

Subset Glauber dynamics on graphs, hypergraphs and matroids of bounded tree-width*

Magnus Bordewich

School of Engineering and Computing Sciences
Durham University
Durham, UK

`m.j.r.bordewich@durham.ac.uk`

Ross J. Kang[†]

IMAPP
Radboud University Nijmegen
Nijmegen, Netherlands

`ross.kang@gmail.com`

Submitted: Mar 16, 2014; Accepted: Oct 12, 2014; Published: Oct 23, 2014

Mathematics Subject Classifications: 05C85, 68R10, 60J10, 05C31, 68W20, 68W25

Abstract

Motivated by the ‘subgraphs world’ view of the ferromagnetic Ising model, we analyse the mixing times of Glauber dynamics based on subset expansion expressions for classes of graph, hypergraph and matroid polynomials. With a canonical paths argument, we demonstrate that the chains defined within this framework mix rapidly upon graphs, hypergraphs and matroids of bounded tree-width.

This extends known results on rapid mixing for the Tutte polynomial, adjacency-rank (R_2 -)polynomial and interlace polynomial. In particular Glauber dynamics for the R_2 -polynomial was known to mix rapidly on trees, which led to hope of rapid mixing on a wider class of graphs. We show that Glauber dynamics for a very wide class of polynomials mixes rapidly on graphs of bounded tree-width, including many cases in which the Glauber dynamics does not mix rapidly for all graphs. This demonstrates that rapid mixing on trees or bounded tree-width graphs does not offer strong evidence towards rapid mixing on all graphs.

Keywords: Markov chain Monte Carlo; graph polynomials; tree-width; canonical paths; approximate counting

*A preliminary version of this work appeared in the Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP 2011) [15]. Amongst other elements, the extensions to hypergraphs and matroids are new. This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/G066604/1].

[†]This work was completed while this author was employed at Durham University.

1 Introduction

We analyse a subset-sampling Markov chain on graphs (and later hypergraphs and matroids) that is derived from subset expansion¹ graph functions, which include many well-known graph polynomials. We show that this chain mixes rapidly on graphs of constant tree-width.

An *edge subset expansion formula* for a graph function \mathcal{P} is written as follows: for any graph $G = (V, E)$,

$$\mathcal{P}(G) = \sum_{S \subseteq E} w((V, S)) \quad (1)$$

for some graph function w , where (V, S) denotes the graph with vertex set V and edge set S . If the function w is non-negative, that is, $w(G) \geq 0$ for all graphs G , we refer to (1) as an *edge subset weighting* for \mathcal{P} and to w as its *weight function*. In fact, we shall need the weight function to be *positive* — from a statistical physics viewpoint, this results in a so-called ‘soft-core model’. One example of such a function that is prominent in statistical physics, theoretical computer science, and discrete probability is the *random cluster model partition function*, which can be defined for any $G = (V, E)$ and parameters q, μ as

$$Z_{RC}(G; q, \mu) := \sum_{S \subseteq E} q^{\kappa(S)} \mu^{|S|}, \quad (2)$$

where $\kappa(S)$ is the number of components in (V, S) . For more on the random cluster model, see an extensive treatise by Grimmett [33]. Notice that, if $q, \mu \geq 0$, then $w((V, S)) := q^{\kappa(S)} \mu^{|S|}$ provides an edge subset weighting for $Z_{RC}(G; q, \mu)$.

A common approach to approximating graph functions expressed by subset expansion formulae is through Markov chain Monte Carlo sampling (MCMC). A Markov chain with state space $\Omega = \{S : S \subseteq E\} (= 2^E)$ and stationary distribution $\pi(\cdot) \propto w((V, \cdot))$ can be used to sample subsets of the edges. In this approach, an efficient approximation scheme relies on rapid convergence of the Markov chain to its stationary distribution — so-called rapid mixing. In our setting, a natural Markov chain is Glauber dynamics [28], the single bond flip chain, in which possible transitions are the addition or removal of one edge from the current subset (state), with transition probabilities designed to produce the desired stationary distribution. In addition to being a principal tool in the design of efficient approximation schemes for counting problems, MCMC has seen widespread use in computational physics, computational biology, machine learning, and statistics. There have been steady advances in our understanding of such random processes and in showing how quickly they generate good approximations of useful probability distributions in huge, complex data sets. See the lecture notes of Jerrum [40] or a survey by Randall [56] for an overview of the application of these techniques in theoretical computer science.

¹The term ‘subset expansion’ was coined by Gordon and Traldi [32], though it is a special type of ‘states model expansion’ which is commonly applied in statistical physics.

We postpone the precise statement of our main result, Theorem 1, as it requires a host of definitions, but here we give a cursory description. We show that when the weight function w of some subset expansion formula is strictly positive and λ -multiplicative (as defined in Subsection 2.1), then Glauber dynamics is rapidly mixing on graphs of bounded tree-width. Roughly speaking, the condition of λ -multiplicativity is that the weight function is multiplicative with respect to the operation of disjoint graph union as well as “nearly multiplicative” with respect to the operation of composition via small vertex cuts. For now, we remark that many important graph polynomials (e.g. $Z_{RC}(G; q, \mu)$) obey it. The results may be extended to cover hypergraphs and matroids, and polynomials expressed by vertex subset expansion rather than edge subset expansion. Our general approach is an extension of work by Ge and Štefankovič [26] (see also an earlier version [27]), which showed that the Markov chain for the (soft-core) random cluster model — i.e. weighted according to (2) — mixes rapidly upon graphs of bounded tree-width.

Our work started with the R_2 -polynomial; Ge and Štefankovič introduced both this polynomial and its associated Glauber dynamics in an attempt to devise an approximation scheme for $\#BIS$, the problem of counting the number of independent sets in a bipartite graph. Their *adjacency-rank polynomial* is defined for any $G = (V, E)$ and parameters q, μ as

$$R_2(G; q, \mu) := \sum_{S \subseteq E} q^{\text{rk}_2(S)} \mu^{|S|}, \quad (3)$$

where $\text{rk}_2(S)$ is the \mathbb{F}_2 -rank of the *adjacency* matrix for (V, S) . Using a combinatorial interpretation of rk_2 applicable only to bipartite graphs, they showed that the edge subset Glauber dynamics (using the weighting in (3)) mixes rapidly on trees. They conjectured that the chain mixes rapidly on all bipartite graphs, cf. Conjecture 1 in [27]. Our initial objective was to show that the chain mixes rapidly not only on trees, but also on all bipartite graphs of bounded tree-width. We achieved this and more: indeed, we have shown that the R_2 -polynomial fits in our framework without recourse to the combinatorial interpretation for bipartite graphs, and hence that the Markov chain for the R_2 -polynomial mixes rapidly upon all graphs of bounded tree-width. However, in extending the approach of Ge and Štefankovič even further to all λ -multiplicative weight functions, we have highlighted its generality with regard to polynomials and its dependence on bounded tree-width. Since the proof approach works for such a great range of chains that do not mix rapidly in general, it seems that the approach will not extend to larger graph classes without significant alteration. Thus the fact that Glauber dynamics associated with the R_2 -polynomial mixes rapidly on trees cannot be taken as support for the conjecture that it mixes rapidly on all graphs. Indeed, the conjectured rapid mixing of their chain on all bipartite graphs has since been disproved by Goldberg and Jerrum [30].

The structure of this paper is as follows. In Subsections 1.1 and 1.2, we present further background to the work, to graph polynomials and their computation, as well as extensions. In Section 2, we give the definitions that are necessary for a detailed description of the main theorem. We give the main theorem in Section 3 and then indicate some of its consequences. We present the proofs in Section 4. In Section 5, we extend our

results to Glauber dynamics on vertex subsets, that is, on induced subgraphs. For edge subset Glauber dynamics, we outline an extension to hypergraphs in Section 6. We give an indication of how to extend the approach to matroids in Section 7.

1.1 Context

The random cluster model has been widely studied both as given in (2), and as the Tutte polynomial of a graph [61]. Under a suitable transformation, $Z_{RC}(G; q, \mu)$ is equivalent to the Tutte polynomial, defined for any $G = (V, E)$ and parameters x, y as

$$T(G; x, y) := \sum_{S \subseteq E} (x - 1)^{r(E) - r(S)} (y - 1)^{|S| - r(S)}, \quad (4)$$

where $r(S)$ is the \mathbb{F}_2 -rank of the incidence matrix for (V, S) . A wealth of combinatorial and structural information can be obtained from evaluations of this function. Indeed, this polynomial has a remarkable universality property, which informally speaking says that it subsumes any graph invariant that can be computed by deletion and contraction of edges [54], cf. [64]. In addition, the Tutte polynomial specialises to several key univariate graph polynomials, including the chromatic polynomial of Birkhoff [6]. It specialises to the Jones polynomial in knot theory [42]. By its connection with the random cluster model, it also generalises the partition functions of the Ising [38] and Potts [55] models². Consult the monograph of Welsh [63] for more on these crucial connections. In addition to $Z_{RC}(G; q, \mu)$ and $T(G; x, y)$, we shall highlight a few other specific polynomials from the literature, but for a broad account of the development of graph polynomials, consult the recent surveys by Makowsky [47] and by Ellis-Monaghan and Merino [23, 24].

