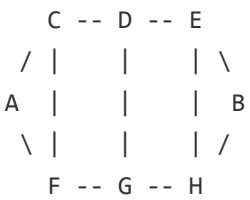
Feedback — Week3A (Advanced)

Help Center

You submitted this quiz on **Thu 19 Feb 2015 4:33 PM PST**. You got a score of **4.00** out of **5.00**. You can attempt again, if you'd like.

Question 1

For the following graph:



Write the adjacency matrix A, the degree matrix D, and the Laplacian matrix L. For each, find the sum of all entries and the number of nonzero entries. Then identify the true statement from the list below.

Your Answer		Score	Explanation
The sum of the entries of A is 8.			
A has 30 nonzero entries.			
The sum of the entries of L is 0.	✔	1.00	
D has 30 nonzero entries.			
Total		1.00 / 1.00	

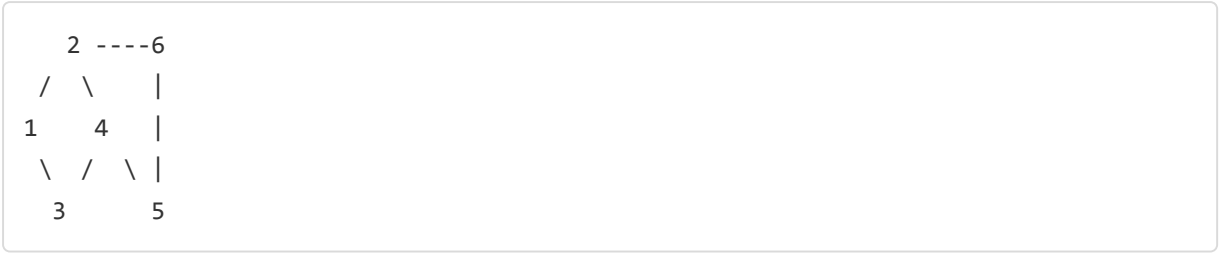
Question Explanation

The Adjacency matrix has 1 in row i and column j if there is an edge between the nodes represented by these two rows. In this graph, all the nodes have degree 3 except A and B, whose degrees are 2. Thus, there are 6*3 + 2*2 = 22 nonzero entries. Each of these entries is 1, so the sum of all entries, including the 0's, is 22.
The degree matrix has the degree of node i in the i-th position along the diagonal. All other entries are 0. Thus, there are 8 nonzero entries, and their sum is 22, as we observed in the paragraph above.

The Laplacian matrix is D-A. Since A has only off-diagonal entries, and D has only diagonal entries, the number of nonzero entries in L is the sum of those numbers for A and D. That is, 22+8 = 30. The sum of all the entries of any Laplacian matrix is 0, because both the positive entries (on the diagonal) and the negative entries (off-diagonal) sum to the sum of the degrees of the nodes.

Question 2

You are given the following graph.



The goal is to find two clusters in this graph using Spectral Clustering on the Laplacian matrix. Compute the Laplacian of this graph. Then compute the second eigen vector of the Laplacian (the one corresponding to the second smallest eigenvalue).

To cluster the points, we decide to split at the mean value. We say that a node is a tie if its value in the eigen-vector is exactly equal to the mean value. Let's assume that if a point is a tie, we choose its cluster at random. Identify the true statement from the list below.

Your Answer	Score	Explanation
◉ 1 and 3 in the same cluster ✔	1.00	
○ 4 and 5 are in the same cluster		
○ 3 and 5 can either be in the same cluster or in different clusters (depending on randomness)		
○ 2 and 5 are in different clusters		
Total	1.00 / 1.00	

Question 3

We wish to estimate the surprise number (2nd moment) of a data stream, using the method of AMS. It happens that our stream consists of ten different values, which we'll call 1, 2,..., 10, that cycle repeatedly. That is, at timestamps 1 through 10, the element of the stream equals the timestamp, at timestamps 11 through 20, the element is the timestamp minus 10, and so on. It is now timestamp 75, and a 5 has just been read from the stream. As a start, you should calculate

the surprise number for this time.

For our estimate of the surprise number, we shall choose three timestamps at random, and estimate the surprise number from each, using the AMS approach (length of the stream times 2*m*-1, where *m* is the number of occurrences of the element of the stream at that timestamp, considering all times from that timestamp on, to the current time). Then, our estimate will be the median of the three resulting values.

You should discover the simple rules that determine the estimate derived from any given timestamp and from any set of three timestamps. Then, identify from the list below the set of three "random" timestamps that give the closest estimate.

