Social Network Analysis Selected solutions

Exercises

1. How does the modularity maximization clustering algorithm define a cluster/community?

2. The term *spectral* clustering implies eigenvectors/eigenvalues are being used in some way. For what matrix are we calculating eigenvectors in these methods? The adjacency matrix? Do we use the eigenvectors associated with the largest eigenvalues, like in PCA?

The Laplacian.
The smallest (typically start w/ second smallest)

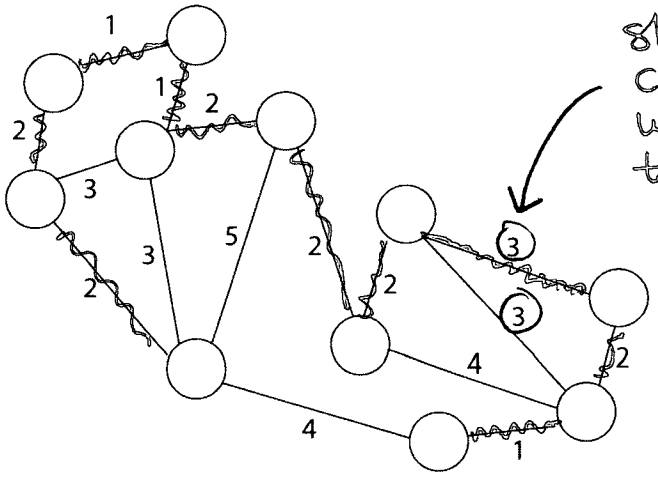
3. Does the modularity maximization algorithm automatically determine the number of clusters? Do spectral methods that use k-means?

Yes.

NO. If you use K-means you must specify K.

4. What clustering method that you've already learned about is the Minimum Spanning Tree method equivalent to?

5. For the following graph, highlight the minimum spanning tree. Is the minimum spanning tree unique? In other words, will everyone draw the same MST?



no. in final step w/ this tree you should find a choice about which "3" edge to add to the tree b/c adding both would create a cycle.

6. For the following hypothesis, indicate whether they are at the individual level, dyad level, or network level.

   a. Degree in an email network is related to tenure at a company.
   node (individual)

   b. Density of a company's employee email network is related to stock price.
   network

   c. Number of friends on facebook is a predictor of introversion or extroversion.
   node

   d. The number of pages that two people both "like" is related to how long they've been friends.
   dyad

   e. The amount of trade between two countries is related to the amount of communication between their highest governmental bodies.
   dyad

   f. The military strength of a country is related to how many countries from whom they import goods.
   node

7. Why do we turn to simulation to compute p-values for hypothesis tests on networks? Why can't we use traditional methods?

8. If we use the QAP approach to simulate a p-value for a dyadic level hypothesis, in essence what are we changing about the network in each iteration of that simulation? In other words, if I have a network and I perform a QAP permutation, what have I essentially done?

   relabel nodes

9. What does network autocorrelation entail? What types of attributes are involved?

   - a relationship between a variable and the network structure.
   - node level attributes - either continuous or categorical.

10. Give an example of a question that could be answered by network autocorrelation.

    Are individuals more likely to associate with/be near others with similar SAT scores?

11. What type of models are Exponential Random Graph Models (ERGMs) similar to?

12. How do we interpret a predicted probability as outputted by an ERGM?

    Prob of an edge between node i and node j if all else held constant.

13. Unlike models you have previously been exposed to, the input variables to an ERGM are called change statistics. What is a change statistic?

14. For an edge level variable, like *number of emails sent* between person i and person j, what is the change statistic associated a given edge?

15. For a network level variable, like *number of triangles in the network*, how would you calculate the change statistic for a given edge?

16. What does homophily mean?

17. What statistics are available to compare different ERGM models for the same data set?

18. What network statistics might you examine on networks simulated by your model to see if your model is accurately re-creating the observed network?

List of Key Terms

Laplacian Matrix

Normalized Laplacian

Modularity

Spectral Clustering

Modularity Maximization

Minimum Spanning Trees

Simulating p-values

Individual Level Hypothesis

Dyadic Level Hypothesis

Network Autocorrelation

Network Regression

ERGM models

Change statistics

AIC/BIC/deviance

Model fit Assessment

Homophily