Final Project:

In this project, you will be analyzing real network data. The goal is for you to gain experience working with network data, by finding something interesting about that data.There are two options for obtaining data, which come with different expectations:

1) Existing network data sources: My website, contains a range of pre-existing data sources. Most are in .net or .gml format, but some are in edge lists or adjacency matrices.
2) Data collection: you might choose to collect your own data using one of the data collection techniques we've discussed in class: surveys, observation, or passive collection from archival/electronic sources. There are also some data sources listed on my website that would require some processing (for example: the Enron data set)

If you choose to use existing data, my expectations for your analysis will be *much* higher. Data collection requires considerable effort, and so I anticipate that most of your work will go into that part of the process. The analysis you do on that data can be as simple as a run down and interpretation of the relevant measures.
The relevant url on my website: www.andrew.cmu.edu/user/andersok/Katharine_A._Anderson/Social_Economic_and_Information_Networks.html
You should select a data set and question that interests you. Groups for this project can be 2-4 people.

Option 1: Existing Data

Part 1: Project Proposal (1 page, 20 points)
Your proposal can be in any format you wish. Somewhere in it, you should include the following:

- The data source(s) that you will be using
- A visualization of the network(s)
- A rundown of some basic measures in that network that you found interesting using whatever format you find most compelling for each (in-text description, table, network visualization)
- A question or questions that you are going to try to answer using this data (this need not be the question you end up answering, but you need a place to start!)
- A plan of attack: what are you going to measure? what comparisons are you going to make? what would the results of those measures tell you?

Part 2: Final Project Report (3-6 pages, but as short as possible) 40 points
Your final report can be in any format you wish. Somewhere in it, you should include the following:

- The question or questions that you set out to answer.
- A description of why you think those questions or questions are interesting or important.
- The data source(s) that you ended up using to address that question (these may be different from the sources you originally thought you would use)
- A visualization of the network(s)
- An analysis of the network: What did you measure in an attempt to answer your question? What did you expect to find? What did you actually find? What does that mean in the context of your question? What are the larger implications of your results?

Part 3: Final Project Presentation (5-10 slides, 20 points)
This is intended to communicate your results to the rest of the class. When you're putting together your presentation, you should consider the following:

- Give your audience a clear statement of the problem you're interested in.
- Give your audience a reason to care about that question. Why should they be interested in hearing your results?
- Make your results easy to understand. Include only the most relevant results, and present those results in the clearest way possible. Often pictures are more compelling than words.
- Consider the visual appearance of your presentation. This is not the most important part of presenting work, but a visually polished slide will encourage your audience to pay closer attention.

Option 2: Data Collection

Part 1: Project Proposal (1 page, 20 points)
Your proposal can be in any format you wish. Somewhere in it, you should include the following:

- The data you plan to collect
- A question or questions that you think could be answered using this data (why are you bothering to collect it)
- How you plan to collect it. If you will be collecting data via survey, this should include a copy of the relevant survey questions. If you are scraping data, provide a url and a plan for obtaining the data from that site.

Part 2: Final Project Report (3-6 pages, but as short as possible) 40 points
Your final report can be in any format you wish. Somewhere in it, you should include the following:

- The question or questions that you set out to answer, and a description of why you think those questions or questions are interesting or important.
- The data you collected that you think you could use to address that question (note that if you've collected the data, you need not have addressed that question)
- A description of how you collected that data, including any pitfalls you discovered (mistakes are part of this process!) What worked well? What would you do differently next time? Include copies of any materials you used in data collection
- A visualization of the network(s)
- A ***basic*** analysis of the network: Report the relevant measures for the network, and interpret them in this particular context.

Part 3: Final Project Presentation (5-10 slides, 20 points)
This is intended to communicate your results to the rest of the class. When you're putting together your presentation, you should consider the following:

- Give your audience a clear statement of the problem you're interested in.
- Give your audience a reason to care about that question. Why should they be interested in hearing your results?
- Make your results easy to understand. Include only the most relevant results, and present those results in the clearest way possible. Often pictures are more compelling than words.
- Consider the visual appearance of your presentation. This is not the most important part of presenting work, but a visually polished slide will encourage your audience to pay closer attention.