Approximating the Tutte polynomial

Mark Jerrum
University of Edinburgh

CLARENDON PRESS · OXFORD
2006

PREFACE

We consider some algorithmic problems associated with matroids. These problems are computationally intractable if we insist on exact solutions, so we concentrate instead on producing approximate solutions within specified relative error. Technically we shall be seeking a "fully polynomial randomised approximation scheme" or "FPRAS". First we consider the problem of counting matroid bases. Feder and Mihail presented a polynomial time algorithm for approximating the number of bases in "balanced" matroids. We describe their algorithm, and present an improved analysis due to Jerrum and Son. Then we widen the discussion to the considerably more general problem of (approximately) evaluating the Tutte polynomial. This is a two-variable polynomial $T(M;x,y)$ associated with a matroid M that encodes much information about M. In particular, the number of bases of M is equal to $T(M;1,1)$. We review what is known about the complexity of approximating the Tutte polynomial, and extend the boundary some way. (This section describes recent joint work with Leslie Goldberg.) We conclude with some speculations.

CONTENTS

1

PRELIMINARIES

Let E be a ground set of size m and $\mathcal{B} \subseteq 2^E$ a collection of subsets of E. We say that $\mathcal{B}$ forms the collection of *bases* of a *matroid* $M = (E, \mathcal{B})$ if the following *exchange axiom* holds:

> For every pair of bases $X, Y \in \mathcal{B}$ and every element $e \in X \setminus Y$, there exists an element $f \in Y \setminus X$ such that $X \cup \{f\} \setminus \{e\} \in \mathcal{B}$.

It is an easy consequence of the exchange axiom that all bases have the same size, and this is the *rank* r of M. The exchange axiom captures the notion of linear independence. Thus if $S = \{u_0, \ldots, u_{m-1}\}$ is a set of n-vectors over a field K, then the maximal linearly independent subsets of S clearly satisfy the exchange axiom, and hence form the bases of a matroid with ground set S. The rank of this matroid is the dimension of the vector space spanned by S. A matroid that arises in this way is *vectorial*, and is said to be *representable over* K. A matroid that is representable over every field is *regular*.

Several other equivalent axiomatisations of matroid are possible, each shedding different light on the notion of linear independence, but the above choice turns out to be the most appropriate for our needs. For other possible axiomatisations, and more on matroid theory generally, consult (Oxley, 1992) or (Welsh, 1976).

The advantage of the abstract viewpoint provided by matroid theory is that it allows us to perceive and exploit formal linear independence in a variety of combinatorial situations. Most importantly, the spanning trees in an undirected graph $G = (V, E)$ form the bases of a matroid of rank $r = |V| - 1$, the *cycle matroid of* G, with ground set E. A matroid that arises as the cycle matroid of some graph is called *graphic*. It is well known that the number of spanning trees of a graph may be computed efficiently, specifically in time $|V|^3$, using Kirchhoff's Matrix-tree Theorem. Perhaps less well know is the fact that the same basic approach extends to counting the bases of a regular matroid. (Regular matroids are a strict superset of graphic matroids.) It can be shown that the bases of a regular matroid are in 1-1 correspondence with the non-singular $r \times r$ submatrices of an $r \times m$ unimodular matrix, and that the number of these can be computed using the Binet-Cauchy formula. Refer to (Dyer and Frieze, 1994, §3.1) for more detail.

For reasons explained in §3.1, we do not expect to find an efficient algorithm for computing the number of bases for much wider classes of matroids. However, there is no reason to suppose that we cannot efficiently *estimate* the number of bases in a sense we now make precise. A *counting problem* is a function f :

$\Sigma^* \to \mathbb{N}$ mapping problem instances (e.g., a particular matroid M, encoded over a finite alphabet Σ) to natural number solutions (e.g., the number of bases of M). A *randomised approximation scheme* for f is a randomised algorithm that takes as input an instance $w \in \Sigma^*$, and an error tolerance $\varepsilon > 0$, and outputs a number $N \in \mathbb{N}$ (a random variable of the "coin tosses" made by the algorithm) such that, for every instance w,

$$\Pr\left[e^{-\varepsilon} f(w) \leq N \leq e^{\varepsilon} f(w)\right] \geq \tfrac{3}{4}\,. \tag{1.1}$$

We speak of a *fully polynomial randomised approximation scheme*, or *FPRAS*, if the algorithm runs in time bounded by a polynomial in $|w|$ and ε^{-1}. The threshold $\frac{3}{4}$ appearing in (1.1) could be replaced by any number in the open interval $(\frac{1}{2}, 1)$ without material change.

A key algorithmic question is: does there exist an FPRAS for estimating the bases of an arbitrary matroid? To make the question precise, it is necessary to specify how matroids are to be encoded as problem instances. In the case of vectorial matroids, a natural convention would be list the vectors forming the ground set. More generally, one could specify a matroid M by providing an "oracle" that, when presented with a subset of the ground set, is able to pronounce on whether or not it is an independent set of M. (A set $I \subseteq E$ is *independent* if it is contained in some basis. There are technical reasons for preferring an independence oracle to a basis oracle.) This is a very liberal setting which admittedly takes us a little outside the formal definition of "counting problem" given earlier. It is the one we shall tacitly adopt in §2, which brings together much of what is known about the basis counting problem.

Later, in §3, we shall widen the scope considerably. The *Tutte polynomial* $T(M; x, y)$ of a matroid M is a two-variable polynomial that encodes much fascinating information about M. For example, $T(M; 1, 1)$ counts bases in M, the subject of the first part of the article. But many other points and curves in the (x, y)-plane are also of interest: for example $T(M; 2, 1)$ counts independent sets of M, while the hyperbola $(x-1)(y-1) = 2$ corresponds to the partition function of the Ising model at varying temperatures.

Dominic Welsh and coauthors initiated the study of the computational complexity of the Tutte polynomial. In a series of papers, starting with (Jaeger, Vertigan and Welsh, 1990), they obtained almost complete information about the computational complexity of evaluating the Tutte polynomial exactly, for various classes of matroids. It transpires that the Tutte polynomial is computational intractable (in a precise sense that will be explained in §3.1) except at a small number of "special points and curves" depending on the class of matroids from which the problem instance is selected. Steven Noble's article in this volume describes this work.