It was shown in 1990 by Jaeger, Vertigan and Welsh [39] that, in general, for fixed (rational) values of x and y , the evaluation of $T(G; x, y)$ is #P-hard, except on a few special points and curves in the (x, y) -plane. As a result, there have been substantial efforts since then to pin down the *approximation* complexity of computing $T(G; x, y)$. For large swaths of the (x, y) -plane, it is now known that the computation of $T(G; x, y)$ either does not admit a fully polynomial-time randomised approximation scheme (FPRAS) unless $\text{RP} = \text{NP}$, or is at least as hard as #BIS (the problem of counting independent sets in bipartite graphs) under approximation-preserving reductions, cf. Goldberg and Jerrum [29]. The sole positive approximation result applicable to general graphs is the breakthrough FPRAS, using MCMC, by Jerrum and Sinclair [41] for the partition function of the ferromagnetic Ising model — this corresponds to computation of $T(G; x, y)$ along the portion of the parabola $(x - 1)(y - 1) = 2$ with $y > 1$. Various approaches have led to efficient approximations in some regions of the Tutte plane for specific classes of graphs — cf. e.g. Alon, Frieze and Welsh [2], Karger [43], and Bordewich [12].

The polynomials and Markov chains that we capture in our framework are defined for all graphs; however, we obtain rapid mixing results only on classes of graphs of bounded (constant) tree-width. For brevity, we do not define tree-width here, but merely say

²If $x, y \geq 1$ or $q, \mu \geq 0$, then, respectively, $T(G; x, y)$ or $Z_{RC}(G; q, \mu)$ generalise the partition functions of the *ferromagnetic* Ising and Potts models.

that it is an essential concept in structural graph theory and parameterised complexity — see modern surveys on the topic by Bodlaender [11] and Hliněný et al. [35]. The restriction of tree-width is commonly used in graph algorithms to reduce the complexity of a computationally difficult problem, usually by way of dynamic programming. For example, it is already known that many of the polynomials covered here can be exactly evaluated efficiently for graphs of bounded tree-width. Independently, Andrzejak [3] and Noble [50] exhibited polynomial-time algorithms to compute the Tutte polynomial of graphs with bounded tree-width. Works of Makowsky and Mariño [48] and Noble [51] have significantly generalised this, in the former case, to a wide array of polynomials under the framework of monadic second order logic (MSOL), and, in the latter case, to the so-called *U*-polynomial [52], a polynomial that includes not only the Tutte polynomial but also a powerful type of knot invariant as special cases.

Even though many of the polynomials we refer to can be computed exactly in polynomial time for graphs of bounded tree-width, it remains of interest to show that their associated Glauber dynamics are rapidly mixing. There have been significant and concerted endeavours by researchers spanning physics, computer science and probability to determine the mixing characteristics of Glauber dynamics on many related Markov chains. Spin systems have been of particular interest; indeed, the main thrust of the work of Jerum and Sinclair [41] was to tackle the partition function for the ‘spins world’ of the ferromagnetic Ising model (using a translation to the rapidly mixing ‘subgraphs world’). For more on the connections among the ‘spins world’, the ‘subgraphs world’ and the ‘random cluster world’, see the recent work of Huber [37]. Much illuminating recent work on the mixing times of Glauber dynamics has been restricted to trees or tree-like graphs, cf. e.g. [5, 19, 21, 31, 49, 59].

1.2 Extensions

The primary focus in our paper is to establish results for graph polynomials defined according to *edge* subset expansion; however, we can also adapt our methodology to polynomials defined according to *vertex* subset expansion, which may be viewed as the ‘induced subgraphs world’. To our knowledge, this form of Markov chain has not been greatly examined. One of our motivations to consider vertex subset expansion is to cover the graph polynomial introduced by Arratia, Bollobás and Sorkin [4]: the *bivariate interlace polynomial*, defined for any graph $G = (V, E)$ and parameters x, y as

$$q(G; x, y) := \sum_{S \subseteq V} (x - 1)^{\text{rk}_2(S)} (y - 1)^{|S| - \text{rk}_2(S)}, \quad (5)$$

where $\text{rk}_2(S)$ is the \mathbb{F}_2 -rank of the adjacency matrix for $G[S]$, the subgraph of G induced by the vertices in S . This polynomial specialises to the independence polynomial and is intimately related to Martin polynomials [1]. Just as for the Tutte polynomial, computation of the bivariate interlace polynomial is $\#P$ -hard in almost the entire plane, as was shown by Bläser and Hoffmann [7]. The multivariate interlace polynomial, a generalisation of the interlace polynomial, can be evaluated efficiently for graphs of bounded

tree-width, cf. Courcelle [18] and Bläser and Hoffmann [8]. Subject to a condition on the vertex subset weightings, which we have called *vertex λ -multiplicativity*, we can establish rapid mixing for vertex subset Glauber dynamics on graphs of constant tree-width.

We also observe that our framework can be extended in a natural way to hypergraphs. Though hypergraph polynomials and their corresponding Markov chains have not been studied nearly as widely as their graph counterparts, we point out some important examples. Goldberg and Jerrum [29] used a multivariate hypergraph Tutte polynomial to prove approximate counting hardness of the ferromagnetic (bivariate graph) Tutte polynomial. There is a significant body of research into Glauber dynamics for counting colourings and independent sets in hypergraphs [13, 14, 16, 22]. Hypergraph variants of the multivariate chromatic polynomial and the edge elimination polynomial have been introduced by White [65]. The study of (counting) complexity for hypergraph parameters continues to grow in importance in relation to constraint satisfaction problems (CSPs). Much of this study involves a restriction of the input hypergraph's width. Unfortunately, there is no unique standard hypergraph analogue of tree-width: consult Hliněný et al. [35] for an overview of several competing notions of width for hypergraphs and the relationship with CSPs. We have made a choice of width parameter that gives the most straightforward extension from our graph framework.

Another possible extension is to matroids, a setting with close connections to the Tutte polynomial, cf. Welsh [62]. As an indication of this possibility, we show that (edge) subset Glauber dynamics associated with the Tutte polynomial for matroids mixes rapidly on matroids of bounded branch-width. The difficulty in dealing with matroids is that the definition of λ -multiplicativity we have used for graphs and hypergraphs involves vertex cuts, which are indefinable in a matroid. So although we have not formed a general condition akin to λ -multiplicativity for matroids, we show that the matroid rank function has suitable properties in matroids of bounded branch-width. We note that Hliněný [34] has shown the existence of a polynomial-time algorithm for the computation of the Tutte polynomial for matroids of bounded branch-width representable over a finite field.

2 Definitions

2.1 λ -multiplicative weight functions

In this subsection, we describe the condition we require on our graph functions \mathcal{P} . This condition prescribes that the weight function is multiplicative with respect to the operation of disjoint graph union as well as “nearly multiplicative” with respect to the operation of composition via small vertex cuts.

We use the notation $\hat{\lambda} := \max\{\lambda, 1/\lambda\}$. For a graph $G = (V, E)$, a vertex cut K is said to separate sets V_1 and V_2 if (V_1, K, V_2) is a partition of V and there is no edge of E that is incident to both a vertex of V_1 and a vertex of V_2 . A partition (E_1, E_2) of E is appropriate (for K) if E_1 has no edge adjacent to a vertex in V_2 and E_2 has no edge adjacent to a vertex in V_1 .

For fixed $\lambda > 0$, we say that the weight function w is *λ -multiplicative*, if for any

$G = (V, E)$, any vertex cut K that separates sets V_1 and V_2 , and any appropriate partition (E_1, E_2) , we have

$$\hat{\lambda}^{-|K|} \leq \frac{w((V_1 \cup K, E_1))w((V_2 \cup K, E_2))}{w(G)} \leq \hat{\lambda}^{|K|}. \quad (6)$$

As mentioned above, if w is λ -multiplicative, then it follows that w is multiplicative with respect to disjoint union (by taking $K = \emptyset$); furthermore, taking $V_2 = \emptyset$ implies that the addition or deletion of a few edges in the graph does not change w wildly.

2.2 Examples of valid polynomials

In this subsection, we emphasise specific examples of edge subset weightings and justify that their weight functions are λ -multiplicative.

Let $G = (V, E)$ be any graph, K be any vertex cut that separates vertex subsets V_1 and V_2 , and (E_1, E_2) be any appropriate partition. We define G' to be the disjoint union of graphs $(V_1 \cup K, E_1)$ and $(V_2 \cup K, E_2)$. We could imagine forming G' from G by splitting each vertex in K , taking incident edges in E_1 with one copy of the vertex and those in E_2 with the other. It is trivial to verify multiplicativity with respect to disjoint union for each of the weight functions considered below. Therefore, to establish λ -multiplicativity for these weight functions, it will suffice to verify that $\hat{\lambda}^{-|K|} \leq w(G')/w(G) \leq \hat{\lambda}^{|K|}$.

First, we observe that the partition function of the *random cluster model* for $q, \mu > 0$ satisfies the condition. Recalling (2), the relevant weight function is $w((V, S)) := q^{\kappa(S)} \mu^{|S|}$. To handle the $\mu^{|S|}$ factor, note that the graphs G and G' have the same number of edges. For the $q^{\kappa(S)}$ factor, the number of components in G' can be at most $\kappa(G) + |K|$ since G' can be obtained by splitting $|K|$ vertices of G . Thus, w is λ -multiplicative if we take $\lambda := q$.

This can also be seen in the context of the *Tutte polynomial* when $x, y > 1$. Recalling (4), the relevant weight function is $w((V, S)) := (x - 1)^{r(E) - r(S)} (y - 1)^{|S| - r(S)}$. As before, it is easy to take care of the $(x - 1)^{r(E)} (y - 1)^{|S|}$ factor. For the remaining $((x - 1)(y - 1))^{-r(S)}$ factor, it is enough to observe that the incidence matrix of G may be obtained from the incidence matrix of G' as follows. The matrix for G' has two rows for each of the vertices in K , one from $(V_1 \cup K, E_1)$ and one from $(V_2 \cup K, E_2)$. If we replace one of these two rows with the sum of the two rows, we do not alter the rank; if we then delete the other of the two rows, we change the rank by at most 1. Repeating this for each vertex in K , we obtain the incidence matrix for G , at a total change in the rank r of the incidence matrix of at most $|K|$. Thus, w is λ -multiplicative if we take $\lambda := (x - 1)(y - 1)$.

