Feedback — Week 4 Quiz
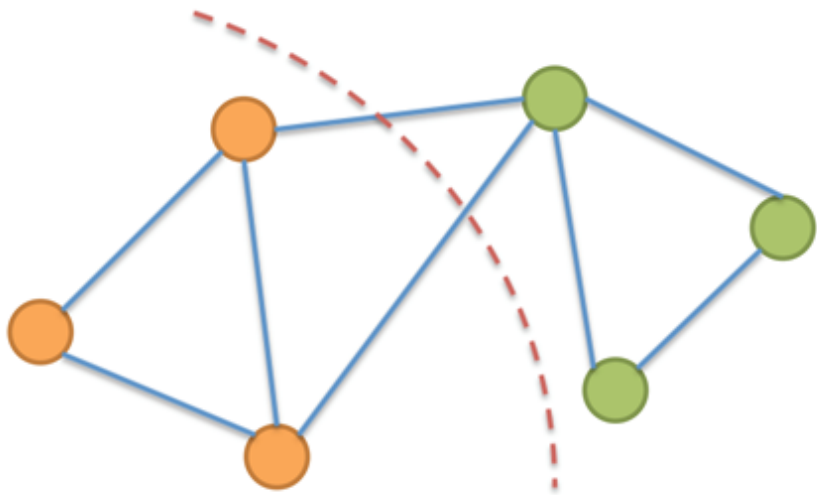
You submitted this quiz on **Sat 23 May 2015 1:28 AM PDT**. You got a score of **6.00** out of **6.00**.

Question 1

In the graph above, every edge has weight = 1. What are the RatioCut and NCut measures for the partitioning shown above where the graph is divided into two clusters with three nodes each?
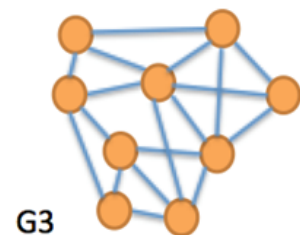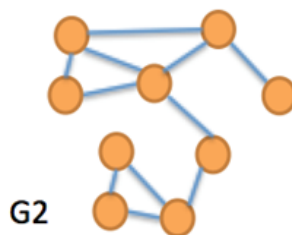
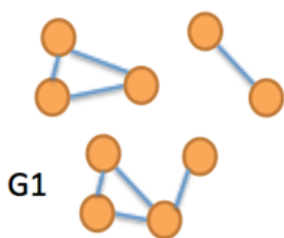| Your Answer		Score	Explanation
○ RatioCut = 2/3, NCut = 4/8			
○ RatioCut = 4/3, NCut = 2/8			
◉ RatioCut = 2/3, NCut = 2/8	✔	1.00	
○ RatioCut = 2/8, NCut = 4/8			
Total		1.00 / 1.00	

Question Explanation

$$RatioCut = \frac{1}{2}\sum_{i=1}^{2}\frac{W(C_i, \overline{C_i})}{|C_i|} = \frac{1}{2}\left(\frac{2}{3}+\frac{2}{3}\right) = \frac{2}{3}$$

$$NCut = \frac{1}{2}\sum_{i=1}^{2}\frac{W(C_i, \overline{C_i})}{d(C_i)} = \frac{1}{2}\left(\frac{2}{3+2+3}+\frac{2}{4+2+2}\right) = \frac{2}{8}$$

Question 2

G1   G2   G3

Based on the graphs given above, which of the following statements is true?