In light of the almost everywhere hardness of the Tutte polynomial, it is natural to consider the computational complexity of approximate computation. Jerrum and Sinclair's approximation algorithm (FPRAS) for the partition function of the ferromagnetic Ising model holds out some hope for positive results (Jerrum

and Sinclair, 1993). Indeed, Dominic Welsh had already made some first steps in the study of the computational complexity of approximating the Tutte polynomial (Welsh, 1994). The existing results are fragmentary; we describe what is known, and make a definite advance. More remains to be done.

2

COUNTING MATROID BASES

In order to estimate the number of bases in a matroid in the FPRAS sense it is enough to be able sample bases almost u.a.r. (uniformly at random) in polynomial time. The notion that the size of many combinatorially defined sets may be inferred with small error from relatively few random samples is quite standard, and in §2.3, we'll sketch how this may be accomplished in the case of matroid bases. For the time being, we'll turn to the problem of sampling bases almost u.a.r. The approach adopted is the by now quite standard one of Markov chain simulation.

2.1 Bases-exchange graph

The exchange axiom presented at the start of this article suggests a natural walk on the set of bases of a matroid M. The *bases-exchange graph* $G(M)$ of a matroid M has vertex set $\mathcal{B}(M)$ and edge set

$$\big\{\{X, Y\} : X, Y \in \mathcal{B}(M) \text{ and } |X \oplus Y| = 2\big\},$$

where $\oplus$ denotes symmetric difference. Note that the edges of the bases-exchange graph $G(M)$ correspond to the transformations guaranteed by the exchange axiom. Indeed, it is straightforward to check, using the exchange axiom, that the graph $G(M)$ is always connected. By simulating a random walk on $G(M)$ it is possible, in principle, to sample a basis (almost) u.a.r. from $\mathcal{B}(M)$. Although it has been conjectured that this random walk is rapidly mixing for all matroids M, the conjecture has never been proved and the circumstantial evidence in its favour seems slight. By "rapidly mixing" we mean that its "mixing time" (roughly, the number of steps taken to converge to near-equilibrium) is bounded by a polynomial in m, the size of the ground set. Precise definitions will be given presently.

Nevertheless, there is an interesting class of matroids, the "balanced" matroids, for which rapid mixing has been established. The definition of balanced matroid is due to (Feder and Mihail, 1992), as is the proof of rapid mixing. In this article, we diverge from their analysis in order to achieve a tighter bound on mixing time.

2.2 Balanced matroids

Two absolutely central operations on matroids are contraction and deletion. An element $e \in E$ is said to be a *loop* (resp., *coloop*) if it occurs in no basis (resp., every basis). If $e \in E(M)$ is an element of the ground set of M then, provided

e is not a coloop, the matroid $M \setminus e$ obtained by *deleting* e has ground set $E(M \setminus e) = E(M) \setminus \{e\}$ and bases $\mathcal{B}(M \setminus e) = \{X \subseteq E(M \setminus e) : X \in \mathcal{B}(M)\}$. Provided e is not a loop, the matroid M/e obtained by *contracting* e has ground set $E(M/e) = E(M) \setminus \{e\}$ and bases $\mathcal{B}(M/e) = \{X \subseteq E(M/e) : X \cup \{e\} \in \mathcal{B}(M)\}$. Any matroid obtained from M by a series of contractions and deletions is a *minor* of M.

Suppose a basis $X \in \mathcal{B}(M)$ is chosen u.a.r. The matroid M is said to possess the *negative correlation property* if the inequality $\Pr(e \in X \wedge f \in X) \leq \Pr(e \in X) \Pr(f \in X)$ holds for all pairs of distinct elements $e, f \in E$. Another way of expressing negative correlation is by writing $\Pr(e \in X \mid f \in X) \leq \Pr(e \in X)$; in other words the knowledge that f is present in X makes the presence of e less likely.[1] Further, the matroid M is said to be *balanced* if all minors of M (including M itself) possess the negative correlation property. Feder and Mihail showed that regular matroids (and hence graphic matroids), are balanced (Feder and Mihail, 1992). So the class is not without interest, even if it does not include all matroids.

2.3 The mixing time of the bases-exchange walk

The mixing time of the bases-exchange walk for balanced matroids was first analysed by (Feder and Mihail, 1992). Here we shall provide a tighter analysis of its mixing time by computing the logarithmic Sobolev ("log-Sobolev") constant. The raw materials are as follows. (Gross, 1975) introduced the log-Sobolev constant. (Diaconis and Saloff-Coste, 1996) pioneered the application of the log-Sobolev constant to bounding the mixing time of finite Markov chains. (Jerrum and Son, 2002) analysed the log-Sobolev constant for the bases-exchange walk, thus obtaining an improved bound on mixing time. All the key steps in the argument are presented here, but for a detailed account refer to (Jerrum, 2004; Jerrum, 2005).

Before proceeding, let's give a precise description of the walk. What we have is a Markov chain on state space $\Omega = \mathcal{B}(M)$ whose transition probabilities $P : \Omega^2 \to [0, 1]$ are given by the following trial, in which the initial state is $X_0 \in \Omega$:

1. Choose e u.a.r. from X_0, and f u.a.r. from E.
2. If $Y = X_0 \cup \{f\} \setminus \{e\} \in \mathcal{B}$ then $X_1 = Y$; otherwise $X_1 = X_0$.

The new state is X_1. It is a easy consequence of the exchange axiom that the Markov chain is ergodic. Note that all non-zero transition probabilities are equal to $p = 1/rm$ where r is the rank of M and $m = |E(M)|$. Note further that the transition probabilities are symmetric, so the (unique) stationary distribution $\pi : \Omega \to [0, 1]$ of the Markov chain is uniform.

Since our goal is to sample a state (basis) u.a.r., we are interested in how quickly the Markov chain converges to stationarity. The most common measure of rate of convergence is the (ℓ_1-) mixing time:

[1] We assume here that $\Pr(f \in X) > 0$, i.e., that f is not a loop.

$$\tau = \max_{x \in \Omega} \min \left\{ t : \|P^t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq e^{-1} \right\},$$

where $P^t(x, \cdot)$ denotes the t-step distribution conditioned on starting in state x, and $\| \cdot \|_{\mathrm{TV}}$ denotes *total variation norm*

$$\|\sigma\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\sigma(x)|.$$

We say that the Markov chain is *rapidly mixing* if τ is bounded by a polynomial in the size of some natural measure of instance size, in our case m, the size of the ground set. Ideally, we would of course like the polynomial to be of low degree.