Next, we see that the *adjacency-rank polynomial* of Ge and Štefankovič [27] satisfies the condition if $q, \mu > 0$. Recalling (3), the relevant weight function is $w((V, S)) := q^{\text{rk}_2(S)} \mu^{|S|}$. As before, it is simple to handle the $\mu^{|S|}$ factor. For the $q^{\text{rk}_2(S)}$ factor, we note that the adjacency matrix of G may be formed from the adjacency matrix of G' by $|K|$ row additions, followed by $|K|$ column additions and finally the deletion of $|K|$ rows and $|K|$ columns. Since we must delete both rows and columns, the rank rk_2 of the adjacency

matrix may change by up to $2|K|$. Thus, in this case, w is λ -multiplicative when taking $\lambda := q^2$.

Now, consider the *multivariate Tutte polynomial* as formulated by Sokal [58], defined for any graph $G = (V, E)$ and parameters $q, \vec{v} = \{v_e\}_{e \in E}$ by

$$Z_{Tutte}(G; q, \vec{v}) := \sum_{S \subseteq E} q^{\kappa(S)} \prod_{e \in S} v_e. \quad (7)$$

Under this expansion, $w := q^{\kappa(S)} \prod_{e \in S} v_e$ is an edge subset weight function if $q > 0$ and $v_e > 0$ for any $e \in E$ are fixed. We can handle the $q^{\kappa(S)}$ factor as we did for the random cluster model partition function. For the $\prod_{e \in S} v_e$ factor, observe that G and G' have the same set of edges. Thus, w is λ -multiplicative when taking $\lambda := q$.

Last, we discuss the *U-polynomial* of Noble and Welsh [52], defined for any graph $G = (V, E)$ and parameters $y, \vec{x} = \{x_i\}_{i=1}^{|V|}$ by

$$U(G; \vec{x}, y) := \sum_{S \subseteq E} (y - 1)^{|S| - r(S)} \prod_{i=1}^{|V|} x_i^{\kappa(i, S)}, \quad (8)$$

where $\kappa(i, S)$ denotes the number of components of order i in (V, S) . If $y > 1$ and $x_i > 0$ for all i , then $w((V, S)) := (y - 1)^{|S| - r(S)} \prod_{i=1}^{|V|} x_i^{\kappa(i, S)}$ gives an edge subset weighting. The $(y - 1)^{|S| - r(S)}$ factor can be handled as above. For the $\prod_{i=1}^{|V|} x_i^{\kappa(i, S)}$ factor, observe that $\sum_i |\kappa(i, G) - \kappa(i, G')|$ is at most $3|K|$, since, if we obtain G' by splitting the vertices in K , each time we split a vertex we either change the size of a single component or split a single component into two smaller components. Thus, taking $x' := \max_i \max\{x_i, x_i^{-1}\}$ and $y' := \max\{y - 1, (y - 1)^{-1}\}$, we see that w is λ -multiplicative when taking $\lambda := y'x'^3$.

2.3 Glauber dynamics for edge subsets

In this subsection, we define the Markov chain associated with the edge subset expansion formula for \mathcal{P} . From the formulation in (1), the *single bond flip chain* \mathcal{M} on a given graph $G = (V, E)$ is defined as follows. We start with an arbitrary subset $X_0 \subseteq E$ and repeatedly generate X_{t+1} from X_t by running the following experiment.

1. Pick an edge $e \in E$ uniformly at random and let $S = X_t \oplus \{e\}$.
2. Set $X_{t+1} = S$ with probability $\frac{1}{2} \min\{1, w((V, S))/w((V, X_t))\}$ and with the remaining probability set $X_{t+1} = X_t$.

By convention, we denote the state space of \mathcal{M} by Ω (i.e. $\Omega = 2^E$) and its transition probability matrix by P . With standard arguments, it can be shown that \mathcal{M} is a reversible Markov chain that has a unique stationary distribution π satisfying $\pi(S) \propto w((V, S))$. Hence, we may use \mathcal{M} as a Markov chain in MCMC sampling for the following problem.

PWE(\mathcal{P}): \mathcal{P} -WEIGHTED EDGE SUBSETS

Input: a graph $G = (V, E)$.

Output: a subset $S \subseteq E$ with probability $w((V, S))/\mathcal{P}(G)$.

The term *rapidly mixing* applies to a Markov chain that quickly converges to its stationary distribution. We make this precise here. The *total variation distance* $\|\nu - \nu'\|_{TV}$ between two probability distributions ν and ν' is defined by $\|\nu - \nu'\|_{TV} = \frac{1}{2} \sum_{H \in \Omega} |\nu(H) - \nu'(H)|$. For $\varepsilon > 0$, the *mixing time* of a Markov chain \mathcal{M} (with state space Ω , transition matrix P and stationary distribution π) is defined as

$$\tau(\varepsilon) := \max_{H \in \Omega} \{ \min\{t \mid \|P^t(H, \cdot) - \pi(\cdot)\|_{TV} \leq \varepsilon\} \}.$$

In this paper, we shall say that a chain \mathcal{M} *mixes rapidly* if, for any fixed ε , $\tau(\varepsilon)$ is (upper) bounded by a polynomial in the number of vertices of the input graph. This definition for rapid mixing is the one more commonly used in theoretical computer science, whereas often in statistical physics or discrete probability a stricter $O(n \log n)$ bound is mandated.

3 Results

We are now prepared to state the main theorem.

Theorem 1. *Let $G = (V, E)$ where $|V| = n$. If w is λ -multiplicative for some $\lambda > 0$, then the mixing time of \mathcal{M} on G satisfies*

$$\tau(\varepsilon) = O \left(n^{4+4(\text{tw}(G)+1)|\log \lambda|} \left(\max_{S \subseteq E} \left\{ \log \frac{\mathcal{P}(G)}{w((V, S))} \right\} + \log \frac{1}{\varepsilon} \right) \right)$$

(where $\text{tw}(G)$ denotes the tree-width of G).

In later sections, we describe extensions of this theorem to induced subgraphs (Theorem 6), to hypergraphs (Theorem 11), and to matroids (Theorem 13).

In Subsection 2.2, we noted some examples of polynomials that have λ -multiplicative weight functions. Thus, Theorem 1 implies that each of their associated Glauber dynamics on edge subsets is rapidly mixing upon graphs of bounded tree-width.

Corollary 2. *Let $G = (V, E)$ where $|V| = n$. In the following list, we state conditions on the parameters which guarantee rapid mixing of the single bond flip chain on G associated with the stated polynomial and weighting. We also state the mixing time bound.*

1. *For fixed $q, \mu > 0$ and the weighting w of $Z_{RC}(G; q, \mu)$ given by (2), the mixing time satisfies*

$$\tau(\varepsilon) = O \left(n^{4+4(\text{tw}(G)+1)|\log q|} \left(n^2 + \log \frac{1}{\varepsilon} \right) \right).$$

Equivalently, for fixed $x, y > 1$ and the weighting w of $T(G; x, y)$ given by (4), the mixing time satisfies

$$\tau(\varepsilon) = O \left(n^{4+4(\text{tw}(G)+1)|\log((x-1)(y-1))|} \left(n^2 + \log \frac{1}{\varepsilon} \right) \right).$$

2. For fixed $q, \mu > 0$ and the weighting w of $R_2(G; q, \mu)$ given by (3), the mixing time satisfies

$$\tau(\varepsilon) = O\left(n^{4+8(\text{tw}(G)+1)|\log q|} \left(n^2 + \log \frac{1}{\varepsilon}\right)\right).$$

3. For fixed $q > 0$ and $v_e > 0$ for all e and the weighting w of $Z(G; q, \vec{v})$ given by (7), the mixing time satisfies

$$\tau(\varepsilon) = O\left(n^{4+4(\text{tw}(G)+1)|\log q|} \left(n^2 + \log \frac{1}{\varepsilon}\right)\right).$$

4. For fixed $y > 1$ and $x_i > 0$ for all i and the weighting w of $U(G; \vec{x}, \mu)$ given by (8), the mixing time satisfies

$$\tau(\varepsilon) = O\left(n^{4+4(\text{tw}(G)+1)|\log(y'x'^3)|} \left(n^2 + \log \frac{1}{\varepsilon}\right)\right),$$

where $x' = \max_i \max\{x_i, x_i^{-1}\}$ and $y' = \max\{y - 1, (y - 1)^{-1}\}$.

Here, we point out that Ge and Štefankovič obtained part 1 above and showed part 2 above in the special case of trees. Parts 2–4 directly extend these findings, and our main theorem considerably broadens the scope of mixing time bounds for subset Glauber dynamics on graphs of bounded tree-width.

4 Proofs

Let us first give an outline of our arguments.

Although our main result is stated in terms of tree-width, we do not treat tree-width directly but instead use linear-width, a more restrictive width parameter introduced by Thomas [60]. This strategy was also employed by Ge and Štefankovič in the two specific cases mentioned above. For any graph $G = (V, E)$, an ordering (e_1, \dots, e_m) of E has linear-width at most ℓ , if, for each $i \in \{2, \dots, m\}$, there are at most ℓ vertices that are incident to both an edge in $\{e_1, \dots, e_{i-1}\}$ and an edge in $\{e_i, \dots, e_m\}$. The *linear-width* $\text{lw}(G)$ of $G = (V, E)$ is the smallest integer ℓ such that there is an ordering of E with linear-width at most ℓ . The motivation for using linear-width is that it implies an ordering of the edges which we can then use to define canonical paths between pairs of edge subsets. Then we show that λ -multiplicativity is the general condition under which we can bound the congestion of these canonical paths. The use of canonical paths is a standard technique for obtaining a bound on the mixing time of MCMC methods — see the lecture notes of Jerrum [40] for an expository account of this approach.

