Social and economic networks. Final exam. 1st part

(It is enough to solve exercises for an amount of 10 points)

1) (**2 points**) Do Exercise 7 in Chapter 3 (pages 95–97) in Easley-Kleinberg's *Networks, Crowds, and Markets* (you have a copy of the book in the course's Virtual Classroom).

2) (**2 points**) Let $\mathcal{V}$ be a partition of an undirected graph G. For every community $X_r \in \mathcal{V}$, let a_r be the fraction of edges incident to nodes in X_r, and for every $X_r, X_s \in \mathcal{V}$, let $e_{r,s}$ be the fraction of edges in G that link nodes in X_r with nodes in X_s.

(a) Write an expression for the modularity $Q(\mathcal{V})$ of the partition $\mathcal{V}$ in terms of these quantities a_r and $e_{r,s}$.

(b) In the AMEN (*AIDS in Multi-Ethnic Neighborhoods*) study (1992), the following frequencies of ethnicities in a large sample of heterosexual couples in San Francisco were recorded

		Women				
		Black	**Hispanic**	**White**	**Other**	**Total**
Men	**Black**	0.258	0.016	0.035	0.013	0.323
	Hispanic	0.012	0.157	0.058	0.019	0.247
	White	0.013	0.023	0.306	0.035	0.377
	Other	0.005	0.007	0.024	0.016	0.053
	Total	0.289	0.204	0.423	0.084	

Consider the undirected network of relationships for the sample studied and the partition defined by the nodes' ethnicity (black, hispanic, white, and other). Compute the modularity of this partition. Assuming that the sample was representative of their community, what can you conclude about ethnic homophily in this community?

3) (**2 points**) Consider the configuration model where a fraction p_1 of nodes have degree 1 and the remaining fraction of nodes $p_3 = 1 - p_1$ have degree 3. Let us consider, to fix ideas, networks of order $n = 10^4$.

(a) Take first $p_1 = 0.5$, $p_3 = 0.5$. Estimate through a simulation the expected fraction of a network of order $n = 10^4$ covered by its largest component.

(b) Now, plot a graph of the expected fraction (estimated through simulations) of a network of order $n = 10^4$ covered by its largest component for values of p_1 from 0 to 1 in 0.01 steps. According to your graph, what is the probability p_1 at which the giant component disappears? Is your estimate consistent with the theory?

4) (**4 points**) Stochastic models can be used to predict links in a network. Now, in a homework you were asked to produce an ERGM of Grey's Anatomy sexual contact network. Using your model (or a friend's, if you did not deliver this homework):

(a) What is the link in the network with the highest probability (or the links, in the case of ties)? Look up somewhere on the Internet whether the corresponding contact appeared early or late in the series.

(b) What is the most probable new (that is, absent from the network) sexual contact between a pair of characters in that network? Look up somewhere on the Internet whether this contact has occurred in the series after season 8, and (if the answer is positive) when.

(c) What about an Stochastic Block Model? Fit an "optimal" SBM to this network, and use the result to give an answer to the previous questions.

(d) What approach do you consider more reasonable to answer questions (a) and (b) for this specific network, the ERGM one or the SBM one? Justify your answer.

5) (**4 points**) I want you to design networks where the most central nodes with respect to some set S of centrality measures are different.

(a) Build, by hand, a small network that has this property with respect to $S = \{\text{degree}, \text{betweenness}\}$. That is, where no node with highest degree is a node with highest betweenness. Plot the network, indicating the nodes of highest degree and of highest betweenness. Explain the strategy you followed to build your network with the desired property.

(b) Build, by hand, a small network that has this property with respect to $S = \{\text{closeness}, \text{betweenness}\}$. Plot the network, indicating the nodes of highest closeness and of highest betweenness. Explain the strategy you followed to build your network with the desired property.

(c) Now, write a program that searches the space of all connected networks of a given size n and either finds a network with this property for $S = \{\text{degree}, \text{betweenness}, \text{closeness}, \text{eigenvector}\}$ or reports that there is no such network for this n. Run your program starting with $n = 4$ and provide a smallest network (with least order, and among those with this order, with least size) with this property. Plot the network, indicating which are the nodes with the highest centrality values. What are the features of the network that make the most central nodes according to the different centrality measures to be different? If necessary, to hint the answer to this question, find some more networks with this property.