

Note: Before you attempt this problem you should read through Section 4.5 in Vershynin's book. We will be analyzing the bound given in Theorem 4.5.6 empirically. Note that this problem set is designed to be quick so that you can work on your final projects.

For $n = 50, 100, 250, 500, 750, 1000$, for a large number of $0 < q < p < 1$ values sample a graph $G \sim G(n, p, q)$, run the spectral clustering algorithm on the associated adjacency matrix, and compute the number of misclassifications the algorithm makes. For each n value, make a separate plot which shows how the number of misclassifications depends on the quantity $\mu = \min(q, p - q)$ used in Theorem 4.5.6. Is the upper bound in the book accurate in the sense that the empirical upper bound decays as it does in the theory?

■