The key property we require that relates the linear-width of G to the more well-studied parameters path-width $\text{pw}(G)$ and tree-width $\text{tw}(G)$ of G is the following set of inequalities, details of which can be found in Bodlaender [9], Chung and Seymour [17],

Fomin and Thilikos [25], Ge and Štefankovič [26], and Korach and Solel [46]. For any graph G on n vertices,

$$\text{pw}(G) \leq \text{lw}(G) \leq \text{pw}(G) + 1 \leq (\text{tw}(G) + 1)(\log_2 n + 1) + 1. \quad (9)$$

We follow a canonical paths strategy to bound the mixing time of \mathcal{M} . Given $G = (V, E)$, let $\sigma = (e_1, \dots, e_m)$ be an ordering of E . Given $I, F \in \Omega$, let $I \oplus F$ denote the symmetric difference of I and F , let $\sigma[I \oplus F] := (e_{i_1}, \dots, e_{i_k})$ denote the restriction of σ to $I \oplus F$ (that is, $\{e_{i_1}, \dots, e_{i_k}\} = I \oplus F$ and $i_1 < \dots < i_k$), and let $\gamma_{\sigma, I \rightarrow F}$ denote the *canonical path* from I to F , defined as

$$\gamma_{\sigma, I \rightarrow F} := (H_0, \dots, H_k),$$

where $H_0 = I$, $H_j = H_{j-1} \oplus \{e_{i_j}\}$ for all $j \in \{1, \dots, k\}$ (and hence $H_k = F$). Let $\Gamma_\sigma = \{\gamma_{\sigma, I \rightarrow F} \mid I, F \in \Omega\}$.

To bound the mixing time of \mathcal{M} , we will, for some appropriately chosen σ , bound the *congestion* $\varrho(\Gamma_\sigma)$ of the canonical paths, which is defined by

$$\varrho(\Gamma_\sigma) := \max_{\substack{(H, H'):\\ P(H, H') > 0}} \left\{ \frac{1}{\pi(H)P(H, H')} \sum_{\substack{I, F:\\ (H, H') \in \gamma_{\sigma, I \rightarrow F}}} \pi(I)\pi(F)|\gamma_{\sigma, I \rightarrow F}| \right\}, \quad (10)$$

where $|\gamma_{\sigma, I \rightarrow F}|$ denotes the length of the path $\gamma_{\sigma, I \rightarrow F}$. The mixing time can then be bounded using the following inequality of Sinclair [57], see also Diaconis and Stroock [20]: for any set Γ of canonical paths,

$$\tau(\varepsilon) \leq \max_{H \in \Omega} \left\{ \varrho(\Gamma) \cdot \left(\log \frac{1}{\pi(H)} + \log \frac{1}{\varepsilon} \right) \right\}. \quad (11)$$

The remainder of the section is devoted to showing the following.

Theorem 3. *Suppose $G = (V, E)$ has linear-width ℓ and let $\sigma = (e_1, \dots, e_m)$ be an ordering of E with linear-width at most ℓ . If w is λ -multiplicative for some $\lambda > 0$, then $\varrho(\Gamma_\sigma) \leq 2m^2\hat{\lambda}^{4\ell}$.*

Theorem 3 immediately implies a good mixing time bound for the Markov chain \mathcal{M} and hence Theorem 1 follows.

Corollary 4. *Let $G = (V, E)$ where $|E| = m$. If w is λ -multiplicative for some $\lambda > 0$, then the mixing time of \mathcal{M} on G satisfies*

$$\tau(\varepsilon) = O \left(m^2 \hat{\lambda}^{4\text{lw}(G)} \left(\max_{S \subseteq E} \left\{ \log \frac{\mathcal{P}(G)}{w((V, S))} \right\} + \log \frac{1}{\varepsilon} \right) \right).$$

Proof. Substitute the congestion bound of Theorem 3 into inequality (11). □

Proof of Theorem 1. Substitute the upper bound on $\text{lw}(G)$ in (9) into Corollary 4. \square

In the proof of Theorem 3, we will need the following lemma.

Lemma 5. *Suppose $G = (V, E)$ has linear-width ℓ and let $\sigma = (e_1, \dots, e_m)$ be an ordering of E with linear-width at most ℓ . Suppose $I, F \in \Omega$ and H is on $\gamma_{\sigma, I \rightarrow F}$. If w is λ -multiplicative for some $\lambda > 0$, then*

$$\frac{w((V, I))w((V, F))}{w((V, H))w((V, C))} \leq \hat{\lambda}^{4\ell},$$

where $C = I \oplus F \oplus H$.

Proof. Since H is on $\gamma_{\sigma, I \rightarrow F}$, we may assume that $H = H_j$ for some $j \in \{0, \dots, k\}$. Let $Q = \{e_1, e_2, \dots, e_{i_j-1}, e_{i_j}\}$ and $\overline{Q} = E \setminus Q$. Then

$$H = (F \cap Q) \cup (I \cap \overline{Q}) \quad \text{and} \quad C = (I \cap Q) \cup (F \cap \overline{Q}). \quad (12)$$

We can partition V into three sets as follows. Let V_1 denote the set of vertices that are incident only to edges in Q ; let V_2 denote the set of vertices that are incident only to edges in \overline{Q} ; let K denote the set of remaining vertices, that is, the set of vertices incident to edges in Q and \overline{Q} . Note that $|K|$ is at most the linear-width ℓ .

No vertex v_1 of V_1 is adjacent to a vertex v_2 of V_2 , as otherwise the edge between them would simultaneously be in Q and \overline{Q} . This implies that K is a vertex cut separating V_1 and V_2 with respect to G , and also with respect to the graphs (V, I) , (V, F) , (V, H) , (V, C) . Furthermore, $(I \cap Q, I \cap \overline{Q})$, $(F \cap Q, F \cap \overline{Q})$, $(H \cap Q, H \cap \overline{Q})$, $(C \cap Q, C \cap \overline{Q})$ are edge partitions that are appropriate for K . Therefore, by the fact that w is λ -multiplicative and $|K| \leq \ell$,

$$\hat{\lambda}^{-\ell} \leq \frac{w((V_1 \cup K, J \cap Q))w((V_2 \cup K, J \cap \overline{Q}))}{w((V, J))} \leq \hat{\lambda}^{\ell} \quad \text{for } J \in \{I, F, H, C\}.$$

By (12), it follows that

$$\begin{aligned} (V_1 \cup K, H \cap Q) &= (V_1 \cup K, F \cap Q), \\ (V_2 \cup K, H \cap \overline{Q}) &= (V_2 \cup K, I \cap \overline{Q}), \\ (V_1 \cup K, C \cap Q) &= (V_1 \cup K, I \cap Q) \quad \text{and} \\ (V_2 \cup K, C \cap \overline{Q}) &= (V_2 \cup K, F \cap \overline{Q}). \end{aligned}$$

Now, letting r be

$$w((V_1 \cup K, I \cap Q))w((V_2 \cup K, I \cap \overline{Q}))w((V_1 \cup K, F \cap Q))w((V_2 \cup K, F \cap \overline{Q})),$$

we obtain

$$\frac{w((V, I))w((V, F))}{r} \leq \hat{\lambda}^{2\ell} \quad \text{and} \quad \frac{r}{w((V, H))w((V, C))} \leq \hat{\lambda}^{2\ell},$$

whereby the lemma easily follows. \square

Proof of Theorem 3. Let $(H, H') \in \Omega \times \Omega$ such that $P(H, H') > 0$. We will bound the expression within the max of the definition for $\varrho(\Gamma_\sigma)$. We let $\hat{H} = H$ if $\pi(H) \leq \pi(H')$ and $\hat{H} = H'$ otherwise. Denote by $\text{cp}(H, H')$ the set of pairs (I, F) such that $(H, H') \in \gamma_{\sigma, I \rightarrow F}$. We define the function $\text{inj} : \text{cp}(H, H') \rightarrow \Omega$ by $(I, F) \rightarrow I \oplus F \oplus \hat{H}$. Observe that inj is an injection, for, given $J \in \Omega$, we can determine the unique (I, F) such that $\text{inj}(I, F) = J$ by first computing $J \oplus \hat{H} = I \oplus F$ and then using the ordering σ to recover I and F . Since w is λ -multiplicative, we have by Lemma 5 that

$$\frac{w((V, I))w((V, F))}{w((V, \hat{H}))w((V, \text{inj}(I, F)))} \leq \hat{\lambda}^{4\ell}. \quad (13)$$

Regardless of whether $\pi(H) \leq \pi(H')$ or $\pi(H) > \pi(H')$, a brief calculation yields that $\pi(H)P(H, H') = \pi(\hat{H})/(2m)$. Thus,

$$\begin{aligned} \frac{1}{\pi(H)P(H, H')} & \sum_{(I, F) \in \text{cp}(H, H')} \pi(I)\pi(F)|\gamma_{\sigma, I \rightarrow F}| \\ &= \frac{2m}{\pi(\hat{H})} \sum_{(I, F) \in \text{cp}(H, H')} \pi(I)\pi(F)|\gamma_{\sigma, I \rightarrow F}| \\ &\leq \frac{2m^2}{\mathcal{P}(G)} \sum_{(I, F) \in \text{cp}(H, H')} \frac{w((V, I))w((V, F))}{w((V, \hat{H}))} \end{aligned} \quad (14)$$

$$\leq \frac{2m^2}{\mathcal{P}(G)} \sum_{(I, F) \in \text{cp}(H, H')} w((V, \text{inj}(I, F)))\hat{\lambda}^{4\ell} \quad (15)$$

$$\leq 2m^2\hat{\lambda}^{4\ell}, \quad (16)$$

where (14) follows from the facts $|\gamma_{\sigma, I \rightarrow F}| \leq m$ and $\pi(S) = w((V, S))/\mathcal{P}(G)$, (15) follows from (13), and (16) follows from the fact that inj is an injection. Then, substituting the bound (16) into (10), we obtain $\varrho(\Gamma_\sigma) \leq 2m^2\hat{\lambda}^{4\ell}$, as claimed. \square

5 Glauber dynamics for vertex subsets

Until now, we have considered edge subsets (subgraphs) and Glauber transitions which change one edge at a time. In this section, we modify our methods to treat vertex subsets (induced subgraphs) and transitions that involve one vertex at a time — each such transition can affect many edges, up to the maximum degree of G . We sketch how to obtain rapid mixing for this process upon graphs of bounded tree-width still with only a modest condition on the base graph polynomials.

