

Sunbelt 2015 statnetWeb Workshop – Guided Network Analysis

Load the faux.mesa.high network and complete the following exercises with your group, discussing and recording your answers to the questions.

Network:

- How many nodes are in the network? How many edges?
- What relationship do edges in the network represent?
- Are the edges in the network valued or binary?
- What attributes are associated with each vertex?

Descriptive analysis:

- Play with color coding and sizing the nodes in the network plot.
- Do you see evidence of homophily by Grade?
- Look at the mixing matrix for Grade, now what would you say about homophily by Grade?
 - How would you quantify the level of homophily?
 - How would you explain that to a student?
- Is there homophily by Race or Sex?
- What percent of nodes in the network are isolates?
- How does that compare to what we would expect from randomly generated networks?
 - What is the difference between the “CUG” and “BRG” null models?
- How does the number of concurrent edges in faux.mesa.high compare to what we would expect by random?
- Which node has the highest betweenness score?

Statistical analysis:

- Fit a model with “edges” as the only ergm term.
 - This is a Bernoulli model. What does this model assume?
 - What is the log-odds of any edge existing under this model?
 - Save this model.
- To add homophily to our model we need a meaningful measurement of the level of homophily in faux.mesa.high
 - Read the documentation for the terms “nodecov”, “nodefactor” and “nodematch”.
 - Which term should we use to add homophily by Grade to our model? How can we include the level of homophily within each grade level?
 - Add your chosen term to the formula and fit this model.
 - Save this model.
- Look at the goodness-of-fit plots for the two models you have saved. In what area do we need to improve our fit to the observed network?
 - Choose a new term to add to the model.
 - [Table of common ergm terms](#)
 - [Cross-reference tables for ergm terms](#)

- Fit and save your model.
- Check the goodness-of-fit plots for this model. Based on the plot for edge-wise shared partners, perhaps we should try to fit the levels of clustering rather than the degree distribution.
- The count of the number of triangles in a network is often used as a measure of clustering. Reset the ergm formula. Try to capture the level of clustering in the network by fitting an “edges + triangle” model.
 - What happens? Raise your hand if you’ve never seen this before or want to discuss what is going on.
 - Look for another term that represents triangle formation and add this to the model.
 - Add homophily by Grade again and save this model. (You should have at least 4 models saved now).
- On the goodness-of-fit page, compare all of your saved models under the default GOF terms.
- Explore more terms that might be useful in this model.
- With one of your saved models, run 50 simulations and check that the simulation statistics hover around the target statistics.