The Tutte Polynomial

Dominic Welsh*
Merton College, University of Oxford, Oxford OX1 4JD, England

Accepted 19 March 1999

ABSTRACT: This is a close approximation to the content of my lecture. After a brief survey of well known properties, I present some new interpretations relating to random graphs, lattice point enumeration, and chip firing games. I then examine complexity issues and concentrate in particular, on the existence of randomized approximation schemes. © 1999 John Wiley & Sons, Inc. Random Struct. Alg., 15, 210–228, 1999

1. INTRODUCTION

The Tutte polynomial is a polynomial in two variables x, y which can be defined for a graph, matrix, or, even more generally, a matroid. Most of the interesting applications arise when the underlying structure is a graph or a matrix, but matroids are an extremely useful vehicle for unifying the concepts and definitions. For example, each of the following is a special case of the general problem of evaluating the Tutte polynomial of a graph (or matrix) along particular curves of the (x, y) plane: (i) the chromatic and flow polynomials of a graph; (ii) the all terminal reliability probability of a network; (iii) the partition function of a Q-state Potts model; (iv) the Jones polynomial of an alternating knot; (v) the weight enumerator of a linear code over GF(q).

In this paper I shall briefly review the standard theory of the Tutte polynomial and list its well known evaluations. I shall then present some new recent evaluations in terms of (a) colorings and flows in random graphs, (b) lattice point enumeration and the Ehrhart polynomial, and (c) configurations arising in chip firing games.

* Supported in part by Esprit Working Group 21726, “RAND2.”

I shall then turn to complexity questions. It has been shown in Vertigan and Welsh [31] that apart from a few special points and two special hyperbolae, the exact evaluation of any such invariant is #P-hard even for the very restricted class of planar bipartite graphs. However the question of which points have a fully polynomial randomized approximation scheme is not well understood and it is this question which is addressed in the closing sections.

The graph terminology used is standard. The complexity theory and notation follow Garey and Johnson [11]. The matroid terminology follows Oxley [23]. Further details of many of the concepts treated here can be found in Welsh [32].

2. BASIC THEORY

First consider the following recursive definition of the function $T(G; x, y)$ of a graph G, and two independent variables x, y.

If G has no edges, then $T(G; x, y) = 1$; otherwise, for any $e \in E(G)$:

(2.1) $T(G; x, y) = T(G'_e; x, y) + T(G''_e; x, y)$, where G'_e denotes the deletion of the edge e from G and G''_e denotes the contraction of e in G;
(2.2) $T(G; x, y) = xT(G'_e; x, y)$, e an isthmus;
(2.3) $T(G; x, y) = yT(G''_e; x, y)$, e a loop.

From this, it is easy to show by induction that T is a 2-variable polynomial in x, y, which we call the *Tutte polynomial* of G.

In other words, T may be calculated recursively by choosing the edges in *any* order and repeatedly using (2.1)–(2.3) to evaluate T. The remarkable fact is that T is well defined in the sense that the resulting polynomial is independent of the order in which the edges are chosen.

Example. If G is the complete graph K_3, then

$$T(G; x, y) = x^2 + x + y.$$

If G is the complete graph K_4, then

$$T(G; x, y) = x^3 + 3x^2 + 2x + 4xy + 2y + 3y^2 + y^3.$$

Alternatively, and this is often the easiest way to prove properties of T, we can show that T has the following expansion.

First recall that if $A \subseteq E(G)$, the *rank* of A, $r(A)$, is defined by

$$r(A) = |V(G)| - k(A), \tag{2.4}$$

where $k(A)$ is the number of connected components of the graph $G : A$ having vertex set $V = V(G)$ and edge set A.

It is now straightforward to prove:

(2.5) The Tutte polynomial $T(G; x, y)$ can be expressed in the form

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

It is easy and useful to extend these ideas to matroids.

A *matroid* M is just a generalization of a matrix and can be defined simply as a pair (E, r), where E is a finite set and r is a submodular *rank function* mapping $2^E \to \mathbf{Z}$ and satisfying the conditions

$$0 \leq r(A) \leq |A|, \qquad A \subseteq E, \tag{2.6}$$

$$A \subseteq B \quad \Rightarrow \quad r(A) \leq r(B), \tag{2.7}$$

$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B), \qquad A, B \subseteq E. \tag{2.8}$$

The edge set of any graph G with its associated rank function as defined by (2.4) is a matroid, but this is just a very small subclass of matroids known as graphic matroids.

A much larger class is obtained by taking any matrix B with entries in a field F and letting E be its set of columns and for $X \subseteq E$ defining the rank $r(X)$ to be the maximum size of a linearly independent set in X. Any abstract matroid which can be represented in this way is called *representable* over F.

A few basic facts which we shall need are the following:

(2.9) Graphic matroids are representable over every field.

(2.10) A matroid M is representable over every field iff it has a representation over the reals by a matrix B which is *totally unimodular*.Such a matroid is called *regular*.

Given $M = (E, r)$ the *dual matroid* $M^* = (E, r^*)$, where r^* is defined by

$$r^*(E \backslash A) = |E| - r(E) - |A| + r(A). \tag{2.11}$$

A set X is *independent* if $r(X) = |X|$; it is a *base* if it is a maximal independent subset of E. An easy way to work with the dual matroid M^* is not via the rank function, but by the following definition.

(2.12) M^* has as its bases all sets of the form $E \backslash B$, where B is a base of M.

