Torpid Mixing of the Wang-Swendsen-Kotecký Algorithm for Sampling Colorings

TOMASZ ŁUCZAK, *Department of Discrete Mathematics, Adam Mickiewicz University, 60-769 Poznań, Poland, tomasz@amu.edu.pl.*

ERIC VIGODA, *Division of Informatics, University of Edinburgh, The King's Buildings, Edinburgh EH1 3JZ, United Kingdom, vigoda@dcs.ed.ac.uk.*

ABSTRACT: We study the problem of sampling uniformly at random from the set of k-colorings of a graph with maximum degree Δ. We focus attention on the Monte Carlo Markov chain method, particularly on a popular Markov chain for this problem, the Wang-Swendsen-Kotecký (WSK) algorithm. The second author recently proved that the WSK algorithm quickly converges to the desired distribution when $k \geq \frac{11}{6}\Delta$. We study how far these positive results can be extended in general. In this note we prove the first non-trivial results on when the WSK algorithm takes exponentially long to reach the stationary distribution and is thus called *torpidly mixing*. In particular, we show that the WSK algorithm is torpidly mixing on a family of bipartite graphs when $3 \leq k < \frac{\Delta}{20 \log \Delta}$, and on a family of planar graphs for any number of colors. We also give a family of graphs for which, despite their small chromatic number, the WSK algorithm is not ergodic when $k \leq \frac{\Delta}{2}$, provided k is larger than some absolute constant k_0.

Keywords: Randomized algorithms, Markov chain Monte Carlo, coloring, torpid mixing

1 Introduction

We consider the problem of sampling uniformly at random from the set Ω of proper k-colorings of a graph $G = (V, E)$ with maximum degree Δ, which corresponds to sampling from the Gibbs distribution of the zero-temperature anti-ferromagnetic Potts model [10]. A proper k-coloring σ is a labelling $\sigma : V \to C = [k]$ such that all neighboring vertices have different colors. Previous work on this problem has centered on the Markov chain Monte Carlo method. The goal of the method is to design a Markov chain whose stationary distribution is uniform over Ω and which has small *mixing time*, the time to get (close) to stationarity. A Markov chain whose mixing time is polynomial in $n = |V|$ is known as *rapidly mixing*, while a *torpidly mixing* chain has a mixing time which is exponential in n^ϵ for some $\epsilon > 0$. Two particular Markov chains are typically considered: the Wang-Swendsen-Kotecký (WSK) algorithm [12, 13] and the simple Glauber dynamics [5]. Recently the second author proved that both of these chains are rapidly mixing when $k > \frac{11}{6}\Delta$ [11]. We study how far these positive results can be extended in general and give examples of when the algorithms are torpidly mixing.

We begin by defining the Markov chains associated with the Glauber dynamics and the WSK algorithm. From a coloring σ, the transitions $\sigma \mapsto \sigma'$ are the following:

- Choose a vertex v and color c uniformly at random.
- In the induced graph on vertices with colors c or $\sigma(v)$ in σ, consider the component S containing v. The new coloring is created in the following manner.
  WSK algorithm: 'Flip' the set S by interchanging colors c and $\sigma(v)$ on it.
  Glauber dynamics: If $|S| = 1$, then flip S, otherwise the coloring remains the same.

As one might expect, rapid mixing of the Glauber dynamics implies rapid mixing of the WSK algorithm. This implication is easily proven (e.g., using the approach of Diaconis and Saloff-Coste [3]), however the reverse implication is not true. For example, on the star with n vertices and $k = n^{1-\epsilon}$ (for $\epsilon > 0$) colors, the Glauber dynamics is torpidly mixing while the WSK algorithm is rapidly mixing (see the remark at the end of section 3). For these reasons we focus our attention on the mixing properties of the WSK algorithm.

The WSK algorithm is particularly appealing since it is ergodic on any bipartite graph with any number of colors (see [2, 4]). Lubin and Sokal [9] have also proven that the WSK algorithm is not ergodic for three colors on a periodic square lattice of size $3m \times 3n$ where m and n are relatively prime (when m or n is odd, as in this case, the graph is not bipartite). Ergodicity in other cases remains an open problem.

