COMM 645, LABS 6 & 7: DUE OCTOBER 24, 2012

Please e-mail your lab to comm645@ognyanova.net before 2pm on Wednesday.

This assignment combines labs 6 and 7. It tests your understanding of CUG and QAP tests (Lab 6), as well as your ability to build, estimate, and interpret exponential random graph models (Lab 7).

For this exercise, we will use network data describing different communication connections in a team of 17 professionals involved in the design of military installations.

The data is stored in the following three files:

- **CRIeq.csv** – A 17x17 directed, binary information **retrieval** network. An edge **i -> j** in this network indicates that team member **i** reports going to team member **j** for information about environmental quality (EQ).

- **CAIeq.csv** – A 17x17 directed, binary information **allocation** network. An edge **i -> j** in this network indicates that team member **i** supplies team member **j** with information or documentation about environmental quality (EQ).

- **EXeq.csv** – A continuous attribute containing each team member's level of expertise in the area of environmental quality (EQ). A person's level of expertise is calculated as the average of their perceived expertise as reported by all other team members.

Your assignment consists of the following:

1) Read the three data files into R. Create a network object for each of the two networks. Add the expertise vector as a vertex attribute to both of the network objects.
2) [Lab 6] Write up, translate in network terms, and test 3 hypotheses based on the following suggestions:
    a. We expect that there will be a small number of key members that the team tends to go to for EQ information retrieval.
    b. We expect that there will be a small number of key members that the team tends to go to for EQ information allocation.
    c. If a team member reports retrieving EQ information from a colleague, we also expect that he/she will allocate EQ information to that colleague.

    Hint (2a,2b): You're supposed to compare observed centralization scores for the two networks to those of graphs with the same size and density drawn from a uniform distribution. Think about the appropriate type of centralization you want to look at here.

3) [Lab 7] Formulate network hypotheses based on the following suggestions. Build and estimate an exponential random graph model based on the information retrieval network to test them.

    a. Contacting people to retrieve information takes time and resources. Team members will not establish this type of connection indiscriminately.
    b. Team member A will feel more comfortable asking team member B for information if B also tends to retrieve information from A.
    c. If team member A retrieves information from B, and B retrieves information from C, then A will likely want to establish a direct information retrieval link to C.
    d. Team members will be more likely to retrieve information from others if they also have an information allocation relationship with them.
    (Remember, information allocation is a separate type of link, stored in another network)
    e. Team members will retrieve information from colleagues perceived as having a high level of expertise.
    f. Team members with high levels of expertise will likely have more information retrieval connections to others (that is probably how they got to be experts in the first place!)

    In your results section, report which hypotheses were supported and which were not. Make sure to also report the goodness of fit for your model with regard to in-degree and out-degree distributions.

Hints:
- In addition to basic structural signatures and node attributes, you can add a second relation/network to the erg model as an edge covariate.
- Instead of just transitivity, you can use the so-called "Geometrically weighted edgewise shared partner" parameter. For a better model fit, try including something like gwesp(0.2, fixed=T). Transitivity suggests that if A and B share a common friend, they are more likely to connect. If having a friend in common contributes x to the log odds of forming a link with someone, two friends contribute 2x, three friends 3x, and so on. In reality, we are more likely to see diminishing returns. The difference between not having a friend in common and having one shared friend is huge. The difference between having 10 friends in common and having 11 is probably much smaller. The gwesp parameter captures the idea that as the number of shared acquaintances increases, each new one matters less. It is more flexible than the transitive triple parameter and often leads to a better model fit.

Your report from this lab should be formatted as a **mini research paper**. Include Methods (what you did), Results (what you discovered) and Discussion (what it all means). Follow APA guidelines in your writing.

In addition to your report, turn in the **r script** you used for the data analysis. No need to include R output, only the code you used to read the data and conduct the analysis.