We now just extend the definition of the Tutte polynomial from graphs to matroids by

$$T(M; x, y) = \sum_{A \subseteq E(M)} (x-1)^{r(E)-r(A)}(y-1)^{|A|-r(A)}. \tag{2.13}$$

Much of the theory developed for graphs goes through in this more general setting and there are many other applications as we shall see. For example, routine checking shows that

$$T(M; x, y) = T(M^*; y, x). \tag{2.14}$$

In particular, when G is a planar graph and G^* is any plane dual of G, (2.14) becomes

$$T(G; x, y) = T(G^*; y, x). \tag{2.15}$$

We close this section with what I call the "recipe theorem" from Oxley and Welsh [24]. Its crude interpretation is that whenever a function f on some class of matroids can be shown to satisfy an equation of the form $f(M) = af(M'_e) + b(M''_e)$ for some $e \in E(M)$, then f is essentially an evaluation of the Tutte polynomial.

Here M'_e is the *restriction* of $M = (E, r)$ to the set $E\backslash\{e\}$ with r unchanged. The *contraction* M''_e can be defined by $M''_e = (M^*)'_e$ and is the exact analogue of contraction in graphs. For matrices it corresponds to *projection* from the column vector e. A *minor* of M is any matroid N obtainable from M by a sequence of contractions and deletions.

The recipe theorem can now be stated as follows:

(2.16) Theorem. *Let $\mathscr{C}$ be a class of matroids which is closed under direct sums and the taking of minors, and suppose that f is well defined on $\mathscr{C}$ and satisfies*

$$f(M) = af(M'_e) + bf(M''_e), \qquad e \in E(M), \tag{2.17}$$

$$f(M_1 \oplus M_2) = f(M_1)f(M_2). \tag{2.18}$$

Then f is given by

$$f(M) = a^{|E|-r(E)}b^{r(E)}T\left(M; \frac{x_0}{b}, \frac{y_0}{a}\right),$$

where x_0 and y_0 are the values f takes on coloops and loops, respectively.

Any invariant f which satisfies (2.17)–(2.18) is called a *Tutte–Grothendieck* (TG) *invariant*.

Thus, what we are saying is that any TG invariant has an interpretation as an evaluation of the Tutte polynomial.

3. RELIABILITY AND FLOWS

We illustrate the use of the recipe theorem with two applications.

Reliability theory deals with the probability of points of a network being connected when individual links or edges are unreliable. It has a huge literature, see, for example, Colbourn [7].

Let G be a connected graph in which each edge is independently *open* with probability p and *closed* with probability $q = 1 - p$. The (*all terminal*) *reliability* $R(G; p)$ denotes the probability that in this random model there is a path between each pair of vertices of G. Thus

$$R(G; p) = \sum_A p^{|A|}(1-p)^{|E\backslash A|}, \tag{3.1}$$

where the sum is over all subsets A of edges which contain a spanning tree of G, and $E = E(G)$.

It is immediate from this that R is a polynomial in p and a simple conditioning argument shows the following connection with the Tutte polynomial.

(3.2) If G is a connected graph and e is not a loop or coloop, then

$$R(G; p) = qR(G'_e; p) + pR(G''_e; p),$$

where $q = 1 - p$.

Using this with the recipe Theorem 2.16, it is straightforward to check the following statement.

(3.3) Provided G is a connected graph,

$$R(G; p) = q^{|E|-|V|+1} p^{|V|-1} T(G; 1, q^{-1}).$$

We now turn to flows. Take any graph G and orient its edges arbitrarily. Take any finite Abelian group H and call a mapping ϕ: $E(G) \to H \backslash \{0\}$ an *H-flow* if Kirchhoff's laws are obeyed at each vertex of G, the algebra of course being that of the group H.

Note. Standard usage is to describe what we call an H-flow a *nowhere zero H*-flow.

A very good survey of the classical problems of nowhere zero flows is given by Jaeger [12]. It also gives a nice explanation of the (not difficult to prove) relationship between flows in a planar graph and colorings of its planar dual. We start our brief treatment here with the following, rather surprising result.

(3.4) The number of H-flows on G depends only on the order of H and not on its structure.

This is an immediate consequence of the fact that the number of flows is a TG-invariant. To see this, let $F(G; H)$ denote the number of H-flows on G. Then a straightforward counting argument shows that the following is true.

(3.5) Provided the edge e is not an isthmus or a loop of G, then

$$F(G; H) = F(G''_e; H) - F(G'_e; H).$$

Now it is easy to see that if C and L represent, respectively, a coloop (=isthmus) and a loop, then

$$F(C; H) = 0, \qquad F(L; H) = o(H) - 1, \tag{3.6}$$

where $o(H)$ is the order of H.

Accordingly we can apply the recipe theorem and obtain:

(3.7) For any graph G and any finite abelian group H,

$$F(G;H) = (-1)^{|E|-|V|+k(G)}T(G;0,1-o(H)).$$

The observation (3.4) is an obvious corollary.

A consequence of this is that we can now speak of G *having a k-flow* to mean that G has a flow over *any* or equivalently *some* Abelian group of order k.

Moreover it follows that there exists a polynomial $F(G;\lambda)$ such that if H is Abelian of order k, then $F(G;H) = F(G;k)$. We call F the *flow polynomial* of G.

The duality relationship (2.15) leads to:

(3.8) If G is planar, then the flow polynomial of G is essentially the chromatic polynomial of G^* in the sense that

$$\lambda^{k(G)}F(G;\lambda) = P(G^*;\lambda).$$

A consequence of this and the four color theorem is that:

(3.9) Every planar graph having no isthmus has a 4-flow.