The natural conjecture is that the WSK algorithm is rapidly mixing for all $k > \Delta$. However, if we restrict the chromatic number of G, assuming perhaps that G is bipartite, can we expect to prove rapid mixing for a much larger range of k? Furthermore, can Δ be replaced by another quantity such as the average degree, and thus obtain results useful, say, for planar graphs. This relates to the following conjecture of Dominic Welsh [14]:

Conjecture *For any fixed $k \geq 4$ there is no fully polynomial randomized approximation scheme (fpras) for counting the number of k-colorings of a planar graph.*

In this note we give counter-examples to some of the previous questions. We begin with a simple example which shows that no lower bound of k, expressed solely in terms of the average degree of G can guarantee that the WSK algorithm is rapidly mixing on G.

THEOREM 1.1
There exists a family of planar graphs F_n such that the WSK algorithm is torpidly mixing on F_n for every fixed $k \geq 3$.

We next show that the WSK algorithm is not ergodic on a class of graphs $H(k, n)$, formally defined in section 3, when $k \leq \Delta/2$, despite their small chromatic number.

THEOREM 1.2
For a sufficiently large constant k_0, there exists a family of graphs $H(k, n)$ with maximum degree Δ such that the WSK algorithm is not ergodic (more precisely, not irreducible) on $H(k, n)$ for every k such that $k_0 \leq k \leq \Delta/2$.

Finally, we show torpid mixing of the WSK algorithm on a class of bipartite graphs $G(k, n)$ defined in section 3.

THEOREM 1.3
There exists a family of bipartite graphs $G(k, n)$ with maximum degree Δ such that the WSK algorithm is torpidly mixing on $G(k, n)$ when $3 \leq k < \frac{\Delta}{20 \log \Delta}$.

2 Background

A discrete-time Markov chain with transition probability matrix P defined on a finite state space Ω is called *ergodic* if it satisfies the following two properties:

- *aperiodicity*: $\gcd\{t : P^t(i, i) > 0\} = 1$ for all $i \in \Omega$;
- *irreducibility*: for all $i, j \in \Omega$, there exists a $t = t_{ij}$ such that there is a positive probability of going from state i to state j after t steps, i.e., $P^t(i, j) > 0$.

An ergodic Markov chain has a unique stationary distribution π. Moreover, if P is symmetric (i.e., $P(i, j) = P(j, i)$ for all i, j), then π is uniform over all states. Our goal is to bound the time until the chain is sufficiently close to the stationary distribution. From an initial state i, the total variation distance from π is

$$d_i(t) = d_{TV}(P^t(i, \cdot), \pi) = \frac{1}{2} \sum_{j \in \Omega} |P^t(i, j) - \pi(j)|.$$

We are interested in the *mixing time*, defined as:

$$\tau = \min\{t : 2d_i(t') \leq 1/e \text{ for all } t' \geq t \text{ and all } i \in \Omega\}.$$

To prove torpid mixing, we rely on the notion of conductance. The conductance Φ_S of a non-empty set S is

$$\Phi_S = \frac{\sum_{i \in S, j \in \Omega \setminus S} \pi(i) P(i, j)}{\pi(S)}.$$

It is well-known (e.g., see [1]) that

$$1/\tau \leq 2 \min_{0 < \pi(S) \leq 1/2} \Phi_S.$$

We use the following consequence of this result. Define the boundary of a set $S \subset \Omega$ as

$$\partial S = \{i \in S : P(i, j) > 0 \text{ for some } j \in \overline{S}\}.$$

Then,

$$\Phi_S \leq \sum_{i \in \partial S} \pi(i)/\pi(S) = \pi(\partial S)/\pi(S).$$

This implies the following theorem.

THEOREM 2.1
If for some $S \subset \Omega$, where $0 < |S| \leq |\Omega|/2$, the ratio $|\partial S|/|S|$ is exponentially small in n^ϵ for some $\epsilon > 0$, then the Markov chain is torpidly mixing. □