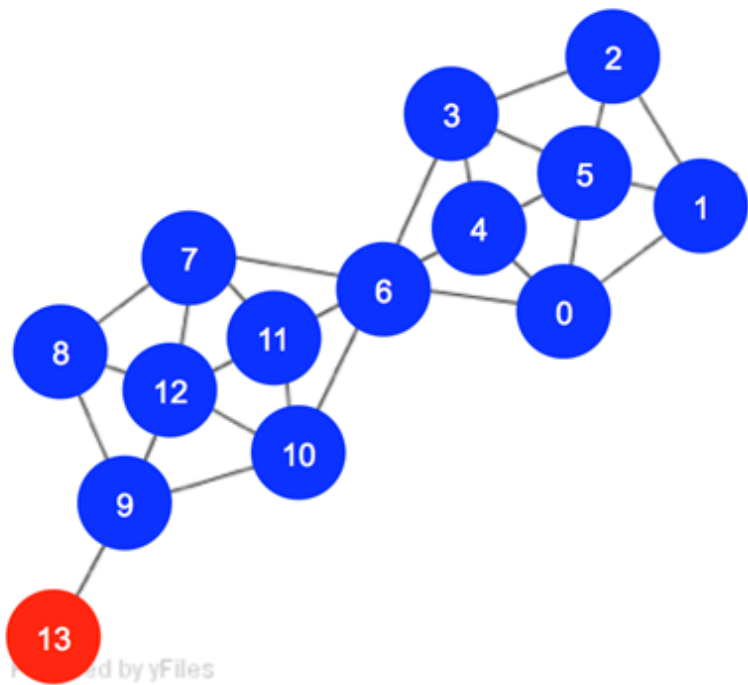| Your Answer	Score	Explanation
○ The second eigenvalue for the Laplacian of G3 is less than that of G1		
○ The second eigenvalue for the Laplacian of G2 is less than that of G1		
○ The second eigenvalue for the Laplacian of G3 is less than that of G2		
◉ The second eigenvalue for the Laplacian of G2 is less than that of G3	✔ 1.00	
Total	1.00 / 1.00	

Question Explanation

The first three eigenvalues for G1 are all 0 since there are three connected components in G1.

G3 is more densely connected than G2, which implies its second eigenvalue is comparatively larger since the second eigenvalue models the connectivity of a connected graph.

Question 3

Using the graph shown above, is Node 0 in Node 4's ε-neighborhood for ε=0.7?

Your Answer	Score	Explanation
○ No		
● Yes	✔	1.00
Total	1.00 / 1.00	

Question Explanation

$$\Gamma(n_0) = \{1, 4, 5, 6, 0\}$$
$$\Gamma(n_4) = \{3, 4, 5, 6, 0\}$$
$$\sigma(n_0, n_4) = \frac{|\Gamma(n_0) \cap \Gamma(n_4)|}{\sqrt{(|\Gamma(n_0)||\Gamma(n_4)|)}} = \frac{4}{\sqrt{(5 * 5)}} = \frac{4}{5} > 0.7$$

Question 4

Which of the following statements about NetClus and RankClus are true?

Your Answer		Score	Explanation
☐ RankClus is only able to perform bipartitioning of the network, whereas NetClus is able to generate more than two clusters.	✔	0.25	This is false because both RankClus and NetClus are able to generate more than two clusters. The limitation of RankClus is that it is only able to handle bi-typed heterogeneous information networks.
☐ Unlike in RankClus, NetClus does not rank the nodes in each cluster during the clustering process.	✔	0.25	This is false because both RankClus and NetClus rely on node rankings in each cluster to refine clusters based on the most representative nodes in each cluster.
☑ Both use an EM framework for determining the conditional rank scores and cluster membership of each node.	✔	0.25	
☐ Both require high quality initial clusters to be able to perform well.	✔	0.25	This is false because the performance of neither algorithm is impacted by the initialization.
Total		1.00 / 1.00	

Question 5

Which of the following similarity measures is better at finding highly visible objects in a network?

Your Answer		Score	Explanation
◉ Random Walk	✔	1.00	
○ SimRank			
○ PathSim			
Total		1.00 / 1.00	

Question Explanation

PathSim is good at finding peers, i.e. objects with strong connectivity and similar visibility; SimRank tends to favor pure objects, i.e. objects with highly skewed distribution in their in-links or out-links.

Question 6

How is user guidance used in PathSelClus?

Your Answer	Score	Explanation
◉ Users are asked to provide a small number of seed examples for each cluster at the beginning, which PathSelClus then uses to automatically select the most optimal meta-paths for clustering.	✔ 1.00	
○ Users are asked to assess the quality of the clusters at the end of the clustering process.		
○ Users are asked to specify which meta-paths are good for clustering.		
Total	1.00 / 1.00	

Question Explanation

We use standard cluster quality measures to determine the goodness of the resulting

clustering. Users are not asked specifically for the meta-paths for clustering. The algorithm is able to figure them out based on input examples from the user.