What is much more surprising is that the following statement is believed to be true:

(3.10) Tutte's 5-Flow Conjecture. *Any graph having no isthmus has a* 5*-flow.*

It is far from obvious that there is any universal constant k such that graphs without isthmuses have a k-flow. However, Seymour [26] showed:

(3.11) Theorem. *Every graph having no isthmus has a* 6*-flow.*

For more on this and a host of related graph theoretic problems, we refer to Jaeger [12].

4. THE RANDOM CLUSTER MODEL

A special case of the Tutte polynomial which is particularly relevant to the theme of this meeting is the following correlated percolation model which was defined by Fortuin and Kasteleyn [10] and which contains as special cases the ferromagnetic versions of the Ising and Potts models of statistical physics. It can be defined as follows.

Given $G = (V, E)$ and parameters $p \in [0,1]$, classical percolation theory, like random graph theory, is concerned with a probability model in which each edge *independently* is *white* or *open* with probability p and *black* or *closed* with probability $q = 1 - p$. In the random cluster model there is an additional parameter Q which is used to set up a probability measure on the edge set of G so that the probability $\mu(A)$ that the set A of edges is white is given by

$$\mu(A) = Z^{-1} p^{|A|}(1-p)^{|E \backslash A|} Q^{k(A)}.$$

Here $k(A)$ is the number of components in the graph $G \parallel A = (V, A)$ and Z is the normalizing constant defined by

$$Z = Z(G; p, Q) = \sum_{A \subseteq E} \mu(A).$$

In other words, Z is the *partition function* of the model.

The relation with T is that

$$Z(G; p, Q) = p^{r(E)} q^{r^*(E)} Q^{k(G)} T\left(G; 1 + \frac{Qq}{p}, \frac{1}{q}\right), \tag{4.1}$$

where r^* is the dual rank, $q = 1 - p$, and $k(G)$ is the number of connected components of G.

It follows that for any given $Q \geq 0$, determining the partition function Z reduces to determining T along the hyperbola H_Q given by $(x-1)(y-1) = Q$. Moreover, since in its physical interpretations, p is a probability, the reparametrization means that Z is evaluated only along the positive branch of this hyperbola. In other words, Z is the specialization of T to the quadrant $x \geq 1$, $y \geq 1$.

The antiferromagnetic Ising and Potts models are contained in T along the negative branches of the hyperbolae H_Q, but do not have representations in the random cluster model. For more on this model and its relation to T, see Welsh [32, Chap. 4].

5. SOME WELL KNOWN INVARIANTS

We now collect together some of the naturally occurring interpretations of the Tutte polynomial. Throughout G is a graph, M is a matroid, and E will denote $E(G)$, $E(M)$, respectively.

- **(5.1)** At $(1, 1)$, T counts the number of bases of M (spanning trees in a connected graph).
- **(5.2)** At $(2, 1)$, T counts the number of independent sets of M (forests in a graph).
- **(5.3)** At $(1, 2)$, T counts the number of spanning sets of M, that is, sets which contain a base, which in the case that G is a connected graph means the number of connected subgraphs.
- **(5.4)** At $(2, 0)$, T counts the number of acyclic orientations of G.
- **(5.5)** Another interpretation at $(2, 0)$, and this for a different class of matroids, was discovered by Zaslavsky [35]. This is in terms of counting the number of different arrangements of sets of hyperplanes in n-dimensional Euclidean space.
- **(5.6)** $T(G; -1, -1) = (-1)^{|E|}(-2)^{d(B)}$, where B, is the bicycle space of G; see Rosenstiehl and Read [25]. When G is planar, it also has interpretations in terms of the Arf invariant of the associated knot.

(5.7) The chromatic polynomial $P(G;\lambda)$ is given by

$$P(G;\lambda) = (-1)^{r(E)}\lambda^{k(G)}T(G; 1-\lambda, 0),$$

where $k(G)$ is the number of connected components.

(5.8) The flow polynomial $F(G;\lambda)$ discussed in Section 3, is given by

$$F(G;\lambda) = (-1)^{|E|-r(E)}T(G; 0, 1-\lambda),$$

and comparing with (5.6) we see the duality between flows and colorings when G is planar.

(5.9) The (all terminal) reliability $R(G : p)$ is given by

$$R(G; p) = q^{|E|-r(E)}p^{r(E)}T(G; 1, 1/q),$$

where $q = 1 - p$.

In each of the above cases, the interesting quantity (on the left-hand side) is given (up to an easily determined term) by an evaluation of the Tutte polynomial. We shall use the phrase "specializes to" to indicate this. Thus, for example, along $y = 0$, T specializes to the chromatic polynomial.

It turns out that the hyperbolae H_α defined by

$$H_\alpha = \{(x, y) : (x-1)(y-1) = \alpha\}$$

seem to have a special role in the theory. We note several important specializations below.

(5.10) Along H_1, $T(G; x, y) = x^{|E|}(x-1)^{r(E)-|E|}$.

(5.11) Along H_2, when G is a graph, T specializes to the partition function of the Ising model.

(5.12) Along H_q, for general positive integer q, T specializes to the partition function of the Potts model of statistical physics.

The statement (5.11) is just the special case of (5.12) when the number Q of states in the Potts model equals 2. The partition function $Z(G; Q, J)$ of the Q-state Potts model is obtained from the expression (4.1) giving the partition function $Z(G; p, Q)$ of the random cluster model by the substitution

$$p = 1 - e^{-\beta J}.$$

Here $\beta > 0$ is inverse temperature and J is the (constant) pairwise interaction.