A *vertex subset expansion formula* for \mathcal{P} is written as follows: for any simple graph $G = (V, E)$,

$$\mathcal{P}(G) = \sum_{S \subseteq V} w(G[S]) \quad (17)$$

for some graph function w , where $G[S]$ denotes the subgraph of G induced by S . If the function w is non-negative, we refer to (17) as a *vertex subset weighting* for \mathcal{P} and to w as its *weight function*. Again, for our results to hold, aside from some other constraints, we need the weight function to be positive on all induced subgraphs.

From the formulation in (17), we define the *single site flip chain* \mathcal{M}' on a given graph $G = (V, E)$ as follows. We start with an arbitrary subset $X_0 \subseteq V$ and repeatedly generate X_{t+1} from X_t by running the following experiment.

1. Pick a vertex $v \in V$ uniformly at random and let $S = X_t \oplus \{v\}$.
2. Set $X_{t+1} = S$ with probability $\frac{1}{2} \min \{1, w(G[S])/w(G[X_t])\}$ and with the remaining probability set $X_{t+1} = X_t$.

We denote the state space of \mathcal{M}' by Ω' (i.e. $\Omega' = 2^V$) and its transition probability matrix by P' . It can be shown that \mathcal{M}' is a reversible Markov chain that has a unique stationary distribution π' satisfying $\pi'(S) \propto w(G[S])$. Hence, we may use \mathcal{M}' as a Markov chain in MCMC sampling for the following problem.

PWV(\mathcal{P}): \mathcal{P} -WEIGHTED VERTEX SUBSETS

Input: a graph $G = (V, E)$.

Output: a subset $S \subseteq V$ with probability $w(G[S])/\mathcal{P}(G)$.

We now describe the condition required of the weight function w in (17). For fixed $\lambda > 0$, we say that the weight function w is *vertex λ -multiplicative*, if for any $G = (V, E)$ and K a vertex cut that separates sets V_1 and V_2 with respect to G , we have

$$\hat{\lambda}^{-|K|} \leq \frac{w(G[V_1])w(G[V_2 \cup K])}{w(G)} \leq \hat{\lambda}^{|K|}. \quad (18)$$

Note that, if w is vertex λ -multiplicative, then it follows that w is multiplicative with respect to disjoint union by taking $K = \emptyset$; furthermore, taking $V_2 = \emptyset$ gives that the addition of a few vertices does not change w wildly.

The main result of this section is the following.

Theorem 6. *Let $G = (V, E)$ where $|V| = n$. If w is vertex λ -multiplicative for some $\lambda > 0$, then the mixing time of \mathcal{M}' on G satisfies*

$$\tau(\varepsilon) = O \left(n^{2+4(\text{tw}(G)+1)|\log \lambda|} \left(\max_{S \subseteq V} \left\{ \log \frac{\mathcal{P}(G)}{w(G[S])} \right\} + \log \frac{1}{\varepsilon} \right) \right).$$

5.1 A sketch of the proof

As before, we do not treat tree-width directly, but instead work with a different width parameter. For any graph $G = (V, E)$, an ordering (v_1, \dots, v_n) of V has vertex-separation at most ℓ , if, for each $i \in \{2, \dots, n\}$, there are at most ℓ vertices in $\{v_1, \dots, v_{i-1}\}$ that are adjacent to a vertex in $\{v_i, \dots, v_n\}$. The *vertex-separation* $\text{vs}(G)$ of $G = (V, E)$ is the smallest integer ℓ such that there is an ordering of V with vertex-separation at most ℓ .

It was shown by Kinnersley [45] that the vertex-separation of G satisfies $\text{vs}(G) = \text{pw}(G)$, and so the inequalities in (9) remain relevant.

To bound the mixing time of \mathcal{M}' , we again follow a canonical paths argument. Given $G = (V, E)$, let $\sigma = (v_1, \dots, v_n)$ be an ordering of V . Given $I, F \in \Omega'$, let $I \oplus F$ denote the symmetric difference of I and F , let $\sigma[I \oplus F] := (v_{i_1}, \dots, v_{i_k})$ denote the restriction of σ to $I \oplus F$ (that is, $\{v_{i_1}, \dots, v_{i_k}\} = I \oplus F$ and $i_1 < \dots < i_k$), and let $\gamma_{\sigma, I \rightarrow F}$ denote the *canonical path* from I to F , defined as

$$\gamma_{\sigma, I \rightarrow F} := (H_0, \dots, H_k),$$

where $H_0 = I$, $H_j = H_{j-1} \oplus \{v_{i_j}\}$ for all $j \in \{1, \dots, k\}$ (and hence $H_k = F$). Let $\Gamma_\sigma = \{\gamma_{\sigma, I \rightarrow F} \mid I, F \in \Omega'\}$. Using inequality (11), our bound on the mixing time again follows from a bound on the *congestion* $\varrho(\Gamma_\sigma)$, which is defined analogously to (10).

Theorem 7. *Suppose $G = (V, E)$ has vertex-separation ℓ . Let $\sigma = (v_1, \dots, v_n)$ be an ordering of V with vertex-separation at most ℓ . If, for some $\lambda > 0$, w is vertex λ -multiplicative, then $\varrho(\Gamma_\sigma) \leq 2n^2 \hat{\lambda}^{4\ell}$.*

Theorem 7 immediately implies a good mixing time bound for the Markov chain \mathcal{M}' and hence Theorem 6 also.

Corollary 8. *Let $G = (V, E)$ where $|V| = n$. If w is vertex λ -multiplicative for some $\lambda > 0$, then the mixing time of \mathcal{M}' on G satisfies*

$$\tau(\varepsilon) = O\left(n^2 \hat{\lambda}^{4\text{vs}(G)} \left(\max_{S \subseteq V} \left\{ \log \frac{\mathcal{P}(G)}{w(G[S])} \right\} + \log \frac{1}{\varepsilon} \right)\right).$$

Proof. Substitute the congestion bound of (11) into Theorem 7. □

Proof of Theorem 6. Substitute the upper bound on $\text{vs}(G) = \text{pw}(G)$ of (9) into Corollary 8. □

We omit the proof of Theorem 7 as it is similar to that of Theorem 3, but give the details for the analogue of Lemma 5.

Lemma 9. *Suppose $G = (V, E)$ has vertex-separation ℓ and let $\sigma = (v_1, \dots, v_n)$ be an ordering of V with vertex-separation at most ℓ . Suppose $I, F \in \Omega'$ and H is on $\gamma_{\sigma, I \rightarrow F}$. If w is vertex λ -multiplicative for some $\lambda > 0$, then*

$$\frac{w(G[I])w(G[F])}{w(G[H])w(G[C])} \leq \hat{\lambda}^{4\ell},$$

where $C = I \oplus F \oplus H$.

Proof. Since H is on $\gamma_{\sigma, I \rightarrow F}$, we may assume that $H = H_j$ for some $j \in \{0, \dots, k\}$. Let $Q = \{v_1, \dots, v_{i_j}\}$, and $\bar{Q} = V \setminus Q$. Then

$$H = (F \cap Q) \cup (I \cap \bar{Q}) \quad \text{and} \quad C = (I \cap Q) \cup (F \cap \bar{Q}). \quad (19)$$

We can partition V into three sets as follows. Let V_1 denote the set of vertices \overline{Q} ; let V_2 denote the subset of Q containing vertices adjacent only to other vertices of Q ; and let K denote the set of remaining vertices, that is, the set of vertices of Q incident to vertices of V_1 . Note that $|K|$ is at most the vertex-separation ℓ .

Clearly, K is a vertex cut separating V_1 and V_2 with respect to G and also with respect to the graphs $G[I]$, $G[F]$, $G[H]$, $G[C]$. Therefore, by the fact that w is vertex λ -multiplicative, and noting that $V_2 \cup K = Q$,

$$\hat{\lambda}^{-\ell} \leq \frac{w(G[Q \cap J])w(G[\overline{Q} \cap J])}{w(G[J])} \leq \hat{\lambda}^{\ell} \quad \text{for } J \in \{I, F, H, C\}.$$

By (19), it follows that $H \cap Q = F \cap Q$, $H \cap \overline{Q} = I \cap \overline{Q}$, $C \cap Q = I \cap Q$ and $C \cap \overline{Q} = F \cap \overline{Q}$. Now, letting $r = w(G[Q \cap I])w(G[\overline{Q} \cap I])w(G[Q \cap F])w(G[\overline{Q} \cap F])$, we obtain that

$$\frac{w(G[I])w(G[F])}{r} \leq \hat{\lambda}^{2\ell} \quad \text{and} \quad \frac{r}{w(G[H])w(G[C])} \leq \hat{\lambda}^{2\ell},$$

whereby the lemma easily follows. \square

5.2 An example of a vertex subset chain

Recalling the bivariate interlace polynomial $q(G; x, y)$ in (5), which, for fixed $x, y > 1$, is given by the vertex subset weighting

$$w(G[S]) := (x - 1)^{\text{rk}_2(S)}(y - 1)^{|S| - \text{rk}_2(S)}.$$

With arguments very similar to those given in Subsection 2.2, it is not difficult to verify that this weight function is vertex λ -multiplicative, as follows. For any graph $G = (V, E)$ and any vertex cut K that separates sets V_1 and V_2 , consider the graph G' consisting of the disjoint union of $G[V_1]$ and $G[V_2 \cup K]$. We note that the adjacency matrix of G may be formed from the adjacency matrix of G' by altering at most $|K|$ rows and $|K|$ columns, changing the rank rk_2 of the adjacency matrix by up to $2|K|$. Thus, w is vertex λ -multiplicative by taking $\lambda := (x - 1)^2/(y - 1)^2$. So, by Theorem 6, it follows that a natural Markov chain derived from the bivariate interlace polynomial — a chain which has not been studied extensively, as far as we are aware — mixes rapidly on tree-width-bounded graphs.