Your Answer		Score	Explanation
○ {37, 46, 55}			
○ {30, 47, 62}			
○ {14, 35, 42}			
◉ {3, 45, 72}	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

First, the surprise number is 5*64 + 5*49 = 565. The reason is that the elements 1 through 5 appear 8 times, so they contribute $5*8^2$, and the elements 6 through 10 appear 7 times, contributing $5*7^2$.
Notice that for this contrived example, the AMS estimate is a nondecreasing function of the timestamp. Thus, of any three timestamps, the middle one will give the median estimate, and we do not have to calculate all three.

At each of the timestamps between 36 and 45, inclusive, the element appearing then appears exactly 4 times, from that time forward. Thus, each of these timestamps generates an estimate of 75*(2*4 - 1) = 525, which is as close to 565 as we can get. Each of the correct answers has a middle timestamp in this range.

Similarly, for the timestamps between 26 and 35, the estimate is 75*(2*5 - 1) = 675 and for the timestamps between 46 and 55 the estimate is 75*(2*3 - 1) = 375. Neither of these groups offer as close an estimate, and the timestamps earlier or later offer even worse estimates.

Question 4

We wish to use the Flagolet-Martin lgorithm of Section 4.4 to count the number of distinct elements in a stream. Suppose that there ten possible elements, 1, 2,..., 10, that could appear in the stream, but only four of them have actually appeared. To make our estimate of the count of

distinct elements, we hash each element to a 4-bit binary number. The element x is hashed to $3x + 7$ (modulo 11). For example, element 8 hashes to 3*8+7 = 31, which is 9 modulo 11 (i.e., the remainder of 31/11 is 9). Thus, the 4-bit string for element 8 is 1001.

A set of four of the elements 1 through 10 could give an estimate that is exact (if the estimate is 4), or too high, or too low. You should figure out under what circumstances a set of four elements falls into each of those categories. Then, identify in the list below the set of four elements that gives the exactly correct estimate.

Your Answer		Score	Explanation
○ {4, 5, 6, 7}			
○ {1, 3, 6, 8}			
◉ { 2, 6, 8, 10}	✔	1.00	
○ {1, 2, 3, 9}			
Total		1.00 / 1.00	

Question Explanation

Here is a table of the hash values and resulting bit strings for each of the ten elements:

x	$3x+7$ (mod 11)	Bit String
1	10	1010
2	2	0010
3	5	0101
4	8	1000
5	0	0000
6	3	0011
7	6	0110
8	9	1001
9	1	0001
10	4	0100

In order to give the correct estimate (4), a set must have at most two 0's at the end of the hash value of any of its members, but must have a member with exactly two 0's at the end. Observe from the table above that 10 is the only element whose hash value has exactly two bits at the end. However, 1, 2, 3, 6, 7, 8, and 9 have zero or one 0 at the end, so the correct answers are any set of four elements that includes 10 and does not include 4 or 5.

Question 5

Suppose we are using the DGIM algorithm of Section 4.6.2 to estimate the number of 1's in suffixes of a sliding window of length 40. The current timestamp is 100, and we have the

following buckets stored:

End Time	100	98	95	92	87	80	65
Size	1	1	2	2	4	8	8

Note: we are showing timestamps as absolute values, rather than modulo the window size, as DGIM would do.

Suppose that at times 101 through 105, 1's appear in the stream. Compute the set of buckets that would exist in the system at time 105. Then identify one such bucket from the list below. Buckets are represented by pairs (end-time, size).

Your Answer		Score	Explanation
○ (80,8)			
○ (100,2)			
◉ (95,4)	✖	0.00	While this bucket exists prior to time 105, it is combined when the 1 enters at time 105.
○ (102,2)			
Total		0.00 / 1.00	

Question Explanation

Time 101: When the first 1 comes in, we construct (101,1). That gives us three 1's, so we combine (100,1) and (98,1) into (100,2). That gives us three 2's, so we combine (95,2) and (92,2) into (95,4).
Time 102: Create (102,1).

Time 103: Create (103,1). Combine (102,1) and (101,1) into (102,2).

Time 104: Create (104,1).

Time 105: Drop (65,8), because it is now completely outside the window of length 40. We have three 1's, so we combine (104,1) and (103,1) into (104,2). We have three 2's, so we combine (102,2) and (100,2) into (102,4). We have three 4's, so we combine (95,4) and (87,4) into (95,8). Note that there are only two 8's, not three, since we already dropped the last 8.