Thus $Z_{\text{Potts}}(G) = Z(G; Q, \beta, J)$ is given by

$$Z_{\text{Potts}}(G) \propto T\left(G; 1 + \frac{Qe^{-\beta J}}{1-e^{-\beta J}}, e^{\beta J}\right).$$

Comparing this with the hyperbola H_Q we see that the ferromagnetic version ($J > 0$) corresponds to the positive branch of H_Q with temperature increasing from 0 at $x = 1$ to ∞ as $x \to \infty$.

The antiferromagnetic Potts model ($J < 0$) is contained in that part of H_Q where $0 < y < 1$.

(5.13) Along H_q, when q is a prime power, for a matroid M of vectors over $GF(q)$, T specializes to the weight enumerator of the linear code over $GF(q)$, determined by M. Equation (2.14) relating $T(M)$ to $T(M^*)$ gives the MacWilliams identity of coding theory.

(5.14) Along H_q, for any positive, not necessarily integer, q, T specializes to the partition function of the random cluster model discussed in Section 4.

(5.15) Along the hyperbola $xy = 1$, when G is planar, T specializes to the Jones polynomial of the alternating link or knot associated with G. This connection was first discovered by Thistlethwaite [30].

Other more specialized interpretations can be found in the survey of Brylawski and Oxley [6] and Welsh [32]. In the next few sections we give some more recent applications.

6. COUNTING COLORINGS AND FLOWS IN RANDOM GRAPHS

Although the theory of random graphs is highly developed, less attention seems to have been paid to counting problems. Here we give some results obtained in Welsh [34] which give new interpretations of the Tutte polynomial as the expected value of classical counting functions.

Given an arbitrary graph G and $p \in [0, 1]$, we denote by G_p the *random subgraph* of G obtained by deleting each edge of G independently with probability $1 - p$.

This is a generalization of the standard random graph model $G_{n,p}$ which corresponds to $(K_n)_p$.

First an easy result to illustrate the notation. If $f(G_p)$ denotes the number of forests in G_p, then, for G and p fixed, this is a random variable and has an expectation which we denote by $\langle f(G_p)\rangle$.

Routine calculation gives:

(6.1) For any connected graph G,

$$\langle f(G_p)\rangle = p^{|V|-1}T\left(G; 1 + \frac{1}{p}, 1\right).$$

Turning now to colorings, we have:

(6.2) Theorem. *For any connected graph G and $0 < p \leq 1$, the random subgraph G_p has a chromatic polynomial whose expectation is given by*

$$\left\langle P(G_p; \lambda)\right\rangle = (-p)^{|V|-1}\lambda T(G; 1 - \lambda p^{-1}, 1 - p).$$

For the flow polynomial there is a similar, but more complicated evaluation, namely:

(6.3) Theorem. *For any graph G, the flow polynomial $F(G_p; \lambda)$ has expectation given by*:

(a) *if* $p \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$, *then*

$$\langle F(G_p; \lambda)\rangle = p^r (q-p)^{r^*} T\left(G; qp^{-1}, 1 + \frac{\lambda p}{q-p}\right),$$

where $q = 1 - p$;

(b) *if* $p = \frac{1}{2}$, *then*

$$\langle F(G_{\frac{1}{2}}; \lambda)\rangle = \lambda^{|E|-|V|+k(G)} 2^{-|E|}.$$

7. CHIP FIRING

A very recent new specialization of T has been obtained by Merino-Lopez [21]. It concerns a version of chip firing as in Bjorner, Lovász, and Shor [5] and gives a specific relationship between evaluations of T along the line $x = 1$ and the generating function of critical configurations in the chip firing game.

First we define the *chip firing game*. Take a graph $G = (V, E)$ with a special vertex q (called the *bank*). Each vertex v of $V \backslash \{q\}$ has an integer quantity of chips $\theta(v)$ and

$$\theta(q) = \sum_{v \neq q} \theta(v).$$

A *firing* of a vertex v is the transfer of one chip from v to each of its neighbors with a consequent decrease in $\theta(v)$. An ordinary vertex v may only fire if $\theta(v)$ is not less than its degree and in this case v is said to be *ready*.

The special vertex q may (must) fire only when no ordinary vertex is ready. In this case, the configuration θ is said to be *stable*.

For any configuration θ, a sequence $v_1, v_2, \ldots, v_k$ of vertices is *q-legal* for θ if v_1 is ready in θ, v_2 is ready in the configuration obtained from θ by v_1 firing, and so on. Furthermore, we *insist* that q occurs in a q-legal sequence if and only if the preceding configuration is stable. In other words, q *must* fire when it is ready.

Thus a q-legal sequence can continue indefinitely. Hence certain configurations are *recurrent* in that there exists a q-legal sequence which starts and ends with it.

A configuration is *critical* if it is recurrent and stable, and it is known that if θ is critical, there is a q-legal sequence which starts and ends with θ and realizes its recurrence and moreover in this sequence each vertex fires exactly once.

The *level* of a configuration θ is defined by

$$\text{level}(\theta) = \sum_{i \neq q} \theta(i) + \deg(q) - |E|.$$

The theorem conjectured by Biggs [4] and proved by Merino-Lopez [21] is the following.

(7.1) Theorem. *If c_i denotes the number of critical configurations of level i in a graph G with special vertex q, then*

$$P_q(G; y) = \sum_{i=0}^{\infty} c_i y^i = T(G; 1, y).$$

A first, nontrivial consequence of this is that it shows $P_q(G; y)$ is independent of choice of q.

Other immediate consequences are that it proves the existence of bijections between the set of critical configurations of different levels and previously known evaluations of T along the line $x = 1$, for example, coefficients of the reliability polynomial.

