

PNet for Dummies

An introduction to
estimating exponential
random graph (p^*)
models with PNet

Version 1.04

Nicholas Harrigan

To download the latest copy of this manual go to:
<http://www.sna.unimelb.edu.au/pnet/pnet.html#download>

INTRODUCTION.....	3
TERMINOLOGY	4
STEP 1: INSTALLING PNET.....	5
STEP 2: PREPARING MATRIX AND ATTRIBUTE FILES.	6
2.1: Preparing Attributes in Excel	6
2.2 Integrating Attribute and Network Data in a VNA file	8
2.3 Transforming VNA File into Raw Matrix and Attribute files	10
STEP 3: ESTIMATION IN PNET.....	12
3.1 Setting up PNet Estimation.....	12
3.2 Preventing Model Degeneracy.....	18
3.3 Running an Estimation.....	21
3.4 Fitting an Estimation	22
STEP 4: GOODNESS OF FIT.....	25
4.1 Running a Goodness of Fit.....	25
4.2 Interpreting GOF statistics	25
APPENDICES.....	27
Appendix 1: Interpreting GOF statistics	27
Appendix 2: Recommended Starting Parameters.....	27
Appendix 3: Running an estimation (Summary).....	28

Introduction

PNet for Dummies is intended to walk the new user through one complete estimation in PNet. It is not a comprehensive guide to PNet. Currently the most comprehensive guide to PNet is the ***PNet Users Manual***.

PNet for Dummies exists to help get the new user started, helping them overcome the most common initial barriers, so that they can begin exploring and experimenting with PNet themselves.

To this end, ***PNet for Dummies*** tries to emphasise solutions to some common problems, through dealing with issues such as:

- synchronising your network and attribute data (using VNA files)
- transforming your data into raw matrix and raw attribute formats.
- deciding which configurations/parameters to select for your model
- preventing degeneracy in your model
- identifying the causes of unreasonable parameter estimates in your model
- fitting your model
- interpreting goodness of fit statistics.

For the most part, we have written ***PNet for Dummies*** as a way of documenting many of the heuristic (that is, rule-of-thumb) solutions which we have come across as we have learnt to use PNet. We hope you will find some of our solutions useful to your work.

The vast majority of the solutions documented in ***PNet for Dummies*** are techniques which have been developed by the staff and students of the Social Networks Laboratory at the University of Melbourne, in particular, Pip Pattison, Garry Robins, Peng Wang, Dean Lusher, Galina Daraganova and Johan Koskinen.

Any suggestions, feedback, comments or corrections would be greatly appreciated, and can be emailed to nick_harrigan@yahoo.com or nick.harrigan@anu.edu.au

Terminology

You will notice certain indented text that is written in `Courier` font, such as this:

```
Data>Import>VNA
```

Such text is intended to emphasize that the text is referring to a piece of semi-programming language text, such as an excel formula, or it is a direction to access a program or menu in a computer program.

The use of the ">" symbol" refers to opening of either folders in a computers desktop, or the opening of menus inside a computer program. For example, the line above refers to a UCINET menu, and asks the reader to select the "Data" menu, then select the "Import" sub-menu, and then the "VNA" sub-menu.

Step 1: Installing PNet

Before running PNet, you will need to install the program. You can do this by going to the downloads part of the PNet section of the MelNet website, which can be found here:

<http://www.sna.unimelb.edu.au/pnet/pnet.html#download>

Simply click on the "PNet setup.exe" link under the heading " PNet for Single Networks"

Follow the prompts, selecting "run", "ok" and "next" as appropriate. The default setting for PNet are all fine EXCEPT that you should not run PNet from the shortcuts. If you do, PNet will leave about 4 or 5 little files on your desktop, containing starting statistics and update files. Instead, you should always run PNet from the folder it is installed in. If you want to set up a short cut, set up a short cut to that folder, not to the PNet file itself.

If you attempt to run PNet and it does not work, it may be because you do not have either Java Platform or the Microsoft .NET framework installed. These can both be downloaded from the PNet website under the heading "Required Environment"

Step 2: Preparing matrix and attribute files.

Before using PNet, we need to prepare matrix and attribute files which PNet will use as inputs. At the moment PNet only accepts one form of input file: Raw matrix and attribute data. Most other network and attribute file types can be transformed to Raw data types by UCINET. This section (Step 2) provides a step by step guide to preparing your data in Excel, creating an integrated network and attribute VNA file, and then transforming this file into the raw data format using UCINET. If you already have your data in raw matrix and raw attribute format (and the rows and columns of these two raw data files match) then you should feel free to skip this section.

2.1: Preparing Attributes in Excel

I tend to prepare my attributes in Excel.

I list each node down the left hand column, and list the attributes across the top row.

I label each attribute according to it's type, either "BIN_" for binary attributes, and "CONT_" for continuous attributes, and "CAT_" for categorical attributes.

I group attributes of the same type in adjacent columns (i.e. Binary attributes next to binary attributes, Continuous next to continuous, etc)

	A	C	D	E	F	L	M
	Company Name	BIN_Think_T ank	BIN_BCAChr msPnl	BIN_BCA_A LL	BIN_Australi an	CONT_Top4 BusAsscsEx ecs	CONT_N_No nExecDir
1	Company Name						
2	AWB Limited	0	0	1	1	0	11
3	APN News & Media Limited	0	0	0	1	0	11
4	National Australia Bank Limited	1	0	1	1	1	10
5	Westfield Group	1	1	1	1	0	10
6	Queensland Sugar Limited	0	0	0	1	0	10
7	Ten Network Holdings Limited	0	0	0	1	0	10
8	ABB Grain Ltd	0	0	0	1	0	10
9	Leighton Holdings Limited	0	0	1	0	1	9
10	Woodside Petroleum Ltd	1	0	1	1	1	9
11	Mitre 10 Australia Ltd	0	0	0	1	0	9
12	Coles Myer Ltd	0	0	1	1	0	8
13	Commonwealth Bank of Austral	0	1	1	1	0	8
14	Rio Tinto Plc - Rio Tinto Limited	0	0	0	1	0	8
15	Qantas Airways Limited	0	1	1	1	1	8
16	Westfarmers Limited	0	0	1	1	0	8
17	Publishing and Broadcasting Lirr	0	0	0	1	0	8
18	BHP Billiton Limited	0	1	1	1	0	7

To attach these attributes to their network, I place the data in a VNA file (which is a file type for UCINET. It is similar to a DL file type).

To do this I create a formula at the end of each row in Excel. The formula I use for the attributes (i.e. all lines except for the first line, which has the attribute names in it) is:

```
=""&A2&""&" "&C2&" "&D2&" "&E2&" "&F2&" "&G2&"
"&H2&" "&I2&" "&J2&" "&K2&" "&L2&" "&M2&" "&N2&"
"&O2&" "&P2&" "&Q2&" "&R2&" "&S2"
```

This basically just says, put double inverted commas around the name of the node, and then list the attribute values one at a time, with a space between each one.

