Problem Set 2

[Your name] and [student ID]
MAT1841-2021

Problem 1 [BHK 4.14] (20 points). Consider a distribution ı over $\{0,1\}^2$ with $p(00) = p(11) = \frac{1}{2}$ and $p(01) = p(10) = 0$. Give a connected graph on $\{0,1\}^2$ that would be bad for running Metroplis-Hastings and a graph that would be good for running Metropolis-Hastings. What would be the problem with Gibbs sampling?

Problem 2 [BHK 4.17] (30 points). Suppose one wishes to generate uniformly at random a regular, degree three, undirected, not necessarily connected multi-graph with 1,000 vertices. A multi-graph may have multiple edges between a pair of vertices and self loops. One decides to do this by a Markov Chain Monte Carlo technique. In particular, consider a (very large) network where each vertex corresponds to a regular degree three, 1,000 vertex multi-graph. For edges, say that the vertices corresponding to two graphs are connected by an edge if one graph can be obtained from the other by a flip of a pair of edges. In a flip, a pair of edges (a, b) and (c, d) are replaced by (a, c) and (b, d).

1. Prove that the network whose vertices correspond to the desired graphs is connected. That is, for any two 1000-vertex degree-3 multigraphs, it is possible to walk from one to the other in this network.
2. Prove that the stationary distribution of the random walk is uniform over all vertices.
3. Give an upper bound on the diameter of the network.
4. How would you modify the process if you wanted to uniformly generate connected degree three multi-graphs?

Problem 3 [BHK 4.47] (20 points). A researcher was interested in determining the importance of various edges in an undirected graph. They computed the stationary probability for a random walk on the graph and let p_i be the probability of being at vertex i. If vertex i was of degree d_i, the frequency that edge (i, j) was traversed from i to j would be $\frac{1}{d_i}p_i$ and the frequency that the edge was traversed in the opposite direction would be $\frac{1}{d_j}p_j$. Thus, they assigned an importance of $\left|\frac{1}{d_i}p_i - \frac{1}{d_j}p_j\right|$ to the edge. What is wrong with their idea?

Problem 4 [BHK 4.55] (30 points). Create a random directed graph with 200 vertices and roughly 8 out-edges per vertex. Add k new vertices and calculate the pagerank with and without directed edges from the k added vertices to vertex 1. How much does adding the k edges change the pagerank of vertices for various values of k and restart frequency? How much does adding a loop at vertex 1 change the pagerank? To do the experiment carefully one needs to consider the initial pagerank of a vertex to which the star of k new vertices is attached.