# Sampling hypergraphs with given degrees

Martin Dyer* Catherine Greenhill† Pieter Kleer‡ James Ross†
Leen Stougie§

22 June 2020



**Abstract**

There is a well-known connection between hypergraphs and bipartite graphs, obtained by treating the incidence matrix of the hypergraph as the biadjacency matrix of a bipartite graph. We use this connection to describe and analyse a rejection sampling algorithm for sampling simple uniform hypergraphs with a given degree sequence. Our algorithm uses, as a black box, an algorithm $\mathcal{A}$ for sampling bipartite graphs with given degrees, uniformly or nearly uniformly, in (expected) polynomial time. The expected runtime of the hypergraph sampling algorithm depends on the (expected) runtime of the bipartite graph sampling algorithm $\mathcal{A}$, and the probability that a uniformly random bipartite graph with given degrees corresponds to a simple hypergraph. We give some conditions on the hypergraph degree sequence which guarantee that this probability is bounded below by a constant.

*Keywords*: hypergraph, degree sequence, sampling, algorithm, Markov chain

## 1 Introduction

Hypergraphs are combinatorial objects which can be used to abstractly represent general dependence structures, with applications in many areas including machine learning [28] and bioinformatics [26]. We consider the problem of efficiently sampling simple, $r$-uniform hypergraphs with a given degree sequence, either uniformly or approximately uniformly.

More precisely, a hypergraph $H = (V, E)$ consists of a finite set $V = V(H) = \{v_1, \ldots, v_n\}$ of nodes, and a multiset $E = E(H)$ of edges, where each edge is a nonempty multisubset of $V$. We say that $H$ is *simple* if there are no repeated edges in $E$ and no edge of $E$ contains a repeated node (so $E$ is a set of subsets of nodes). For any node $v_i \in V$, we define the degree of $v_i$ by

$$d_i = \deg_H(v_i) = |\{e \in E(H) : v_i \in e\}|,$$

*School of Computing, University of Leeds, Leeds LS2 9JT, UK. m.e.dyer@leeds.ac.uk

†School of Mathematics and Statistics, UNSW Sydney, Sydney NSW 2052, Australia. c.greenhill@unsw.edu.au, james.ross@unsw.edu.au

‡Max-Planck-Institut für Informatik, Saarland Informatics Campus, 66123 Saarbrücken, Germany. pkleer@mpi-inf.mpg.de

§CWI, INRIA-Erable, Vrije Universiteit, Amsterdam, The Netherlands Leen.Stougie@cwi.nl

and write $\boldsymbol{d} = (d_1, \ldots, d_n)$ for the degree sequence of $H$. For a positive integer $r$, we say that $H$ is *$r$-uniform* if every edge contains exactly $r$ nodes, counting multiplicities when $H$ is not simple. We then write $\mathcal{H}_r(\boldsymbol{d})$ for the set of $r$-uniform simple hypergraphs with degree sequence $\boldsymbol{d}$. If there is a positive integer $d$ such that $d_i = d$ for all $i \in [n]$, then we write $\mathcal{H}_r(n, d)$ for the set of all $r$-uniform $d$-regular hypergraphs on $n$ nodes, instead of $\mathcal{H}_r(\boldsymbol{d})$.

Recently, Deza, Levin, Meesum & Onn [13] proved that the construction problem for simple 3-uniform hypergraphs is NP-hard. That is, given $\boldsymbol{d}$ it is NP-hard to decide whether there exists a 3-uniform hypergraph with degree sequence $\boldsymbol{d}$. This implies that it is not possible to approximate $|\mathcal{H}_r(\boldsymbol{d})|$ efficiently in general, since approximate counting can distinguish 0 from a positive number. Moreover, hardness of construction also directly implies that (approximate) uniform sampling is a difficult problem in general.

Arafat et al. [2] recently gave an algorithm to construct a non-simple hypergraph with given degrees and edge sizes. Chodrow [9] considers Markov Chain Monte Carlo approaches for generating such hypergraphs. We emphasize that throughout this work, we only consider simple hypergraphs. To the best of our knowledge, the only rigorously-analysed algorithm for this problem in the literature is the configuration model, see Section 2.2.

Our approach is based on the well-known connection between hypergraphs and bipartite graphs. We first fix some notation for bipartite graphs and then explain this relation. A bipartite graph $B = (X \cup Y, A)$ consists of a bipartition $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ of nodes, and an edge set $A \subseteq X \times Y = \{\{x, y\} : x \in X, y \in Y\}$. For a pair of nonnegative integer sequences $\boldsymbol{d} = (d_1, \ldots, d_n)$ and $\boldsymbol{r} = (r_1, \ldots, r_m)$, let $\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$ be the set of all simple bipartite graphs $B$ such that

$$\deg_B(x_i) = d_i \text{ for all } x_i \in X, \quad \text{and} \quad \deg_B(y_j) = r_j \text{ for all } y_j \in Y.$$

We say that $(\boldsymbol{d}, \boldsymbol{r})$ is the (bipartite) degree sequence of $B$. Note that $\mathcal{B}(\boldsymbol{d}, \boldsymbol{r}) = \emptyset$ unless $\sum_{i=1}^{n} d_i = \sum_{j=1}^{m} r_j$. If there is a fixed integer $r$ such that $r_j = r$ for all $j \in [m]$, then we write $\mathcal{B}(\boldsymbol{d}, r)$ instead of $\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, and we call such bipartite graphs *half-regular*. If in addition $\boldsymbol{d} = (d, d, \ldots, d)$ is regular then we write $\mathcal{B}(n, d, r)$. For any node $v \in X \cup Y$, let $\mathcal{N}_B(v) = \{w \in X \cup Y : \{v, w\} \in A\}$ be the neighbourhood of node $v$ in $B$.

Every hypergraph $H = (V, E)$ can be represented as a bipartite graph $B_H$, as follows. Fix a labelling of the edges of $H$, say $E = \{e_1, \ldots, e_m\}$, then let

$$X = V, \quad Y = E \quad \text{and} \quad A = \Big\{\{v_i, e_j\} \in X \times Y \mid v_i \in e_j\Big\}.$$

If $H \in \mathcal{H}_r(\boldsymbol{d})$ then $B(H) \in \mathcal{B}(\boldsymbol{d}, r)$. Conversely, every bipartite graph $B = (X \cup Y, A)$ corresponds to a hypergraph $H_B = (V, E)$, where $V = X$ and

$$E = \{\mathcal{N}_B(y) \mid y \in Y\}.$$

Furthermore, $H_B$ is simple if and only if every node in $Y$ has a distinct set of neighbours in $B$; that is, if $N_B(y_i) = N_B(y_j)$ implies $i = j$. If $H_B$ is simple then we say that the bipartite graph $B$ is *H-simple*.

Write $\mathcal{B}^*(\boldsymbol{d}, r)$ for the set of all H-simple half-regular bipartite graphs, and define $\varphi : \mathcal{B}^*(\boldsymbol{d}, r) \to \mathcal{H}_r(\boldsymbol{d})$ as the canonical mapping that maps $B$ to $H_B$, as described above. We can use rejection sampling to turn any sampling algorithm for $\mathcal{B}(\boldsymbol{d}, r)$ into a sampling algorithm for $\mathcal{H}_r(\boldsymbol{d})$, as follows:

```
HypergraphSampling(d, r, A)
Input: Parameters (d, r);
        algorithm A for sampling from B(d, r).
Begin
    repeat
        sample B from B(d, r) using A
        until B is H-simple
    output φ(B)
end.
```

Note that for all $H \in \mathcal{H}_r(\boldsymbol{d})$ we have $|\varphi^{-1}(H)| = m!$, as there are $m!$ distinct ways to label the edges of $H$ when $H$ is simple. Hence, if $\mathcal{A}$ samples *uniformly* from $\mathcal{B}(\boldsymbol{d}, r)$ then the output of HypergraphSampling is uniform over $\mathcal{H}_r(\boldsymbol{d})$.

The expected number of times HypergraphSampling draws a sample from $\mathcal{B}(\boldsymbol{d}, r)$, using algorithm $\mathcal{A}$, depends on the the proportion of bipartite graphs in $\mathcal{B}(\boldsymbol{d}, r)$ which are H-simple: that is, the ratio $|\mathcal{B}^*(\boldsymbol{d}, r)|/|\mathcal{B}(\boldsymbol{d}, r)|$. The goal of our work is to identify pairs $(\boldsymbol{d}, r)$ for which $|\mathcal{B}(\boldsymbol{d}, r)|/|\mathcal{B}^*(\boldsymbol{d}, r)|$ is bounded above by a polynomial in $n$. For such pairs, if the output of $\mathcal{A}$ is close to uniform then this implies that the expected number of times we run $\mathcal{A}$ before an H-simple element of $\mathcal{B}(\boldsymbol{d}, r)$ is found is at most polynomial. This is made more specific in the next subsection.