Write this formula out for the first line of data (not the first row with headings) and then copy and paste it down the rest of the row.

For the top line, which contains the attribute labels, I use this formula:

```
=""&A1&""&"", "&C1&", "&D1&", "&E1&", "&F1&",
"&G1&", "&H1&", "&I1&", "&J1&", "&K1&", "&L1&",
"&M1&", "&N1&", "&O1&", "&P1&", "&Q1&", "&R1&", "&S1"
```

This formula says, put the first column heading in double inverted commas, and then list the rest of the variable names, with a comma between each one. NOTE: TO USE THIS FORUMULA YOUR ATTRIBUTE NAMES MUST NOT HAVE ANY SPACES IN THEM. If you want to have spaces in your attribute names then you will have to place double inverted commas around each one of them (as I have done with the first column in my attribute heading list).

You should end up with a column in your dataset that looks like this:

	R	S	T	U	V	W	X	Y	Z
	CONT_Prop Execs_on_G ovtBds	CONT_Reve nue_\$bil	"Company Name", BIN_Think_Tank, BIN_BCACHrmsPnl, BIN_BCA_ALL, BIN_Au						
1			"AWB Limited" 0 0 1110 0 0 0 0 113 0.727272727272727 0.666666666666667 0 0 5.15						
2	0	5.156545	"APN News & Media Limited" 0 0 0 110 0 0 0 0 113 0.818181818181818 10 0 1.282793						
3	0	1.282793	"National Australia Bank Limited" 10 11110 0 0 110 3 110.2 0 37.406						
4	0	37.406	"Westfield Group" 111110 0 110 10 4 110.3 0 1.8766						
5	0	1.8766	"Queensland Sugar Limited" 0 0 0 10 0 0 0 0 10 3 10.666666666666667 0 0 1.551998						
6	0	1.551998	"Ten Network Holdings Limited" 0 0 0 110 0 110 10 3 110 0 0.954586						
7	0	0.954586	"ABB Grain Ltd" 0 0 0 110 0 0 0 0 10 3 110 0 0.795892						
8	0	0.795892	"Leighton Holdings Limited" 0 0 10 10 0 10 19 3 110.111111111111 0.333333333333333						
9	0.3333333333	6.381376	"Woodside Petroleum Ltd" 10 1110 0 10 19 3 0.666666666666667 10.111111111111 0 3						
10	0	3.003311	"Woodside Petroleum Ltd" 10 1110 0 10 19 3 0.666666666666667 10.111111111111 0 3						

2.2 Integrating Attribute and Network Data in a VNA file

Attribute Data

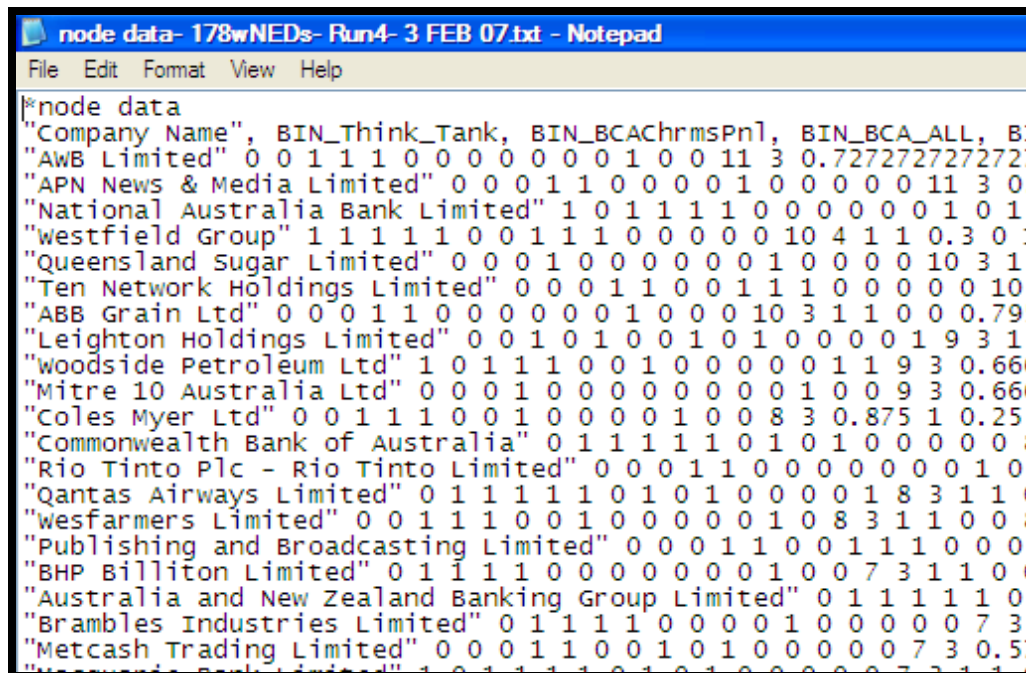
Select this column and press copy.

Open Notepad, and press paste.

Type the follow text at the top of the file:

```
*node data
```

You should have a screen that looks like this:



```
node data- 178wNEDs- Run4- 3 FEB 07.txt - Notepad
File Edit Format View Help
*node data
"Company Name", BIN_Think_Tank, BIN_BCAChrmsPnl, BIN_BCA_ALL, BI
"AwB Limited" 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 11 3 0.7272727272727
"APN News & Media Limited" 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 11 3 0.
"National Australia Bank Limited" 1 0 1 1 1 1 0 0 0 0 0 0 0 1 0 1
"westfield Group" 1 1 1 1 1 0 0 1 1 1 0 0 0 0 0 10 4 1 1 0.3 0 1
"Queensland Sugar Limited" 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 10 3 1
"Ten Network Holdings Limited" 0 0 0 1 1 0 0 1 1 1 0 0 0 0 0 10
"ABB Grain Ltd" 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 10 3 1 1 0 0 0.795
"Leighton Holdings Limited" 0 0 1 0 1 0 0 1 0 1 0 0 0 0 0 1 9 3 1
"woodside Petroleum Ltd" 1 0 1 1 1 0 0 1 0 0 0 0 0 0 1 1 9 3 0.666
"Mitre 10 Australia Ltd" 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 9 3 0.666
"Coles Myer Ltd" 0 0 1 1 1 0 0 1 0 0 0 0 1 0 0 8 3 0.875 1 0.25
"Commonwealth Bank of Australia" 0 1 1 1 1 1 0 1 0 1 0 0 0 0 0 8
"Rio Tinto Plc - Rio Tinto Limited" 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0
"Qantas Airways Limited" 0 1 1 1 1 1 0 1 0 1 0 0 0 0 1 8 3 1 1 0
"wesfarmers Limited" 0 0 1 1 1 0 0 1 0 0 0 0 0 1 0 8 3 1 1 0 0 8
"Publishing and Broadcasting Limited" 0 0 0 1 1 0 0 1 1 1 0 0 0
"BHP Billiton Limited" 0 1 1 1 1 0 0 0 0 0 0 0 1 0 0 7 3 1 1 0 0
"Australia and New Zealand Banking Group Limited" 0 1 1 1 1 1 0
"Brambles Industries Limited" 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 7 3
"Metcash Trading Limited" 0 0 0 1 1 0 0 1 0 1 0 0 0 0 0 7 3 0.57
"Metcash Trading Limited" 1 0 1 1 1 1 0 1 0 1 0 0 0 0 0 7 3 1 1
```