Corollary 10. *Let $G = (V, E)$ where $|V| = n$. If $x, y > 1$ are fixed, then for the single site flip chain on G associated with the weighting w of $q(G; x, y)$ given by (5), the mixing time satisfies*

$$\tau(\varepsilon) = O\left(n^{2+8(\text{tw}(G)+1)|\log((x-1)/(y-1))|} \left(n + \log \frac{1}{\varepsilon}\right)\right).$$

We believe that it would be of wider interest to study further properties of this single site flip chain on general graphs, in particular to compare it with known results on the random cluster, Potts and Ising models.

6 Subset Glauber dynamics for hypergraphs

In this section, we outline an extension of our results to hypergraphs. Recall that a *hypergraph* H is a pair (V, E) where V is the vertex set and E is a hyperedge set that consists of (arbitrary) vertex subsets. We let $\eta(H)$ denote the maximum size of a hyperedge in H . To each hypergraph $H = (V, E)$, we associate a graph $G_p(H)$ called the *primal graph* or *Gaifman graph* defined by $V(G_p(H)) = V$ and $E(G_p(H)) = \{\{u, v\} \subseteq V \mid \{u, v\} \subseteq e \text{ for some } e \in E\}$.

We remark at the outset of this section that the generalisation to hypergraphs we give here could be seen as routine. We do not include any details of the proof. Nonetheless, we include the necessary elements for the statement of the hypergraph version of Theorem 1, even if these are straightforward generalisations of the graph versions. We also provide some justification for the hypergraph analogue of (9). It is important to observe that the running time of the hypergraph Markov chain includes an adjustment to take into account the maximum hyperedge size $\eta(H)$.

For a hypergraph polynomial \mathcal{P} , a *hyperedge subset expansion formula* is written as follows: for any hypergraph $H = (V, E)$,

$$\mathcal{P}(H) = \sum_{S \subseteq E} w((V, S)) \quad (20)$$

for some graph function w , where (V, S) denotes the hypergraph with vertex set V and hyperedge set S . Notice that if every hyperedge has order two, then this formula is the same as in (1). If the function w is non-negative, that is, $w(H) \geq 0$ for all hypergraphs H , we refer to (20) as a *hyperedge subset weighting* for \mathcal{P} and to w as its *weight function*. For our purposes, we require the weight function to be positive.

For any hypergraph $H = (V, E)$, an ordering (e_1, \dots, e_m) of E has linear-width at most ℓ , if, for each $i \in \{2, \dots, m\}$, there are at most ℓ vertices that are contained in both a hyperedge in $\{e_1, \dots, e_{i-1}\}$ and a hyperedge in $\{e_i, \dots, e_m\}$. The *linear-width* $\text{lw}(H)$ of $H = (V, E)$ is the smallest integer ℓ such that there is an ordering of E with linear-width at most ℓ .

The tree-width $\text{tw}(H)$ and path-width $\text{pw}(H)$ of a given a hypergraph H are defined similarly to the tree- and path-width of graphs, cf. Hliněný et al. [35]. Since all the vertices in a clique appear in the same bag of any tree decomposition, it follows that $\text{tw}(H)$ is equal to the tree-width $\text{tw}(G_p(H))$ of the primal graph associated with H and, similarly, $\text{pw}(H) = \text{pw}(G_p(H))$. Noting that, just as for graphs, $\text{lw}(H) \leq \text{pw}(H)$, we obtain

$$\text{lw}(H) \leq \text{pw}(H) \leq (\text{tw}(H) + 1)(\lceil \log_2 n \rceil + 1) + 1. \quad (21)$$

For a hypergraph $H = (V, E)$, a vertex cut K is said to separate sets V_1 and V_2 if (V_1, K, V_2) is a partition of V and there is no hyperedge of E that contains both a vertex of V_1 and a vertex of V_2 . A partition (E_1, E_2) of E is appropriate (for K) if E_1 has no hyperedge containing a vertex in V_2 and E_2 has no hyperedge containing a vertex in V_1 .

For fixed $\lambda > 0$, we say that the weight function w is λ -multiplicative, if for any $H = (V, E)$, any vertex cut K that separates sets V_1 and V_2 , and any appropriate partition

(E_1, E_2) , we have

$$\hat{\lambda}^{-|K|} \leq \frac{w((V_1 \cup K, E_1))w((V_2 \cup K, E_2))}{w(H)} \leq \hat{\lambda}^{|K|}.$$

From the formulation in (20), the *single bond flip chain* \mathcal{M} on a given hypergraph $H = (V, E)$ is defined as follows. We start with an arbitrary subset $X_0 \subseteq E$ and repeatedly generate X_{t+1} from X_t by running the following experiment.

1. Pick a hyperedge $e \in E$ uniformly at random and let $S = X_t \oplus \{e\}$.
2. Set $X_{t+1} = S$ with probability $\frac{1}{2} \min \{1, w((V, S))/w((V, X_t))\}$ and with the remaining probability set $X_{t+1} = X_t$.

We denote the state space of \mathcal{M} by Ω (i.e. $\Omega = 2^E$) and its transition probability matrix by P . It can be shown that \mathcal{M} is a reversible Markov chain that has a unique stationary distribution π satisfying $\pi(S) \propto w((V, S))$.

We have obtained the following hypergraph generalisation of Theorem 1.

Theorem 11. *Let $H = (V, E)$ be a hypergraph with $|V| = n$. If w is λ -multiplicative for some $\lambda > 0$, then the mixing time of \mathcal{M} on H satisfies*

$$\tau(\varepsilon) = O \left(n^{2\eta(H)+4(\text{tw}(H)+1)|\log \lambda|} \left(\max_{S \subseteq E} \left\{ \log \frac{\mathcal{P}(H)}{w(S)} \right\} + \log \frac{1}{\varepsilon} \right) \right).$$

The proof of this theorem closely follows the pattern set in the case of edge subset Glauber dynamics for graphs. We follow a canonical paths argument that uses an edge ordering by linear-width, to obtain a mixing time bound in terms of linear-width, immediately combined with (21) and the fact that $|E| = O(n^{\eta(H)})$ to obtain Theorem 11. For brevity, we omit the full details; however a rereading of Section 4 with the hypergraph definitions in mind will confirm that the same proof is valid.

7 Subset Glauber dynamics for matroids

In this section, we outline how to apply our methodology to the Tutte polynomial for matroids. We do not attempt to give the fullest framework for matroids, though it could be distilled from the proof sketches below. A significant difference in treating matroids compared to graphs or hypergraphs is that they are not defined with vertices. Since there are then no direct analogues of a vertex cut, this affects the choice of width parameter and the potential definition for matroidal λ -multiplicativity. Fortunately, there is an existing, well-studied notion of connectivity in matroids, which leads to a notion of matroidal path-width, which we have found useful here.

Recall that a matroid M is a pair (E, \mathcal{I}) , where E is a ground set of edges and \mathcal{I} is a collection of subsets of E (referred to as the independent sets of M) satisfying the following properties:

- $\emptyset \in \mathcal{I}$;
- for any $A' \subseteq A \subseteq E$, we have $A \in \mathcal{I} \implies A' \in \mathcal{I}$; and
- for any $A \subseteq E$, all maximal members of \mathcal{I} contained in A have the same cardinality.

For any $A \subseteq E$, the *rank* $r(A)$ of A is the cardinality of a maximal member of \mathcal{I} contained in A , and the *connectivity* $\lambda(A)$ of A is equal to $r(A) + r(E \setminus A) - r(E)$. Both r and λ are non-negative, submodular functions. For further reading on matroid theory, consult a standard reference, by e.g. Oxley [53] or Welsh [62].

The Tutte polynomial naturally extends from graphs to matroids as follows: for any matroid $M = (E, \mathcal{I})$ and parameters x, y , it is defined by

$$T(M; x, y) := \sum_{S \subseteq E} (x - 1)^{r(E) - r(S)} (y - 1)^{|S| - r(S)}. \quad (22)$$

For our purposes, we shall assume $x, y > 1$. Note that if M is a graphic matroid and hence M is the cycle matroid $M(G)$ for some graph G , then the expressions in (4) and (22) are equal.

For any matroid $M = (E, \mathcal{I})$, an ordering (e_1, \dots, e_m) of E has path-width at most ℓ , if, for each $i \in \{2, \dots, m\}$, $\lambda(\{e_1, \dots, e_{i-1}\}) \leq \ell$. The *path-width* $\text{pw}(M)$ of $M = (E, \mathcal{I})$ is the smallest integer ℓ such that there is an ordering of E with path-width at most ℓ . For more on this width parameter, consult Kashyap [44].

The path-width of a matroid provides an upper bound on a more closely-studied matroidal width parameter, branch-width. For any matroid $M = (E, \mathcal{I})$, a *branch-decomposition* of M is a cubic tree T with exactly $|E|$ leaves, labelled in one-to-one correspondence with the elements of E ; such a branch-decomposition has width at most ℓ , if, for any $e \in T$, letting X_e denote a subset of E corresponding to the set of labels contained in one of the two components of $T \setminus e$, then $\lambda(X_e) \leq \ell$. The *branch-width* $\text{bw}(M)$ of $M = (E, \mathcal{I})$ is the smallest integer ℓ such that there is a branch-decomposition of M with width at most ℓ . The branch-width of M can be used to upper bound the path-width of M in the following way.