Tight upper bounds on mixing time can sometimes be derived through consideration of the log-Sobolev constant of the Markov chain. The ingredients for this are the *Dirichlet form*

$$\mathcal{E}_P(\varphi, \varphi) = \frac{1}{2} \sum_{x,y \in \Omega} \pi(x) P(x, y) (\varphi(x) - \varphi(y))^2,$$

and the entropy-like quantity

$$\mathcal{L}_\pi(\varphi) = \mathbb{E}_\pi \left[\varphi^2 \left(\ln \varphi^2 - \ln(\mathbb{E}_\pi \, \varphi^2) \right) \right].$$

A logarithmic Sobolev inequality has the form

$$\mathcal{E}_P(\varphi, \varphi) \geq \alpha \, \mathcal{L}_\pi(\varphi), \quad \text{for all } \varphi : \Omega \to \mathbb{R}, \tag{2.1}$$

where $\alpha > 0$ is the *logarithmic Sobolev constant*.

For time-reversible Markov chains, the mixing time τ is related to α by

$$\tau \leq (4 + \ln \ln \pi^*)/4\alpha, \tag{2.2}$$

where $\pi^* = \min_x \pi(x)$ (Diaconis and Saloff-Coste, 1996, Cor. 3.11). Those authors actually prove a stronger inequality, since their τ is defined relative to the more demanding ℓ_2-norm. In our case, $\ln \ln(1/\pi^*) \leq \ln \ln \binom{m}{r} \leq \ln m$.

We compute a bound on the log-Sobolev constant of the bases-exchange walk via an inductive argument. Recall that the state space $\Omega = \mathcal{B}$ is the set of all bases. Let $e \in E$ be arbitrary element of the ground set that is not a loop or a co-loop. (Loops and coloops may be eliminated by performing a deletion or contraction, respectively.) Partition the state space into two sets $\Omega = \Omega_0 \cup \Omega_1$, where Ω_0 (resp, Ω_1) are the bases (states) that exclude (resp., include) the element e. Denote by π the stationary distribution of the MC (which is just the uniform distribution on Ω) and by π_0 (resp, π_1) the distributions induced by π on Ω_0 (resp., Ω_1), which are themselves of course uniform.

With respect to this partition of the state space, we have the following decomposition of the Dirichlet form:

$$\mathcal{E}_P(\varphi, \varphi) = \pi(\Omega_0) \mathcal{E}_{P_0}(\varphi, \varphi) + \pi(\Omega_1) \mathcal{E}_{P_1}(\varphi, \varphi) + \mathcal{C}, \tag{2.3}$$

where

$$\mathcal{E}_{P_b}(\varphi,\varphi) = \frac{1}{2}\sum_{x,y\in\Omega_b} \pi_b(x)P(x,y)(\varphi(x)-\varphi(y))^2$$

and

$$\mathcal{C} = \sum_{x\in\Omega_0, y\in\Omega_1} \pi(x)P(x,y)(\varphi(x)-\varphi(y))^2.$$

Likewise, for the entropy-like quantity $\mathcal{L}_\pi(\varphi)$:

$$\mathcal{L}_\pi(\varphi) = \pi(\Omega_0)\mathcal{L}_{\pi_0}(\varphi) + \pi(\Omega_1)\mathcal{L}_{\pi_1}(\varphi) + \mathcal{L}_\pi(\bar{\varphi}), \tag{2.4}$$

where

$$\mathcal{L}_{\pi_b}(\varphi) = \mathbb{E}_{\pi_b}\left[\varphi^2\big(\ln\varphi^2 - \ln(\mathbb{E}_{\pi_b}\varphi^2)\big)\right]$$

and

$$\mathcal{L}_\pi(\bar{\varphi}) = \sum_{b=0,1} \pi(\Omega_b)\big[(\mathbb{E}_{\pi_b}\varphi^2)\big(\ln(\mathbb{E}_{\pi_b}\varphi^2) - \ln(\mathbb{E}_\pi\varphi^2)\big)\big]. \tag{2.5}$$

The use of the notation $\mathcal{L}_\pi(\bar{\varphi})$ for the expression on the right hand side of (2.5) is justified, provided we interpret $\bar{\varphi}:\Omega\to\mathbb{R}^+$ as the function that is constant $\sqrt{\mathbb{E}_{\pi_b} f^2}$ on Ω_b, for $b=0,1$.

Our aim to exploit (2.3) and (2.4) to synthesise an inequality of the form $\mathcal{E}_P(\varphi,\varphi) \geq \alpha\,\mathcal{L}_\pi(\varphi)$ from ones of the form

$$\mathcal{E}_{P_b}(\varphi,\varphi) \geq \alpha_b\,\mathcal{L}_{\pi_b}(\varphi) \quad \text{and} \quad \mathcal{C} \geq \bar{\alpha}\,\mathcal{L}_\pi(\bar{\varphi}). \tag{2.6}$$

The inequalities $\mathcal{E}_{P_b}(\varphi,\varphi) \geq \alpha_b\,\mathcal{L}_{\pi_b}(\varphi)$, for $b=0,1$, will naturally be our inductive hypotheses. (Note that Ω_0 and Ω_1 can be regarded as the sets of bases of the matroids $M\setminus e$ and M/e formed by deletion and contraction of/along e.) The derivation of $\mathcal{C} \geq \bar{\alpha}\,\mathcal{L}_\pi(\bar{\varphi})$ is by way of algebraic manipulation, for which we need the following.

Lemma 2.1 *With $\mathcal{L}_\pi(\bar{\varphi})$ defined as in (2.5),*

$$\mathcal{L}_\pi(\bar{\varphi}) \leq \big(\sqrt{\mathbb{E}_{\pi_0}\varphi^2} - \sqrt{\mathbb{E}_{\pi_1}\varphi^2}\big)^2.$$

Lemma 2.1 can be viewed as a statement about the log-Sobolev constant of a two-point space. It is a weakening of (Diaconis and Saloff-Coste, 1996, Thm. A.2); but since it is much easier to prove than the sharp inequality, we provide a short derivation at the end of this section.