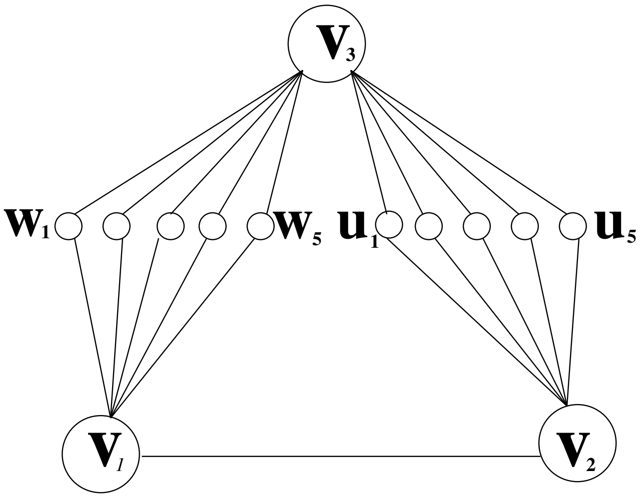

FIG. 1. Graph F_5

3 Proofs

We begin by introducing some simple notation. For a coloring $\sigma \in \Omega$, graph G, and pair of colors $c, c' \in C$, let $G_\sigma(c, c')$ denote the induced subgraph of G on the set of vertices that have colors c or c' in σ.

Let F_n denote the planar graph on $2n + 3$ vertices, specifically the vertex set is $\{v_1, v_2, v_3, w_1, \ldots, w_n, u_1, \ldots, u_n\}$, and the following edges: v_1 is adjacent to v_2; the vertices $w_1, \ldots, w_n$ are adjacent to both v_1 and v_3; finally, the vertices $u_1, \ldots, u_n$ are adjacent to both v_2 and v_3.

Proof of Theorem 1: Let S denote the set of colorings in which vertices v_1 and v_3 have the same color. By symmetry there is an equal number of colorings where v_2 and v_3 have the same color, thus $|S| \leq |\Omega|/2$. For $\sigma \in S$, where $\sigma(v_1) = \sigma(v_3) = c$, consider whether there exists a transition which recolors either, but not both, of the vertices v_1 or v_3 to a new color c' and thus leaves the set S. In order for such a transition to exist, v_1 and v_3 must lie in different components in $G_\sigma(c, c')$; this is the case if none of the vertices $w_1, \ldots, w_n$ have color c'. Therefore the boundary ∂S consists of all colorings in which $\{w_1, \ldots, w_n\}$ are colored by at most $k - 2$ colors. This implies that

$$\frac{|\partial S|}{|S|} \leq \frac{(k-1)(k-2)^n}{(k-1)^n} \leq \exp\Big(-\frac{n-1}{k-1}\Big).$$

The assertion now follows from Theorem 2.1. □

For natural numbers k, n, let $H(k, n)$ denote a graph with vertex set $V_1 \cup \cdots \cup V_k$ such that the induced subgraph on each set of vertices V_i is an independent set of size n, and for every $1 \leq i < j \leq k$ the induced subgraph on vertices $V_i \cup V_j$ is a cycle of length $2n$. Moreover, we assume that $H(k,n)$ contains no cycles on fewer than five vertices. One can show that, for a given k and sufficiently large n, such a graph exists using either an elementary probabilistic argument, or the following simple "switching" procedure. Take any graph H of the above structure which, perhaps, contains some short cycles of length at most five, and let $\{u,w\}$, $u \in V_i$, $w \in V_j$, be an edge which belongs to such a cycle. The number of vertices of H which are within distance six from u is at most $(2k)^7$, so there exist $u' \in V_i$, $w' \in V_j$, such that $\{w',u'\}$ is an edge of the graph, both w' and u' are at distance at least seven from u, and, finally, if we replace the edges $\{u,w\}$ and $\{u',w'\}$ by $\{u,w'\}$ and $\{u',w\}$ the subgraph induced by $V_i \cup V_j$ in the graph obtained in this way is a cycle. Note that such a switch destroys one short cycle not creating any new ones, so in this way one can eliminate all short cycles from the graph one by one.

Proof of Theorem 2: Consider a coloring σ in which each set V_i is monochromatic with a distinct color. A transition of the WSK algorithm only swaps colors among the sets V_i while maintaining the property that each set V_i is monochromatic with a distinct color. Any coloring accessible from σ must use exactly k colors, but there do exist colorings which use fewer than k colors. Since $H(k,n)$ is a graph of girth five with maximal degree $2k$, its chromatic number is bounded from above by $(1+o(1))2k/\log(2k) < k$ for sufficiently large k (see Kim [8]). This covers the case when $k = \Delta/2+1$, but the maximum degree can be arbitrarily increased by fixing a vertex v and adding extra vertices each with an edge only to v. □

For $r = \lfloor 20k\log k \rfloor \geq 65$, let $G(k,n)$ be an r-regular bipartite graph with vertex set $V_1 \cup V_2$, where $|V_1| = |V_2| = n$, such that:

(A) Each set W, such that $W \subset V_i$ for some $i = 1,2$ and $|W| \leq n/6r$, has at least $2r|W|/3$ neighbors;

(B) Each pair of sets W_1, W_2, such that $W_i \subseteq V_i$ and $|W_i| \geq (3n\log r)/r$ for $i = 1,2$, is joined by an edge.