8. THE EHRHART POLYNOMIAL OF A UNIMODULAR ZONOTOPE

Let $\mathbb{Z}^n$ denote the n-dimensional integer lattice in $\mathbb{R}^n$ and let P be an n-dimensional lattice polytope in R^n; that is, a convex polytope whose vertices have integer coordinates. Consider the function $i(P; t)$, which, when t is a positive integer, counts the number of lattice points which lie inside the dilated polytope tP. Ehrhart [9] initiated the systematic study of this function by proving that it is always a polynomial in t, and that, in fact,

$$i(P, t) = \chi(P) + c_1 t + \cdots + c_{n-1} t^{n-1} + \operatorname{vol}(P) t^n.$$

Here

$$c_0 = \chi(P) \quad \text{is the Euler characteristic}$$

of P and $\operatorname{vol}(P)$ is the volume of P.

Until recently the other coefficients of $i(P, t)$ remained a mystery, even for simplices; see, for example, Stanley [29] and Diaz and Robins [8].

However, in the special case that P is a unimodular zonotope, there is a nice interpretation of these coefficients. First recall that if A is an $r \times n$ matrix, written in the form $A = [a_1, \ldots, a_n]$, then it defines a *zonotope* $Z[A]$ which consists of those points p of $\mathbb{R}^r$ which can be expressed in the form

$$p = \sum_{i=1}^{n} \lambda_i a_i, \qquad 0 \leq \lambda_i \leq 1.$$

In other words, $Z[A]$ is the *Minkowski sum* of the line segments $[0, a_i]$, $1 \leq i \leq n$.

It is a convex polytope which, when A is a totally unimodular matrix, has all integer vertices and in this case it is described as a *unimodular zonotope*. For these polytopes, a result from Stanley [28] shows that

$$i(Z(A); t) = \sum_{k=0}^{r} i_k t^k,$$

where i_k is the number of subsets of columns of the matrix A which are linearly independent and have cardinality k.

In other words, the Ehrhart polynomial $i(Z(A); t)$ is the generating function of the number of independent sets in the matroid $M(A)$. But from (2.3) we know that for any matroid M, the evaluation of $T(M; x, y)$ along the line $y = 1$ also gives this generating function. Hence, combining these observations we have the following result.

(8.1) Theorem. *If M is a regular matroid and A is any totally unimodular representation of M, then the Ehrhart polynomial of the zonotope $Z[A]$ is given by*

$$i(Z(A);\lambda) = \lambda^r T\left(M; 1+\frac{1}{\lambda}, 1\right),$$

where r is the rank of M.

Another new interpretation of T follows from what is sometimes known as the Ehrhart–Macdonald reciprocity law. This states that for any convex polytope P with integer vertices in $\mathbb{R}^n$ and for any positive integer t, the function $k(P;t)$ counting the number of lattice points lying strictly inside tP is given by

$$k(P;t) = (-1)^n i(P;t).$$

This gives:

(8.2) Corollary. *If A is an $r \times n$ totally unimodular matrix of rank r, then for any positive integer λ, the number of lattice points of $\mathbb{R}^r$ lying strictly inside the zonotope $\lambda Z[A]$ is given by*

$$k[Z[A];\lambda] = \lambda^r T\left(M(A); 1-\frac{1}{\lambda}, 1\right).$$

In particular we have the following new interpretations

(8.3) The number of lattice points strictly inside $Z[A]$ is $T(M;0,1)$.

9. THE COMPLEXITY OF THE TUTTE PLANE

We have seen that along different curves of the x, y plane, the Tutte polynomial evaluates many diverse quantities. Since it is also the case that for particular curves and at particular points the computational complexity of the evaluation can vary from being polynomial time computable to being #P-hard, a more detailed analysis of the complexity of evaluation is needed in order to give a better understanding of what is and is not computationally feasible for these sorts of problems. The main result of Jaeger, Vertigan, and Welsh [13] is the following:

(9.1) Theorem. *The problem of evaluating the Tutte polynomial of a graph at a point (a,b) is #P-hard except when (a,b) is on the special hyperbola*

$$H_1 \equiv (x-1)(y-1) = 1$$

or when (a,b) is one of the special points $(1,1)$, $(-1,-1)$, $(0,-1)$, $(-1,0)$, $(i,-i)$, $(-i,i)$, (j,j^2), and (j^2,j), where $j = e^{2\pi i/3}$. In each of these exceptional cases the evaluation can be done in polynomial time.

As far as planar graphs are concerned, there is a significant difference. The technique developed using the Pfaffian to solve the Ising problem for the plane square lattice by Kasteleyn [18] can be extended to give a polynomial time algorithm for the evaluation of the Tutte polynomial of any planar graph along the special hyperbola

$$H_2 \equiv (x-1)(y-1) = 2.$$

A good reference and treatment of this extension can be found in Chapter 8 of Lovász and Plummer [19]. Thus this hyperbola is also "easy" for planar graphs. However, it is easy to see that H_3 cannot be easy for planar graphs since it contains the point $(-2, 0)$ which counts the number of 3-colorings and since deciding whether a planar graph is 3-colorable is NP-hard, this must be at least NP-hard. However, it does not seem easy to show that H_4 is hard for planar graphs. The decision problem is after all trivial by the four color theorem. The fact that it is $\#P$-hard is just part of the following extension of Theorem 9.1 from Vertigan and Welsh [13].

(9.2) Theorem. *The evaluation of the Tutte polynomial of bipartite planar graphs at a point (a, b) is $\#P$-hard except when*

$$(a, b) \in H_1 \cup H_2 \cup \{(1, 1), (-1, -1), (j, j^2), (j^2, j)\},$$

when it is computable in polynomial time.