A key consequence of balance, observed by Feder and Mihail, is that the transitions of the Markov chain that span Ω_0 and Ω_1 support a certain kind of fractional matching. Precisely:

Lemma 2.2 *Suppose P, Ω_0, Ω_1, π_0 and π_1 are as above. Then there is a function $w:\Omega_0\times\Omega_1\to\mathcal{R}^+$ such that (i) $\sum_{y\in\Omega_1} w(x,y) = \pi_0(x)$, for all $x\in\Omega_0$; (ii) $\sum_{x\in\Omega_0} w(x,y) = \pi_1(y)$, for all $y\in\Omega_1$; and (iii) $w(x,y)>0$ entails $P(x,y)>0$, for all $(x,y)\in\Omega_0\times\Omega_1$.*

Proof See (Feder and Mihail, 1992, Cor. 3.3). □

Observe that $\sum_{x,y} w(x,y) = 1$, so $w(\cdot,\cdot)$ can be regarded as a probability distribution on edges. Those familiar with coupling arguments will immediately see that Lemma 2.2 can be interpreted as guaranteeing a coupling of certain random variables (r.v's) that is supported on the edges of the bases-exchange graph. Specifically, let $(G_0, G_1) \in \mathbb{R}^2$ be the r.v. defined on $(\Omega_0 \times \Omega_1, w)$ as follows: select $(x,y) \in \Omega_0 \times \Omega_1$ according to the distribution $w(\cdot,\cdot)$ and return $(G_0, G_1) = (\varphi(x)^2, \varphi(y)^2)$. Then, using $\mathbb{E}_w$ to denote expectations with respect to the sample space just described,

$$\begin{aligned}
\mathcal{L}_\pi(\bar{\varphi}) &\le \big(\sqrt{\mathbb{E}_{\pi_0}\,\varphi^2} - \sqrt{\mathbb{E}_{\pi_1}\,\varphi^2}\,\big)^2 \\
&= \big(\sqrt{\mathbb{E}_w\, G_0} - \sqrt{\mathbb{E}_w\, G_1}\,\big)^2 \\
&\le \mathbb{E}_w\left[\big(\sqrt{G_0} - \sqrt{G_1}\,\big)^2\right] && \text{by Jensen's inequality} \\
&= \sum_{(x,y)\in\Omega_0\times\Omega_1} w(x,y)\big(\varphi(x) - \varphi(y)\big)^2 \\
&\le \sum_{(x,y):\, w(x,y)>0} \frac{\pi(x)}{\pi(\Omega_0)}\big(\varphi(x) - \varphi(y)\big)^2 && \text{by Lemma 2.2(i)} \\
&\le \frac{1}{p\,\pi(\Omega_0)} \sum_{(x,y)\in\Omega_0\times\Omega_1} \pi(x)P(x,y)\big(\varphi(x) - \varphi(y)\big)^2 \\
&\le \frac{2}{p}\,\mathcal{C}, && (2.7)
\end{aligned}$$

where we have assumed, by symmetry, that $\pi(\Omega_0) \ge \pi(\Omega_1)$ and hence $\pi(\Omega_0) \ge \frac{1}{2}$. Here, $p = 1/rm$ is the uniform transition probability for the bases-exchange walk.

Comparing (2.7) with (2.6), we see that we may take $\bar{\alpha} = 2/p$. Then, substituting (2.6) into (2.3) and (2.4) we arrive at the (trivial) recurrence

$$\alpha_{m,p} \ge \min\{\alpha_{m-1,p},\, p/2\},$$

for the log-Sobolev constant $\alpha_{m,p}$ of the bases-exchange walk of a balanced matroid on a ground set of size m, and uniform transition probability p. Thus, by a trivial induction, $\alpha_{m,p} \ge p/2$. We have thus established.

Theorem 2.3 *The logarithmic Sobolev constant of the bases-exchange walk of a balanced matroid M is bounded below by $1/2rm$, where r is the rank of M and m the size of its ground set.*

This bound is tight, up to a constant factor, as can be seen by taking the function φ that is constant -1 on Ω_0 and constant 1 on Ω_1.

Corollary 2.4 *The mixing time of the bases-exchange walk of a balanced matroid is $O(rm \log m)$.*

Proof Apply inequality (2.2). □

We now have an efficient procedure for sampling bases of a balanced matroid almost u.a.r: simulate the bases-exchange walk for $O(rm \log m)$ steps and return the current state (basis). Given this sampling procedure, the construction of an FRPAS for counting bases of a balanced matroid is now routine. The strategy, in brief, is the following. By collecting random samples from Ω, estimate the ratio $|\Omega_0| : |\Omega_1|$ to sufficient accuracy. Suppose, w.l.o.g, that $|\Omega_0| \geq |\Omega_1|$. Now recall that Ω_0 is isomorphic to $\mathcal{B}(M \setminus e)$. Recursively, compute an estimate for $|\Omega_0| = |\mathcal{B}(M \setminus e)|$ and multiply it by the previously obtained estimate for $|\Omega|/|\Omega_0|$. The result is an estimate for $|\Omega| = |\mathcal{B}(M)|$. (The reason for recursing on the larger subset of the partition is to control the variance of the estimator for $|\Omega|/|\Omega_0|$. If the two sets are of roughly equal size it doesn't matter which we choose.) An analysis of the sample sizes required to achieve sufficient accuracy yields:

Corollary 2.5 *There is an FPRAS for estimating the number of bases of a balanced matroid, with running time* $O(rm^3 \log m)$.

In this corollary and the previous one we have suppressed the dependence of the running time on the parameter controlling "accuracy": closeness to uniformity in the case of Corollary 2.4, and permitted relative error in the result in the case of Corollary 2.5. A complete analysis would obviously need to track these dependencies. See, e.g., (Jerrum, 2003, Prop. 3.4) for details. We have also tacitly assumed that M is presented as an "independence oracle", so that each step of the bases-exchange walk can be simulated in constant time. This assumption may not always be realistic.

We close the section with the promised:

Proof of Lemma 2.1 Let r and s be positive numbers with $r + s = 1$. To prove Lemma 2.1 is is enough to establish the inequality

$$r\xi^2 \ln \frac{\xi^2}{r\xi^2 + s\eta^2} + s\eta^2 \ln \frac{\eta^2}{r\xi^2 + s\eta^2} \leq (\xi - \eta)^2,$$

for all $\xi, \eta \in \mathbb{R}$.

