1.022 Introduction to Network Models **Fall 2018**

Homework 2

2.1 Centrality measures

[10 points] Given the following adjacency matrix A of a network, compute the degree and the eigenvector centrality, using matrix computations (you may use a numerical software, but provide the formulae and reasoning for what you are computing).

$$A = \begin{bmatrix} 0 & 0.1 & 0.1 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0.1 & 0.1 & 0 \end{bmatrix}$$

[5 points] Comment and try to explain the difference in the centrality measure you observe (5 or less sentences)

2.2 The Laplacian matrix and spectral clustering

[10 points] Show that the Laplacian matrix $L = D - A$ for an undirected, unweighted graph can be decomposed as $L = BB^T$. Here B is the $N \times E$ node-to-edge incidence matrix of the graph, defined as $B_{ie} = -1$ if node i is the tail of edge e, $B_{ie} = +1$ if node i is the head of edge e and $B_{ie} = 0$ otherwise, where each edge has been assigned an arbitrary orientation.

Hint: Look at the individual entries L_{ij} and express them in terms of components of B_{ij} (write out the matrix multiplication in index notation). Consider the case of an off-diagonal and diagonal entry separately.

[10 point] Conclude from the above that the Laplacian is positive semidefinite, by considering the quadratic form $x^T Lx$, i.e., show that $x^T Lx \geq 0 \quad \forall x$.

Hint: using the decomposition found above, is it possible to write the quadratic form as the squared norm of another vector? If yes, then you would only have to argue why this implies that the expression is non-negative.

[10 point] We define the cut as the number of edges between two non-overlapping sets of nodes $\mathcal{V}_1, \mathcal{V}_2$ such that $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}$ and $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$, i.e., $\mathcal{V}_1, \mathcal{V}_2$ are a partition of the node set $\mathcal{V}$ (see lecture slides). Let us define the partition indicator vector $s \in \{-1, 1\}^n$ with the entries $s_i = -1$ if node i belongs to the first set $\mathcal{V}_1$ and $s_i = 1$ if node i belongs to the second set $\mathcal{V}_2$.

Show that the cut can be written as:

$$\text{cut} = \frac{1}{4} s^T Ls$$

2.3 Centrality measures and spectral clustering

[10 points] Consider the Karate club network. Compute the degree, eigenvector, closeness, betweeness and Katz centrality (with the default value provided by networkx) of each node. Use networkx functions! Produce one plot per centrality measure

[10 points] Compute the second eigenvector of the Laplacian matrix of the Karate club network. Assign nodes into clusters according to their signs of this vector, and plot the nodes in each cluster with a different color.

2.4 Network creation and visualization

This exercise serves as a brief introduction to networkx, a computational graph analysis tool in python that we will use in this course for analysing networks. You will have to construct a static network based on the temporal contact patterns among children and teachers in a provide a link. The dataset has been originally collected by the SocioPatterns consortium. For more information, see here: `http://www.sociopatterns. org/datasets/primary-school-temporal-network-data/`

The data set contains the temporal network of contacts between the children and teachers used in the study "Mitigation of infectious disease at school: targeted class closure vs school closure", BMC Infectious Diseases 14:695 (2014).

The file contains a tab-separated list representing the active contacts during 20-second intervals of the data collection. Each line has the form `t i j Ci Cj`, where `i` and `j` are the anonymous IDs of the persons in contact, `Ci` and `Cj` are their classes, and the interval during which this contact was active is $[t - 20s, t]$, where the time t is measured in seconds. If multiple contacts are active in a given interval, you will see multiple lines starting with the same value of t.

[15 points] Read in the dataset and create a static network according to the following specifications: The edge weight between two nodes should be proportional to the number of measured contacts.

[15 points] Compute and plot the degree distribution of the network. Do the same for the clustering coefficient and the betweenness centrality.

[10 points] Use the spring force layout to plot the graph, with node sizes proportional to the node degree, and node colors given by the class labels.