Save this file. As you can see from the title of my file (at the top of the Notepad screen) I tend to label my files "node data ..." and then some descriptive information about the data set.

Go back to your Excel file and count the number of Binary, Continuous, and Categorical attributes you have in your dataset. Note the order (they should be grouped into their attribute types). Write down these three numbers and the order (we will use them later).

Matrix data

I prepare my matrix data in VNA file format as well.

If you have your matrix data in any other format, you can convert it to VNA file format by opening it in NETDRAW (not UCINET), and then selecting:

```
File>Save Data As>VNA>Complete
```


Export the Attribute File as EXCEL data:

Data>Export>Excel, Select your UCINET Attribute file, press Ok.

Note: if you have any unexplained problems with exporting to the Excel file or any other export file, try deleting the entire path of the "Output file" or "Output dataset", and then just retyping in a name you want to call the file, without the rest of the path (i.e. without the bit that says "C:\MyDocuments\etc."

Open the Excel file you have created.

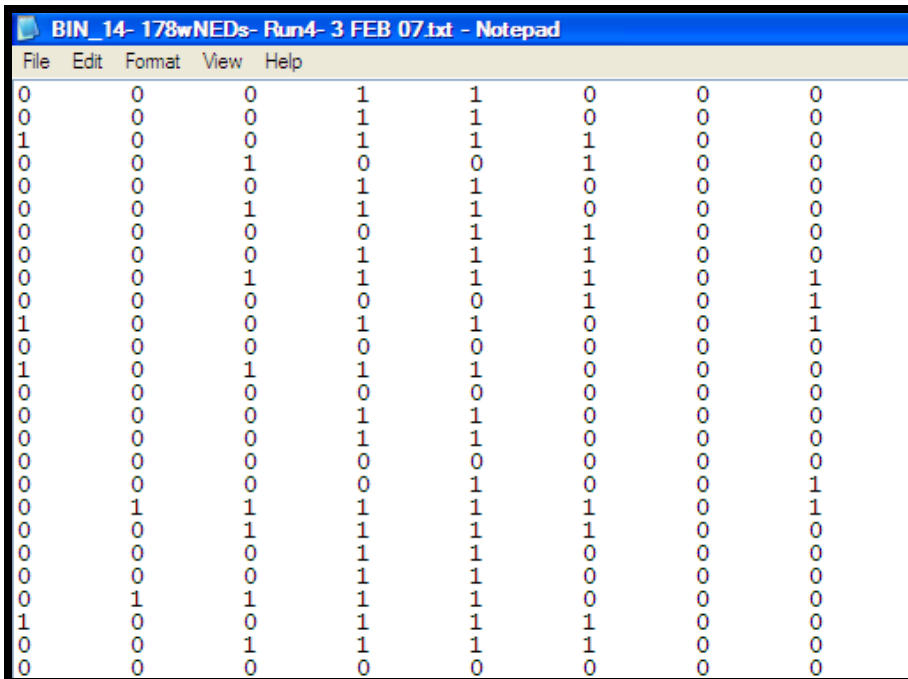
Select the cells containing the binary data. Do not select the column or row headings. Press copy.

Open Notepad. Press paste.

Press save. Save as a file with the title "BIN_" and then the number of binary variables in this dataset, and then some name you will remember (eg. "BIN_14-NonExec250nodes.txt"). This is your final Binary attribute file.

Repeat this process for your Continuous and Categorical attributes.

They should each look something like this:



File	Edit	Format	View	Help										
0	0	0	1	1	0	0	0	0						
0	0	0	1	1	0	0	0	0						
1	0	0	1	1	1	0	0	0						
0	0	1	0	0	1	0	0	0						
0	0	0	1	1	0	0	0	0						
0	0	1	1	1	0	0	0	0						
0	0	0	0	1	1	1	0	0						
0	0	0	1	1	1	0	0	0						
0	0	1	1	1	1	0	0	1						
0	0	0	0	0	1	0	0	1						
1	0	0	1	1	0	0	0	1						
0	0	0	0	0	0	0	0	0						
1	0	1	1	1	0	0	0	0						
0	0	0	1	1	0	0	0	0						
0	0	0	1	1	0	0	0	0						
0	0	0	1	1	0	0	0	0						
0	1	1	1	1	0	0	0	0						
1	0	0	1	1	1	0	0	0						
0	0	1	1	1	1	0	0	0						
0	0	0	0	0	0	0	0	0						

Step 3: Estimation in PNet

3.1 Setting up PNet Estimation

Before you open PNet, take the raw matrix, and attribute files that you have created and put them in a new folder on their own. Make a copy of this folder and call it something like "Original Network and Attribute files"