## 1.1 Our contributions

The total variation distance between two probability distributions $\sigma$, $\pi$ on a set $\Omega$ is given by

$$d_{TV}(\sigma, \pi) = \frac{1}{2} \sum_{x \in \Omega)} |\sigma(x) - \pi(x)| = \max_{S \subseteq \Omega} |\sigma(S) - \pi(S)|. \tag{1.1}$$

Suppose that the distribution of the output of algorithm $\mathcal{A}$ on $\mathcal{B}(\boldsymbol{d}, r)$ is $\sigma_{\mathcal{B}}$, and let $\sigma_{\mathcal{H}}$ be the output of the algorithm HypergraphSampling. Then $\sigma_{\mathcal{H}}$ is a distribution on $\mathcal{H}_r(\boldsymbol{d})$ which is obtained by setting $H = \varphi(B)$ where $B$ has distribution $\sigma_{\mathcal{B}}$, conditioned on $B \in \mathcal{B}^*(\boldsymbol{d}, r)$. To make it clear which distribution we are using, we write $\mathbb{P}_\sigma$ for the probability mass function of the distribution $\sigma$. Let $\pi_{\mathcal{B}}$ be the uniform distribution on $\mathcal{B}(\boldsymbol{d}, r)$, and let $\pi_{\mathcal{H}}$ be the uniform distribution on $\mathcal{H}_r(\boldsymbol{d})$.

A *fully-polynomial almost uniform sampler (FPAUS)* for sampling from a set $|\Omega|$ is an algorithm that, with probability at least $\frac{3}{4}$, outputs an element of $\Omega$ in time polynomial in $\log |\Omega|$ and $\log(1/\varepsilon)$, such that the output distribution is *$\varepsilon$-close* to the uniform distribution $\pi$ on $\Omega$ in *total variation distance*: that is, $d_{TV}(\sigma, \pi) \le \varepsilon$. If $\Omega = \mathcal{B}(\boldsymbol{d}, r)$ or $\Omega = \mathcal{H}_r(\boldsymbol{d})$ then $\log |\Omega| = O(M \log M)$: this follows from [18, Theorem 1.3], restated below as Theorem 4.1. So an FPAUS for $\mathcal{H}_r(\boldsymbol{d})$ or $\mathcal{B}(\boldsymbol{d}, r)$ must have running time bounded above by a polynomial in $d_{\max}$, $n$ and $\log(1/\varepsilon)$.

The following result summarises the properties of HypergraphSampling $(\boldsymbol{d}, r, \mathcal{A})$ in terms of the output distribution and runtime of $\mathcal{A}$. The proof, which is fairly standard, is presented in Section 3.

**Theorem 1.1.** *Suppose that $n$ is a positive integer, $\boldsymbol{d} = (d_1, \ldots, d_n)$ is a sequence of positive integers, and $r$ is a positive integer such that $\mathcal{B}^*(\boldsymbol{d}, r)$ is non-empty.*

(i) *The output distribution $\sigma_{\mathcal{H}}$ of* HyperGraphSampling *satisfies*

$$d_{TV}(\sigma_{\mathcal{H}}, \pi_{\mathcal{H}}) \leq \frac{3}{2} \cdot \frac{d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}})}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))}.$$

(ii) *The expected runtime of* HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ *is $q(\boldsymbol{d}, r) \cdot \tau(\boldsymbol{d}, r)$, where $\tau(\boldsymbol{d}, r)$ is the (expected) runtime of algorithm $\mathcal{A}$ on $\mathcal{B}(\boldsymbol{d}, r)$ and $q(\boldsymbol{d}, r)^{-1} = \mathbb{P}_{\sigma_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))$. Furthermore, the probability that* HyperGraphSampling *needs more than $t\, q(\boldsymbol{d}, r)$ iterations of $\mathcal{A}$ before finding an element of $\mathcal{B}^*(\boldsymbol{d}, r)$ is at most $\exp(-t)$ for any $t > 0$.*

(iii) *Suppose that $d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}}) \leq \varepsilon$ and $\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r)) \geq 1 - c_0$ for some $\varepsilon \in (0, 1)$ and $c_0 \in (0, 1 - \varepsilon)$. Then*

$$d_{TV}(\sigma_{\mathcal{H}}, \pi_{\mathcal{H}}) \leq \frac{3\varepsilon}{2(1 - c_0)}$$

*and the expected runtime of* HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ *is at most*

$$(1 - c_0 - \varepsilon)^{-1}\, \tau(\boldsymbol{d}, r).$$

(iv) *If $\mathcal{A}$ is an FPAUS for $\mathcal{B}(\boldsymbol{d}, r)$ and the assumptions of (iii) hold, then we can transform* HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ *into an FPAUS for $\mathcal{H}_r(\boldsymbol{d})$ by terminating after $\lceil 2(1 - c_0 - \varepsilon)^{-1} \rceil$ calls to $\mathcal{A}$ and reporting* `FAIL`.

We see from Theorem 1.1 that $d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}})$ and $\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))$ are the two crucial quantities which control both the expected runtime of HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$, and how far the output varies from uniform. The first of these quantities is determined by the choice of algorithm $\mathcal{A}$. In our next two theorems, we provide a lower bound on $\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))$ when $\boldsymbol{d} = (d, d, \ldots, d)$ is regular, and give an asymptotic lower bound on this probability which holds when $\boldsymbol{d}$ is irregular but sparse.

*Remark* 1.2. For Theorem 1.1(iv) to provide an FPAUS with an *explicit* upper bound on the runtime, explicit bounds are needed on $\varepsilon$ and $c_0$. However, for Theorem 1.1(iii) it is enough to know that sufficiently small values of $c_0$, $\varepsilon$ exist.

**Theorem 1.3.** *Let $n \in \mathbb{N}$, $d \in \mathbb{N}$ and $r \geq 3$ so that $m = nd/r \in \mathbb{N}$ and $\binom{m}{2} < \binom{n}{r}$. Then*

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(n, d, r)) \geq 1 - \binom{m}{2}\binom{n}{r}^{-1}.$$

*Hence Theorem 1.1(iii) applies when $\binom{m}{2} \leq c_0\binom{n}{r}$ for some $c_0 \in (0, 1-\varepsilon)$, where $d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}}) \leq \varepsilon$.*

*Remark* 1.4. When $r \geq 3$ is a fixed constant, the lower bound in Theorem 1.3 is $1 - o(1)$ if $d = o(n^{r/2-1})$.

**Theorem 1.5.** *Assume that for each positive integer $n$ we have an integer $r = r(n) \geq 3$ and a sequence $\boldsymbol{d} = \boldsymbol{d}(n) = (d_1, \ldots, d_n)$ of positive integers such that $M = \sum_{i=1}^n d_i$ tends to infinity with $n$. Suppose that $r$ divides $M$ for infinitely many values of $n$, and let $m = M/r$. Assume that $r^2 d_{\max}^2 = o(M)$ and let $\pi_{\mathcal{B}}$ be the uniform distribution on $\mathcal{B}(\boldsymbol{d}, r)$. Then*

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r)) \geq 1 - \frac{n^r\, d_{\max}^r}{M^r} \cdot \binom{M/r}{2} \binom{n}{r}^{-1} \cdot (1 + o(1)).$$

*Writing $m = M/r$ and $d = M/n$, wee see that Theorem 1.1(iii) applies when*

$$\left(\frac{d_{\max}}{d}\right)^r \binom{m}{2} \leq c_0 \binom{n}{r} \tag{1.2}$$

*for some $c_0 \in (0, 1 - \varepsilon)$, where $d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}}) \leq \varepsilon$.*

In the hypergraph setting, $m$ is the number of edges and $d$ is the average degree of any hypergraph in $\mathcal{H}_r(\boldsymbol{d})$.

*Remark* 1.6. When $r \geq 3$ is a fixed constant, the lower bound in Theorem 1.5 is $1 - o(1)$ if $d_{\max} = o(M^{1-2/r})$. Similarly, if $r \geq 3$ is a fixed constant and $d = O(d_{\max})$ then the lower bound in Theorem 1.5 is $1 - o(1)$ whenever $d^2 = o(n^{r-2})$, as in the regular case.

There are several algorithms $\mathcal{A}$ in the literature for efficiently sampling from $\mathcal{B}(\boldsymbol{d}, r)$, either uniformly or almost uniformly, under various conditions on $\boldsymbol{d}$ and $r$. These will be reviewed in Section 2.1, together with the properties of the resulting algorithm HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$. In Section 2.2 we discuss the configuration model for hypergraphs, which can be used as an expected polynomial-time sampling algorithm when $rd_{\max} = O(\log n)$.

Then in Section 3 we provide a general framework which we use to analyse the algorithm HypergraphSampling. In Section 4 we fill in the details for the regular regime (Theorem 1.3) and the irregular, sparse regime (Theorem 1.5).

# 2 Related work

## 2.1 Various bipartite sampling algorithms, and implications

In this section we describe several algorithms for efficient sampling from $\mathcal{B}(\boldsymbol{d}, r)$, uniformly or almost uniformly, under various conditions on $\boldsymbol{d}$, $r$. We also apply Theorem 1.1 to describe when HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ is an efficient algorithm for sampling from $\mathcal{H}_r(\boldsymbol{d})$ (uniformly or near-uniformly), for each choice of $\mathcal{A}$.

The first two algorithms mentioned below perform *exactly uniform* sampling from $\mathcal{B}(\boldsymbol{d}, r)$.

