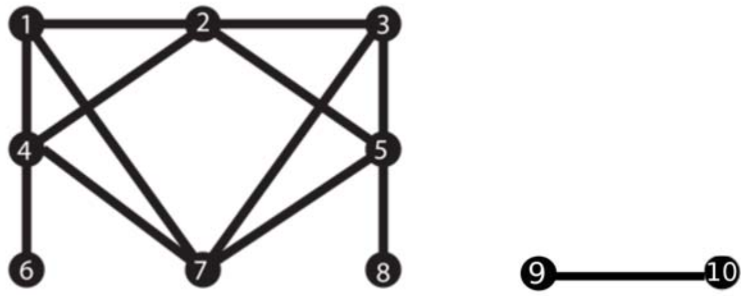
Homework exercises for Bioinformatics I, Bio390
Biological networks, Andreas Wagner

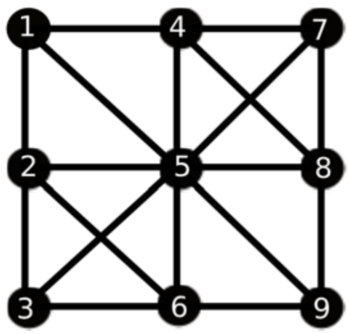
Note: These exercises are for you to solve on your own. You do not have to turn them in and they will not be graded. Even though solutions are provided at the end of this document, we highly recommend that you solve them and do so before looking at the solutions, because similar (not necessarily identical) problems will occur on the final exam.

Exercise 1: (Graph Representation)



a) A graph is a collection of nodes together with the edges that connect the nodes. List the degrees of all nodes in the above graph.
b) All information about a graph can be encapsulated in an adjacency matrix that describes the relationship between nodes and edges of the graph. Write down the adjacency matrix for the above graph.
c) Graphs can be represented by shortest path matrices that describe the minimum number of edges needed for linking pairs of nodes. For the above graph, write down the shortest path matrix. [Note: if there is no path connecting two nodes, the shortest path length between them will be infinity.]
d) How many connected components does the above graph have? List the set of nodes belonging to each component.

Exercise 2: (Network Properties)



a) **(Clustering Coefficients)**

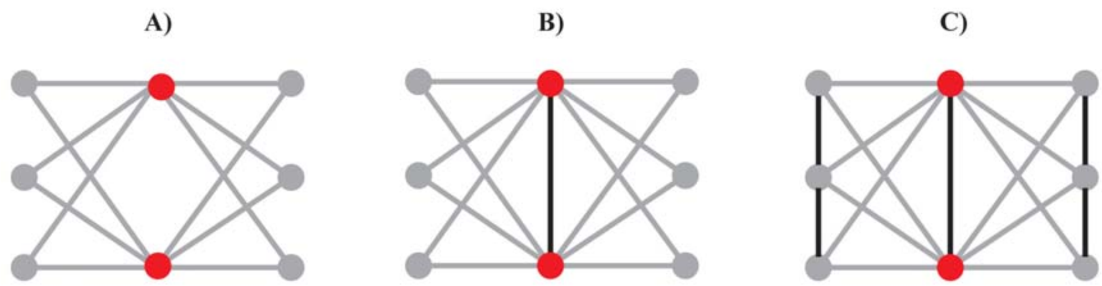
Identify the node(s) with the lowest and the node(s) with the highest clustering coefficient in the above network and write down their clustering coefficients. How does the clustering coefficient relate to the degree of a node? Is it true that a node with higher degree (e.g. a hub) always has a higher clustering coefficient in comparison to other nodes?

b) (Clique)

A clique is a graph in which every pair of nodes is connected. What is the degree of a node belonging to a clique of size n? How many edges does a clique of size n have? What is the clustering coefficient of a clique of size n? Write down the size of the largest clique in the above graph. How many cliques of size 3 can you see in the above graph?

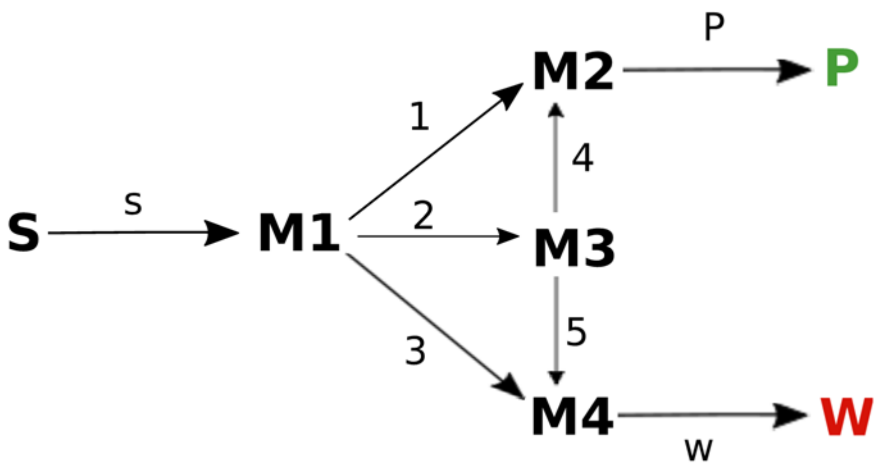
c) (Assortativity)

Among the following networks identify the most assortative and the most disassortative networks. (NOTE: no calculation is required, just briefly explain your reasoning.).
Note that in each of the following networks, two nodes are colored in red. These are hub nodes, which have especially high degree. Use these networks to hypothesize how the clustering coefficient of a hub in an assortative network compares in magnitude to the clustering coefficient of a hub in a disassortative network? Is it higher or lower? Briefly explain your answer.