Notice that in the above there is no mention of the more general question about the complexity of evaluating $T(M; x, y)$, where M is a matroid. This is because for general M, the size of input is huge. To describe a typical matroid on a set of n elements will need time $\Omega(2^n)$. The complexity of such questions is therefore not of current practical importance.

The interesting, practical problems are when M has a natural succinct description and then it seems to be the case that this description is by a graph or matrix.

It follows immediately from the fact that any graph can be represented as a totally unimodular matrix that if a problem is hard (in any formal sense) for graphs, then it will be at least as hard for matrices.

10. APPROXIMATING TO WITHIN A RATIO

For positive numbers a and $r \geq 1$, we say that a third quantity $\hat{a}$ *approximates a within ratio r* or is an *r-approximation* to a, if

$$r^{-1}a \leq \hat{a} \leq ra. \tag{10.1}$$

In other words, the ratio $\hat{a}/a$ lies in $[r^{-1}, r]$.

Now consider what it would mean to be able to find a polynomial time algorithm which gave an approximation within r to the number of 3-colorings of a graph. We would clearly have a polynomial time algorithm which would decide whether or not a graph is 3-colorable. But this is NP-hard. Thus no such algorithm can exist unless $NP = P$.

The same argument can be applied to any function which counts objects whose existence is NP-hard to decide. Hence:

(10.2) Proposition. *Unless $NP = P$, there can be no polynomial time approximation to $T(G, 1-k, 0)$ for integer $k \geq 3$.*

We now consider a randomized approach to counting problems and make the following definition.

An ϵ-δ *approximation scheme* for a counting problem f is a Monte Carlo algorithm which on every input $\langle x, \epsilon, \delta \rangle$, $\epsilon > 0$, $\delta > 0$, outputs a number $\tilde{Y}$ such that

$$\Pr\{(1-\epsilon)f(x) \leq \tilde{Y} \leq (1+\epsilon)f(x)\} \geq 1-\delta.$$

Now let f be a function from input strings to the natural numbers. A *randomized approximation scheme* for f is a probabilistic algorithm that takes as an input a string x and a rational number ϵ, $0 < \epsilon < 1$, and produces as output a random variable Y, such that

$$\Pr\left\{1-\epsilon \leq \frac{Y}{f(x)} \leq 1+\epsilon\right\} \geq \frac{3}{4}. \tag{10.3}$$

A *fully polynomial randomized approximation scheme* (fpras) for a function $f\colon \Sigma^* \to \mathbf{N}$ is a randomized approximation scheme which runs in time which is a polynomial function of n *and* ϵ^{-1}.

Suppose now we have such an approximation scheme and suppose further that it works in polynomial time. Then we can boost the success probability up to $1-\delta$ for any desired $\delta > 0$, by using the following trick of Jerrum, Valiant, and Vazirani [16]). This consists of running the algorithm $O(\log \delta^{-1})$ times and taking the median of the results.

We make this precise as follows:

(10.4) Proposition. *If there exists an fpras for computing f, then there exists an ϵ-δ approximation scheme for f which on input $\langle x, \epsilon, \delta \rangle$ runs in time which is bounded by $O(\log \delta^{-1})\mathrm{poly}(|x|, \epsilon^{-1})$.*

The existence of an fpras for a counting problem is a very strong result see for example [14] and [27]; it is the analogue of an RP algorithm for a decision problem and corresponds to the notion of tractability. However we should also note:

(10.5) Proposition. *If $f\colon \Sigma^* \to \mathbf{N}$ is such that deciding if f is nonzero is NP-hard, then there cannot exist an fpras for f unless NP is equal to random polynomial time RP.*

Since this is thought to be unlikely, it makes sense only to seek out an fpras when counting objects for which the decision problem is not NP-hard.

Hence we have immediately from the NP-hardness of k-coloring for $k \geq 3$ that:

(10.6) Proposition. *Unless $NP = RP$, there cannot exist an fpras for evaluating $T(G; -k, 0)$ for any integer $k \geq 2$.*

Recall now from (5.12) that along the hyperbola, H_Q, for positive integer Q, T evaluates the partition function of the Q-state Potts model.

In an important paper, Jerrum and Sinclair [15], have shown that there exists an fpras for the ferromagnetic Ising problem. This corresponds to the $Q = 2$ Potts model and, thus, their result can be restated in the terminology of this paper as follows.

(10.7) Theorem. *There exists an fpras for estimating T along the positive branch of the hyperbola H_2.*

However it seems to be difficult to extend the argument to prove a similar result for the Q-state Potts model with $Q > 2$, and this remains one of the outstanding open problems in this area.

A second result of Jerrum and Sinclair [15] is the following:

(10.8) Theorem. *There is no fpras for estimating the antiferromagnetic Ising partition function unless $NP = RP$.*

In the context of its Tutte plane representation, this can be restated as follows.

(10.9) Theorem. *Unless $NP = RP$, there is no fpras for estimating T along the curve*

$$\{(x, y) : (x-1)(y-1) = 2, 0 < y < 1\}.$$

The following extension of the result is proved in Welsh [33].

(10.10) Theorem. *On the assumption that $NP \neq RP$, the following statements are true.*

- **(a)** *Even in the planar case, there is no fully polynomial randomized approximation scheme for T along the negative branch of the hyperbola H_3.*
- **(b)** *For $Q = 2, 4, 5, \ldots,$ there is no fully polynomial randomized approximation scheme for T along the curves*

$$H_Q^- \cap \{x < 0\}.$$

The reader will notice the curious difference between the results for $Q = 3$ and $Q \neq 3$. The reason for this is that the proof that the whole negative branch of the hyperbola H_3 is hard to approximate hinges on the fact that if we could approximate it, we could use duality and decide whether a planar graph is 3-colorable, and this is NP-hard. I believe that the result stated in (a) for $Q = 3$ is also true for $Q \neq 3$, but I have been unable to prove it.