Applying the inequality $\ln a \leq a - 1$, which is valid for all $a > 0$:

$$\begin{aligned} r\xi^2 \ln \frac{\xi^2}{r\xi^2 + s\eta^2} + s\eta^2 \ln \frac{\eta^2}{r\xi^2 + s\eta^2} &\leq r\xi^2 \frac{s(\xi^2 - \eta^2)}{r\xi^2 + s\eta^2} + s\eta^2 \frac{r(\eta^2 - \xi^2)}{r\xi^2 + s\eta^2} \\ &= \frac{rs(\xi^2 - \eta^2)^2}{r\xi^2 + s\eta^2} \\ &= \frac{rs(\xi + \eta)^2}{r\xi^2 + s\eta^2} (\xi - \eta)^2 \\ &\leq (\xi - \eta)^2. \end{aligned}$$

To verify the final inequality, first note that by scaling one may assume that $\xi + \eta = 1$; it is then easy to see (by calculus) that the extremal case is when $\xi = s$ and $\eta = r$. □

2.4 Matroids in general

Not all matroids are balanced. However, it has on occasion been conjectured that the bases-exchange graph of any matroid has edge expansion 1. If true — and it would be a remarkable result in its generality — then the bases-exchange walk would be rapidly mixing for all matroids, and there would be an FPRAS for counting bases of unrestricted matroids in the independence-oracle model. On the other hand, if there are classes of matroids that do not admit an FPRAS, then a proof of this fact (modulo some reasonable complexity-theoretic hypothesis) seems some way off. One thing that is known for certain is that any efficient approximation algorithm based on an independence oracle must necessarily be randomised (Azar, Broder and Frieze, 1994).

3

THE TUTTE POLYNOMIAL

The *Tutte polynomial* of a matroid $M = (E, \mathcal{B})$ is a two-variable polynomial T defined by

$$T(M; x, y) = \sum_{A \subseteq E} (x-1)^{r(E)-r(A)} (y-1)^{|A|-r(A)}, \tag{3.1}$$

where $r(A)$ denotes the *rank* of A, i.e., the size of the largest independent set contained in A. Evaluations of the Tutte polynomial at various points and along various curves in $\mathbb{R}^2$ yield much interesting information about M.[2] For example, $T(M; 1, 1)$ is equal to the number of bases of M and $T(M; 2, 1)$ to the number of independent sets. In the case when M is the cycle matroid of a graph G, and q a positive integer, $T(M; 1-q, 0)$ is the number of q-colourings of G. More generally, along the hyperbola $H_q = \{(x, y) : (x-1)(y-1) = q\}$, the Tutte polynomial $T(M; x, y)$ specialises to the partition function of the q-state Potts model, up to some easily computable normalising factor. The positive branch of the hyperbola H_q corresponds to the *ferromagnetic* Potts model, and the negative branch (at least the part above the x-axis) to the *antiferromagnetic* Potts model. For much more on this fascinating subject, refer to (Welsh, 1993).

3.1 Rudiments of computational complexity

Recall our view of counting problems as functions $f : \Sigma^* \to \mathbb{N}$. We have already seen one formalisation of the notion of tractability of a counting problem, namely the FPRAS. A more a direct and demanding notion is simple polynomial-time computability. A function f is said to belong to the class FP if there is an algorithm for computing $f(w)$ that runs in time polynomial in the length $|w|$ of the instance w.[3]

Of course, we don't expect every counting problem to admit an FPRAS, let alone be a member of FP. Just as with decision problems, we can gain evidence that a counting problem is intractable by showing it to be complete (or at least hard) for an appropriate complexity class. The appropriate class in this instance is #P (Valiant, 1979). Briefly, #P contains those functions $f : \Sigma^* \to \mathbb{N}$ that can be expressed in the form

$$f(w) = \big|\{z \in \Sigma^* : |z| \leq p(|w|) \text{ and } \Pi(w, z)\}\big|,$$

[2] In contrast to the notation used in the previous section, x and y will used to denote coordinates in $\mathbb{R}^2$. This change in notation is unlikely to cause any confusion.

[3] The "F" in FP stands for "Function". The complexity class P is formally restricted to decision problems (predicates on Σ^*).

where p is a polynomial and $\Pi \subseteq \Sigma^* \times \Sigma^*$ a polynomial-time-computable two-place predicate. We can think of $\Pi(w, z)$ as being a "witness-checking predicate": if w encodes a graph G, and z a subset $A \subseteq E(G)$ of the edges of G, then $\Pi(w, z)$ might decide whether A is a spanning tree of G. In that instance, $f(w)$ counts spanning trees in the graph encoded by w. Informally, then, #P contains counting problems associated with easily recognised combinatorial structures. A problem is #P-hard (resp., #P-complete) if it is hard (resp., complete) for #P with respect to polynomial-time Turing reducibility. Counting perfect matchings in a bipartite graph is the archetypal #P-complete problem.

Computing the Tutte polynomial at positive integer lattice points falls precisely into the above setting. At rational points, one has to bend the framework a little, but this is a technical point. For much more on computational complexity in general, and #P in particular, consult (Papadimitriou, 1994).

3.2 What is known

The computational complexity of exact evaluation of $T(M; x, y)$ has been extensively studied by Welsh and other authors, beginning with (Jaeger, Vertigan and Welsh, 1990). The situation for matroids in general is not very interesting. The hyperbola H_1 is trivial, the Tutte polynomial evaluating to $(x-1)^r y^{|E|}$ there. Elsewhere, evaluation of the Tutte polynomial is #P-hard. In particular, evaluating $T(M; 1, 1)$ is #P-hard for the class of transversal matroids (Colbourn, Provan and Vertigan, 1995).

For restricted classes of matroids the picture is more complex and more interesting. For graphical matroids, in addition to the hyperbola H_1, the Tutte polynomial may be computed in polynomial time at the special points $(1, 1)$, $(-1, 0)$, $(-1, -1)$ and $(0, -1)$. (We consider here only evaluations at points in $\mathbb{R}^2$: some other special points emerge if the scope is widened to $\mathbb{C}^2$.) For example, $T(G, -1, 0)$ counts 2-colourings of G, and hence is 0 if G is non-bipartite and $2^{\kappa(G)}$ otherwise, where $\kappa(G)$ denotes the number of connected components of G. Aside from H_1 and the special points, evaluating Tutte polynomial of a general graph is #P-hard (Jaeger, Vertigan and Welsh, 1990).

Restricting further to planar graphs, the hyperbola H_2 must be added to the set of polynomial-time evaluations. As we noted earlier, along this hyperbola the Tutte polynomial is, up to an easily computable normalising factor, equal to the partition function of the Ising model, which is the special case $q = 2$ of the Potts model. In contrast to the hyperbola H_1, and the special points, H_2 is tractable for a distinctly non-trivial reason. From classical work of Fisher, Kasteleyn and Temperley it is known that the partition function of a Ising system with n sites or vertices may be expressed as an $n \times n$ determinant, which may be evaluated in polynomial time.

