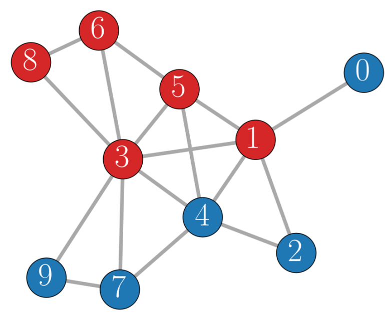
Assignment 4 (Graph Theory and Networks)

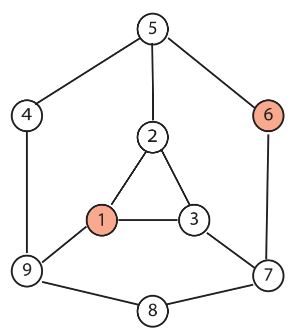
Discussed in class on March 26

---

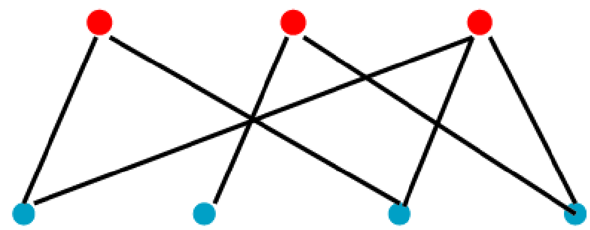
(1) For the graph below, determine the similarity between the red group of nodes (numbers 1, 3, 5, 6, 8) and the blue group (numbers 0, 2, 4, 7, 9), based on structural equivalence. First determine the similarity (S_{ij}) of nodes i,j with i from the red group and j from the blue group. Then compute the average of the similarity values to give the group similarity.



(2) Do one complete iteration of the Girvan-Newman divisive method on the graph below (there is nothing special about the colored nodes). Which edges are removed?



(3) Consider the bipartite network below, consisting of two classes of nodes, A (red) and B (blue). Determine the modularity of the partition of the network into two communities, A and B. If you did a different bisection of the network, so that a community could have both red and blue nodes, would the modularity increase, decrease, or could it do either?



(4) Suppose that A and B are two of the $q > 2$ clusters of a network partition, with degrees k_A and k_B, respectively. Also suppose that these are approximately the same, so $k_A \approx k_B = k$. Let L_A, L_B, and L_{AB} be the number of links inside cluster A, inside cluster B, and between A and B, respectively. Compute the difference in modularity between this partition and the one in which A and B are merged, called cluster C. What condition on k makes the partition with A and B merged have larger modularity than the one in which they are separated?

(5) The complementary cumulative distribution function $\mathbf{P}_k$ of a discrete probability distribution p_k is defined as $\mathbf{P}_k = \sum_{j=k}^{\infty} p_j$. Show that the complementary cumulative distribution function of a power-law distribution $p_k = Ck^{-\alpha}$ with exponent $\alpha > 1$ and $k = 1, 2, ...$ is itself a power-law distribution. What is the exponent of the $\mathbf{P}_k$ distribution? (Hint: approximate summation with integration.)

(6) Consider the following process. First, start with a collection of N nodes and no edges. Then, one by one, add an edge between two nodes not already connected to each other. Continue until you have a complete network.

(a) How many steps are there in the process?

(b) Which of the following tends to be true about the sequence of largest component sizes as you add edges? Explain your answer.

(i) Increases slowly at the beginning of the sequence, increases very fast at the end.

(ii) Increases slowly until some threshold, increases very quickly for a short time, and then increases slowly afterwards.

(iii) Increases at a constant rate.

(iv) Increases very fast at the beginning of the sequence, then tapers off and increases slowly until the end.

(v) Increases and decreases randomly.