HW3

HW 3: Sampling and online data

Note: June 7

First problem, required Take at look at data from the HIV Transmission Network Metastudy Project, available from ICPSR. We'll use data from the Egodyads file, which represents the connections between individuals.

1. Choose one of the networks. Repeat the following (steps 1-3) five times:

- Take a sample (say 10%) of the nodes in the network.

- Fit *latentnet* to the 10% sample. Use the parameters from the model fit using *latentnet*.

2. Fill in the following table using the observed subgraphs. Also add some statistics of your own.

Network/statistic	Clustering coeff	Avg Degree
Full graph		
Sample 1		
Sample 2		
Sample 3		
Sample 4		
Sample 5		

3. Fill in the following table using the average statistics from your simulated graphs using the fit from *latentnet*. Also add some statistics of your own.

Network/statistic	Clustering coeff	Avg Degree
Sample 1		
Sample 2		
Sample 3		
Sample 4		
Sample 5		

4. If computationally feasible, fit a model to the full graph using either *latentnet* or *VBLPCM*. Simulate several graphs based on the fitted parameters and compute the statistics used in your tables above. How do they compare?

5. Repeat 1-3, but this time let's simulate using sampling weights. Choose an individual-level covariate and over-sample individuals with that covariate in your subgraph. Do the results change?

Second problem, optional Set up an account with the Twitter API to scrape data from Twitter. Tutorial here. Once you do that, you can explore sentament and text analysis using the *tm* package.