(I) If $rd_{\max} \leq C \log n$ then the bipartite configuration model gives rise to an algorithm for sampling (exactly) uniformly from $\mathcal{B}(\boldsymbol{d}, r)$. But the bipartite configuration model is equivalent to the configuration model for hypergraphs, described in Section 2.2, and so there is no advantage to working in the bipartite graph setting when $rd_{\max} \leq C \log n$. (See Lemma 2.3).

(II) Next suppose that $(d_{\max} + r)^4 = O(M)$. Building on the work of [17, 24], Arman, Gao and Wormald [3, Theorem 4] describe an algorithm which samples uniformly from $\mathcal{B}(\boldsymbol{d}, r)$ with expected runtime $O(M)$. Note that $d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}}) = 0$ as the output is exactly uniform. Using this algorithm as $\mathcal{A}$ and applying Theorem 1.1, we see that HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ performs exact sampling from $\mathcal{H}_r(\boldsymbol{d})$ with expected runtime $O(M)$ whenever (1.2) holds for some constant $c_0 \in (0, 1)$. In particular, this holds whenever $(\boldsymbol{d}, r)$ are as described in Remark 1.6.

The next algorithm produces output which is *asymptotically uniform*, meaning that the output distribution is only $o(1)$ from uniform in total variation distance.

(III) If $d_{\max}+r = O(M^{1/4-\tau})$ for some positive constant $\tau$ then the algorithm from [3] can be applied, as described in (II). An alternative is to use the sampling algorithm of Bayati, Kim and Saberi [4, Theorem 1], which has expected runtime $O(d_{\max}M)$ (see the proof of [4, Theorem 1]). The output of this algorithm satisfies $d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}}) = o(1)$, where this vanishing term depends only on $n$ and cannot be made smaller by increasing the runtime of the algorithm. Hence we can take $\varepsilon = o(1)$ in Theorem 1.1, and conclude that for this choice of $\mathcal{A}$, the algorithm HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ has expected runtime $O(d_{\max}M)$ whenever (1.2) holds for some constant $c_0 \in (0, 1)$, and the distribution of the output is within $o(1)$ of uniform: that is, $d_{TV}(\sigma_{\mathcal{H}}, \pi_{\mathcal{H}}) = o(1)$.

*Remark* 2.1. Although the Arman, Gao and Wormald algorithm applies for a slightly wider range of values of $(\boldsymbol{d}, r)$, has a better expected runtime bound and performs exactly uniform sampling, the Bayati, Kim and Saberi algorithm has one advantage: it is much easier to implement. Indeed, Bayati, Kim and Saberi [4, Theorem 3] used sequential importance sampling to give an algorithm which is close to an FPAUS, except that the runtime is polymial in $1/\varepsilon$, while in an FPAUS the dependence should be on $\log(1/\varepsilon)$. However, this algorithm is valid only when $d_{\max} = O(M^{1/4-\tau})$ for some $\tau > 0$ and no longer has the advantage of simplicity, and so it is surpassed by the fast, precisely uniform sampling algorithm of Arman, Gao and Wormald [3], described in (II) above. (Other authors, such as Chen et al. [8], have used sequential importance sampling to sample bipartite graphs with given degrees, but without rigorous analysis.)

Now we survey some algorithms which are FPAUSs for $\mathcal{B}_r(\boldsymbol{d})$. Each can be used as $\mathcal{A}$ to give an FPAUS for $\mathcal{H}_r(\boldsymbol{d})$, using Theorem 1.1(iv), so long as (1.2) holds (or $\binom{m}{2} \leq c_0\binom{n}{r}$ when $\boldsymbol{d}$ is regular) for some $c_0 \in (0, 1-\varepsilon)$. In all cases, the polynomial bound on the runtime is quite a high-degree polynomial and is not believed to be tight. We do not always state the runtime.

(IV) Jerrum, Sinclair and Vigoda [21] described and analysed a simulated annealing algorithm which gives an FPAUS for sampling perfect matchings from a given bipartite graph. As a corollary, they obtained an FPAUS for sampling bipartite graphs for given degrees [21, Corollary 8.1]. Bezáková, Bhatnagar and Vigoda [5] adapted the algorithm from [21] to provide a simplified FPAUS for $\mathcal{B}(\boldsymbol{d}, r)$, valid for any $(\boldsymbol{d}, r)$, with running time

$$O((nm)^2M^3\Delta \log^4(nm/\varepsilon)),$$

where $n$ and $m$ are the number of nodes in each part of the bipartition, and $\Delta = \max\{d_{\max}, r\}$.

(V) Another well-studied Markov chain for sampling (bipartite) graphs with given degrees is the switch Markov chain. It is the the simplest Markov chain which walks on the set of all (bipartite) graphs with a given degree sequence, as it deletes and replaces only two edges at a time. The chain has been used in many contexts, including contingency tables [14], and was first applied to bipartite graphs by Kannan, Tetali and Vempala [22]. The mixing time (runtime) of the switch chain has been shown to be polynomial for various bipartite and general degree sequences, see for example [1, 10, 16, 19, 25]. If a Markov chain leads to an FPAUS then we say that the Markov chain is rapidly mixing. In particular:

- Miklos, Erdős and Soukup [25] show that the switch Markov chain is rapidly mixing for half-regular bipartite degree sequences (in which one part of the bipartition has regular degrees). An explicit polynomial bound is not clearly stated.
- Cooper, Dyer and Greenhill [10] considered regular graphs and showed that the switch chain is rapidly mixing on the set of all $d$-regular graphs, for any $d = d(n)$. Their analysis was extended by Greenhill and Sfragara [19] who adapted the proof to sparse, irregular degree sequences. Neither of these works explicitly treated bipartite graphs, though the arguments in both papers are simpler when restricted to bipartite graphs. In Corollary A.3 we state an upper bound on the mixing time of the switch chain on $\mathcal{B}(\boldsymbol{d}, r)$ which arises from the arguments of [10, 19] when $3 \le d_{\max} \le \frac{1}{3}\sqrt{M}$. Specifically we show that in this range, the switch chain gives an FPAUS with running time

$$\Delta^{10} M^7 \big(\tfrac{1}{2} M \log(M) + \log(\varepsilon^{-1})\big)$$

  where $\Delta = \max\{d_{\max},\, r\}$.
- Jerrum and Sinclair [20] defined a notion of P-stability for degree sequences. Roughly speaking, a degree sequence $\boldsymbol{d}$ is P-stable if small perturbations to $\boldsymbol{d}$ only change the number of realisations (graphs with degree sequence $\boldsymbol{d}$) by a small amount. The notion of P-stable can be adapted to bipartite graphs. Amanatidis and Kleer [1] defined a possibly stronger notion, *strong stability*, and showed that the switch chain is rapidly mixing for any strongly stable degree sequence, and for any strongly stable bipartite degree sequence. Erdős et al. [16] proved that the switch chain is rapidly mixing for any P-stable class of bipartite degree sequences.

(VI) The Curveball chain [27] is another Markov chain for sampling bipartite graphs with given degrees, in which multiple switches are performed simultaneously. Carstens and Kleer [7] showed that the Curveball chain is rapidly mixing whenever the switch chain is rapidly mixing.

*Remark* 2.2. We have focussed on uniform hypergraphs, but our approach can be adapted to non-uniform hypergraphs. Given a vector $\boldsymbol{r} = (r_1, \ldots, r_m)$ which stores the desired edge sizes, let $m_\ell$ be the number of edges of size $\ell$, that is,

$$m_\ell = |\{j \in [m] \,:\, r_j = \ell\}|$$

for $\ell \geq 2$. Each hypergraph $H$ on $[n]$ with edge sizes given by $\boldsymbol{r}$ and with degree sequence $\boldsymbol{d}$ corresponds to exactly $\prod_{\ell=2}^{n} m_\ell!$ bipartite graphs from $\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, as now we must restrict to edge labellings $e_1, \ldots, e_m$ so that $|e_j| = r_j$ for $j = 1, \ldots, m$. All of the bipartite graph sampling algorithms mentioned in this section generalise to $\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, with the exception of the result by Miklos, Erdős and Soukup regarding the switch chain for half-regular bipartite graphs [25].

## 2.2 Sampling hypergraphs using the configuration model

To the best of our knowledge, the only rigorously-analysed algorithm for sampling hypergraphs with given degrees is the *configuration model*. The analogue of the configuration model for hypergraphs has been used by various authors, for example, in the study of random hypergraphs [12] and for asymptotic enumeration [15]. In the configuration model corresponding to $\mathcal{H}_r(\boldsymbol{d})$ there are $n$ objects, called cells, and the $i$th cell contains $d_i$ (labelled) points. A *configuration* is a partition of the $M = \sum_{i=1}^{n} d_i$ points into $M/r$ parts, each containing $r$ points. A random configuration can be chosen in $O(M)$ time. Shrinking each cell to a node gives an $r$-uniform hypergraph which may contain loops (that is, an edge containing a node more than once) or repeated edges. If the resulting hypergraph is not simple then the configuration is rejected and a new random configuration is sampled. We say that a configuration is simple if the corresponding hypergraph is simple.