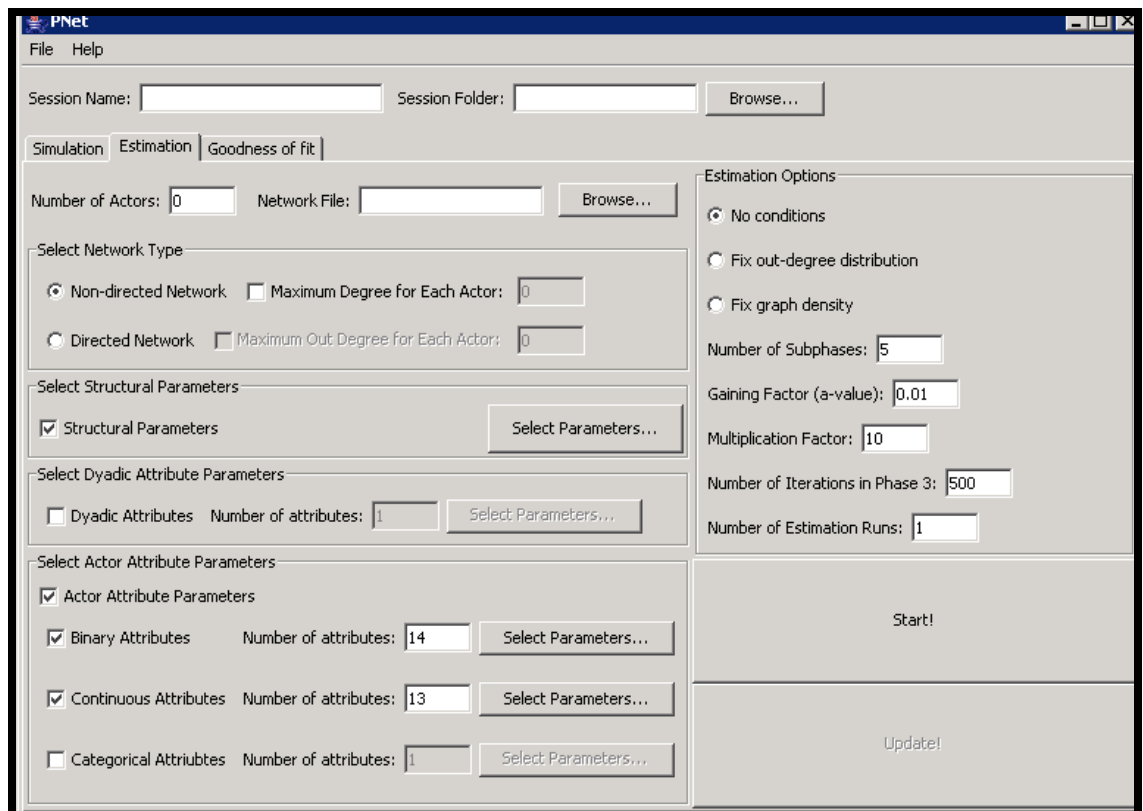
Then name the other folder the name of the session which you are going to run in PNet. I like to name my folders/sessions experiments, so I give them a name like "Exp 1 –Top250_27Att- 6 FEB 07", which means it's experiment 1, and it's of my top 250 corporations network, with 27 attributes, and it's on the 6th Feb 2007. I am a bit obsessive with naming files. You might prefer something simpler.

Make a mental (or physical) note of where this folder is.

Open PNet

Click on the tab labelled "Estimation"

You should have screen that looks like this:



Type a session name in the top left hand corner box.

Select a session folder by clicking Browse, then finding the folder where your raw matrix and attribute files are saved, and then pressing open. I tend to name my session the same name as my session folder.

Type in the number of actors in the "Number of Actors" box.

Select your "Network file". This is your raw matrix file and should be in the session folder you just selected.

Select your Network Type. Generally you won't be fixing the maximum degree for each actor, so leave this blank.

Tick the "Structural Parameters" box, then press "Select Parameters".

In the new form that opens, click on the border in the bottom left hand corner and open the box so that you can see as many of the parameters as possible. This will make it easier to navigate.

Note: *In each of these "Select Parameter" forms, you will be given a range of options depending on the network type and whether the form lists structural parameters or attribute parameters. This 'PNet for Dummies' guide includes suggestions about which parameters will be useful for many standard estimations. However, it doesn't include a complete list of parameters, nor does it explain how to interpret each of these parameters. For a complete list of the definitions of all the parameters used in PNet, see Appendix B in the PNet User Manual. For explanations of the meaning and interpretation of these parameters the reader will need to consult specific papers which interpret these parameters, many of which are included on the MelNet website: <http://sna.unimelb.edu.au>*

Structural Parameters: Non-directed graphs

If you have a non-directed network you should have a screen that looks something like this:

The screenshot shows a window titled "Structural Parameter Selection" with three columns of parameters. Each parameter has a checkbox and a text input field. The "Markov Parameters" column includes Edge, 2-star, 3-star, Triangle, and Isolates. The "New Parameters" column includes p1, p2, p3, q10, q20, q30, d3, d41, d42, d6, and c. The "High-Order Parameters" column includes K-star, K-star-2nd, K-triangle, K-triangle-2nd, and K-2-path, each with a "lambda" input field set to 2. At the bottom, there are three buttons: "Clear All", "Select All", and "OK".

If it is a non-directed graph, then generally you will want to select the following structural parameters:

- Edge
- K-star
- K-triangle
- K-2-path

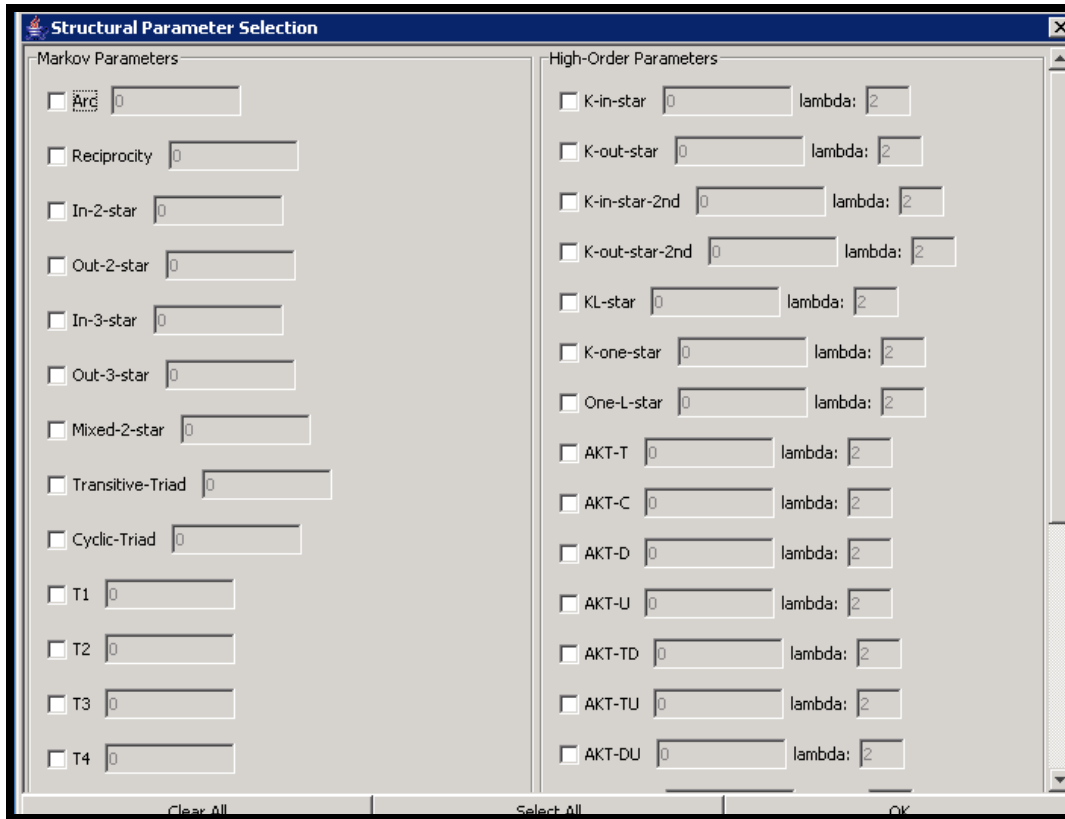
For the higher order parameters (the K- parameters) the default of lambda of 2 is fine.