Exercise 3: (Metabolic Networks)

For the metabolic network shown below, let $v=(v_s, v_1, v_2, v_3, v_4, v_5, v_p, v_w)$ be the flux vector corresponding to the labeled reactions. P is an essential biomass product of the network, and v_p thus corresponds to the biomass growth flux of this network. W is a waste by-product whose synthesis rate should be minimized.



a) Write down the differential equations for each of the four metabolites M_1 through M_4 as a function of the metabolic fluxes v_i.

b) Draw the stoichiometry matrix of the metabolic network.

c) Write down the biomass growth flux of the above metabolic network, if the flux producing m_1, v_s=10 and the flux producing W, v_w=2.

d) Show the vector v of all feasible steady-state solutions, assuming that the flux producing m_1, v_s=10 and the fluxes of v_1, v_2 are equal. (Hint: This flux vector will contain two variables, because it is not uniquely determined by the constraints just mentioned.)

e) Give the feasible steady-state solution that maximize the biomass growth flux of the above metabolic network in steady state, if the flux producing m_1, v_s=10 and the fluxes of v_1, v_2 are equal.

Solutions to Exercises

Exercise 1

a) node1: 3, node2: 4, node3: 3, node4: 4, node5: 4, node6: 1, node7: 4 and node8: 1, node9: 1. Node10: 1.

b)

$$A=\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

c)

$$S=\begin{bmatrix} 0 & 1 & 2 & 1 & 2 & 2 & 1 & 3 & \infty & \infty \\ 1 & 0 & 1 & 1 & 1 & 2 & 2 & 2 & \infty & \infty \\ 2 & 1 & 0 & 2 & 1 & 3 & 1 & 2 & \infty & \infty \\ 1 & 1 & 2 & 0 & 2 & 1 & 1 & 3 & \infty & \infty \\ 2 & 1 & 1 & 2 & 0 & 3 & 1 & 1 & \infty & \infty \\ 2 & 2 & 3 & 1 & 3 & 0 & 2 & 4 & \infty & \infty \\ 1 & 2 & 1 & 1 & 1 & 2 & 0 & 2 & \infty & \infty \\ 3 & 2 & 2 & 3 & 1 & 4 & 2 & 0 & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 & 1 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 1 & 0 \end{bmatrix}$$

d) there are 2 connected components. The first component harbors nodes 1 to 8, and the second component contains nodes 9 and 10.

Exercise 2

a) Node 5 has the lowest clustering coefficient (0.3571), and nodes 3 and 7 have the highest clustering coefficient (1).

The clustering coefficient is not necessarily proportional to the number of neighbors a node has (i.e. its degree). It depends on how the neighbors of a node are connected to each other.

No, higher degree does not imply a higher clustering coefficient. For example node 5 with the highest degree in this network has the lowest clustering coefficient.

b) Degree: n-1; number of edges: $\binom{n}{2}$; clustering coefficient equals 1. The largest clique in the above graph has 4 nodes. There are 12 cliques of size 3.

c) In network A, high-degree nodes only connect to low-degree nodes. In network B, high-degree nodes also connect to other high-degree nodes. This increases assortativity in comparison to network A. Finally, in network C, low-degree nodes are also connected to other low-degree nodes. This further increases the assortativity of the network C in comparison to network B. Network A is the most disassortative and network C is the most assortative network.

Hubs in disassortative networks are preferentially connected to low-degree nodes (e.g. network A), and so they have lower clustering coefficient as compared with hubs in assortative networks that are connected to low-degree nodes, which may themselves be connected to each other. Therefore, the clustering coefficients of hubs in assortative networks will tend to be higher than in disassortative networks.

Exercise 3

a) $dm_1/dt=v_s-v_1-v_2-v_3$
$dm_2/dt=v_1+v_4-v_P$
$dm_3/dt=v_2-v_4-v_5$
$dm_4/dt=v_3+v_5-v_w$

b) $$S=\begin{pmatrix} 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 \end{pmatrix}$$

c) $v_w + v_p= v_s$ and $v_s=10$, $v_w=2$ therefore $v_p=8$

d) v=(10, x, x, 10-2x, y, (x-y),x+y,10-x-y), for {x, y: 0≤x≤5, 0≤y≤x}.

e) To maximize biomass growth, the « waste » flux v_w must be zero. It follows then that v_3 and v_5 must be equal to zero as well. This implies the following flux vector $v=(v_s, v_1, v_2, v_3, v_4, v_5, v_p, v_w)$ =(10, x, 10-x, 0, 10-x, 0,10,0). In addition, because the fluxes through reactions 1 and 2 are constrained to be equal, we have x=10-x, and thus x=5, such that the final flux vector becomes $v=(v_s, v_1, v_2, v_3, v_4, v_5, v_p, v_w)$ =(10, 5, 5, 0, 5, 0,10,0)