Exercise

Estimating Bernoulli and Markov ERGMs

Swinburne University of Technology, Australia ©

To estimate parameters, select the **Estimation** tab in MPNet. Make sure that the datafile (in this case, The Corporation data – comm_undirected.txt) is in the session folder, enter the number of actors (38) and the name of the Network datafile (using Browse...). Select structural parameters.

Network data: **comm_undirected.txt**

This is in **Data > The Corporation** folder.
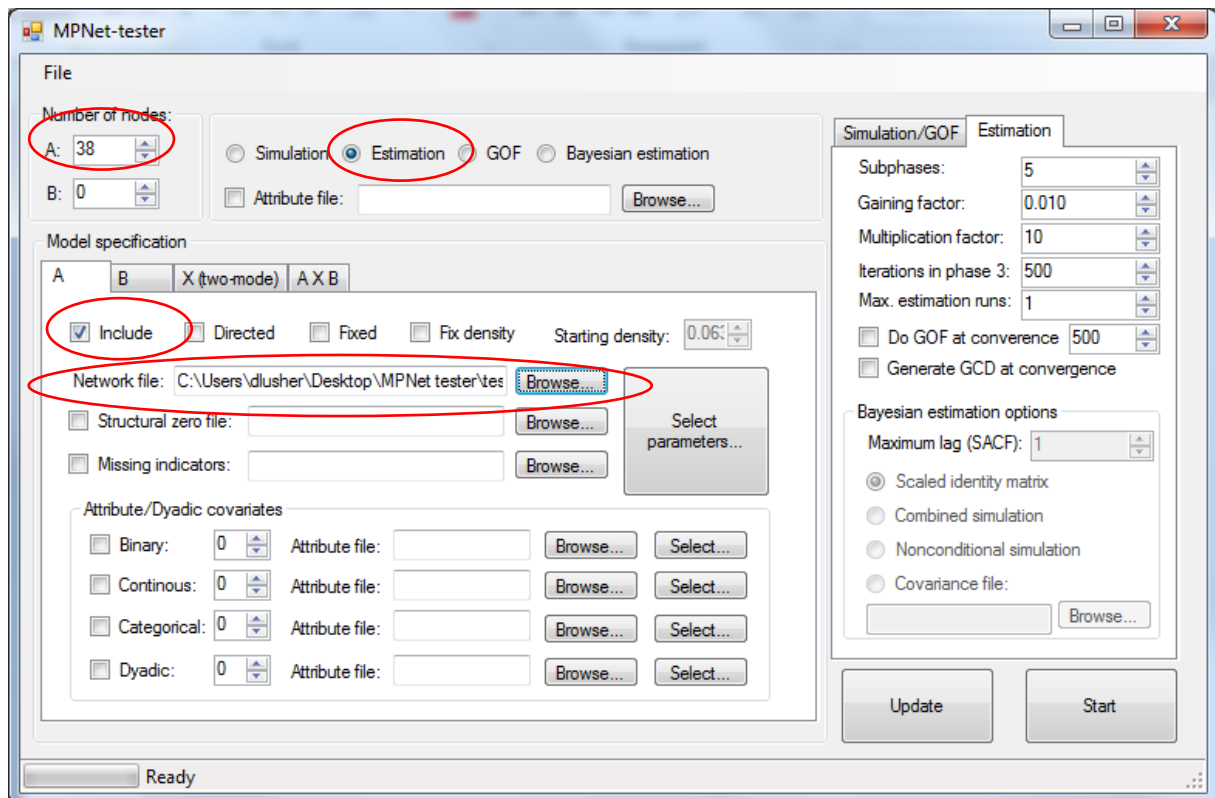
11.1 Estimating a Bernoulli model

The only parameter that is to be estimated in a Bernoulli model is the Edge parameter. When we *simulated* a Bernoulli models in the previous exercies, we selected star and triangle parameters under the simulation tab but kept their values at 0, so that MPNet would count the numbers of triangles and 2-stars. When *estimating* here, however, we ONLY select the Edge parameter. If we selected the star and triangle parameters, they would be estimated as well and it would no longer be a Bernoulli model (it would be a Markov model – later.)

Go back to the main window and click Start! to commence estimation.
At the bottom of the output, you will have something like:

Effects	Lambda	Parameter	Stderr	t-ratio	SACF
EdgeA	2.0000	-2.6935	0.152	-0.009	0.278 *

<u>Effects</u>: *The parameter name*

<u>Lambda:</u> *More on this later.*

<u>Parameter:</u> *The parameter estimate*

<u>Stderr (Estimated standard error):</u> *Each parameter estimate has a standard error (SE) which is a measure of the precision or how certain we are of the parameter estimate. A small SE indicates greater precision and certainty, while a large SE indicates less certainty.*

<u>t-ratio (the convergence statistic)</u>: *a measure of how stable the parameter estimates are (NOT whether a parameter is significant – see below). For a quick convergence check, you want the convergence t-statistic to be less than 0.1 (in absolute value).*

> *If the convergence statistic is not less than 0.1, you can repeat the estimation but starting this time from the last finishing estimate. Go back to the MPNet window, click **Update** and check what has happened to the starting parameter value by clicking on **Select parameters...** Then click **Start** again. The new estimate will again appear at the bottom of the output file.*

<u>SACF (*sample autocorrelation function*)</u>: *More on this later*

<u>Asterisk</u>: *indicates a significant effect (Absolute value of the estimate is more than twice the standard error.)*

11.2 Estimating a Markov model

We already know that the Bernoulli model is not good for The Corporation data because it underestimates the number of observed triangles in the data.