If you have a significant number of isolates in your model, you may want to select the structural parameter:

- Isolates

Structural parameters: directed graphs

If you have a directed graph, then you should get a screen that looks like this:



If it is a directed graph, and it is of low density, then generally you will want to select the following structural parameters:

Markov:

- Arc
- Reciprocity
- Isolates (if you have a number of isolates in your dataset)

High-Order Parameters:

- K-in-star
- K-out-star
- AKT-T (this is an alternating K-triangle – transitive)
- A2p-T (alternating K-2-path-transitive)

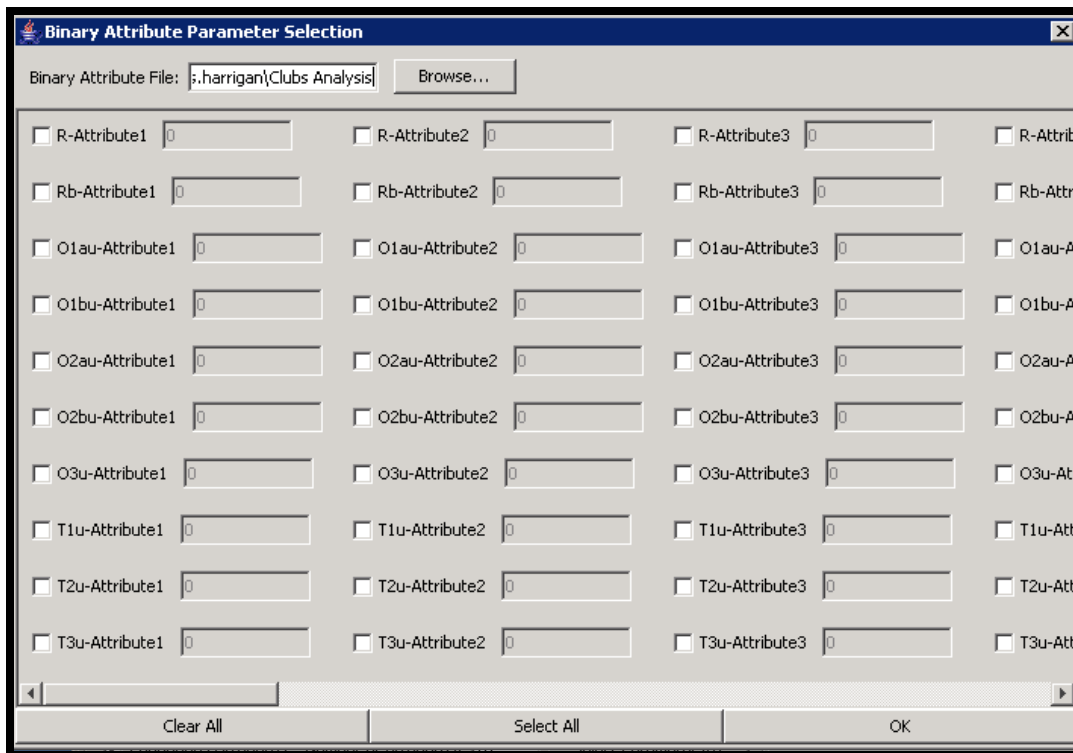
For more dense networks, you may want to begin increasing the complexity of your model by start adding in more higher order triangle effects, such as:

- AKT-D
- AKT-U
- AKT-C
- and so on.

For the moment we will assume you have only one network you want to model, so we will ignore the dyadic attributes parameter.

If you have attributes, select the box next to "Actor Attributes Parameter", and then select the boxes for the types of variables (binary, continuous, categorical) and then type the number of each type of variables you have into the designated box. After doing this, press "Select Parameters". [If this does not open, it is because you have not chosen a "Session Folder" (at the top of the page). Select a 'session folder' and then the "Select Parameters" button should work]

These "Select Parameters" forms vary depending on the type of network (directed or non-directed) and also the number of variables you have for your actors.



Press Browse and select your binary attribute file.

Non-directed binary attributes

For non-directed binary attributes, it is generally enough to select two parameters for each attribute:

- R
- Rb

Select the box next to each of these parameters for each variable, and press OK.

Directed binary attributes

For directed, binary attributes, it is generally sufficient to select the following three parameters for each attribute:

- Rb
- Rs
- Rr

Non-directed continuous attributes

For non-directed continuous attributes, it is generally enough to select the following parameters for each attribute:

- sum
- difference

Directed continuous attributes

For directed continuous attributes, it is generally enough to select the following parameters for each attribute:

- sender
- receiver

Of the "Estimation Options" the only ones that you will generally change are:

- Multiplication Factor
- Number of Runs

Select all the text after "****This graph contains:****" and before "*Graph Density:". Press copy.

Open a new Excel document, select the first column, and press paste. . A warning will appear, but press OK. The values should paste down the left hand column, with one line of text per row.

Now read through the records one by one. Delete the blank row that is between the structural parameters and the attribute parameters.

What you will notice is that some of the attribute parameters (and even perhaps some of the structural parameters) which you had selected to be in your model will be missing. So, in my example, in the image above, you can see that there is no "Rb for Attribute1", or for Attribute 7 or Attributes 11 or 12. This is because there are no instances of these configurations in my network.

If these configurations are left in your model, then the model will most likely be degenerate. Thus the challenge is to clearly identify the configurations that are not in your network, and remove them from your model.

I do this by creating this Excel file. I generally save the file with the name "Excluded Configurations [experiment name]"

Go through the file, and where you find one of your configurations that is missing, insert a row. After creating rows for each missing configuration, go back and type in the name of the missing configuration in the empty row and the value "0" next to it. In the column next to the missing configuration place a large "X". Select this row and press "Bold". Insert a row above the first row, and then in the first row of the second column type the heading "Excluded Configurations". Save this file. It should look something like this:

	A	B	C
1		Excluded Configurations	
2	* No. of vertices: 178		
3	* No. of edges: 114		
4	* No. of Isolates: 83.0		
5	* No. of K-star(2.00): 179.3		
6	* No. of K-Triangel(2.00): 58.5		
7	* No. of AltTwoPath(2.00): 251.0		
8	* No. of Rb for Attribute1: 0	X	
9	* No. of Rb for Attribute2: 4		
10	* No. of Rb for Attribute3: 44		
11	* No. of Rb for Attribute4: 94		
12	* No. of Rb for Attribute5: 99		
13	* No. of Rb for Attribute6: 14		
14	* No. of Rb for Attribute7: 0	X	
15	* No. of Rb for Attribute8: 11		
16	* No. of Rb for Attribute9: 3		
17	* No. of Rb for Attribute10: 47		
18	* No. of Rb for Attribute11: 0	X	
19	* No. of Rb for Attribute12: 0	X	
20	* No. of Rb for Attribute13: 15		
21	* No. of R for Attribute1: 32		

There are two reasons for taking such care with creating this Excluded Configurations file. Firstly, we may need to exclude further configurations if there are other problems with the remaining parameters (for example, the problem of 'separation'). Secondly, we may need to repeat the experiment, on another day, or after restarting PNet, and in this case, it is necessary to keep a log of which configurations have been excluded.