Hence, the configuration model can be used for efficient sampling when the probability that a randomly chosen configuration is simple is bounded below by the inverse of some polynomial in $n$. This implies that the expected number of trials before a simple configuration is found is at most polynomial.

It follows from asymptotic results of Dudek, Frieze, Ruciński and Šileikis [15] that when $\boldsymbol{d} = (d, d, \ldots, d)$ is regular, the expected number of trials before a simple configuration is sampled is

$$\exp\left(\tfrac{1}{2}(r-1)(d-1) + o(1)\right)$$

assuming that $r = 3$ and $d = d(n) = o(n^{1/2})$, or $r \geq 4$ and $d = d(n) = o(n)$. (Asymptotics are as $n \to \infty$, restricted to values of $n$ such that $dn$ is divisible by $r$.) For irregular degrees, let $M_2 = M_2(\boldsymbol{d}) = \sum_{j=1}^{n} d_j(d_j - 1)$. It follows from [6, Corollary 2.3] that the expected number of trials before a configuration is sampled is

$$\exp\left(\frac{(r-1)M_2}{2M} + o(1)\right) \leq \exp\left(\tfrac{1}{2}(r-1)(d_{\max} - 1) + o(1)\right)$$

whenever $r^4 d_{\max}^3 = o(M)$. Here $r = r(n)$ and $\boldsymbol{d} = \boldsymbol{d}(n)$ are such that $r$ divides $M$ for infinitely many values of $n$. We collect these facts together into the following lemma.

**Lemma 2.3.** *The configuration model gives an efficient algorithm for sampling uniformly from $\mathcal{H}_r(\boldsymbol{d})$ whenever $rd_{\max} = O(\log n)$. If $rd_{\max} \le C \log n$ for some constant $C > 0$ then the expected runtime of this algorithm for $\mathcal{H}_r(\boldsymbol{d})$ is $O(n^C M) = O(d_{\max} n^{C+1})$. If $rd = o(\log n)$ then the expected runtime of this algorithm is $O(M) = O(d_{\max} n)$. Note $d_{TV}(\sigma_{\mathcal{H}}, \pi_{\mathcal{H}}) = 0$ as the output is exactly uniform.*

Gao and Wormald [17] built on earlier work of McKay and Wormald [24] to give a fast algorithm for exactly uniform sampling of $d$-regular graphs. Using a recent improvement of Arman, Gao and Wormald [3], a uniformly random $d$-regular graph on $n$ vertices can be generated in expected runtime $O(dn + d^4)$ whenever $d = o(n^{1/2})$, and a random graph with degree sequence $\boldsymbol{d}$ can be generated in runtime $O(M)$ whenever $d_{\max}^4 = O(M)$. It is likely that this approach could be adapted to the problem of sampling hypergraphs uniformly.

# 3 Analysis of HypergraphSampling

First we prove Theorem 1.1.

*Proof of Theorem 1.1.* To prove (i), observe that by definition,

$$\sigma_{\mathcal{H}}(H) = \sum_{B \in \varphi^{-1}(H)} \mathbb{P}_{\sigma_{\mathcal{B}}}(B \mid B \in \mathcal{B}^*(\boldsymbol{d}, r)) = \frac{1}{\mathbb{P}_{\sigma_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} \cdot \sum_{B \in \varphi^{-1}(H)} \sigma_{\mathcal{B}}(B).$$

This equality also holds with $\sigma_{\mathcal{H}}$, $\sigma_{\mathcal{B}}$ replaced by $\pi_{\mathcal{H}}$, $\pi_{\mathcal{B}}$, respectively. Since the set of all preimages $\{\varphi^{-1}(H) \;:\; H \in \mathcal{H}_r(\boldsymbol{d})\}$ forms a partition of $\mathcal{B}^*(\boldsymbol{d}, r)$, and using the triangle inequality, we have

$$\begin{aligned} d_{TV}(\sigma_{\mathcal{H}}, \pi_{\mathcal{H}}) &= \tfrac{1}{2} \sum_{H \in \mathcal{H}_r(\boldsymbol{d})} |\, \sigma_{\mathcal{H}}(H) - \pi_{\mathcal{H}}(H) \,| \\ &\le \tfrac{1}{2} \sum_{H \in \mathcal{H}_r(\boldsymbol{d})} \sum_{B \in \varphi^{-1}(H)} \left| \frac{\sigma_{\mathcal{B}}(B)}{\mathbb{P}_{\sigma_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} - \frac{\pi_{\mathcal{B}}(B)}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} \right| \\ &\le \tfrac{1}{2} \sum_{B \in \mathcal{B}^*(\boldsymbol{d}, r)} \left| \frac{\sigma_{\mathcal{B}}(B)}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} - \frac{\pi_{\mathcal{B}}(B)}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} \right| \\ &\qquad + \tfrac{1}{2} \sum_{B \in \mathcal{B}^*(\boldsymbol{d}, r)} \left| \frac{\sigma_{\mathcal{B}}(B)}{\mathbb{P}_{\sigma_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} - \frac{\sigma_{\mathcal{B}}(B)}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} \right| \\ &\le \frac{d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}})}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} + \tfrac{1}{2} \cdot \left| \frac{1}{\mathbb{P}_{\sigma_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} - \frac{1}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))} \right| \sum_{B \in \mathcal{B}^*(\boldsymbol{d}, r)} \sigma_{\mathcal{B}}(B) \\ &\le \tfrac{3}{2} \cdot \frac{d_{TV}(\sigma_{\mathcal{B}}, \pi_{\mathcal{B}})}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))}, \end{aligned}$$

The final inequality follows from applying (1.1) with $S = \mathcal{B}^*(\boldsymbol{d}, r)$.

Next, (ii) is immediate as the number of times that HyperSampling calls $\mathcal{A}$ has a geometric distribution with mean $1/q(n)$. Then, (1.1) implies that

$$q(\boldsymbol{d}, r)^{-1} = \mathbb{P}_{\sigma_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r)) \ge \mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r)) - \varepsilon,$$

and (iii) follows.

Finally, suppose that $\mathcal{A}$ is an FPAUS for $\mathcal{B}(\boldsymbol{d}, r)$ and (1.2) holds for some $c_0 \in (0, 1-\varepsilon)$. It follows from (ii) and (iii) that the probability that HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ performs more than $\lceil 2(1-c_0-\varepsilon)^{-1}\rceil$ iterations of $\mathcal{A}$ is at most $e^{-2} \leq \frac{1}{4}$. Therefore, terminating HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$ after this many calls to $\mathcal{A}$ gives an FPAUS for $\mathcal{H}_r(\boldsymbol{d})$. To achieve a total variation of $\varepsilon$ from HypergraphSampling$(\boldsymbol{d}, r, \mathcal{A})$, the algorithm $\mathcal{A}$ should be given input $\varepsilon' = 2\varepsilon\,(1-c_0)/3$, by (i). ☐

A general approach for bounding $\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r))$ is given by the following lemma. The constant $c_1$ in Lemma 3.1 captures the maximum edge probability relative to the uniform case (in which every neighborhood is equally likely). If $c_1$ is large then some neighbourhoods are much more likely under $\sigma_{\mathcal{B}}$ than they would be under the uniform distribution. The extent to which the events "$\mathcal{N}_B(y) = \mathcal{W}$" are negatively correlated, as $y$ varies over $Y$ for fixed $\mathcal{W} \subseteq X$, is described by the constant $c_2$. Intuitively, if the degree sequence is close to regular then we expect both $c_1$ and $c_2$ to be close to one.

**Lemma 3.1.** *Suppose that $\boldsymbol{d}$ is a sequence of nonnegative integers such that $\mathcal{B}(\boldsymbol{d}, r)$ is non-empty, and that $B = (X \cup Y, A) \in \mathcal{B}(\boldsymbol{d}, r)$ is a random bipartite graph according to the uniform distribution $\pi_{\mathcal{B}}$. Then, suppose that there are constants $c_1$ and $c_2$ such that for any $y, y' \in Y$ and any subset $\mathcal{W} \subseteq X$ of size $r$,*

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y) = \mathcal{W}) \leq c_1 \cdot \binom{n}{r}^{-1}$$

*and*

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y') = \mathcal{W} \mid \mathcal{N}_B(y) = \mathcal{W}) \leq c_2 \cdot \mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y') = \mathcal{W}).$$

*Then*

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r)) \geq 1 - c_1 c_2 \binom{m}{2} \binom{n}{r}^{-1}.$$

*Proof.* Let $B$ be an element from the set $\mathcal{B}(\boldsymbol{d}, r)$ drawn uniformly at random, and let $X^{\{r\}}$ be the set of all $r$-subsets of $X$. We omit the subscript $\pi_{\mathcal{B}}$ on all following probabilities. For $1 \leq k < \ell \leq m$, we define the random variable

$$Z_{k\ell} = \begin{cases} 1 & \text{if } \mathcal{N}_B(y_k) = \mathcal{N}_B(y_\ell) \\ 0 & \text{otherwise,} \end{cases}$$

for $y_k, y_\ell \in Y$, and we let

$$Z = \sum_{1 \leq k < \ell \leq m} Z_{k\ell}$$

be the random variable denoting the number of pairs of nodes $(y_k, y_\ell)$ that have the same neighborhood in $B$. Note that

$$\mathbb{P}\left(B \in \mathcal{B}^*(\boldsymbol{d}, r)\right) = 1 - \mathbb{P}(Z \geq 1). \tag{3.1}$$