One can show, using the first moment method (cf. Janson, Łuczak, Ruciński [6]) that for any fixed $r \geq 65$ and sufficiently large n, that with probability at least $1/2$ the random r-regular graph has both of the above properties. Thus, for a given k and sufficiently large n, there exists a graph $G(k,n)$ which has both properties (A) and (B).

Proof of Theorem 3: For $\sigma \in \Omega$, let $V_1^c = V_1^c(\sigma)$ (similarly V_2^c) denote the set of vertices $v \in V_1$ ($v \in V_2$, respectively) such that $\sigma(v) = c$. We focus attention on the dominant color of V_1. For $\sigma \in \Omega$, let $c_* = c_*(\sigma)$ denote a color where $|V_1^{c_*}| \geq |V_1^c|$ for all $c \in C$. We denote the dominant color class by $V_1^* = V_1^{c_*}$ and the remaining portion by $V_1^{rest} = V_1 \setminus V_1^*$. We will bound the conductance of the set $S = \{\sigma \in S : |V_1^*| \geq n(1-1/6r)\}$. A necessary condition for a coloring σ to lie in the boundary ∂S is that there must exist a color c and a component X in the induced graph $G_\sigma(c,c_*)$ such that $|X \cap V_1^*| \geq |V_1^*| - n(1-1/6r)$, i.e., after flipping X we would have $|V_1^*| < n(1-6/r)$. We consider the set $S' \subset S$ consisting of colorings σ which satisfy any of the following conditions:

(i) $|V_1^{rest}| > n/12r$,
(ii) there exists a color $c \neq c_*$ such that $|V_2^c| < n/2k$,
(iii) there exists a color $c \neq c_*$ such that at least $n/12r$ vertices of V_1^* do not belong to the largest component in the graph $G_\sigma(c_*, c)$.

Note that the conditions (i) and (iii) imply that $\partial S \subseteq S'$ and thus it is sufficient to bound $|S'|/|S|$. We begin by estimating the number of colorings S_1' in which condition (i) holds, i.e., at least $n/12r$ vertices of V_1 do not have the dominant color c_*. In order to build such a coloring we first need to choose V_1^{rest}, where $j = |V_1^{rest}|$ and $n/12r \leq j \leq n/6r$. There are k choices for c_*, $\binom{n}{j}$ ways of choosing the set V_1^{rest}, and $(k-1)^j$ ways of coloring the chosen set. From property (A) of the graphs $G(k,n)$, we know that the number of neighbors of the set V_1^{rest} is at least $2jr/3$ and that any set of at least $n/4r^2$ vertices from V_2 has some neighbor in V_1^*. Thus, at least $2jr/3 - n/4r^2$ vertices of V_2 have at least two distinct colors in their neighborhood, and the number of colorings for which (i) holds is bounded from above by

$$
\begin{aligned}
|S_1'| &\leq k \sum_{j=n/12r}^{n/6r} \binom{n}{j} (k-1)^{j+n-2jr/3+n/4r^2} (k-2)^{2jr/3-n/4r^2} \\
&\leq k(k-1)^n \sum_{j=\frac{n}{12r}}^{\frac{n}{6r}} \left(\frac{ne(k-1)}{j} \left(1 - \frac{1}{k-1}\right)^{2r/3} \right)^j \left(1 - \frac{1}{k-1}\right)^{-n/4r^2} \\
&\leq nk(k-1)^n \left(12er^2(k-1) \left(1 - \frac{1}{k-1}\right)^{2r/3-3/2r} \right)^{n/12r} \\
&\leq n(k-1)^n \left(12er^3 \left(1 - \frac{1}{k-1}\right)^{2r/3} \right)^{n/12r} .
\end{aligned}
$$

