Name	Jackson Wakefield
ASU ID Number	1216414420 - jswakefi

CSE 472: Social Media Mining

Homework I - Linear Algebra, Graph Essentials, Network Measures

Prof. Huan Liu
Due at 2021, September 14^{th}, 11:59 PM

This is an ***individual*** homework assignment. Please submit a digital copy of this homework to **Gradescope**. This is a fillable PDF and you are able to type into answer boxes provided for each question.

1. **[Linear Algebra]** Consider 2-dimensional data points of [0, -1], [1, 0], [2, 1], [1, 1], [-1, 1], [-1, -1], [-2, -1].

   (a) All the data points can be gathered together and shown with one matrix. Let's assume $[X]_{2\times 7}$ is that matrix. Fill the following matrix.

$$X = \begin{pmatrix} \boxed{0} & \boxed{1} & \boxed{2} & \boxed{1} & \boxed{-1} & \boxed{-1} & \boxed{-2} \\ \boxed{-1} & \boxed{0} & \boxed{1} & \boxed{1} & \boxed{1} & \boxed{-1} & \boxed{-1} \end{pmatrix}_{2\times 7}$$

   (b) What is the point showing the center of these points? [*Hint:* Calculate the mean of the values in each dimension].

$$\mu = \begin{pmatrix} \boxed{0} \\ \boxed{0} \end{pmatrix}_{2\times 1}$$

(c) Calculate $Y = (X - \mu)(X - \mu)^T$ in which X^T is the transpose of X. To calculate $(X - \mu)$, easily subtract the μ from all the data points.

$$Y = \begin{pmatrix} \boxed{12} & \boxed{5} \\ \boxed{5} & \boxed{6} \end{pmatrix}_{2\times 2}$$

(d) Solve $|Y - \lambda I| = 0$ to extract the values of λ. $|\cdot|$ is the determinant and I is the identity matrix. λ values are called eigenvalues. **Round the values to the nearest integer**.

$\lambda_1 = \boxed{3}$, $\lambda_2 = \boxed{15}$

(e) Calculate the corresponding eigenvector to the **largest** eigenvalue.

$$v = \begin{pmatrix} \boxed{2} \\ \boxed{1} \end{pmatrix}_{2\times 1}$$

(f) Compute $\hat{X} = v^T X$.

$$\hat{X} = \begin{pmatrix} \boxed{-1} & \boxed{2} & \boxed{5} & \boxed{3} & \boxed{-1} & \boxed{-3} & \boxed{-5} \end{pmatrix}_{1\times 7}$$

Congratulations you performed Principle Component Analysis (PCA) procedure, a well known dimensionality reduction method in machine learning. In other words, you projected your 2-dimesional data into 1-dimensional one such that you preserve the variance as much as possible (i.e. the least information has been lost).

2. **[Special Graphs, Graph Traversal]** A *complete bipartite graph* $K_{m,n}$ is a graph that its vertices can be partitioned into two subsets of m and n vertices such that every pair of graph vertices in the two sets are adjacent.

(a) For which values of m and n, $K_{m,n}$ would be regular?

When m = n, K(m, n) would be regular.

(b) The *complement graph* $\overline{G}$ of a simple graph G has the same vertices as G. Two vertices are adjacent in $\overline{G}$ if and only if they are not adjacent in G. What is the number of edges in the complement graph of $K_{m,n}$?

This one took a lot of thinking

The number of edges in the complement of K(m, n) is:

((m * (m-1)) / 2) + ((n * (n-1)) / 2))

(c) Describe the shape of the complement graph of $K_{m,n}$?

In the context of this class, the shape of the complement to this graph is very important. In the standard K(m, n), it is possible to traverse the entire network by starting at any given point, whereas in the graph's complement, the graph is separated into two entirely unique networks - meaning it is only ever possible to traverse a fraction of the overall graph when starting at any given point.

(d) For which value of m and n, $K_{m,n}$ could have an Euler path?

When the values of m and n are even, K(m, n) is guarunteed to have a euler path. In addition, if either m or n are equal to 2, K(m, n) is guarunteed to have a euler path. This is because the definition of a Euler path provides that a graph has a euler circuit if the degree of every vertex is even, or if there are at most two vertices with an odd degree. This means that every K(m, n) graph with m or n = 2 is guarunteed to only have at most two vertices with an odd degree, and every graph where m and n are even is guarunteed to have even degrees for every vertex.

(e) For which value of m and n, $K_{m,n}$ are trees?

K(m, n) is a tree if and only if the value of m or n is 1.

(f) We will denote the nodes in the subset with m vertices as $V = \{v_1, v_2, ...v_m\}$ and the nodes in the subset with n vertices as $U = \{u_1, u_2, ...u_n\}$. Let's assume $m \geq n$, starting from v_1, what would be the sequence of visited nodes in BFS?

The sequence of nodes visited would be:

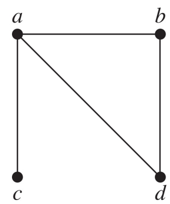
v1 -> all vertices in U [U(1) -> ... -> u(n)] -> all remaining vertices in V [V(2) -> ... -> V(m)]

(g) What would be the sequence of visited nodes in DFS?

The sequence of nodes visited would be:

V(1) -> U(1) -> ... -> V(n) -> U(n), after which all remaining nodes in V would be traversed [V(n) -> ... -> V(m)]

(h) The *degree sequence* of a graph is a sequence of the degrees of the vertices of the graph in nonincreasing order. For example, the degree sequence of the following graph is $\{3, 2, 2, 1\}$



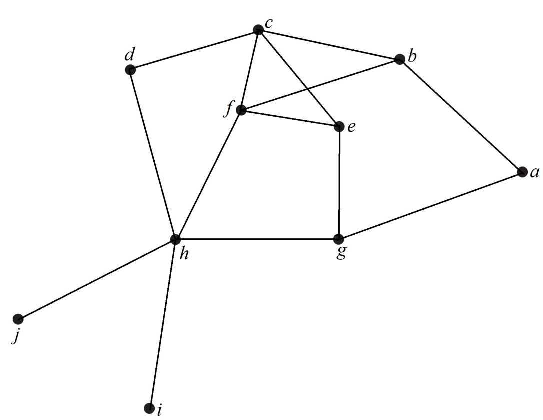
What is the degree sequence of $K_{m,n}$, if $m \geq n$?

The degree sequence would be: {n m's + m n's} for example, if m was 5 and n was 3, the sequence would be 3 5's followed by 5 3's -> {5, 5, 5, 3, 3, 3, 3, 3}

(i) Research the *recommender systems*. If we visualize a recommender system network as a bipartite graph with set of nodes V and U, what would be the possible nodes and edges in this network?

V	
U	
Edges	

3. **[Network Measures]** For the given graph compute the following measures.



(a) What is the normalized degree centrality for node f?

$$C_d^{norm}(f) = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

(b) What is the normalized betweenness centrality for node h?

$$C_b^{norm}(h) = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

(c) What is the closeness centrality for node c?

$$C_c(c) = \frac{1}{\boxed{}} = \boxed{}$$

(d) In the case of subgraphs how can you know if a cycle is socially balanced?

(e) What is the maximum value of local clustering coefficient and when can it be observed for all the nodes in a graph?

Maximum value of local clustering coefficient = $\boxed{}$