Using the union bound over all possible pairs of nodes $(y_k, y_\ell)$,

$$\mathbb{P}(Z \geq 1) \leq \sum_{1 \leq k < \ell \leq m} \mathbb{P}(Z_{k\ell} = 1).$$

Likewise, for a fixed pair of nodes $(y_k, y_\ell)$, applying the union bound over $X^{\{r\}}$ shows us that

$$\mathbb{P}(Z_{k\ell} = 1) \leq \sum_{\mathcal{W} \in X^{\{r\}}} \mathbb{P}(\mathcal{N}_B(y_k) = \mathcal{N}_B(y_\ell) = \mathcal{W}).$$

Now, using the law of total probability,

$$\begin{aligned}
\mathbb{P}(Z \geq 1) &\leq \sum_{1 \leq k < \ell \leq m} \sum_{\mathcal{W} \in X^{\{r\}}} \mathbb{P}(\mathcal{N}_B(y_k) = \mathcal{N}_B(y_\ell) = \mathcal{W}) \\
&= \sum_{1 \leq k < \ell \leq m} \sum_{\mathcal{W} \in X^{\{r\}}} \mathbb{P}(\mathcal{N}_B(y_k) = \mathcal{W} \mid \mathcal{N}_B(y_\ell) = \mathcal{W}) \cdot \mathbb{P}(\mathcal{N}_B(y_\ell) = \mathcal{W}) \\
&\leq c_1 c_2 \binom{m}{2} \binom{n}{r}^{-1}. \qquad (3.2)
\end{aligned}$$

The proof is completed by combining (3.1) and (3.2). □

# 4 Probability of H-simplicity

The expected running time of HypergraphSampling is governed by the runtime of algorithm $\mathcal{A}$ and the probability that a randomly chosen element of $\mathcal{B}(\boldsymbol{d}, r)$ is H-simple. In Section 4.1 we provide an asymptotic estimate which holds when $\boldsymbol{d}$ is irregular and sparse. In Section 4.2 we give a combinatorial argument for the case of $d$-regular $r$-uniform hypergraphs. In particular, these sections yield Theorem 1.5 and Theorem 1.3, respectively.

## 4.1 Irregular, sparse degrees

In this section we prove a lower bound on the probability that a uniformly random graph from $\mathcal{B}(\boldsymbol{d}, r)$ is H-simple, using an asymptotic formula for irregular, sparse bipartite graphs. Given a bipartite degree sequence $(\boldsymbol{d}, \boldsymbol{r})$, define $M_k = \sum_{i=1}^n (d_i)_k$ and $R_k = \sum_{j=1}^m (r_j)_k$, where $(a)_b = a(a-1)\cdots(a-b+1)$ denotes the falling factorial. Let $R = R_1$ and $M = M_1$, and note that $R = M$ for any graphical bipartite degree sequence. We also let $d_{\max} = \max_i d_i$ and $r_{\max} = \max_j r_j$.

The following asymptotic enumeration result is a simpler, but weaker restatement of the main theorem from [18], which is slightly stronger than that of McKay [23]. (It follows from Theorem 4.1 that $\mathcal{B}(\boldsymbol{d}, \boldsymbol{r}) \neq \emptyset$ when $r_{\max} d_{\max} = o(M^{2/3})$.)

**Theorem 4.1.** ([18, Theorem 1.3]) *Suppose that $M \to \infty$, and that for $\boldsymbol{d} = (d_1, \ldots, d_n)$, $\boldsymbol{r} = (r_1, \ldots, r_m)$ are sequences of nonnegative integer functions of $M$ which both sum to $M$. If $r_{\max} d_{\max} = o(M^{2/3})$ then*

$$|\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})| = \frac{M!}{\prod_{i=1}^n d_i! \prod_{j=1}^m r_j!} \exp\left( -\frac{M_2 R_2}{2M^2} + O\left( \frac{d_{\max}^2 r_{\max}^2}{M} \right) \right).$$

Using this enumeration result, we can prove the main result of this section. First some useful identities which will be used in the proof without further comment: if $|\eta| < 1$ then $\exp(\eta) = 1 + O(\eta)$ and $(1+\eta)^{-1} = 1 + O(\eta)$. Also observe that if $r^2 = o(M)$ then

$$\frac{M^r}{(M)_r} \le \frac{M^r}{(M-r)^r} = \big(1 - r/M\big)^{-r} = \exp\big(O(r^2/M)\big) = 1 + O(r^2/M). \tag{4.1}$$

**Theorem 4.2.** *Assume that for each positive integer $n$ we have an integer $r = r(n) \ge 3$ and a sequence $\boldsymbol{d} = \boldsymbol{d}(n) = (d_1, \dots, d_n)$ of positive integers such that $M = \sum_{i=1}^n d_i$ tends to infinity with $n$. Assume that $r^2 d_{\max}^2 = o(M)$ and let $\pi_{\mathcal{B}}$ be the uniform distribution on $\mathcal{B}(\boldsymbol{d}, r)$. Then*

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r)) \ge 1 - \frac{n^r\, d_{\max}^r}{M^r} \cdot \binom{M/r}{2} \binom{n}{r}^{-1} \cdot (1 + o(1)).$$

*Proof.* Let $m = M/r$, which by assumption is an integer. Then, suppose that we have some neighbourhood $\mathcal{W} \in X^{\{r\}}$, and two integers $k, \ell \in [m]$. We will prove the desired result by conditioning on the neighbourhoods of $y_k$ and $y_\ell$ being equal to $\mathcal{W}$, at which point we can apply Lemma 3.1. To do this, we first define two bipartite degree sequences $(\boldsymbol{d}', \boldsymbol{r}')$ and $(\boldsymbol{d}'', \boldsymbol{r}'')$, as follows:

$$d_i' = \begin{cases} d_i - 1 & \text{if } x_i \in \mathcal{W}, \\ d_i & \text{if } x_i \in X \setminus \mathcal{W}, \end{cases} \qquad \text{and} \quad r_j' = \begin{cases} 0 & \text{if } j = k, \\ r & \text{if } j \in [m] \setminus \{k\}, \end{cases}$$

and

$$d_i'' = \begin{cases} d_i - 2 & \text{if } x_i \in \mathcal{W}, \\ d_i & \text{if } x_i \in X \setminus \mathcal{W}, \end{cases} \qquad \text{and} \quad r_j'' = \begin{cases} 0 & \text{if } j \in \{k, \ell\}, \\ r & \text{if } j \in [m] \setminus \{k, \ell\}. \end{cases}$$

We also extend the notation for $M$ and $M_2$ to $\boldsymbol{d}'$ and $\boldsymbol{d}''$ by appending one or two dashes, and likewise for $R$ and $R_2$. By assumption, $r^2 d_{\max}^2 = o(M)$, which implies that $r d_{\max} = o(M^{1/2})$. Hence, the conditions of Theorem 4.1 are satisfied, and we can approximate both $|\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})|$ and $|\mathcal{B}(\boldsymbol{d}', \boldsymbol{r}')|$. Considering the ratio of these, since $d_i' = d_i$ whenever $x_i \notin \mathcal{W}$, and $r_j' = r_j$ whenever $j \neq k$, many terms cancel, leading to

$$\frac{|\mathcal{B}(\boldsymbol{d}', \boldsymbol{r}')|}{|\mathcal{B}(\boldsymbol{d}, r)|} = \frac{r!}{(M)_r} \cdot \prod_{x_i \in \mathcal{W}} d_i \cdot \exp\left( \frac{M_2 R_2}{2M^2} - \frac{M_2' R_2'}{2(M')^2} + O(r_{\max}^2 d_{\max}^2 / M) \right). \tag{4.2}$$

Next, by definition of $M_2'$, we see that

$$M_2' = M_2 - \sum_{x_i \in \mathcal{W}} (d_i)_2 + \sum_{x_i \in \mathcal{W}} (d_i - 1)_2 = M_2 - 2 \sum_{x_i \in \mathcal{W}} (d_i - 1) = M_2 \left(1 - O\left(r d_{\max} M_2^{-1}\right)\right).$$

Similarly,

$$M' = M - r = M\left(1 - O\left(r M^{-1}\right)\right), \qquad R_2' = (r-1)\, M' = R_2 \left(1 - O\left(r M^{-1}\right)\right).$$

Then

$$\frac{M_2 R_2}{2M^2} - \frac{M_2' R_2'}{2(M')^2} \le \frac{M_2 R_2}{2M^2} - \frac{M_2 R_2}{2M^2} \cdot \left(1 - O\big(r d_{\max} M_2^{-1} + r M^{-1}\big)\right)$$

$$
\begin{aligned}
&= O\big(rd_{\max}R_2M^{-2} + rM_2R_2M^{-3}\big)\\
&= O\bigg(\frac{r^2d_{\max}}{M}\bigg).
\end{aligned}
$$

