Overview
The project will be to undertake an analysis of a social network that you find interesting. You can select any network dataset you like, but preferable related to your graduate work or thesis area.

I do not want a quick and routine analysis; a good project will show understanding of the problem and possible solutions and techniques to consider. The technical conclusions should be stated clearly.

Your solution to the problem should be submitted in the form of a data analysis report. The report should be typed and at most 20 pages of 12 pt or greater type, including any appendices, tables, and figures. Please include margins of at least one inch on all sides of the paper. It should be submitted on the class website under the “Homework” section.

The report should be self-contained and suitable for a non-statistician with a M.S. level of knowledge of statistics.

Advanced network modeling methods should be defined briefly in the text to the extent necessary for understanding of the results, along with a reference. It is recommended that the following outline be followed in preparing the report:

i. **Abstract.** This should consist of a brief statement of the results of your network analysis. This should be like that of a research paper analysing the network.

ii. **Introduction.** Here include a clear statement of the scientific questions addressed by the network data. The goal of the statistical analysis and the scientific context of the problem should be clear to all who read the introduction.

iii. **Analysis and Results.** Describe your analysis and its results clearly and concisely. If necessary, use graphical displays and tables to convey the results. Describe methods used, approaches taken to examine the underlying assumptions and so on. Explain why the methodology is appropriate. Avoid using highly technical language.

iv. **Discussion.** This section should describe the scientific and statistical issues raised by the results described in the previous section. Limitations of the study and of the analysis should be discussed here. There may be scientific issues that you would have liked to discuss with the investigator if this had been a real collaboration with a non-statistical scientist. If so, describe these issues and why they would be relevant to the analysis and/or interpretation of results. If appropriate, provide suggestions for further analysis or collection of additional network data. Summarize your conclusions about the issues of scientific concern.
v. **References.** List all books and articles you consulted that are reflected in your report, as well as references that might be useful to the reader if they want to know more.

vi. **Tables and Figures.** These can be included in the text or in a separate section at the end. All should have clear titles/captions, and figures should have explanatory legends.

vii. **Appendices.** There can be one or more appendices. You may include more technical discussion of your methodology or any theory developed to implement your models. Appendices are not required (in fact they are discouraged), and should only be included only if you feel they add something important to your report.

You should spend the majority of your time thinking about the scientific and statistical issues and writing the report rather than spending all your time carrying out the network analysis. You may do a wonderful job of analysis, but it is of no use unless you can communicate the results to your audience.

**Evaluation**

Criteria used to judge performance will include the following three factors, each given equal weight:

i. **Statistical Appropriateness.** Appropriateness of the analyses and models for the network data and questions. Technical execution of the analysis.

ii. **Scientific appropriateness.** Thoughtfulness and simplicity of your analysis. Does your analysis really answer the scientific questions of interest?

iii. **Quality of the written report.** The report will be judged based upon its organization, clarity, and accuracy. Simple, concise sentences are preferred to sentences which are convoluted or otherwise confusing.

**Choice of your data set**

The data set should contain at least 20 nodes, and at least two variable measured for each node. Do not merely use data from a textbook – the world is an interesting place! All data sources must be cited, and described.

A good place to start is the taxonomy in Homework 1. What sort of network data do you have?

Your statistical task is to model the structure in this network and describe it. To do this you should consider the various forms of social structure considered in the course.

**Suggestions for your Analysis**

Here are some suggestions:

- Start by presenting visual summaries of the networks, followed by numerical summary measures.
• Describe the impact of the covariates on the patterns of ties.
• Describe any sub-group cohesion you see in the networks.
• Describe the pattern of centralization of the actors.
• Fit latent position and latent position cluster models to the network. Is there evidence of clustering?
• If you have directed data, fit a triad census model to the networks. Is there evidence of balance in the patterns of ties?
• You can fit any ERGM you like to represent the structure.

This is an open ended question, so feel free to experiment.