Thursday Quiz

1. [0.5 point] After applying the Girvan-Newman algorithm, we delete the edge that has the smallest betweenness value.
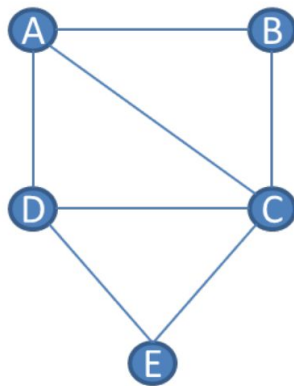   Answer- False
2. [0.5 point] In Affiliation Graph Model (AGM), if two nodes u,v share no community then the edge probability P(u,v) is very high.
   Answer - False
3. [2.5 points] For node E, use the Girvan-Newman algorithm to calculate the betweenness of each edge (do this for node E ONLY). Write down the edges and their betweenness values in the format below:
   (Edge1, Edge2) = Betweenness Value

   Answer-
   (A,C) - 0.5 [0.5 points]
   (A,D) - 0.5 [0.5 points]
   (B,C) - 1 [0.5 points]
   (C,E) - 2.5 [0.5 points]
   (D,E) -1.5 [0.5 points]
4. [2 point] How is edge betweenness used to find communities in Girvan-Newman?
   Answer -
* Edge betweenness:Number of shortest paths passing over the edge.
* Betweenness scores for edges of a graph behave something like a distance metric.
  * Not a true distance metric.
* Could cluster by taking edges in increasing order of betweenness and adding to graph one at a time.
  * At each step, connected components of graph form clusters.
* Girvan-Newman: Start with the graph and all its edges and remove edges with highest betweenness
5. Consider the following matrix F that represents the Node community membership strengths. For each of the possible pairs of edges, what is the probability of that edge? Here, u, v and w are your nodes.

Fu	0	1.3	0	0.8

| Fv | 0.9 | 0 | 0 | 1.4 |
| Fw | 0 | 1.6 | 1 | 0 |

6. [1 point] Differentiate between overlapping and non-overlapping communities.
Answer-
In overlapping communities, communities can share nodes with other communities where as in non-overlapping communities, communities do not share nodes with other Communities. Edge Density in overlapping community is higher

7. [1+1 points] What do you mean by modularity when talking about network communities. What is its use?
Modularity is measure of how well a network is partitioned into communities. It takes the value between the range [-1,1]. It is positive when the number of edges within groups exceeds the expected number. It is useful for selecting the number of clusters