The final equality follows as $M_2 \le d_{\max}M$ and $R_2 = (r-1)M$. Therefore, combining the above identities with (4.1) and (4.2) implies that

$$
\frac{|\mathcal{B}(\boldsymbol{d}',\boldsymbol{r}')|}{|\mathcal{B}(\boldsymbol{d},\boldsymbol{r})|} = \frac{r!}{M^r}\cdot\prod_{x_i\in\mathcal{W}} d_i \cdot \big(1 + O(r^2d_{\max}^2/M)\big). \tag{4.3}
$$

Next, observe that there is a bijective relationship between bipartite graphs $B_0 \in \mathcal{B}(\boldsymbol{d}',\boldsymbol{r}')$, and bipartite graphs $B \in \mathcal{B}(\boldsymbol{d},\boldsymbol{r})$ such that $\mathcal{N}_B(y_k) = \mathcal{W}$, using the map $B \mapsto B_0 = B \setminus \{y_k\}$ which deletes vertex $y_k$ and reduces the degrees of each neighbour of $y_k$ by 1. Hence,

$$
\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_k) = \mathcal{W}) = \frac{|\mathcal{B}(\boldsymbol{d}',\boldsymbol{r}')|}{|\mathcal{B}(\boldsymbol{d},\boldsymbol{r})|}.
$$

By assumption, $r^2d_{\max}^2 = o(M)$, so using (4.3) we find that

$$
\binom{n}{r}\cdot\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_k) = \mathcal{W}) \le \frac{n^r d_{\max}^r}{M^r}\cdot(1+o(1))\,. \tag{4.4}
$$

For future use, observe that (4.4) also holds with $y_k$ replaced by any $y \in Y$.

A similar result holds for the conditional edge probability from Lemma 3.1. First observe that if $\boldsymbol{d}$ and $\boldsymbol{r}$ satisfy the conditions of Theorem 4.1 then so do $\boldsymbol{d}''$ and $\boldsymbol{r}''$. Hence, the same argument that led to (4.3) gives

$$
\begin{aligned}
\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_\ell) = \mathcal{W} \mid \mathcal{N}_B(y_k) = \mathcal{W}) &= \frac{|\mathcal{B}(\boldsymbol{d}'',\boldsymbol{r}'')|}{|\mathcal{B}(\boldsymbol{d}',\boldsymbol{r}')|}\\
&= \frac{r!}{(M-r)^r}\cdot\prod_{x_i\in\mathcal{W}}(d_i-1)\cdot\big(1+O(r^2d_{\max}^2/M)\big).
\end{aligned}
$$

We will divide this expression by the result of replacing $y_k$ with $y_\ell$ in (4.4), to obtain

$$
\begin{aligned}
\frac{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_\ell) = \mathcal{W} \mid \mathcal{N}_B(y_k) = \mathcal{W})}{\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_\ell) = \mathcal{W})} &= \frac{(M)_r}{(M-r)_r}\cdot\prod_{x_i\in\mathcal{W}}\bigg(1-\frac{1}{d_i}\bigg)\cdot\big(1+O(r^2d_{\max}^2/M)\big)\\
&\le \frac{(M)_r}{(M-r)_r}\cdot\big(1+O(r^2d_{\max}^2/M)\big)\\
&= 1+o(1),
\end{aligned}
$$

using (4.1) and the assumption that $r^2d_{\max}^2 = o(M)$. From the above inequality and (4.4), we can apply Lemma 3.1 with

$$
c_1 = \frac{n^r d_{\max}^r}{M^r}\cdot(1+o(1)) \qquad \text{and} \qquad c_2 = 1+o(1),
$$

to complete the proof. □

## 4.2 Regular degrees

In this section we present a combinatorial argument to establish a lower bound on the probability that a uniformly random graph from $\mathcal{B}(\boldsymbol{d}, r)$ is H-simple, when $\boldsymbol{d} = (d, d, \ldots, d)$ is regular. We first prove a 'sensitivity result' for the set of all bipartite graphs with given degrees. We show that adjusting the degrees on one side of the bipartition, to make them closer to regular, can only increase the number of bipartite graphs. It is possible that this result is known, though we could not find it in the literature. We give a proof in Section 4.2.1 for completeness.

**Proposition 4.3.** *Let $n, m \in \mathbb{N}$ and let $(\boldsymbol{d}, \boldsymbol{r})$ be a bipartite degree sequence for the bipartition $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$. Suppose that we have integers $k, \ell \in [n]$ such that $d_k \geq d_\ell + 2$ and define $\boldsymbol{d}'$ by*

$$d'_i = \begin{cases} d_i - 1 & \text{if } i = k \\ d_i + 1 & \text{if } i = \ell \\ d_i & \text{if } i \in [n] \setminus \{k, \ell\}. \end{cases}$$

*Then*

$$|\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})| \leq |\mathcal{B}(\boldsymbol{d}', \boldsymbol{r})|.$$

Using this proposition, we now prove Theorem 1.3.

*Proof of Theorem 1.3.* Let $B = (X \cup Y, A) \in \mathcal{B}(n, d, r)$ so that all nodes in $X$ are $d$-regular and all nodes in $Y$ are $r$-regular. Throughout the proof we let $\mathcal{W} \in X^{\{r\}}$ be a fixed neighbourhood of size $r$, and we consider two fixed nodes $y_k, y_\ell \in Y$.

We will prove the desired result by conditioning on the neighbourhoods of $y_k$ and $y_\ell$ being equal to $\mathcal{W}$, at which point we can apply Lemma 3.1. To do this, let $\mathcal{U} \in X^{\{r\}}$ be any $r$-subset of $X$. We will analyse

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_k) = \mathcal{U} \ \mid \ \mathcal{N}_B(y_\ell) = \mathcal{W}). \tag{4.5}$$

Our goal will be to show that (4.5) achieves a minimum at $\mathcal{U} = \mathcal{W}$. Let $\triangle$ denote the symmetric difference operator. Given $\mathcal{W} = \mathcal{N}_B(y_\ell)$, for any subset $\mathcal{U} \subseteq X^{\{r\}}$ of size $r$, we define a new bipartite degree sequence $(\boldsymbol{d}^{\mathcal{U}}, \boldsymbol{r}^{\mathcal{U}})$ by

$$d_i^{\mathcal{U}} = \begin{cases} d_i - 2 & \text{if } x_i \in \mathcal{U} \cap \mathcal{W} \\ d_i - 1 & \text{if } x_i \in \mathcal{U} \triangle \mathcal{W} \\ d_i & \text{if } x_i \in X \setminus (\mathcal{U} \cup \mathcal{W}) \end{cases} \qquad \text{and} \qquad r_j^{\mathcal{U}} = \begin{cases} 0 & \text{if } y_j \in \{y_k, y_\ell\} \\ r & \text{if } y_j \in Y \setminus \{y_k, y_\ell\} \end{cases}.$$

Now, when $\mathcal{U} \in X^{\{r\}}$ and $|\mathcal{U} \triangle \mathcal{W}| > 0$, we can select a node $x_k \in \mathcal{U} \setminus \mathcal{W}$ and $x_\ell \in \mathcal{W} \setminus \mathcal{U}$, and create a new neighbourhood $\mathcal{U}' = (\mathcal{U} \cup \{x_\ell\}) \setminus \{x_k\}$. Then, we see that $d^{\mathcal{U}'}$ is equal to $d^{\mathcal{U}}$ whenever $i \notin \{k, \ell\}$, and that $d^{\mathcal{U}}$ is a more locally balanced degree sequence than $d^{\mathcal{U}'}$. Hence, $d^{\mathcal{U}'}$ and $d^{\mathcal{U}}$ satisfy the conditions of Proposition 4.3, and applying Proposition 4.3 we conclude that

$$\left|\mathcal{B}(\boldsymbol{d}^{\mathcal{U}'}, \boldsymbol{r}^{\mathcal{U}'})\right| \leq \left|\mathcal{B}(\boldsymbol{d}^{\mathcal{U}}, \boldsymbol{r}^{\mathcal{U}})\right|.$$

Since $|\mathcal{U} \triangle \mathcal{W}| > 0$ is at most $2r$, we can repeat the above process a finite number of times to conclude that for any $\mathcal{U} \in X^{\{r\}}$,

$$\big|\mathcal{B}(\boldsymbol{d}^{\mathcal{W}}, \boldsymbol{r}^{\mathcal{W}})\big| \le \big|\mathcal{B}(\boldsymbol{d}^{\mathcal{U}}, \boldsymbol{r}^{\mathcal{U}})\big|. \tag{4.6}$$

Since

$$\frac{\big|\mathcal{B}(\boldsymbol{d}^{\mathcal{U}}, \boldsymbol{r}^{\mathcal{U}})\big|}{|\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})|} = \mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_\ell) = \mathcal{W}) \cdot \mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_k) = \mathcal{U} \ \mid\ \mathcal{N}_B(y_\ell) = \mathcal{W}, )$$

the inequality in (4.6) implies that for any $\mathcal{U}, \mathcal{W} \in X^{\{r\}}$, we have

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_k) = \mathcal{W} \ \mid\ \mathcal{N}_B(y_\ell) = \mathcal{W}) \le \mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_k) = \mathcal{U} \ \mid\ \mathcal{N}_B(y_\ell) = \mathcal{W}). \tag{4.7}$$

But $X^{\{r\}}$ has size $\binom{n}{r}$, so summing over all possible choices for $\mathcal{U} \in X^{\{r\}}$ in (4.7) shows us that (since the probabilities on the right must sum to unity)

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_k) = \mathcal{W} \ \mid\ \mathcal{N}_B(y_\ell) = \mathcal{W}) \le \binom{n}{r}^{-1}.$$

Since $\boldsymbol{d}$ and $\boldsymbol{r}$ are both regular, by symmetry every possible $\mathcal{W} \in X^{\{r\}}$ is equally likely; hence

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{N}_B(y_\ell) = \mathcal{W}) = \binom{n}{r}^{-1}.$$

Thus we can apply Lemma 3.1 with $c_1 = c_2 = 1$ to conclude that

$$\mathbb{P}_{\pi_{\mathcal{B}}}(\mathcal{B}^*(\boldsymbol{d}, r)) \ge 1 - \binom{m}{2}\binom{n}{r}^{-1}.$$

This completes the proof. □

### 4.2.1 Sensitivity result for bipartite degree sequences

In this section we prove Proposition 4.3.