Since $|S| \geq (k-1)^n$, we conclude that

$$
\begin{aligned}
\frac{|S_1'|}{|S|} &\leq n\Big(12er^3 \Big(1 - \frac{1}{k-1}\Big)^{2r/3} \Big)^{n/12r} \\
&\leq n\Big(12er^3 \exp\Big(- \frac{2r}{3(k-1)} \Big) \Big)^{n/12r} \\
&\leq e^{-n/12r} .
\end{aligned}
$$

Now consider those colorings S_2' for which condition (ii) holds, but condition (i) does not hold. Fix a coloring of the set V_1 where $|V_1^*| \geq n(1 - 1/12r)$. For each vertex of V_2, choose a color uniformly at random from the set of available colors, i.e., from those colors which do not appear in its neighborhood in V_1. Clearly each element from S with the prescribed coloring of V_1 is equally likely to emerge as the result of this procedure. Furthermore at most $n/12$ vertices from V_2 have a neighbor in the set V_1^{rest}, i.e., at least $11n/12$ vertices of V_2 have $k-1$ available colors. Thus the probability that there exists a color $c \neq c_*$ such that $|V_2^c| < n/2k$ is bounded from above by $k \Pr\{B(\lfloor 11n/12 \rfloor, 1/(k-1)) \le n/2k\}$, where $B(\lfloor 11n/12 \rfloor, 1/(k-1))$ is a binomial random variable. Using the well known Chernoff's bounds for the tail of the binomial distribution, for all sufficiently large n,

$$\frac{|S_2'|}{|S'|} \le k \Pr\{B(\lfloor 11n/12 \rfloor, 1/(k-1)) \le n/2k\} \le \exp(-n/20k) \; .$$

Finally, we deal with the colorings S_3' from S which fulfill condition (iii) but neither conditions (i) or (ii). Recall that we now have that $|V_1^{rest}| \le n/12r$. For a color c and coloring $\sigma \in \Omega$, let $X_1^* = X(c,\sigma)$ denote the largest component in the graph $G_\sigma(c, c_*)$. From condition (ii) and property (B), we may assume that for every color c the component X_1^* is a giant component in the sense that $|V_1 \setminus X_1^*| \le (3n \log r)/r$. Thus, $X_1^* \subset V_1^*$ and let $Y_1^* = V_1^* \setminus X_1^*$, i.e., Y_1^* denotes the set of vertices with the dominant color that are not in the largest component. We can now estimate the probability that if we randomly color the vertices of V_2 then there exists a color c such that $|Y_1^*| = m \ge n/12r$.

For a fixed set Y_1^*, from property (A) we know that the number of vertices of V_2 which have some neighbor in the set Y_1^* and no neighbor in the set V_1^{rest} is at least

$$\frac{2r}{3}(|V_1^{rest}| + |Y_1^*|) - r|V_1^{rest}| = \frac{2r}{3}|Y_1^*| - \frac{r}{3}|V_1^{rest}|.$$

Moreover, also by property (A), at most $3|Y_1^*|/2r$ such vertices have no neighbor in the set X_1^*, i.e., in the largest component of the graph $G_\sigma(c_*, c)$. Let $Z_2 = Z_2(c, Y_1^*) \subset V_2$ denote the set of vertices which have no neighbors in V_1^{rest} and at least one neighbor in each of Y_1^* and X_1^*. Then,

$$|Z_2| \ge \frac{2r}{3}|Y_1^*| - \frac{r|V_1^{rest}|}{3} - \frac{3}{2r}|Y_1^*| \ge \frac{4r^2-9}{6r}|Y_1^*| - \frac{r}{3}\frac{n}{12r} \ge \frac{12r}{19}|Y_1^*| - \frac{n}{36}.$$

Vertices of Z_2 only have color c_* in their neighborhood, but if any vertex of Z_2 has color c then it joins some vertex of the set Y_1^* to the largest component of the graph $G_\sigma(c^*, c)$, which would contradict our earlier assumption. Now the probability that, for a given color c and set Y_1^*, no vertex of the set Z_2 is given color c is bounded from above by

$$\left(1 - \frac{1}{k-1}\right)^{\frac{12r}{19}|Y_1^*| - \frac{n}{36}} \le \exp\left(-\frac{1}{k-1}\left(\frac{12r}{19}|Y_1^*| - \frac{n}{36}\right)\right) .$$