Proposition 12. *For any matroid $M = (E, \mathcal{I})$, $\text{pw}(M) \leq \text{bw}(M) \log_2 |E|$.*

Proof. Let $M = (E, \mathcal{I})$ be a matroid. We would like to exhibit an ordering of E with path-width at most $\text{bw}(M) \log_2 |E|$. Let $T = (V_T, E_T)$ be a branch-decomposition of M with width at most $\text{bw}(M)$. For any vertex $v \in V_T$, let T_v be the subtree rooted at v and let E_v denote the subset of E corresponding to the set of labels contained in T_v . (Note that T_v has $(|T_v| + 1)/2$ leaves and hence the same quantity of labels.) To obtain an edge ordering, we perform a stack-based depth-first search (DFS) of T , such that smaller subtrees are explored first. Let σ be the ordering of E induced by the order in which their corresponding leaves in T are explored by the DFS³. We now show that

³In other words, the DFS algorithm proceeds as follows. Put the root of T onto the stack. While the stack is non-empty, pop an element v off of the stack and, if v is a leaf of T , add the edge of E corresponding to v at the end of σ ; otherwise, first push the child of v with the bigger subtree onto the stack, then push the other child of v onto the stack.

$\sigma := (e_1, \dots, e_m)$ has path-width at most $\text{bw}(M) \log_2 m$. Let $i \in \{1, \dots, m\}$ and suppose that (v_1, \dots, v_k) are the vertices of T (ordered from bottom to top) on the stack just after the leaf corresponding to e_i was explored. Note that the region of T unexplored by the DFS up to this point is precisely $\bigcup_{j=1}^k T_{v_j}$. The exploration order used by DFS (prioritising smaller subtrees) implies for $j \in \{1, \dots, k-1\}$ that $|T_{v_{j+1}}| < |T_{v_j}|/2$ and hence $|E_{v_{j+1}}| < |E_{v_j}|/2$. Since there are m labels in T , we see that $k \leq \log_2 m$. It then follows, using submodularity of λ and the fact that the width of the branch-decomposition T is at most $\text{bw}(M)$, that

$$\lambda(\{e_1, \dots, e_i\}) = \lambda\left(\bigcup_{j=1}^k E_{v_j}\right) \leq \sum_{j=1}^k \lambda(E_{v_j}) \leq k \text{bw}(M) \leq \text{bw}(M) \log_2 m,$$

as required. \square

It is worth noting here that matroidal path-width corresponds more closely to graphic linear-width than graphic path-width. Although tree-width can be defined for matroids, it is not a straightforward extension; furthermore, the parameter has not gained as much traction in matroid theory as has branch-width, cf. Hliněný and Whittle [36]. On the other hand, Hliněný and Whittle showed that a matroid's tree-width is bounded if and only if its branch-width is bounded.

From the formulation in (22), the *single bond flip chain* \mathcal{M}_{Tutte} for the Tutte polynomial on a given matroid $M = (E, \mathcal{I})$ is defined as follows. For any $A \subseteq E$, let $w_{Tutte}(A) := (x-1)^{r(E)-r(A)}(y-1)^{|A|-r(A)}$; in other words, w_{Tutte} is the weight function associated with $T(M; x, y)$. We start with an arbitrary subset $X_0 \subseteq E$ and repeatedly generate X_{t+1} from X_t by running the following experiment.

1. Pick an element $e \in E$ uniformly at random and let $S = X_t \oplus \{e\}$.
2. Set $X_{t+1} = S$ with probability $\frac{1}{2} \min\{1, w_{Tutte}(S)/w_{Tutte}(X_t)\}$ and with the remaining probability set $X_{t+1} = X_t$.

We denote the state space of \mathcal{M}_{Tutte} by Ω_{Tutte} (i.e. $\Omega_{Tutte} = 2^E$) and its transition probability matrix by P_{Tutte} . It can be shown that \mathcal{M}_{Tutte} is a reversible Markov chain that has a unique stationary distribution π_{Tutte} satisfying $\pi_{Tutte}(S) \propto w_{Tutte}(S)$.

For any matroid $M = (E, \mathcal{I})$, note that $\text{pw}(M) \leq \ell$ gives an ordering (e_1, \dots, e_m) of E such that, for each $i \in \{2, \dots, m\}$,

$$0 \leq r(\{e_1, \dots, e_{i-1}\}) + r(\{e_i, \dots, e_m\}) - r(E) \leq \ell,$$

and so, letting $c(x, y) = \max\{(x-1)(y-1), ((x-1)(y-1))^{-1}\}$,

$$c(x, y)^{-\ell} \leq \frac{w_{Tutte}(\{e_1, \dots, e_{i-1}\})w_{Tutte}(\{e_i, \dots, e_m\})}{w_{Tutte}(E)} \leq c(x, y)^\ell.$$

Combining the above inequalities with an adaptation of the canonical paths proof given in Section 4 — using appropriate substitutions, such as $\text{pw}(M)$ in the place of $\text{lw}(G)$ and

Proposition 12 in the place of (9), together with the well-known fact that the connectivity function of a matroid is closed under the minor relation — we obtain the following matroidal analogue of Theorem 1. The full details of the proof are left to the reader.

Theorem 13. *Let $M = (E, \mathcal{I})$ where $|E| = m$. The mixing time of \mathcal{M}_{Tutte} on M satisfies*

$$\tau(\varepsilon) = O \left(m^{2+4 \operatorname{bw}(G) \log c(x,y)} \left(\max_{S \subseteq E} \left\{ \log \frac{T(M; x, y)}{w_{Tutte}(S)} \right\} + \log \frac{1}{\varepsilon} \right) \right).$$

8 Conclusion

In this work, we have studied a general framework of graph polynomials and Markov chains defined via subset expansion formulae for these polynomials. We have demonstrated that the system mixes rapidly for graphs of bounded tree-width. On a graph G with n vertices, we have shown a mixing time of order $n^{O(1)} e^{O(\operatorname{pw}(G))} = n^{O(\operatorname{tw}(G))}$. Our results apply to many of the most prominent and well-known polynomials in the field. The mixing times of our processes have, respectively, exponential and super-exponential dependencies upon path-width and tree-width. It would be interesting to investigate if this could be improved, in particular, to achieve something akin to fixed-parameter tractability in terms of tree-width. We consider it unlikely that this could be accomplished using only the approaches from this paper.

For all of our results, we need that the weight function is strictly positive for all (induced) subgraphs. Many of the classical enumeration polynomials such as the matching, independence, clique and chromatic polynomials are captured by the general polynomials that we mention as examples throughout this work. However, these are ‘hard-core models’ — in which some (induced) subgraphs have a zero weighting — and hence are not included in our approach. Many of these are evaluations that fall at the boundary of the regions that we can handle. For example, the Tutte polynomial evaluated at the point $(2, 1)$ counts the number of forests of the graph. We have shown rapid mixing at all fixed points $(2, 1 + \delta)$, for $\delta > 0$, with a mixing time that depends on δ . It would be interesting to consider whether the chains associated with these boundary points mix rapidly for graphs of bounded tree-width.

After a presentation of these results in Zürich, Thore Husfeldt suggested that our approximation schemes may be of practical use for computing partition functions. The aforementioned dynamic programming algorithms for exact computation all require an explicit tree-width decomposition. Although a decomposition can theoretically be computed in linear time [10], the existing decomposition-generation algorithms have running times hindered by large constants and thus are impractical. On the other hand, an explicit decomposition is not a prerequisite for rapid mixing in our framework, so an efficient randomised approximation may turn out to be more practical than an exact deterministic algorithm.

Lastly, we have demonstrated, with appropriate modifications, several extensions of our results to other settings. We have adapted our framework to include induced subgraphs, hypergraphs and matroids, all using an appropriate choice of width parameter

and “ λ -multiplicativity”. These results are indicative; this aspect of our work has not been fully explored and we hope that further generalisations of Theorem 1 are available.

Acknowledgement

We are very grateful for the helpful referee comments.

References

- [1] Martin Aigner and Hein van der Holst. Interlace polynomials. *Linear Algebra Appl.*, 377:11–30, 2004.
- [2] Noga Alon, Alan M. Frieze, and Dominic J. A. Welsh. Polynomial time randomized approximation schemes for Tutte-Gröthendieck invariants: the dense case. *Random Structures Algorithms*, 6(4):459–478, 1995.
- [3] Artur Andrzejak. An algorithm for the Tutte polynomials of graphs of bounded treewidth. *Discrete Math.*, 190(1-3):39–54, 1998.
- [4] Richard Arratia, Béla Bollobás, and Gregory B. Sorkin. A two-variable interlace polynomial. *Combinatorica*, 24(4):567–584, 2004.
- [5] Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probab. Theory Related Fields*, 131(3):311–340, 2005.
- [6] George D. Birkhoff. A determinant formula for the number of ways of coloring a map. *Ann. of Math. (2)*, 14(1-4):42–46, 1912/13.
- [7] Markus Bläser and Christian Hoffmann. On the complexity of the interlace polynomial. In Susanne Albers and Pascal Weil, editors, *25th International Symposium on Theoretical Aspects of Computer Science (STACS 2008)*, volume 1 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 97–108, Dagstuhl, Germany, 2008. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [8] Markus Bläser and Christian Hoffmann. Fast evaluation of interlace polynomials on graphs of bounded treewidth. *Algorithmica*, 61(1):3–35, 2011.
- [9] Hans L. Bodlaender. Some classes of graphs with bounded treewidth. *Bulletin of the EATCS*, 36:116–125, 1988.
- [10] Hans L. Bodlaender. A linear-time algorithm for finding tree-decompositions of small treewidth. *SIAM J. Comput.*, 25(6):1305–1317, 1996.
- [11] Hans L. Bodlaender. Treewidth: Characterizations, applications, and computations. In Fedor V. Fomin, editor, *WG*, volume 4271 of *Lecture Notes in Computer Science*, pages 1–14. Springer, 2006.
- [12] Magnus Bordewich. Approximating the number of acyclic orientations for a class of sparse graphs. *Combin. Probab. Comput.*, 13(1):1–16, 2004.
- [13] Magnus Bordewich, Martin E. Dyer, and Marek Karpinski. Stopping times, metrics and approximate counting. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone,

