

1. Networks and degree distributions. Consider a network with N total vertices and E total undirected, unweighted edges.
 - (a) What is the probability f that an edge connects two vertices connected at random?
 - (b) What is the average number of neighbors J per protein?
 - (c) The probability p_n that a protein has exactly n neighbors follows a binomial distribution. Provide this distribution.
 - (d) Show that in the limit that $f \rightarrow 0$, $N \rightarrow \infty$, and J a finite constant, that the binomial probability distribution for p_n approaches a Poisson distribution. Provide the Poisson distribution.
 - (e) Based on the Poisson distribution, for what value of J do we expect on average only one vertex in the entire network to have no neighbors?

2. Network motifs. Again consider a network with N total vertices and E total edges. The edge probability f and the average degree J can be calculated from N and E and can also be used in your answers. You can assume the thermodynamic limit that $f \rightarrow 0$, $N \rightarrow \infty$, and $fN \rightarrow J$.
 - (a) What is the probability that three vertices, selected at random, have all edges (a 3-clique)? How many 3-cliques exist in the network?
 - (b) How many edges are possible for a group of k vertices? Given a set of k vertices, what is the probability that all these edges exist, giving a k -clique? How many k -cliques are expected in a random network?
 - (c) Now suppose that the N vertices are not randomly organized. Instead, there are M modules, each with n vertices within the module. Edges occur only within modules, with no edges between modules. For example, in a social network, modules might be families, workplaces, or university classes. In a biological network, modules might be protein complexes. In this case, f for edges between modules is 0. What is f within a module? What is the expected number of 3-cliques as a function of M ? What is the clustering coefficient as a function of M , defined as the number of 3-cliques for the given M relative to the number for a non-modular network with $M = 1$?

3. Disease spread, the giant component, and network diameter. Suppose a network has N total vertices, representing people. Each infected person transmits the disease to an additional J people.
 - (a) Define $C_n(J)$ as the number of people infected through a path of length n , with $C_0(J) = 1$ and $C_1(J) = J$. Provide an expression for the general case, $C_n(J)$.
 - (b) Define the total number of infected people as $T(J) = \sum_{n=0}^{\infty} C_n(J)$. Provide $T(J)$ for $J < 1$ and for $J > 1$. Remember that $T(J)$ must be $\leq N$.
 - (c) For $J > 1$, the network radius ρ is roughly defined as $C_\rho(J) = N$, the number of steps to infect the entire network. Provide ρ in terms of N and J .

4. Graph diffusion. A network has 4 vertices, numbered 1, 2, 3, 4. The network has 4 undirected edges: $1 \sim 2$, $1 \sim 3$, $2 \sim 3$, $3 \sim 4$.
- Provide the adjacency matrix \mathbf{A} and the degree matrix \mathbf{D} .
 - Define the matrix $\mathbf{A}(n)$ as having matrix elements $a(n)_{ij}$ that give the number of paths of length n starting at j and ending at i . Provide $\mathbf{A}(n)$ for $n = 0, 1, 2$. Describe how to calculate $\mathbf{A}(n)$ for other values of n .
 - Suppose that a random walk leaving a vertex chooses each of the outgoing edges with equal probability. Let the matrix $\mathbf{P}(n)$ have matrix elements $p(n)_{ij}$ giving the probability that an n -step random walk that started at j ends at i . Provide the matrix $\mathbf{P}(n)$ for $n = 0, 1, 2$. Describe how to calculate $\mathbf{P}(n)$ for other values of n .
5. Graph diffusion. Challenge problem that will not be on the exam. Now consider a general graph with adjacency matrix \mathbf{A} and degree matrix \mathbf{D} .
- The graph Laplacian \mathbf{L} is defined as $\mathbf{D} - \mathbf{A}$. Prove the the column vector $\mathbf{1}$, a column vector with each element 1, is an eigenvector of \mathbf{L} with eigenvalue 0.
 - Suppose that random walks are sampled over many lengths, with the probability of n steps following a Poisson distribution with mean λ . Define the matrix $\mathbf{P}(\lambda)$ as having matrix elements $p(\lambda)_{ij}$ giving the probability that a random walk following this process and starting at j ends at i . Provide a closed-form expression for $\mathbf{P}(\lambda)$.
 - Your answer to the previous question should have had the form $e^{-\lambda\mathbf{M}}$, where \mathbf{M} is a matrix. It is possible to prove that \mathbf{M} is positive semidefinite: it has at least one eigenvector with eigenvalue 0, and all other eigenvectors have eigenvalues ≥ 0 . Suppose that random walks are started at time 0 according to column vector β , with β_i equal to the probability that the random walk starts at vertex i . Prove that in the limit $\lambda \rightarrow 0$, the steady-state distribution is proportional to the eigenvector with eigenvalue 0.
 - Provide a closed-form expression for an eigenvector of \mathbf{M} with eigenvalue 0.
6. Flux-balance analysis of metabolic networks. Suppose that an organism has X muscle cells and Y neural cells. The organism can generate 1000 units of energy per minute. Muscle cells require 1 unit/min and neural cells require 2 units/min. The organism needs at least 200 muscle cells and 100 neural cells to survive.
- What are the coordinates defining the vertices of the feasible region in (X, Y) space?

(b) Suppose that the fitness $\phi(X, Y)$ of the organism is a linear function,

$$\phi(X, Y) = aX + bY,$$

with a and b positive. What condition on (a, b) permits multiple optimal solutions? What condition gives $X > Y$ at the optimum? What condition gives $X < Y$ at the optimum?