Go back to PNet.

Go through the structural and attribute parameters, and deselect the excluded configurations/parameters, as listed on in your Excel table.

You are now ready to run your first estimation.

3.3 Running an Estimation

Reset the number of sub-phases to 5.

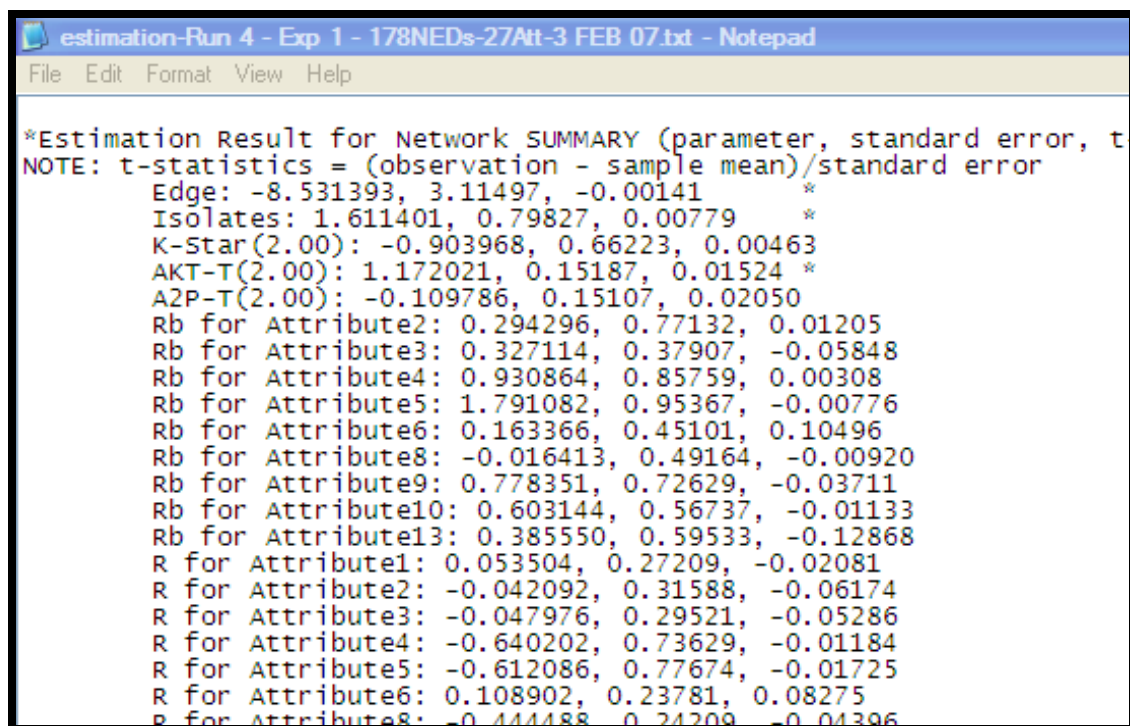
The other settings should be fine for a first run. Press Start.

After a period of time (most small networks will be almost instantaneous. For one of my networks with 250 nodes and 88 ties, a run with a MF of 10 takes about 30 seconds), a dialogue box will appear "Estimation Finished". Press OK.

The estimation file will open in Notepad. Scroll down to the very bottom of the file. The part you are interested is the list of parameter values next to configuration names, under the title:

*Estimation Result for Network SUMMARY

You should see something like this:



```
estimation-Run 4 - Exp 1 - 178NEDs-27Att-3 FEB 07.txt - Notepad
File Edit Format View Help

*Estimation Result for Network SUMMARY (parameter, standard error, t
NOTE: t-statistics = (observation - sample mean)/standard error
Edge: -8.531393, 3.11497, -0.00141 *
Isolates: 1.611401, 0.79827, 0.00779 *
K-star(2.00): -0.903968, 0.66223, 0.00463
AKT-T(2.00): 1.172021, 0.15187, 0.01524 *
A2P-T(2.00): -0.109786, 0.15107, 0.02050
Rb for Attribute2: 0.294296, 0.77132, 0.01205
Rb for Attribute3: 0.327114, 0.37907, -0.05848
Rb for Attribute4: 0.930864, 0.85759, 0.00308
Rb for Attribute5: 1.791082, 0.95367, -0.00776
Rb for Attribute6: 0.163366, 0.45101, 0.10496
Rb for Attribute8: -0.016413, 0.49164, -0.00920
Rb for Attribute9: 0.778351, 0.72629, -0.03711
Rb for Attribute10: 0.603144, 0.56737, -0.01133
Rb for Attribute13: 0.385550, 0.59533, -0.12868
R for Attribute1: 0.053504, 0.27209, -0.02081
R for Attribute2: -0.042092, 0.31588, -0.06174
R for Attribute3: -0.047976, 0.29521, -0.05286
R for Attribute4: -0.640202, 0.73629, -0.01184
R for Attribute5: -0.612086, 0.77674, -0.01725
R for Attribute6: 0.108902, 0.23781, 0.08275
R for Attribute8: -0.444188, 0.24209, -0.04396
```

After the title information you have a list of configurations and three values next to each one. The first value is the **parameter estimate**, the second is the **standard error of that parameter estimate**, the third is the **t-statistic**, which compares the observed number of this configuration in your graph with the mean number of configurations in a sample of 500 graphs generated with these parameter estimates. On some rows you will notice an **asterisk (*)**. This means that the parameter estimate is at least 1.96 (essentially 2) standard errors away from zero, and is thus indicates that there is a 95% or more chance that the parameter estimate is statistically significant (commonly expressed as $p < 0.05$).

3.4 Fitting an Estimation

What does a fitted estimation look like?