- and Ingo Wegener, editors, *ICALP (1)*, volume 4051 of *Lecture Notes in Computer Science*, pages 108–119. Springer, 2006.
- [14] Magnus Bordewich, Martin E. Dyer, and Marek Karpinski. Path coupling using stopping times and counting independent sets and colorings in hypergraphs. *Random Structures Algorithms*, 32(3):375–399, 2008.
 - [15] Magnus Bordewich and Ross J. Kang. Rapid mixing of subset glaufer dynamics on graphs of bounded tree-width. In Luca Aceto, Monika Henzinger, and Jiri Sgall, editors, *ICALP (1)*, volume 6755 of *Lecture Notes in Computer Science*, pages 533–544. Springer, 2011.
 - [16] Russ Bubley. *Randomized algorithms: approximation, generation, and counting*. CPHC/BCS Distinguished Dissertations. Springer-Verlag London Ltd., London, 2001.
 - [17] Fan R. K. Chung and Paul D. Seymour. Graphs with small bandwidth and cutwidth. *Discrete Math.*, 75(1-3):113–119, 1989. Graph theory and combinatorics (Cambridge, 1988).
 - [18] Bruno Courcelle. A multivariate interlace polynomial and its computation for graphs of bounded clique-width. *Electron. J. Combin.*, 15(1):#69, 36 pp., 2008.
 - [19] Amir Dembo and Andrea Montanari. Ising models on locally tree-like graphs. *Ann. Appl. Probab.*, 20(2):565–592, 2010.
 - [20] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.
 - [21] Jian Ding, Eyal Lubetzky, and Yuval Peres. Mixing time of critical Ising model on trees is polynomial in the height. *Comm. Math. Phys.*, 295(1):161–207, 2010.
 - [22] Martin Dyer and Catherine Greenhill. On Markov chains for independent sets. *J. Algorithms*, 35(1):17–49, 2000.
 - [23] Joanna A. Ellis-Monaghan and Criel Merino. Graph polynomials and their applications I: The Tutte polynomial. In Matthias Dehmer, editor, *Structural Analysis of Complex Networks*, pages 219–255. Birkhäuser Boston, 2011.
 - [24] Joanna A. Ellis-Monaghan and Criel Merino. Graph polynomials and their applications II: Interrelations and interpretations. In Matthias Dehmer, editor, *Structural Analysis of Complex Networks*, pages 257–292. Birkhäuser Boston, 2011.
 - [25] Fedor V. Fomin and Dimitrios M. Thilikos. A 3-approximation for the pathwidth of Halin graphs. *J. Discrete Algorithms*, 4(4):499–510, 2006.
 - [26] Qi Ge and Daniel Štefankovič. A graph polynomial for independent sets of bipartite graphs. *Combin. Probab. Comput.*, 21(5):695–714, 2012.
 - [27] Qi Ge and Daniel Štefankovič. A graph polynomial for independent sets of bipartite graphs. *CoRR*, [arXiv:0911.4732](https://arxiv.org/abs/0911.4732), 2009.
 - [28] Roy J. Glauber. Time-dependent statistics of the Ising model. *J. Mathematical Phys.*, 4:294–307, 1963.

- [29] Leslie Ann Goldberg and Mark Jerrum. Approximating the partition function of the ferromagnetic Potts model. *J. ACM*, 59(5):#25, 31 pp., 2012.
- [30] Leslie Ann Goldberg and Mark Jerrum. A counterexample to rapid mixing of the Ge-Štefankovič process. *Electron. Commun. Probab.*, 17:#5, 6 pp., 2012.
- [31] Leslie Ann Goldberg, Mark Jerrum, and Marek Karpinski. The mixing time of Glauber dynamics for coloring regular trees. *Random Structures Algorithms*, 36(4):464–476, 2010.
- [32] Gary Gordon and Lorenzo Traldi. Generalized activities and the Tutte polynomial. *Discrete Math.*, 85(2):167–176, 1990.
- [33] Geoffrey Grimmett. *The random-cluster model*, volume 333 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2006.
- [34] Petr Hliněný. The Tutte polynomial for matroids of bounded branch-width. *Combin. Probab. Comput.*, 15(3):397–409, 2006.
- [35] Petr Hliněný, Sang-il Oum, Detlef Seese, and Georg Gottlob. Width Parameters Beyond Tree-width and their Applications. *The Computer Journal*, 51(3):326–362, 2008.
- [36] Petr Hliněný and Geoff Whittle. Matroid tree-width. *European J. Combin.*, 27(7):1117–1128, 2006.
- [37] Mark Huber. Simulation reductions for the Ising model. *J. Stat. Theory Pract.*, 5(3):413–424, 2011.
- [38] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31:253–258, 1925.
- [39] François Jaeger, Dirk L. Vertigan, and Dominic J. A. Welsh. On the computational complexity of the Jones and Tutte polynomials. *Math. Proc. Cambridge Philos. Soc.*, 108(1):35–53, 1990.
- [40] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.
- [41] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993.
- [42] Vaughan F. R. Jones. A polynomial invariant for knots via von Neumann algebras. *Bull. Amer. Math. Soc. (N.S.)*, 12(1):103–111, 1985.
- [43] David R. Karger. A randomized fully polynomial time approximation scheme for the all-terminal network reliability problem. *SIAM Rev.*, 43(3):499–522, 2001.
- [44] Navin Kashyap. Matroid pathwidth and code trellis complexity. *SIAM J. Discrete Math.*, 22(1):256–272, 2008.
- [45] Nancy G. Kinnersley. The vertex separation number of a graph equals its path-width. *Inform. Process. Lett.*, 42(6):345–350, 1992.
- [46] Ephraim Korach and Nir Solel. Tree-width, path-width, and cutwidth. *Discrete Appl. Math.*, 43(1):97–101, 1993.

- [47] Johann A. Makowsky. From a zoo to a zoology: towards a general theory of graph polynomials. *Theory Comput. Syst.*, 43(3-4):542–562, 2008.
- [48] Johann A. Makowsky and Julian P. Mariño. Farrell polynomials on graphs of bounded tree width. *Adv. in Appl. Math.*, 30(1-2):160–176, 2003. Formal power series and algebraic combinatorics (Scottsdale, AZ, 2001).
- [49] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Fast mixing for independent sets, colorings, and other models on trees. *Random Structures Algorithms*, 31(2):134–172, 2007.
- [50] Steven D. Noble. Evaluating the Tutte polynomial for graphs of bounded tree-width. *Combin. Probab. Comput.*, 7(3):307–321, 1998.
- [51] Steven D. Noble. Evaluating a weighted graph polynomial for graphs of bounded tree-width. *Electron. J. Combin.*, 16(1):#64, 14 pp., 2009.
- [52] Steven D. Noble and Dominic J. A. Welsh. A weighted graph polynomial from chromatic invariants of knots. *Ann. Inst. Fourier (Grenoble)*, 49(3):1057–1087, 1999. Symposium à la Mémoire de François Jaeger (Grenoble, 1998).
- [53] James G. Oxley. *Matroid theory*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1992.
- [54] James G. Oxley and Dominic J. A. Welsh. The Tutte polynomial and percolation. In *Graph theory and related topics (Proc. Conf., Univ. Waterloo, Waterloo, Ont., 1977)*, pages 329–339. Academic Press, New York, 1979.
- [55] Renfrey B. Potts. Some generalized order-disorder transformations. *Proc. Cambridge Philos. Soc.*, 48:106–109, 1952.
- [56] Dana Randall. Rapidly mixing Markov chains with applications in computer science and physics. *Computing in Science Engineering*, 8(2):30–41, March-April 2006.
- [57] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1(4):351–370, 1992.
- [58] Alan D. Sokal. The multivariate Tutte polynomial (alias Potts model) for graphs and matroids. In *Surveys in Combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 173–226. Cambridge Univ. Press, Cambridge, 2005.
- [59] Prasad Tetali, Juan Carlos Vera, Eric Vigoda, and Linji Yang. Phase transition for the mixing time of the Glauber dynamics for coloring regular trees. In Moses Charikar, editor, *SODA*, pages 1646–1656. SIAM, 2010.
- [60] Robin Thomas. Tree-decompositions of graphs, 1996. Lecture notes; available at <http://people.math.gatech.edu/~thomas/PAP/treenotes.pdf>. Accessed 10/2014.
- [61] William T. Tutte. A contribution to the theory of chromatic polynomials. *Canadian J. Math.*, 6:80–91, 1954.
- [62] Dominic J. A. Welsh. *Matroid theory*. Academic Press, London, 1976. L. M. S. Monographs, No. 8.

- [63] Dominic J. A. Welsh. *Complexity: knots, colourings and counting*, volume 186 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1993.
- [64] Dominic J. A. Welsh and Criel Merino. The Potts model and the Tutte polynomial. *J. Math. Phys.*, 41(3):1127–1152, 2000. Probabilistic techniques in equilibrium and nonequilibrium statistical physics.
- [65] Jacob A. White. On multivariate chromatic polynomials of hypergraphs and hyper-edge elimination. *Electron. J. Combin.*, 18(1):#160, 17 pp., 2011.