Math 4740: Homework 6 Solutions: Additional Problem

(a) Let Q be the transition matrix of the proposal chain. For any two distinct colorings $f, g : V \to C$,

$$Q(f,g) = \begin{cases} \frac{1}{|V|\cdot|C|} & \text{if } f \text{ and } g \text{ differ at exactly one vertex,} \\ 0 & \text{if } f \text{ and } g \text{ differ at more than one vertex.} \end{cases}$$

Therefore $Q(f,g) = Q(g,f)$, so the Metropolis transition matrix P has the following formula: if $f \neq g$, then $P(f,g) = Q(f,g)R(f,g)$ where

$$R(f,g) = \min\left\{1, \frac{\pi(g)}{\pi(f)}\right\}.$$

We compute

$$\frac{\pi(g)}{\pi(f)} = \frac{\alpha^{N(g)}/Z}{\alpha^{N(f)}/Z} = \alpha^{N(g)-N(f)}.$$

Suppose f and g differ at a single vertex v. It is actually quicker to compute the difference $N(g)-N(f)$ than to compute either $N(f)$ or $N(g)$ individually. Call an edge of the graph "good" if the two vertices that it connects have different colors, and "bad" if its two vertices have the same color. Then $N(f)$ counts the number of bad edges of the graph when the vertices are colored according to f. If we start with f and recolor the vertex v to make the new coloring g, the only edges that can change between bad and good are the edges between v and its neighbors. Based on this insight, we define $N_v(f)$ to be the number of vertices w that are neighbors of v for which $f(v) = f(w)$, that is, $N_v(f)$ is the number of bad edges coming out of v. Then $N(g)-N(f) = N_v(g)-N_v(f)$. If the graph is very large, it takes much less time to compute $N_v(g) - N_v(f)$ than to compute $N(f)$ or $N(g)$.

Here is a description of the transition rule. Say that (X_n) is the Metropolis chain on graph colorings. Starting from $X_n = f$, define the proposal coloring g by choosing a vertex v uniformly at random and assigning a new color to v also uniformly at random from C. If $g = f$ (because the new color of v is the same as its original color), then set $X_{n+1} = g = f$. If $g \neq f$, then compute the difference $N_v(g) - N_v(f)$. If $N_v(g) - N_v(f) \leq 0$, so that g has the same number of bad edges as f or fewer, then set $X_{n+1} = g$. If $N_v(g) - N_v(f) > 0$, then flip a coin with heads probability of $\alpha^{N_v(g)-N_v(f)}$. Set $X_{n+1} = g$ if heads and $X_{n+1} = f$ if tails.

(b) When $\alpha = 1$, the rule described above always sets $X_{n+1} = g$. That is, the proposal from Q is always accepted, so $P = Q$.

(c) Sending $\alpha \to 0$ in the transition rule from (a) results in the following: If $N_v(g) - N_v(f) \leq 0$, then set $X_{n+1} = g$. If $N_v(g) - N_v(f) > 0$, then set $X_{n+1} = f$. Therefore every proposal that decreases the number of bad edges or keeps the same number of bad edges is accepted, and every proposal that increases the number of bad edges is rejected.

Note that the *derivation* of the transition rule makes no sense when $\alpha = 0$, since π is not defined. Nevertheless the transition rule *itself* is well-defined, and given in the previous paragraph. In fact, it will be seen in part (d) that the Metropolis chain with $\alpha = 0$ may have more than one communicating class of recurrent states, so it may not have a unique stationary distribution.

(d) The transition rule from part (c) rejects all proposals that increase the number of bad edges. In the given example coloring, every proposal that changes the color of a single vertex must increase the number of bad edges, so it will be rejected. Therefore the given coloring is an absorbing state for the Metropolis chain.

To go into more detail: The only bad edge in the example is the one connecting the two red vertices. If one of the red vertices is changed to a different color, then that edge becomes good, but two good edges become bad. If one of the other vertices is changed, then the bad edge does not become good, and at least one previously good edge turns bad.

(e) We'll answer the second and third questions before the first question. When α is very close to 1, the Metropolis chain barely distinguishes between colorings with many versus few bad edges, so it may take a long time to stumble on a proper coloring. When α is very close to 0, the Metropolis chain may spend a long time "trapped" at suboptimal colorings such as the example from part (d), where the probability of accepting any proposal is at most α. Therefore a value of α somewhere in the middle is probably optimal.

Adding more colors diminishes the effectiveness of a "trap" like the one in part (d). In that example, if a fifth color were available, even when $\alpha = 0$ the Metropolis chain would accept proposals using that color. Likely it would not take too long to find a proper coloring. A reasonable guess is that the more colors available, the lower α can safely be set, and the faster the chain will find a proper coloring.

Finally, the first question. Consider the "greedy algorithm" for finding a proper coloring: Start with one vertex, color it with any color. Move to the next vertex and give it a color that is different from all its neighbors, if available. Keep going until all vertices are colored. Let Δ be the highest degree of a vertex in the graph. If the total number of colors $|C|$ is at least $\Delta + 1$, then the greedy algorithm will always give a proper coloring. If $|C| \leq \Delta$ then the greedy algorithm can get trapped. I personally would guess that in this scenario, the Metropolis chain will find a proper coloring reasonably quickly if it exists. But if anyone could prove such a statement, it would be worthy of a published paper!

The existing research on this algorithm focuses on the case where $|C| \geq \Delta+2$ and $\alpha = 0$. In that situation, all the non-proper colorings are transient states of the Markov chain, and there is a single communicating class that contains all the proper colorings. The stationary distribution is uniform on the set of proper colorings. Since $|C| > \Delta$, finding a proper coloring is easy via the greedy algorithm. The question is whether the Metropolis chain quickly finds a *random* proper coloring, that is, how long it takes for the chain to converge to its stationary distribution. It is conjectured that if the graph has n vertices and $|C| \geq \Delta + 2$, the number of steps until the chain is roughly stationary should be on the order of $n \log n$. This has been proved when $|C| > 2\Delta$, and there are partial results for $|C| = \alpha\Delta$ when $1 < \alpha < 2$, but the original conjecture seems quite difficult. For more information, see:

http://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf (Ch.14)

http://www.cc.gatech.edu/~vigoda/survey.pdf