It is extremely *unlikely* that your first estimation run will be an adequate fit to your observed data. An adequate fit is generally defined as:

1. The parameter estimates and standard errors are within the bounds of a reasonable model.
2. The t-statistic for all the configurations in the model are less than 0.1

1. Reasonable parameter estimates and standard errors

The first of these conditions is not a scientific test, but nonetheless it is important. Below is a list of the main problems that give 'unreasonable' parameter estimates:

High values for all estimates, SEs and t-statistics: If you get high values for all, or almost all, of your parameter estimates, standard errors and t-statistics, this is generally because your model has 'wandered into parameter space wilderness' and can not get itself back. This is not a major problem. The best thing to do is to return all the parameter estimates to zero (that is, if you have updated them and they are not already zero), and run the model again. You may have many first runs (up to perhaps 5 or 10?) that end up in the wilderness, but this is not generally a serious problem.

Rare configurations: In the section on *preventing model degeneracy*, we removed all of the configurations which were not found in the observed graph. However, we left all other configurations in the model. Some of those configurations will have had values of just 1 or 2, and in a large graph, these values may be very difficult to statistically analyse, and give unrealistically large parameter values. In this case, it is best to remove these configurations from your model (and to update your "Excluded Configurations" table accordingly)

Separation: Another problem which can occur is *complete separation* or *quasicomplete separation*, where the independent variables, in this case, one of the configurations, completely predicts the dependent variable, in this case the formation of a tie. When this occurs you will get large or very large parameter estimates with high standard errors. How you treat these will depend on your dataset, but in general it may be better to take them out, or to find a better way of specifying this attribute in your model. Make sure you mark this configuration on your "Excluded Configurations" table.

Reference category for a set of dummy variables: A further problem which can occur is when you put in a set of dummy variables which are actually different values for a categorical variable (for example, variables that represent each state in a country). In this case, you must remove one of the variables, which will then be used as a base or reference category. For those who have done logistic regressions before, this will be a familiar concept. Make sure you mark this configuration on your "Excluded Configurations" table.

Interactions: Occasionally you can get strange interactions between different configurations in your model. A classic sign of this is when you have two variables with large, equal but opposite values (eg. 12.1275... and minus-12.1275...). In this type of situation, it is generally the case that the two variables are interacting strongly. One solution is to look at the observed value of these configurations in your graph, and the theoretical meaning of the two configurations, and then choose one of the configurations to drop out of your model. Make sure you mark this configuration on your "Excluded Configurations" table.

Other problems: There are a number of other problems which will occur when attempting to specify a ERGM model. If you are having trouble fitting a model, or it is not having the expected results, look over the results of the estimations and look for patterns in the movements of the parameter values. Try to see which variables are moving together, and see if these movements of the parameter values make sense, or if instead they might be errors created by a poorly specified model, or some form of interaction within the model itself. If you find something like this, choose the theoretically least important configuration and drop it out of your model, and run it again. Make sure you mark this configuration on your "Excluded Configurations" table.

2. Reducing the t-statistics to less than 0.1

Once you have got reasonable numbers for your parameter estimates and standard errors, the major challenge is to reduce the t-statistic for each configuration to less than 0.1. The general process for this is:

1. Run an estimate
2. If this is:
 - **one of your first runs, before you first press 'update', then**, if the estimate has t-statistics lower than 2 for most values, and lower than 4 for all values (for your first run) then press "update".
 - **after your first update, then**, if the estimate has better t-statistics (that is, lower t-statistics, on average) than your current parameter estimates, then press "update".
3. Repeat.

For small datasets (nodes < ~40) with few configurations then the default settings should be adequate.

For models with more configurations and a larger numbers of nodes, the best way to get convergence is to slowly increase the multiplication factor.

The multiplication factor reflects the amount which PNet is able to experiment with different values for your parameter values. When PNet runs and estimation, it spends most of its time 'walking' around the parameter space – meaning that it is slowly changing the parameter values for each of the different configurations in your dataset, and measuring whether these new estimates fit your model better or worse. When you increase the multiplication factor, you give PNet more time to explore different values for

your parameter values, and thus more time to make a more accurate parameter estimate.

The problem with increasing the multiplication factor (MF) is that it increases the processing time. At the moment, we think that the processing time increases linearly with the increase in the multiplication factor (so if a MF of 10 take 30 seconds, a MF of 20 will take 1minute, and a MF of 200 will take 20minutes).

So far, we have experimented with MF of up to 1000, and have noticed dramatically improved results with higher MFs.

Our suggestion is, however, that you only increase the MF slowly, from one estimate to the next. We generally start with the default setting of 10, and if that gives an estimate that has no t-statistics more than 2, then we increase the MF to 20. If the estimates keep getting better, we continually update our parameter values, and approximately double the MF for each estimation run. Generally 200-600 is high enough to fit most models.

Sometimes, however, you may have difficulty getting some of the t-statistics to go below 0.1, even with a MF of 500 or 1000.

In this case, an option is to get a model that almost fits (most t-stats below 0.1, and 1-5 below ~ 0.3), and then set the number of runs to, say 20, and then leave the computer to run over night, or for several days. While this is running, you can check the estimation file at any time to see the results of the estimations that PNet has run. Checking the estimation file will not interfere with the running of PNet.

Step 4: Goodness of Fit

4.1 Running a Goodness of Fit

Once you have estimated a model which has t-statistics for all configurations of less than 0.1, then you are ready to check the goodness of fit of the model.

To do this you press the "Goodness of Fit" tab in PNet. You then need to re-enter all the data which you have entered for the Estimation, including the exact parameter values obtained in your fitted estimate.

The main difference with running a goodness of fit and an estimation, is that when you "Select Parameters", you not only select the configurations which you have values for in your estimation procedure, but **you also need to select the configurations which you want to use to check the goodness of fit for your model**. In practical terms this generally means pressing "Select All" when you are in the various "Select Parameters" windows. You then enter the values for the parameters you have estimates for, and leave the remained at zero. The exception to this is when you are selecting structural parameters for larger networks ($n > 40$). In this case you generally want to leave out all or most of the parameters labelled "New Parameters", since they take an incredibly long time to calculate, and the goodness of fit may take days or weeks or months to finish (especially the new parameter "C" for 5-cliques!).

For smaller networks ($n < 40$) the default settings for the "Simulations Options" are fine. For larger networks you will need to adjust the "Number of Iterations". For a network of 250 nodes with 88 ties, instead of the normal 1,000,000 iterations we ran 88,000,000 iterations (and left the number of samples to pick up as 1000).

4.2 Interpreting GOF statistics

When PNet has finished the GOF a dialogue box will open stating this. Press OK. the GOF statistics will open in Notepad.