In the other direction, if we generalise from graphic to regular matroids, the special point $(1, 1)$ survives (Dyer and Frieze, 1994), but the others do not. For transversal matroids we even lose the point $(1, 1)$ (Colbourn, Provan and

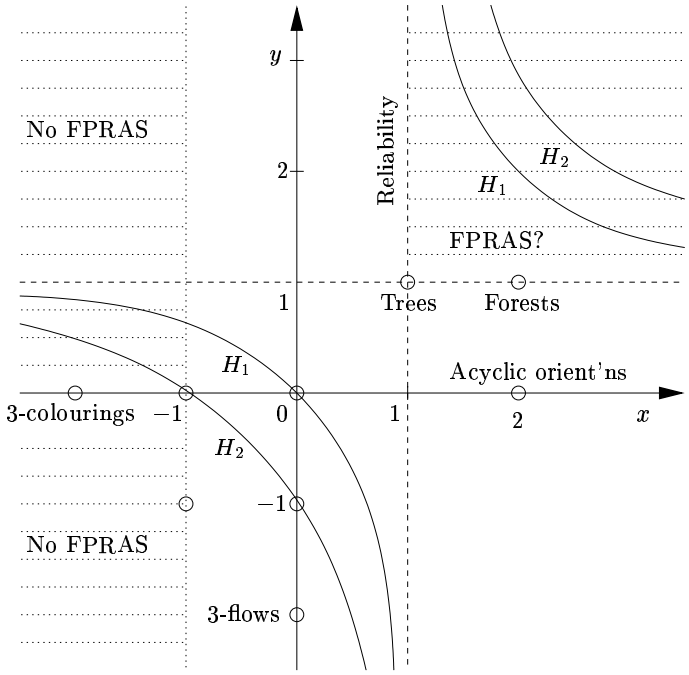

FIG. 3.1. The Tutte plane

Vertigan, 1995). See Noble's article in this volume for more on the complexity of exact evaluation of the Tutte polynomial on graphs.

So much for exact evaluation; what about the possibilities for an FPRAS? For general matroids, we know of no points that aren't already exactly computable in polynomial time. For graphic matroids we know that an FPRAS exists for the hyperbola H_2 in the positive quadrant, i.e., for the ferromagnetic Ising model with no applied field (Jerrum and Sinclair, 1993). Despite our best efforts, nothing more is known for sure, except for restricted classes of graphs.

On the negative direction, there are isolated points $(x, y) \in \mathbb{R}^2$ at which $T(G; x, y)$ is hard to approximate owing to a specific combinatorial interpretation. Take, for example, the point $(-2, 0)$ at which the Tutte polynomial of a graph G counts (proper) 3-colourings of G. An FPRAS for $T(G; -2, 0)$ would, in particular, have to decide whether G has some 3-colourings or none. But determining whether a graph is 3-colourable is an NP-complete decision problem, so no FRPAS can exist unless every problem in NP admits a polynomial-time randomised algorithm. (Technically, we have shown that no FPRAS for $T(G; -2, 0)$ exists, under the assumption for RP $\neq$ NP, a slight strengthening of the celebrated P $\neq$ NP conjecture.)

Welsh has taken this further and shown that, for all positive integers $q \geq 2$, there is no FPRAS for the Tutte polynomial along the segments of the negative

branches of the hyperbolas H_q, lying in the infinite strip $y \in (-1,1)$ (Welsh, 1994). This result again is modulo the complexity theoretic assumption RP $\neq$ NP. These segments correspond to the *antiferromagnetic* q-state Potts model.

As far as I am aware, there is no subset of $\mathbb{R}^2$ of positive measure that has been shown to be immune to an FPRAS.[4] This state of affairs can be corrected, and we now do so.

3.3 Regions of the plane that do not admit an FPRAS

The tensor product of matroids was introduced by (Brylawski, 1982). We define it here in the special case of graphs. Let G be a graph, and K another graph with a distinguished edge f with endpoints u and u'. The tensor product $G \otimes K$ is obtained from G by performing a "2-sum" operation on each edge e of G in turn: Let the endpoints of e be v and v'. Take a copy of K and identify vertex u (resp. u') of K with v (resp. v') of G, and then delete edges e and f. (Since G and K are undirected graphs, there are two ways of performing the 2-sum. This lack of uniqueness is a artefact of viewing a matroid operation in terms of graphs, which have additional structure. In particular, the Tutte polynomial is insensitive to which of the two possible identifications is made.) We are particularly interested in the case where K is a cycle on $k+1$ vertices (this is known as a *k-stretch* in the literature) or a two-vertex graph with $k+1$ parallel edges (a *k-thickening*). Informally, a k-stretch of G replaces each edge of G by a path of length k, while a k-thickening replaces each edge by a bundle of k parallel edges.

A key fact about the tensor product (Welsh, 1993, eq. (6.2.7)) is:

$$T(G \otimes K; x, y) = \alpha\, T(G; x', y'), \tag{3.2}$$

where α is an easily computable number and x' and y' depend only on x, y and K (and are easily computable from them). Specifically,

$$(x', y') = \begin{cases} \big(x^k, q/(x^k-1)+1\big) & \text{for a } k\text{-stretch;} \\ \big(q/(y^k-1)+1, y^k\big) & \text{for a } k\text{-thickening,} \end{cases} \tag{3.3}$$

where $q = (x-1)(y-1)$. We'll say that the point $(x,y) \in \mathbb{R}^2$ may be *shifted to* (x',y') if there is a graph K such that (x,y) and (x',y') are in relation (3.2). Some explicit shifts are provided by (3.3). Observe that $q = (x-1)(y-1)$ is an invariant for these particular shifts and, in fact, for shifts in general. It is this limitation that gives the hyperbolas H_q a special place in the complexity theory of the Tutte polynomial.

Proposition 3.1. (Goldberg and Jerrum) *Suppose $(x,y) \in \mathbb{Q}^2$ satisfies $q = (x-1)(y-1) \notin \{0,1,2\}$. Suppose also that it is possible to shift the point (x,y) to the point (x',y') with $y' \in (-1,1)$, and to (x'',y'') with $y'' \notin [-1,1]$. Then there is no FPRAS for the function $G \mapsto T(G;x,y)$ unless* RP = NP.