*Proof of Proposition 4.3.* We first define an equivalence relation on the graphical realizations $B \in \mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$. Let

$$L_B = \{y \in Y : \{x_k, y\}, \{x_\ell, y\} \in E(B)\}, \quad \text{that is,} \quad L_B = \mathcal{N}_B(x_k) \cap \mathcal{N}_B(x_\ell).$$

We say that two bipartite graphs $B, B'$ are equivalent, denoted by $B \sim B'$, if $\mathcal{N}_B(x_i) = \mathcal{N}_{B'}(x_i)$ for all $x_i \in X \setminus \{x_k, x_\ell\}$ and $L_B = L_{B'}$. One interpretation of the equivalence class of $B$ is as follows. We first write $B = D + B_0$ where $B_0$ contains precisely the edges

$$\bigcup_{i=k,\ell} \{\{x_i, y\} \in E(B) : y \in Y \setminus L_B\},$$

that is, all the edges adjacent to $x_k$ or $x_\ell$ for which the endpoint in $Y$ is not connected to both $x_k$ and $x_\ell$. We emphasize that $x_k$ and $x_\ell$ are fixed throughout the proof. The so-called *base graph* $D$ then contains the remaining edges. The equivalence class $S_B \subseteq \mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$ of $B$ is now given by $B$ itself and all graphs $B'$ of the form

$$B' = D + B'_0$$

where $B'_0$ has the same degree sequence as $B_0$. Note that the degree of $x_k$ in $B_0$ equals $\deg_{B_0}(x_k) = d_k - |L_B|$ and $\deg_{B_0}(x_\ell) = d_\ell - |L_B|$. All nodes in $B_0$ with non-zero degree in $Y$ have degree 1.

**Example 4.4.** Consider the graph $B$ in Figure 1 with degree sequence $\boldsymbol{d} = (1, 3, 5, 2)$ and $\boldsymbol{r} = (3, 2, 1, 1, 2, 2)$. The decomposition into $D + B_0$ is given in Figure 2. Here we have $L_B = \{1, 2\}$.

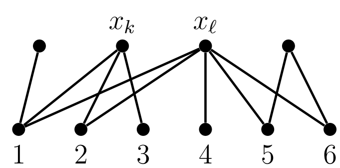


Figure 1: The graph $B$.


Figure 2: The graphs $D$ (left) and $B_0$ (right) for $B$ as in Figure 1.

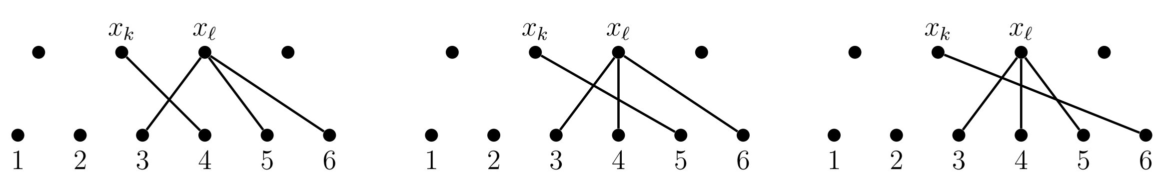


Figure 3: The three choices of $B'_0$ for $B_0$ as in Figure 2.

There are three choices of $B'_0$ in this case, which are given in Figure 3 (depending on which node gets connected to $x_k$). This means that $|S_B| = \binom{4}{1} = 4$. ◀

We next make some simple observations. Again remember that the definition of the base graph and the equivalence class is depending on the fixed choice for $k$ and $\ell$.

**Observation 4.5.** *Given a bipartite degree sequence $(\boldsymbol{d}, \boldsymbol{r})$ and base graph $D$ of some $B \in \mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, we can uniquely determine the equivalence class of the graph $B$ (without knowing $B$). Moreover, given an equivalence class $S$, we can find the common base graph $D$ of all its elements by taking the intersection of all graphs in $S$.*

*Proof.* Given the base graph $D$ and sequence $(\boldsymbol{d}, \boldsymbol{r})$, we can uniquely determine the bipartite degree sequence of $B_0$ (but not $B_0$ itself), and precisely for all graphs $B_0'$ with this bipartite degree sequence, $D + B_0'$ is an element of $S_B$. The second claim follows from the fact that for any two graphs $B, B' \in S$, we have $L_B = L_{B'}$ and $\mathcal{N}_B(x_i) = \mathcal{N}_{B'}(x_i)$ for all $i \in [n] \setminus \{k, \ell\}$, which implies that $D \subseteq \cap_{B \in S} B$. To see equality, note that if some edge $\{x_k, y\}$ or $\{x_\ell, y\}$, with $y \notin L_B$, is contained in some $B \in S$, there is always a realization $B' \in S$ that does not contain it. $\square$

Because of Observation 4.5, we will label an equivalence class $S$ according to its base graph, and write $S = S(D)$. Moreover, we write $|L_{S(D)}|$ for the common value of $|L_B|$ for all $B \in S(D)$.

**Observation 4.6.** *If $D$ is a base graph for some $B \in \mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, then it is also a base graph for some $B \in \mathcal{B}(\boldsymbol{d}', \boldsymbol{r})$.*

*Proof.* This follows from the fact that, for a given $B = D + B_0 \in \mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, we can remove an edge of the form $\{x_k, y\}$ from $B_0$, and add the edge $\{x_\ell, y\}$ to $B_0$. An edge of the form $\{x_k, y\}$ always exists, as $d_k \geq d_\ell + 2$. The edge $\{x_\ell, y\}$ does not exist in $B$ by construction. To see this, note that if also $\{x_\ell, y\} \in E(B)$, then we would have had $y \in L_B$, which implies that $\deg_{B_0}(y) = 0$, as $y$ would then have been connected to both $x_k$ and $x_\ell$. It follows directly that $B' = B - \{x_k, y\} + \{x_\ell, y\} \in \mathcal{B}(\boldsymbol{d}', \boldsymbol{r})$. In particular, $B'$ has the same base graph as $B$. $\square$

Now, in order to show that $|\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})| \leq |\mathcal{B}(\boldsymbol{d}', \boldsymbol{r})|$, we consider a fixed base graph $D$ of some graph $B \in \mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, and its equivalence class $S(D)$. Let $S'(D)$ be the equivalence class in $\mathcal{B}(\boldsymbol{d}', \boldsymbol{r})$ of the base graph $D$.

**Example 4.4 (continued).** For the base graph $D$ as in Figure 2, the equivalence class $S'(D) \subseteq \mathcal{B}(\boldsymbol{d}', \boldsymbol{r})$, with $\boldsymbol{d}' = (1, 4, 4, 2)$ and $\boldsymbol{r} = (3, 2, 1, 1, 2, 2)$, is given by the graphs in Figure 4. Note that $|S'(D)| = \binom{4}{2} = 6$. ◀

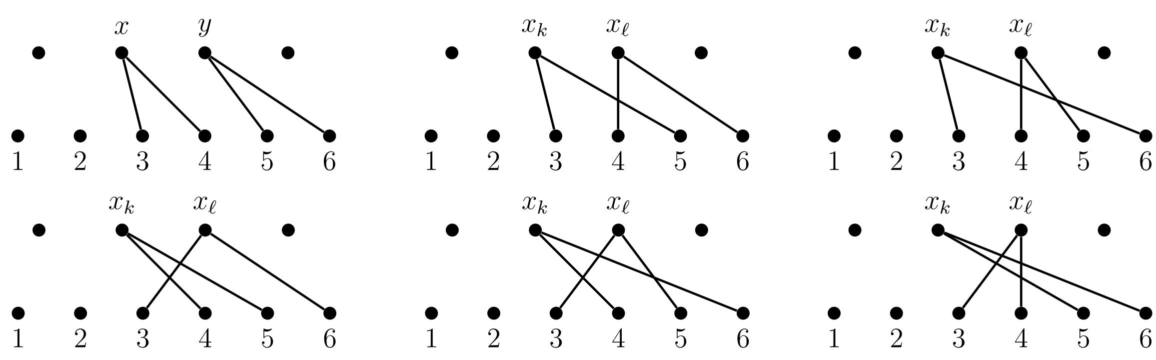


Figure 4: The six graphs $(B_0')'$ so that $D + (B_0')' \in S'(D)$ with $D$ as in Figure 2.