Scroll down the document until you get to the following lines:

```
* There are 10000000 proposed digraphs.  
* Statistic samples are picked up at 1 per 10000 digraphs.  
* Accepted 27932 proposed digraphs.  
  
observation, sample mean (standard error), t-statistic  
t-statistics = (observation - sample mean)/standard deviation
```

Under these lines are the main GOF statistics you will be interested in. The most important statistic is the t-statistic for each configuration. The rules for a good model are these:

1. If the parameter was estimated and specified in your model (that is, if you had to type in the parameter estimate when you were setting up the GOF) then the t-statistic needs to be below 0.1 (as it was in the estimation).
2. If the parameter was not estimated and specified in your model (that is, if you **didn't** type the parameter estimate when you were setting up the

GOF, but rather just selected the configuration and left the parameter value as zero) then the t-statistic should be below 2 for the model to not be a bad fit.

If your model doesn't fit these criteria then it's not a perfect fit. There should be no reason why if the first criteria shouldn't be met if the estimation was run correctly (since the estimation should have made sure that the t-statistic is below 0.1 for those parameters). However, for the parameters that were not-estimated, it is highly likely that some of the t-statistics will be greater than 2. Given that it is highly likely that one or more of these t-statistics will be greater than 2 (because no ERGM model is perfect!) then the decision as to whether to accept the model or to attempt to find another model is one which is up to the individual researcher and their interpretation of the nature of the poorly fit configurations in your model.

```
gof-Exp11Top250ExsNEx_9ATT.txt - Notepad
File Edit Format View Help
# Receiver of Continuous Attr2: 0.01481
# Receiver of Continuous Attr3: 0.18761
* There are 1000000 proposed digraphs.
* Statistic samples are picked up at 1 per 10000 digraphs.
* Accepted 27932 proposed digraphs.

observation, sample mean (standard error), t-statistic
t-statistics = (observation - sample mean)/standard deviation

# Arc: 88.0000 Mean= 86.9550 ( 12.6092 ) t = 0.0829
# Reciprocity: 4.0000 Mean= 3.7170 ( 1.9659 ) t = 0.1440
# 2-In-Star: 44.0000 Mean= 41.9170 ( 14.4555 ) t = 0.1441
# 2-Out-Star: 34.0000 Mean= 31.3210 ( 10.1916 ) t = 0.2629
# 3-In-Star: 18.0000 Mean= 18.4280 ( 13.6730 ) t = -0.0313
# 3-Out-Star: 8.0000 Mean= 9.7030 ( 8.1168 ) t = -0.2098
# Mixed-2-Star: 62.0000 Mean= 66.2060 ( 20.2322 ) t = -0.2079
# T1: 0.0000 Mean= 0.0110 ( 0.1044 ) t = -0.1054
# T2: 0.0000 Mean= 0.1170 ( 0.6766 ) t = -0.1729
# T3: 0.0000 Mean= 0.2440 ( 0.8539 ) t = -0.2857
# T4: 0.0000 Mean= 0.2080 ( 0.5666 ) t = -0.3671
# T5: 0.0000 Mean= 0.1780 ( 0.5163 ) t = -0.3447
# T6: 1.0000 Mean= 0.5640 ( 1.0015 ) t = 0.4354
# T7: 4.0000 Mean= 9.1370 ( 6.8828 ) t = -0.7463
# T8: 7.0000 Mean= 7.0340 ( 5.4117 ) t = -0.0063
# T9(O30T): 2.0000 Mean= 1.2920 ( 1.7555 ) t = 0.4033
# T10(O30C): 0.0000 Mean= 0.2350 ( 0.5235 ) t = -0.4489
```

Appendices

These appendices are designed to act as a quick reference guides, compiling the most useful information in easy to read tables and check lists.

Appendix 1: Interpreting GOF statistics

1. If the parameter was estimated: t-statistic needs to be below 0.1
2. If the parameter was not estimated: t-statistic of less than 1 is a good fit, more than 2 is a bad fit.

Appendix 2: Recommended Starting Configurations

Structural Configurations	Attribute Configurations
<p>Non-directed graph</p> <ul style="list-style-type: none"> • Edge • K-star • K-triangle • K-2-path • Isolates 	<p>Non-directed graph</p> <p>Binary attributes</p> <ul style="list-style-type: none"> • R • Rb <p>Continuous attributes</p> <ul style="list-style-type: none"> • sum • difference
<p>Directed graph</p> <p>Low density graph:</p> <ul style="list-style-type: none"> • Arc • Reciprocity • Isolates • K-in-star • K-out-star • AKT-T • A2p-T <p>Higher density graph also include:</p> <ul style="list-style-type: none"> • AKT-D • AKT-U • AKT-C • and so on. 	<p>Directed graph</p> <p>Binary attributes</p> <ul style="list-style-type: none"> • Rb • Rs • Rr <p>Continuous attributes</p> <ul style="list-style-type: none"> • sender • receiver

Appendix 3: Running an estimation (Summary)

1) Preparing matrix and attribute files.

Attribute Data

- Create Excel file with attributes
- Use Excel formulas to create VNA file

Matrix Data

- Make VNA file of matrix data
- Paste into attribute data VNA file

Creating Raw Matrix and Attribute files

- Import VNA file
- Export as RAW Matrix and EXCEL Attribute
- Cut and paste Binary, Continuous, etc attributes into own RAW files.
- put files into their own folder and make a copy called "originals..."

2) Make an excluded configurations file (to prevent degeneracy):

- Select parameters and set up estimation
- Set "Subphases" to zero. Press "Start".
 - Open the file labelled "start-statistics-[your session name]".
 - Select all the text after "****This graph contains:****" and before "**Graph Density:".
 - Copy into new Excel document,
 - Delete the blank row
 - Save as "Excluded Configurations [experiment name]"
 - Insert a row where configuration missing and type name of missing configuration.
 - In PNet: deselect the excluded configurations

3) Run an Estimation

- Reset the number of sub-phases to 5. Press Start.
- If has t-statistics lower than 2 for most values, and lower than 4 for all values AND it is better that your current parameter estimates, then press "update".
- Repeat, while:
 - doubling the MF for each estimation run.
 - Generally 200-600 is high enough to fit most models.
 - If it is not fitting with runs of 500 or 1000, but is getting close (say all numbers less that 0.2-0.5, then set the number of runs to, say 20, and then leave the computer to run over night.

4) Causes of unreasonable parameter estimates and standard errors

- High values for all estimates, SEs and t-statistics
- Rare configurations
- Separation
- Reference category for a set of dummy variables
- Interactions
- Other problems

5) Goodness of Fit

- re-enter all the data in the "Goodness of Fit" tab in PNet

For large number of nodes:

- select all other parameters, except "new parameters"

For low density networks:

- adjust number of iterations to approx number of edges x 1,000,000

To download the latest copy of this manual go to:
<http://www.sna.unimelb.edu.au/pnet/pnet.html#download>

Any suggestions, feedback, comments or corrections would be greatly appreciated, and can be emailed to nick_harrigan@yahoo.com or nick.harrigan@anu.edu.au