[4] Since it is not possible to represent arbitrary real numbers, we should really restrict attention to rational points. So, technically, we are looking for a set of rational points whose closure has positive measure.

Proof See (Goldberg and Jerrum, 2006). Note that we restrict attention to rational x and y to avoid representational issues. □

Since the notion of "shift" is defined for any class of matroids closed under tensor product, it should be possible to frame statements similar to Proposition 3.1 for classes of matroids other than graphic.

Corollary 3.2 *Suppose (x, y) is a point lying in the open half-plane $x < -1$ but not on the hyperbolas H_0 or H_1. Under the assumption* RP $\neq$ NP *there can be no FPRAS for the function $G \mapsto T(G; x, y)$.*

Proof Let $(x, y) \in \mathbb{R}^2$ be a point not on H_0 or H_1 that satisfies $x < -1$. At the outset, we'll assume further that $(x, y) \notin H_2$ and that $y \neq -1$. There are three cases, depending on y. First assume $y > 1$, and observe that $q = (x-1)(y-1) < 0$. Using a k-stretch, we may shift the point (x, y) to the the point $(x'', y'') = \big(x^k, q/(x^k - 1) + 1\big)$. Now $y'' \in (-1, 1)$ for all sufficiently large even k so Proposition 3.1 applies. (The trivial shift, taking (x, y) to itself, provides the point $(x', y') \notin [-1, 1]$.) A similar argument, but setting k to be large and odd deals with the situation $y < -1$. Finally, when $y \in (-1, 1)$, a 2-stretch shifts (x, y) to the point $(x', y') = \big(x^2, q/(x^2 - 1) + 1\big) = \big(x^2, (y-1)/(x+1) + 1\big)$, with $y' > 1$.

The additional condition $y \neq -1$ may be removed by noting that a 3-stretch shifts $(x, -1)$ to a point $(x', y') = \big(x^3, 1 - 2/(x^2 + x + 1)\big)$ with $x' < -1$ and $y' \in (-1, +1)$, and we have already seen how to deal with such a point. Finally, the hyperbola H_2 was treated by (Welsh, 1994). □

Various other non-FPRASable regions of the Tutte plane may be mapped using the basic proof technique of Corollary 3.2. Refer to (Goldberg and Jerrum, 2006) for recent results.

3.4 Speculations (optimistic)

First, matroids in general. If we were very optimistic, we might speculate that there is an FPRAS for $T(M; 1, 1)$, for a general matroid M specified by an independence oracle. If that were the case, then we would have an FPRAS for the whole of the positive branch of the (degenerate) hyperbola H_0. The reasoning is simple. Assume that $y = 1$ and $x > 1$. (The symmetric case follows by duality.) The Tutte polynomial in this case simplifies to

$$\begin{aligned} T(M; x, 1) &= \sum_{A \in \mathcal{I}(M)} (x-1)^{r(E)-|A|} \\ &= \sum_{k=0}^{r(E)} T(M^{[k]}; 1, 1), \end{aligned}$$

where we have used $\mathcal{I}(M)$ to denote the independent sets of M, and $M^{[k]}$ the truncation of M to rank k (Welsh, 1976, §4.1). In particular, we would have

FPRASes for the number of forests in a graph, and for the reliability polynomial of a graph.

When $x, y \geq 1$, the weight function $w : E \to \mathbb{R}$

$$w(A) = (x-1)^{r(E)-r(A)}(y-1)^{|A|-r(A)}$$

is non-negative, and hence determines a probability distribution π on E. The normalisation factor is of course $\sum_A w(A) = T(M; x, y)$, so, explicitly, $\pi(A) = w(A)/T(M; x, y)$. Specialising to a graph $G = (V, E)$, this probability distribution is the one arising from the *random cluster model*, which may be written

$$\pi(A) = p^{|A|}(1-p)^{|E|-|A|}q^{\kappa(A)}/Z, \tag{3.4}$$

where $\kappa(A)$ denotes the number of connected components of (V, A), and Z is normalising factor (partition function of the random cluster model). The translation between parameters is

$$p = \frac{y-1}{y} \qquad \text{and} \qquad q = (x-1)(y-1).$$

The random cluster model generalises the q-state Potts model to non-integer q.

Consider a Markov chain on state space $\Omega = 2^E$ with transition probabilities defined by the following trial, where A denotes the current state.

1. Select $e \in E$ uniformly at random.
2. Let $w_0 = w(A \setminus \{e\})$ and $w_1 = w(A \cup \{e\})$. Then set

$$A' = \begin{cases} A \setminus \{e\}, & \text{with probability } w_0/(w_0+w_1); \\ A \cup \{e\}, & \text{with complementary probability } w_1/(w_0+w_1). \end{cases}$$

The new state is A'. (This is the single-site heat-bath dynamics applied to the terms of the Tutte polynomial.) Provided $x, y > 1$ (equivalently, $q > 0$ and $p \in (0,1)$) this Markov chain is irreducible and aperiodic, and is time-reversible with stationary distribution π, given by (3.4). If $x > 1$ and $y = 1$ then the Markov chain is still ergodic, but on a subset of 2^E, namely the independent sets of M; a dual statement covers $x = 1$ and $y > 1$. When $x = y = 1$ the single-site dynamics is frozen, which is why we must use the slightly more complex dynamics provided by the bases-exchange walk in that case.

Consider the region

$$R = \big\{(x, y) : x \geq 1, y \geq 1 \text{ and } q = (x-1)(y-1) \leq 2\big\} \setminus (1, 1).$$

This is the closed region bounded by the branches of the hyperbolas H_0 and H_2 in the positive quadrant. There is no known obstacle to rapid mixing of the single-site heat-bath dynamics for $(x, y) \in R$ in the graphical case, so it is reasonable to conjecture that this dynamics provides an FPRAS covering the whole of R. Although I know of no barrier to rapid mixing for matroids in general, it seems

a little rash to conjecture an FPRAS for general matroids over R, since we do not even know the status of the point $(1,1)$. (I assume here that the matroid is presented in terms of an independence oracle.)

(Gore and Jerrum, 1999) exhibit a counterexample to mixing which applies to a variety of dynamics in the case $q = 3$. Indeed the first-order phase transition on which the counterexample is based exists for all $q > 2$ (Bollobás, Grimmett and Janson, 1996). So the region beyond the hyperbola H_2 in the positive quadrant (i.e., the points with $(x-1)(y-1) > 2$) provides somewhat less scope for optimism, though there is no particular reason to doubt that an FPRAS exists. The somewhat daring conjecture of Welsh is that there is an FPRAS covering the region in the upper quadrant bounded by H_0, at least in the graphic case (Welsh, 1993). Again, the conjecture can be strengthened even further, by widening the class of matroids.

