if you wish. A simple/fast way to select all statistics is to click on the **Select All** button.

Click OK. Then click **Update!** This will enter the last parameter estimates as the value of the model parameters. Then click **Start!**

Using the criterion of an absolute t-value of 2 as indicating an extreme result, the output should show that the model does not fit the higher order stars (4- and 5-stars), and the skew of the degree distribution. In other respects the model does a

reasonably good job of reproducing the observed network.

13.3 ADVANCED (Try this at home!) Trying to improve the model

A model is never perfect. Sometimes it is desirable to add further parameters. However, if you want to improve the model for this data, the need to do better in fitting higher order cliques suggests there may be advantage in increasing the lambda value for AT. Larger lambdas for triangles permit the creation of more large cliques.

Go back to Estimation, **Select parameters**...>**Clear All**. Select EdgeA, ASA, ATA, and A2PA parameters as before, and change the lambda value for ATA to 2.5. Run the estimation and then apply Goodness of Fit to converged parameter estimates (Note that in GoF, you have to manually update the lambda value for ATA to 2.5, and similar to ASA for the next model). You should see improvement in the t-values for stars and cliques. However, the skew on the degree distribution is still too low compared to the data.

To increase the dispersion of a positive ASA parameter, an increased lambda can be used. Here the parameter estimate is negative, so we need to decrease lambda. Repeat the estimation with a lambda of 1.8 for ASA and 2.5 for ATA. Goodness of fit will now suggest that most or all of the troublesome effects in the data are plausibly reproduced.

How did we decide on the values of 1.8 and 2.5? A little experimentation with fitting models.

13.4 Try GOF of the social selection model on the Corporation directed communication network

Follow the instructions of the Exercise on Social Selection Models to obtain a model for the directed communication network for the Corporation data set.

Test the model GOF by selecting ALL structural statistics, and ALL attribute statistics. Do you have an adequate model for the dataset?