It is worth emphasizing that the above statements do not rule out the possibility of there being an fpras at *specific points* along the negative hyperbolae. For example:

- T can be evaluated exactly at $(-1, 0)$ and $(0, -1)$, which both lie on H_2^-.
- There is no inherent obstacle to there being an fpras for estimating the number of k-colorings of a planar graph for any $k \geq 4$.

Similarly, since by Seymour's theorem, every bridgeless graph has a nowhere zero 6-flow, there is no obvious obstacle to the existence of an fpras for estimating the number of k-flows for $k \geq 6$. A natural question, in the same spirit is the following.

(10.11) Problem. *Show that there does not exist an fpras for estimating T at* $(0, -5)$. *More generally, show that there is no fpras for estimating the number of k-flows for* $k \geq 6$.

11. POSITIVE RESULTS

Mihail and Winkler [22] have shown that there exists an fpras for counting the number of distinct Eulerian orientations of any Eulerian graph. When G is 4-regular, each Eulerian orientation must have exactly two edges directed in and two out. In other words, Eulerian orientations correspond bijectively to ice configurations or equivalently to the collection of $(Z_2 \times Z_2)$-flows. This is equivalent to the statement:

(11.1) There is an fpras for computing T at $(0, -2)$ for 4-regular graphs.

The reader will note that all the "negative results" are about evaluations of T in the region outside the quadrant $x \geq 1$, $y \geq 1$. In Welsh [32] it is conjectured that the following is true:

(11.2) Conjecture. *There exists an fpras for evaluating T at all points of the quadrant* $x \geq 1$, $y \geq 1$.

Some evidence in support of this is the following.

If we let $\mathcal{G}_\alpha$ be the collection of graphs $G = (V, E)$ such that each vertex has at least $\alpha |V|$ neighbors, then we call a class $\mathcal{C}$ of graphs *dense* if $\mathcal{C} \subseteq \mathcal{G}_\alpha$ for some fixed $\alpha > 0$.

Annan [2] showed that:

(11.3) There exists an fpras for counting forests in any class of dense graphs.

Now the number of forests is just the evaluation of T at the point $(2, 1)$ and a more general version of (11.3) is the following result, also by Annan.

(11.4) For any class of dense graphs, there is an fpras for evaluating $T(G; x, 1)$ for positive integer x.

The natural question suggested by (11.4) is about the matroidal dual—namely, does there exist an fpras for evaluating T at $(1, x)$? This is the reliability question, and in particular, the point $(1, 2)$ enumerates the number of connected subgraphs. Within the class of graphs, duality only applies to planar graphs, so Annan's methods do not work since planar graphs cannot belong to any dense class. What can be proved is the following. Alon, Frieze, and Welsh [1] showed:

(11.5) Theorem. **(a)** *There exists a fully polynomial randomized approximation scheme for evaluating* $T(G; x, y)$ *for all* $x \geq 1$, $y \geq 1$ *for any dense class of graphs.*

(b) *For any class of strongly dense graphs, meaning $G \in \mathcal{G}_\alpha$ for $\alpha > \frac{1}{2}$, there is also such a scheme for $x < 1$, $y \geq 1$.*

Even more recently, Karger [17] has proved the existence of a similar scheme for the class of graphs with no small edge cut set. This can be stated as follows.

For $c > 0$ define the class $\mathcal{G}^c$ by $G \in \mathcal{G}^c$ iff its edge connectivity is at least $c \log |V(G)|$. A class of graphs is *well connected* if it is contained in $\mathcal{G}^c$ for some fixed c.

(11.6) Theorem. *For any fixed (x, y), $y > 1$, there exists c, depending on (x, y), such that for any class $\mathcal{C} \subseteq \mathcal{G}^c$, there is an fpras for evaluating $T(G; x, y)$.*

Notice that though the properties of being well connected and dense are very similar, neither property implies the other.

12. CONCLUSION

As we have seen, Conjecture 11.2 has been proved for classes of dense and well connected graphs. There is also no "natural impediment" to it being true for all graphs. However, for the d-dimensional hypercubical lattice, it is known that there exists $Q(d)$ such that the random cluster model has a first-order discontinuity for $Q > Q(d)$. Indeed it is believed that

$$Q(d) = \begin{cases} 4 & d = 2, \\ 2 & d \geq 6. \end{cases}$$

It is not unreasonable to associate a first-order discontinuity with an inability to approximate. There is no proof of such a general statement, but there are persuasive arguments to suggest that such discontinuities would prevent an approximation scheme based on sampling by the Markov chain method. Hence, a major open question which was first promoted by discussions at this meeting must be whether or not there exists an fpras for the ferromagnetic random cluster model for hypercubical lattices. In other words, can $T(G; x, y)$ be approximated for all $x \geq 1$, $y \geq 1$ when G is a d-dimensional lattice?

Note also that hypercubical lattices are neither dense nor well connected, so the positive results of Section 11 do not apply. Key points with respect to this question appear to be $(2, 1)$, at which T counts forests, and $(2, 0)$, at which T counts the number of acyclic orientations. A new approach to approximation at these points is proposed in Bartels, Mount, and Welsh [3]. This is based on the interpretation of T as the Ehrhart polynomial of a unimodular zonotope $Z[A]$ as described in Section 8. Counting the number of forests is the problem of counting lattice points contained in the zonotope $Z[A]$. Counting the number of acyclic orientations is the problem of counting the vertices of this zonotope. The latter is a much more difficult problem and goes some way to explaining the total lack of success with it. What is even more discouraging is that a completely new approach from [3] to counting lattice points in the zonotope using the same random walk method as used by Lovász and Simonovits [20] and others in their successful attack on approximating volume hits exactly the same barrier—namely, a density condition, as in [1].