We can now bound $|S_3'|$ by summing over the possible choices for c and Y_1^*,

$$\begin{aligned}
\frac{|S_3'|}{|S|} &\le (k-1) \sum_{m=n/12r}^{3n\log r/r} \binom{n}{m} \exp\left(-\frac{1}{k-1}\left(\frac{12r}{19}m - \frac{n}{36}\right)\right) \\
&\le kn \binom{n}{n/12r} \exp\left(-\frac{r}{k-1}\left(\frac{12}{19} - \frac{1}{3}\right)\right)^{n/12r} \\
&\le \left(12er \exp\left(-\frac{16r}{56(k-1)}\right)\right)^{n/12r} \\
&\le e^{-n/12r}.
\end{aligned}$$

Therefore, we arrive at

$$\frac{|\partial S|}{|S|} \leq \frac{|S'|}{|S|} \leq e^{-n/12r} + e^{-n/20k} + e^{-n/12r} \leq 3e^{-n/12r}\,.$$

By Theorem 2.1, this completes the proof of Theorem 1.3. $\square$

Remark. For completeness we explain our comparison of the mixing times of the Glauber dynamics and the WSK algorithm on the star. The star on $n+1$ vertices is the graph composed of vertex set $\{v, w_1, \ldots, w_n\}$ with an edge between vertices v and w_i for all $1 \leq i \leq n$. To see that the Glauber dynamics is torpidly mixing when $k \leq n^{1-\epsilon}$ for fixed $\epsilon > 0$, let the set S denote the set of colorings with a fixed color c for vertex v. In order to recolor v (and thus exit the set S) there must be some color which does not appear in its neighborhood. Therefore $|\partial S|/|S| \leq ((k-2)/(k-1))^n \leq \exp(-n^\epsilon)$, which, by Theorem 2.1, implies that the Glauber dynamics is torpidly mixing. Meanwhile proving rapid mixing of the WSK algorithm is straight-forward using the canonical paths approach of Jerrum and Sinclair (see [7] for a explanation of the technique).

Acknowledgements: This paper was written while we were both visiting Microsoft Research. We thank Microsoft Research, especially the theory group, for providing this opportunity. We also wish to thank Marek Biskup, Christian Borgs, and Jennifer Chayes for a stimulating discussion on the subject, and Mark Jerrum and Alan Sokal for their insightful comments on the first version of the manuscript. Finally, the first author acknowledges the support of KBN grant 2 P03A 021 17.

References

[1] D.J. Aldous and J.A. Fill. Reversible Markov chains and random walks on graphs. Some chapters available at http://stat-www.berkeley.edu/pub/users/aldous/book.html. In preparation.

[2] J. K. Burton Jr. and C. L. Henley. A constrained Potts antiferromagnet model with an interface representation. *Journal of Physics A*, 30(24):8385–8413, 1997.

[3] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.

[4] S. J. Ferreira and A.D. Sokal. Antiferromagnetic Potts models on the square lattice: A high-precision Monte Carlo study. *Journal of Statistical Physics*, 96(3), 1999.

[5] R. Glauber. Time-dependent statistics of the Ising model. *J. Math. Phys.*, 4:294–307, 1963.

[6] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley, New York, 2000.

[7] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In D. Hochbaum, editor, *Approximation algorithms for NP-hard problems*. PWS Publishing Company, 1996.

[8] J.H. Kim. On Brooks' theorem for sparse graphs. *Combin. Probab. Comput.*, 4:97–132, 1995.

[9] M. Lubin and A. D. Sokal. Comment on 'antiferromagnetic Potts models'. *Phys. Rev. Lett.*, 71:1778, 1993.

[10] R. Potts. Some generalized order-disorder transformations. *Proc. Camb. Phil. Soc.*, 48:106–109, 1952.

[11] E. Vigoda. Improved bounds for sampling colorings. In *40th Annual Symposium on Foundations of Computer Science*, pages 51–59, New York City, October 1999.

[12] J. S. Wang, R. H. Swendsen, and R. Kotecký. Antiferromagnetic Potts models. *Phys. Rev. Lett.*, 63, 1989.

[13] J. S. Wang, R. H. Swendsen, and R. Kotecký. Three-state antiferromagnetic Potts models: A Monte Carlo study. *Phys. Rev. B*, 42, 1990.

[14] D. Welsh. *Complexity:Knots, Colourings, and Counting*. London Mathematical Society Lecture Note Series 186, 1993.

Received 10 February 2000.