Problem 1.

Complete Exercise 6.1 in Newman.

Problem 2.

Complete Exercise 6.4 in Newman.

Problem 3.

The *node-edge incidence matrix* of a graph with n nodes and m edges is a matrix $\mathbf{B} \in \mathbb{R}^{2m \times n}$. There are two rows of $\mathbf{B}$ for each edge e. If e links nodes j and ℓ, then there is a row for the $j \to \ell$ "direction" and a row for the $\ell \to j$ "direction". So, we can write an individual entry of $\mathbf{B}$ as $b_{(j \to \ell),i}$. These entries are given by:

$$b_{(j \to \ell),i} = \begin{cases} -1 & i = j \\ +1 & i = \ell \\ 0 & \text{otherwise.} \end{cases}$$

Part (a)

The Laplacian matrix $\mathbf{L}$ of a graph is defined in eq. (6.29) of Newman. Prove using direct matrix multiplication that $\mathbf{L}$ can be computed using one of the two formulae below (and figure out which one):

$$\mathbf{L} = \frac{1}{2}\mathbf{B}^T\mathbf{B} \quad \text{or} \quad \mathbf{L} = \frac{1}{2}\mathbf{B}\mathbf{B}^T \,.$$

Part (b)

Use your result from Part (a) to give a very short proof that $\mathbf{L}$ is a positive-semidefinite matrix.

Problem 4.

In section 6.14.1, Newman considers the role of the Laplacian $\mathbf{L}$ in partitioning or "cutting" graphs into groups. Let's focus on the two-group case. In eq. (6.37), Newman defines an objective function

$$R(\mathbf{s}) = \frac{1}{4}\mathbf{s}^T\mathbf{L}\mathbf{s}\ , \tag{1}$$

where $\mathbf{s} \in \mathbb{R}^n$ is the vector with entries

$$s_i = \begin{cases} +1 & \text{node } i \text{ is in group 1,} \\ -1 & \text{node } i \text{ is in group 2.} \end{cases} \tag{2}$$

The idea is that a choice of $\mathbf{s}$ corresponds to a choice of groups for the nodes. Newman writes—somewhat uncarefully—that "... our goal is to find the vector $\mathbf{s}$ that minimizes the cut size (6.37) for given $\mathbf{L}$."

Assume throughout this problem that we are considering the Laplacian matrix $\mathbf{L}$ of a connected graph.

Part (a)

Find the vector $\mathbf{s}$ that minimizes $R(\mathbf{s})$.

Hint. There are multiple ways to do this, but carefully reading Chapter 6, section 14 of Newman is one.

Part (b)

Comment briefly (2-3 sentences is fine) on whether this vector is useful in the context of the graph partitioning problem.

Part (c)

Suggest a **heuristic** idea for how you might modify Newman's framework in order to generate more useful solutions. What kinds of condition could you impose on $\mathbf{s}$ in order to make it more useful for the partitioning problem?

Please **draw** a simple network to accompany your discussion, and show how your proposed modification might partition the network. **You are not responsible for proving or exactly calculating anything for this part**.

Problem 5.

In an upcoming part of the course, we will spend a lot of time working with *random graphs*. In this problem, you will begin an analysis of the famous Erdős-Rényi random graph model, which was first studied by Solomonoff and Rapoport.

A random graph model is a *probability distribution over graphs*. Many random graph models are specified as recipes for randomly generating a graph. Here's the recipe for the Erdős-Rényi model $G(n, p)$:

- Start with n nodes and no edges.
- Between each pair of distinct nodes, place an edge with probability p. Each edge placement event is independent from each other edge placement event.

Since $G(n, p)$ is a random graph, many of its properties can be described as random variables. Here are a few random variables:

- M, the total number of edges in the graph.
- K_i, the degree of node i.
- The number of components in the graph.

In this problem, you'll show a few things about some of these random variables.

Part (a)

Consider the degree K_i of node i. Argue that, if $p = c/n$ for some constant $c > 0$, then the degree of node i is approximately distributed according to a Poisson distribution **when n is very large** (i.e. $n \to \infty$). Carefully justify each step.

Hint. Homework 0.

Part (b)

Using your result from Part (a), compute the approximate mean and variance of K_i in terms of c, again in the limit $n \to \infty$.

We often report an estimate of a mean in the form $\mu \pm \sigma$, where σ is the standard deviation. Calculate the standard deviation σ. In a large Erdős-Rényi random graph in which each node has expected degree 100, what is the estimate $100 \pm \sigma$?

Note. This is usually considered to be a small amount of variation in the node degrees when compared against real-world network data.

Part (c)

Using your result from Part (a), compute the approximate probability mass function of M, the total number of edges in the graph. Find the approximate mean and variance of M. Here, you should assume that n is large but finite.

Problem 6.

The *degree distribution* of a random graph is an empirical probability distribution p such that p_k is the proportion of nodes with degree k. It is possible to prove that the degree distribution of an Erdős-Rényi random graph is approximately the same as the Poisson distribution for an individual node degree.

Write a computer program to produce two annotated plots.

- Generate an Erdős-Rényi random graph with 10^6 nodes and expected degree 100. Compute its degree distribution. Plot this distribution. Also plot (in a different color or symbol) a Poisson distribution with mean 100. Comment briefly on your findings.
- Acquire, in any way you choose, a network data set containing at least 10^3 nodes. Some data sets are available at the sources listed here, but you're free to look anywhere else as well. Compute and plot the degree distribution of your network. Again plot a Poisson distribution with mean equal to the mean degree of the network. Comment briefly on your findings.

Note. You may find it helpful to plot the degree distributions on log-log axes.