This together with the physical arguments based on first-order discontinuities are reasons for pessimism about the existence of an fpras, and certainly lend additional weight to a growing belief that there can be no fpras for the number of acyclic orientations in a general graph. Proving such a negative result would be a major advance.

REFERENCES

[1] N. Alon, A.M. Frieze, and D.J.A. Welsh, Polynomial time randomised approximation schemes for Tutte–Grothendieck invariants: the dense case, Random Struct Alg, 6 (1995), 459–478.

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

[3] E. Bartels, J. Mount, and D.J.A. Welsh, The win polytope of a graph, Ann Combin, 1 (1997), 1–15.

[4] N. Biggs, The Tutte polynomial as a growth function, preprint, 1997.

[5] A. Bjorner, L. Lovász, and P. Shor, Chip-firing games on graphs, Eur J Combin, 12 (1991), 283–291.

[6] T.H. Brylawski and J.G. Oxley, The Tutte polynomial and its applications, Matroid Applications, N. White, (Editor), Cambridge Univ. Press, Cambridge, UK, 1992, pp. 123–225.

[7] C.J. Colbourn, The Combinatorics of Network Reliability, Oxford Univ. Press, London, 1987.

[8] R. Diaz and S. Robins, The Ehrhart polynomial of a lattice n-simplex, Electron Res Announc Amer Math Soc, 2(1) (1996), 1–6.

[9] E. Ehrhart, Sur un problème de géometrie diophantienne linéaire I, II, J Reine Angew Math, 226 (1967), 1–29; 227 (1967), 25–49; correction 231 (1968), 220.

[10] C.M. Fortuin and P.W. Kasteleyn, On the random cluster model. I. Introduction and relation to other models, Physica, 57 (1972), 536–564.

[11] M.R. Garey and D.S. Johnson, Computers and Intractability—A Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.

[12] F. Jaeger, Nowhere zero flow problems. Selected Topics in Graph Theory L. Beineke and R. J. Wilson (Editors), Academic, New York, 1988, Vol. 3, pp. 71–92.

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

[14] M.R. Jerrum and A. Sinclair, Approximating the permanent, SIAM J Comput, 18 (1989), 149–178.

[15] M.R. Jerrum and A. Sinclair, Polynomial-time approximation algorithms for the Ising model, Proc 17th ICALP, EATCS, 1990, pp. 462–475.

[16] M.R. Jerrum, L.G. Valiant, and V.V. Vazirani, Random generation of combinatorial structures from a uniform distribution, Theoret Comput Sci 43 (1986), 169–188.

[17] D.R. Karger, A randomised fully polynomial time approximation scheme for the all terminal network reliability problem, Proc 36th Annual IEEE Symp on Foundations of Computer Science, 1995, pp. 328–337.

[18] P.W. Kasteleyn, The statistics of dimers on a lattice, Physica, 27 (1961), 1209–1225.

[19] L. Lovàsz and M.D. Plummer, Matching Theory, North-Holland, Amsterdam, 1986.

[20] L. Lovàsz and M. Simonovits, Random walks in a convex body and an improved volume algorithm, Random Struct Alg, 4 (1993), 359–412.

[21] C. Merino-Lopez, Chip-firing and the Tutte polynomial, Ann Combin, 1 (1997), 253–259.

[22] M. Mihail and P. Winkler, On the number of Eulerian orientations of a graph, Proc 3rd Annual ACM-SIAM Symp on Discrete Algorithms (SODA), 1992, pp. 138–145.

[23] J.G. Oxley, Matroid Theory, Oxford Univ. Press, London, 1992.

[24] J.G. Oxley and D.J.A. Welsh, The Tutte polynomial and percolation, Graph Theory and Related Topics, J. A. Bondy and U. S. R. Murty (Editors) Academic, London, 1979, pp. 329–339.

[25] P. Rosenstiehl and R.C. Read, On the principal edge tripartition of a graph, Ann Discrete Math, 3 (1978), 195–226.

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

[27] A.J. Sinclair and M.R. Jerrum, Approximate counting, uniform generating and rapidly mixing Markov chains, Inform Comput, 82 (1989), 93–133.

[28] R.P. Stanley, Decompositions of rational convex polytopes, Ann Discrete Math, 6 (1980), 333–342.

[29] R. Stanley, Enumerative Combinatorics, Wadsworth, Monterey, CA, (1986).

[30] M.B. Thistlethwaite, A spanning tree expansion of the Jones polynomial, Topology, 26 (1987), 297–309.

[31] D.L. Vertigan and D.J.A. Welsh, The computational complexity of the Tutte plane: the bipartite case, Probab Combin Comput Sci, 1 (1992), 181–187.

[32] D.J.A. Welsh, Complexity: Knots, Colourings and Counting, London Mathematical Society Lecture Note Series, Cambridge Univ. Press, Cambridge, UK, 1993, Vol. 186.

[33] D.J.A. Welsh, Randomised approximation in the Tutte plane, Combin Probab Comput, 3 (1993), 137–143.

[34] D.J.A. Welsh, Counting, colorings and flows in random graphs, Bolyai Soc Math Stud, 2 (1996), 491–506.

[35] T. Zaslavsky, Facing up to arrangements: face count formulas for partitions of spaces by hyperplanes, Mem Amer Math Soc, 154 (1975).