Although there is no space to go into the matter here, a number of authors have presented FPRASes for "dense" instances, particularly dense graphs. Refer to (Alon, Frieze and Welsh, 1995; Annan, 1994; Karger, 1999).

3.5 Speculations (pessimistic)

When $x < 1$ or $y < 1$ the terms in expression (3.1) for the Tutte polynomial will vary in sign, which does not auger well the existence of an FPRAS. However, this concern may be illusory, at least in the graphical case, as (Tutte, 1984) has shown that all monomials in the expansion of $T(G;x,y)$ have positive coefficients. In the light of this surprising fact, it would be rash to speculate on non-FPRASable regions in the positive quadrant $x, y \geq 0$. Thus the status of points such as $(2,0)$, where the Tutte polynomial of G counts the number of acyclic orientations of G, seems wide open.

When either $x < 0$ or $y < 0$, "real" cancellation occurs, and in this region there seems to be no plausible general approach to constructing an FPRAS. Even here, the combinatorial interpretation of the Tutte polynomial at certain points provides hope for an FPRAS tailored to those points. For example, $T(G;0,-5)$ may be interpreted as the number of nowhere-zero 6-flows in G, and Seymour has shown that every bridgeless graph has at least one 6-flow (Seymour, 1981). So not only is $T(G;0,-5)$ non-negative; it is even the case that deciding whether $T(G;0,-5) \neq 0$ is trivially polynomial-time solvable. In other words, the task of approximating $T(G;0,-5)$ does not contain within it some NP-hard decision problem. Even so, it has recently been shown that that no FPRAS exists for $T(G;0,-5)$, unless RP = NP. (This is a consequence of a more general result of (Goldberg and Jerrum, 2006).)

Perhaps it is not too ridiculous to conjecture that no FPRAS exists in the union of open halfspaces $x < 0$ and $y < 0$ except for the hyperbola H_1 and a countable number of "special points".

REFERENCES

Alon, Noga, Frieze, Alan, and Welsh, Dominic (1995). Polynomial time randomized approximation schemes for Tutte-Gröthendieck invariants: the dense case. *Random Structures Algorithms*, **6**(4), 459–478.

Annan, J. D. (1994). A randomised approximation algorithm for counting the number of forests in dense graphs. *Combin. Probab. Comput.*, **3**(3), 273–283.

Azar, Y., Broder, A. Z., and Frieze, A. M. (1994). On the problem of approximating the number of bases of a matroid. *Inform. Process. Lett.*, **50**(1), 9–11.

Bollobás, B., Grimmett, G., and Janson, S. (1996). The random-cluster model on the complete graph. *Probab. Theory Related Fields*, **104**(3), 283–317.

Brylawski, Thomas (1982). The Tutte polynomial. I. General theory. In *Matroid theory and its applications*, pp. 125–275. Liguori, Naples.

Colbourn, Charles J., Provan, J. Scott, and Vertigan, Dirk (1995). The complexity of computing the Tutte polynomial on transversal matroids. *Combinatorica*, **15**(1), 1–10.

Diaconis, P. and Saloff-Coste, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, **6**(3), 695–750.

Dyer, Martin and Frieze, Alan (1994). Random walks, totally unimodular matrices, and a randomised dual simplex algorithm. *Math. Programming*, **64**(1, Ser. A), 1–16.

Feder, Tomás and Mihail, Milena (1992). Balanced matroids. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 26–38. ACM Press.

Goldberg, Leslie Ann and Jerrum, Mark (2006). Inapproximability of the Tutte polynomial. In preparation.

Gore, Vivek K. and Jerrum, Mark R. (1999). The Swendsen-Wang process does not always mix rapidly. *J. Statist. Phys.*, **97**(1-2), 67–86.

Gross, Leonard (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.*, **97**(4), 1061–1083.

Jaeger, F., Vertigan, D. L., and Welsh, D. J. A. (1990). On the computational complexity of the Jones and Tutte polynomials. *Math. Proc. Cambridge Philos. Soc.*, **108**(1), 35–53.

Jerrum, Mark (2003). *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel.

Jerrum, Mark (2004, September). Inductive bounds, cubes, trees and matroids. `http://homepages.inf.ed.ac.uk/mrj/pubs.html`. Leture notes.

Jerrum, Mark (2005, February). Logarithmic Sobolev inequalities. `http://homepages.inf.ed.ac.uk/mrj/pubs.html`. Leture notes.

Jerrum, Mark and Sinclair, Alistair (1993). Polynomial-time approximation

algorithms for the Ising model. *SIAM J. Comput.*, **22**(5), 1087–1116.

Jerrum, Mark and Son, Jung-Bae (2002). Spectral gap and log-Sobolev constant for balanced matroids. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS'02)*, pp. 721–729. IEEE Computer Society Press.

Karger, David R. (1999). A randomized fully polynomial time approximation scheme for the all-terminal network reliability problem. *SIAM J. Comput.*, **29**(2), 492–514 (electronic).

Oxley, James G. (1992). *Matroid theory.* Oxford Science Publications. The Clarendon Press Oxford University Press, New York.

Papadimitriou, Christos H. (1994). *Computational Complexity.* Addison-Wesley Publishing Company, Reading, MA.

Seymour, P. D. (1981). Nowhere-zero 6-flows. *J. Combin. Theory Ser. B*, **30**(2), 130–135.

Tutte, W. T. (1984). *Graph theory*, Volume 21 of *Encyclopedia of Mathematics and its Applications.* Addison-Wesley Publishing Company Advanced Book Program, Reading, MA. With a foreword by C. St. J. A. Nash-Williams.

Valiant, Leslie G. (1979). The complexity of enumeration and reliability problems. *SIAM J. Comput.*, **8**(3), 410–421.

Welsh, D. J. A. (1976). *Matroid theory.* Academic Press [Harcourt Brace Jovanovich Publishers], London. L. M. S. Monographs, No. 8.

Welsh, D. J. A. (1993). *Complexity: knots, colourings and counting*, Volume 186 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge.

Welsh, D. J. A. (1994). Randomised approximation in the Tutte plane. *Combin. Probab. Comput.*, **3**(1), 137–143.