Now we will estimate a Markov model with 2-star, 3-star and triangle parameters.

NB: Markov models are not usually successful models either, except perhaps for small datasets like this one. It is presented here only for illustrative purposes.

Select parameters again, and now select Edge, 2-star, 3-star and Triangle parameters. Run the estimation. Check at the bottom of the output. **If any convergence statistic is not less than 0.1**, run the estimation again - after **Updating!** - until you have convergence FOR ALL FOUR PARAMETERS. The **Update!** process sets the previous non-converged estimates as the starting point for a new estimation run.

If you have trouble getting a converged model, then try increasing the **Multiplication Factor** (under the Estimation Tab to the right of the MPNet window) to 30.

You should have results like:

Effects	Lambda	Parameter	Stderr	t-ratio	SACF	
EdgeA	2.0000	-3.1643	0.737	-0.027	0.249	*
Star2A	2.0000	0.1747	0.281	-0.033	0.440	
Star3A	2.0000	-0.1895	0.132	-0.051	0.579	
TriangleA	2.0000	1.8762	0.249	-0.034	0.696	*

Interpretation: For these results we have significant effects for edge and triangle parameters, but not the 2-star or 3-star parameters.
- The <u>negative and significant edge</u> 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 <u>positive and significant triangle</u> effect indicates that we see more triangles than we would expect to see by chance, given the other effects in the model.

Social selection models and Dyadic Covariates

15.1 Social selection models

To accompany the communication network, *communication.txt*, there are three attribute datafiles:

- **office.txt**
  - This is a categorical variable that places the managers in three divisions of the organization
- **seniority.txt**
  - This is a binary variable that describes whether the managers are junior (0) or senior (1).
- **projects.txt**
  - The third is a continuous variable that describes the number of projects in which the managers have worked and hence is a measure of experience.

For MPNet, these three different types of variables need to be in separate files but there can be multiple variables of the one type in each file.
- For instance, there could be a second continuous variable in a second column separated by a tab, with variable names at the top, separated by tabs.

You can estimate models with a variety of attribute effects with the different types of attribute variables (binary, continuous, categorical). The common starting model

specifications for directed networks would include **sender** and **receiver** effects, and **interaction/homophily**. Homophily can be implemented in various ways, as will be seen below.

Select Actor Attribute parameters, and indicate the numbers of attributes in each of the attribute files (one of each).

Select parameters for each of the types of attributes. Specify the file that contains the attribute variables.

- For the **binary attribute** (**seniority**), select *sender*, *receiver* and "*interaction*", a homophily variable that in this case will be 1 if both actors are senior (i.e. both score 1).

- For the **continuous attribute** (**projects**), select *sender*, *receiver* and *difference* effects. For continuous variables, a popular way to implement homophily is the absolute difference between the attribute values of the sender and receiver of the tie. In that case, a **negative** parameter estimate indicates the presence of homophily.

- For the **categorical attribute** (**office**), select the *Matching-attribute*. This is a homophily effect that comes into play when both sender and receiver are in the same category.

  - **NB!!!!!!** For categorical variables, **Do Not** select effects for **both** <u>Matching</u> and <u>Mismatch</u> *in the same model for the same attribute*, as one is the complement of the other (and choosing both will result in the model producing degenerate results due to 100% collinearity).

The output will then include both structural and attribute parameters in the model. So, now you will:

1. Do an estimation including the actor attributes. If you have not closed MPNet from before, the structural parameters will still be there so you will not need to select them again.

Interpretation of results

From the **estimation** you should see a number of significant effects. There is a homophily effect for seniority, and also for projects (remember, a negative "difference" effect means there is little difference, hence homophily). There is also a negative sender effect for seniority, indicating that executives who are not senior (binary variable = 0) are more likely to send ties.

You can also use a dyadic covariate measure as a predictor of network ties. An example of a dyadic covariate is another network (e.g., trust, advice, friendship) or a network of geographic distances between all pairs of people in the network.

For this example, the dataset *advice.txt* is an advice network for the 38 managers. To use this as an exogenous predictor of communication ties, in addition to the existing structural and actor attribute effects, select **Dyadic Attributes**, and then when Selecting parameters, enter the dyadic covariate file and select Covariate-Arc.

Network Data: **communication.txt**

Attribute Data: **seniority.txt**
                         projects.txt
                         office.txt

Covariate Network: **advice.txt**
NB: Currently in MPNet, dyadic covariates matrices need values to be separated by Tabs. Covariate matrices can contain values (e.g. strength of ties) other than binary.

Run the estimation. In this final model, you should see that various structural, actor attribute and dyadic covariate effects all play a role in explaining the structure of the communication network.

There is a very strong and significant effect for the covariate network "advice" which indicates that communication links are very likely associated with advice.