For any base graph $D$ corresponding to some equivalence class $S(D) \subseteq \mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$, we can enumerate the equivalence class by a simple combinatorial argument. Recall that $d_k - |L_{S(D)}|$

and $d_\ell - |L_{S(D)}|$ are the degrees of $x_k$ and $x_\ell$ respectively, for any $B_0$ with $D + B_0 \in S(D)$. Then

$$\begin{aligned} |S(D)| &= \binom{d_k + d_\ell - 2|L_{S(D)}|}{d_\ell - |L_{S(D)}|} \\ &\le \binom{d_k + d_\ell - 2|L_{S(D)}|}{d_\ell + 1 - |L_{S(D)}|} \\ &= \binom{(d_k - 1) + (d_\ell + 1) - 2|L_{S(D)}|}{(d_\ell + 1) - |L_{S(D)}|} \\ &= \binom{d'_k + d'_\ell - 2|L_{S'(D)}|}{d'_\ell - |L_{S'(D)}|} \\ &= |S'(D)|. \end{aligned}$$

To produce the inequality, we have used that

$$\binom{\alpha + \beta}{\beta} \le \binom{\alpha + \beta}{\beta + 1}$$

for any $\alpha, \beta \in \mathbb{N}$ such that $\alpha \ge \beta + 2$. The given constants satisfy this inequality by assumption (see statement of Proposition 4.3). The penultimate equality follows from the fact that $|L_{S(D)}| = |L_{S'(D)}|$. It then follows that

$$|\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})| = \sum_{S(D)} |S(D)| \le \sum_{S'(D)} |S'(D)| = |\mathcal{B}(\boldsymbol{d}', \boldsymbol{r})|$$

because of Observation 4.6 and the fact that the mapping that sends $S(D)$ to $S'(D)$ is injective. This last claim can be seen from the fact that given an equivalence class $S'$, there is a unique $D$ so that $S' = S'(D)$, which is given by the intersection of all graphs in $S'$, as shown in Observation 4.5. This completes the proof of Proposition 4.3. $\square$

## Acknowledgements

Research was supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation Programme Networks 024.002.003. Martin Dyer is supported by the EPSRC research grant EP/S016562/1 "Sampling in hereditary classes". Catherine Greenhill is supported by the Australian Research Council Discovery Project DP190100977.

# References

[1] G. Amanatidis and P. Kleer, Rapid mixing of the switch Markov chain for strongly stable degree sequences and 2-class joint degree matrices, in Proceedings of the 30th Annual ACM–SIAM Symposium on Discrete Algorithms, 2019, ACM–SIAM, New York–Philadelphia, pp. 966–985.

[2] N. A. Arafat, D. Basu, L. Decreusefond and S. Bressan, Construction and random generation of hypergraphs with prescribed degree and dimension sequences. Preprint (2020). arXiv:2004.05429

[3] A. Arman, P. Gao and N. Wormald, Fast Uniform Generation of Random Graphs with Given Degree Sequences, Proceedings of IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS), Baltimore, MD, USA, 2019, pp. 1371-1379.

[4] M. Bayati, J.-H. Kim and A. Saberi, A sequential algorithm for generating random graphs, *Algorithmica* **58** (2010), 860–910.

[5] I. Bezáková, N. Bhatnagar, E. Vigoda, Sampling binary contingency tables with a greedy start, *Random Structures & Algorithms* **30(1-2)** (2007), 168–205.

[6] V. Blinovsky and C. Greenhill, Asymptotic enumeration of sparse uniform linear hypergraphs with given degrees, *European Journal of Combinatorics* **63** (2017), 6–25.

[7] C. J. Carstens and P. Kleer, Speeding up switch Markov chains for sampling bipartite graphs with given degree sequence, *Leibniz International Proceedings in Informatics (LIPIcs)* **116** (2018), 36:1–36:18.

[8] Y. Chen, P. Diaconis, S.P. Holmes and J.S. Liu. Sequential Monte Carlo methods for statistical analysis of tables, *Journal of the American Statistical Association* **100** (2005), 109–120.

[9] P. S. Chodrow, Configuration models of random hypergraphs. Preprint (2019). arXiv:1902.09302

[10] C. Cooper, M. Dyer and C. Greenhill. Sampling regular graphs and a peer-to-peer network. *Combinatorics, Probability and Computing* **16** 557–593, (2007).

[11] C. Cooper, M.E. Dyer and C. Greenhill, Corrigendum: Sampling regular graphs and a peer-to-peer network. arXiv:1203.6111

[12] C. Cooper, A. Frieze, M. Molloy and B. Reed, Perfect matchings in random $r$-regular, $s$-uniform hypergraphs, *Electronic Journal of Combinatorics* **7** (2000), #R57.

[13] A. Deza, A. Levin, S.M. Meesum and S. Onn, Hypergraphic degree sequences are hard, *Bulletin of the EATCS* **127** (2019).

[14] P. Diaconis and A. Gangolli, Rectangular arrays with fixed margins, in D. Aldous, P. Diaconis, J. Spencer and J. M. Steele (eds.), *Discrete Probability and Algorithms*, Springer–Verlag, New York, 1995, pp. 15–41.

[15] A. Dudek, A. Frieze, A. Ruciński, M. Šileikis, Approximate counting of regular hypergraphs, *Information Processing Letters* **113(19-21)** (2013), 785–788

[16] P.L. Erdős, C. Greenhill, T.R. Mezei, I. Miklos, D. Soltész, L. Soukup, The mixing time of the swap (switch) Markov chains: a unified approach. Preprint (2019). arXiv:1903.06600

[17] P. Gao and N. Wormald, Uniform generation of random regular graphs, *SIAM J. Comput.*, 46(4): 1395–1427.

[18] C. Greenhill, B.D. McKay and X. Wang, Asymptotic enumeration of sparse 0-1 matrices with irregular row and colum sums, *Journal of Combinatorial Theory (Series A)* **113** (2006), 291–324.

[19] C. Greenhill and M. Sfragara, The switch Markov chain for sampling irregular graphs and digraphs, *Theoretical Computer Science* **719** (2018), 1–20.

[20] M. Jerrum and A. Sinclair, Fast uniform generation of regular graphs, *Theoretical Computer Science* **73** (1990), 91–100.

[21] M. Jerrum, A. Sinclair and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries, *Journal of the ACM* **51** (2004), 671–697.

[22] R. Kannan, P. Tetali and S. Vempala, Simple Markov chain algorithms for generating random bipartite graphs and tournaments, *Random Structures & Algorithms* **14** (1999), 293–308.

[23] B. D. McKay, Asymptotics for 0-1 matrices with prescribed line sums, in *Enumeration and Design*, Academic Press, Canada, 1984, pp. 225–238.

[24] B.D. McKay and N.C. Wormald, Uniform generation of random regular graphs of moderate degree, *Journal of Algorithms* **11** (1990), 52–67.

[25] I. Miklos, P. L. Erdős and L. Soukup, Towards random uniform sampling of bipartite graphs with given degree sequence, *Electronic Journal of Combinatorics* **20(1)** (2013), #P16.

[26] R. Patro, C. Kingsford. Predicting protein interactions via parsimonious network history inference. *Bioinformatics* **29(10-11)** (2013), 237–246.

[27] N. D. Verhelst, An efficient MCMC algorithm to sample binary matrices with fixed marginals, *Psychometrika* **73(4)** (2008), 705.

[28] D. Zhou, J. Huang, B. Schölkopf, Learning with hypergraphs: Clustering, classification, and embedding. *Proceeding of Advances in neural information processing systems* (2007), 1601–1608.

# A Mixing time bounds for the switch Markov chain

The switch Markov chain was analysed by Cooper, Dyer and Greenhill [10, 11] for regular graphs, and by Greenhill and Sfragara [19] for irregular graphs that are relatively sparse. Some situations which lead to additional factors in the mixing time bounds from [10, 11, 19] do not arise in bipartite graphs, and so it is possible to improve the mixing time bounds for bipartite graphs. To the best of our knowledge, these bounds have not been presented elsewhere, so we write them down here. The proofs from [10, 11, 19] use the multicommodity flow method, and are quite long and technical. We do not give full details, but rather explain how the proofs from [10, 11, 19] can be adapted to the bipartite setting, and give the resulting mixing time bounds. For all notation that is not defined, and all other missing details, we refer the reader to [10, 11, 19].

We begin with regular bipartite graphs, where there are $n$ nodes in each side of the bipartition and all nodes have degree $d$. (We stress that this case is not particularly relevant for the problem of sampling hypergraphs, unless the hypergraph is $d$-regular and $d$-uniform.) For regular bipartite graphs, the bound on the bipartite switch chain is a factor of $\frac{1}{32}\,d^6n^2$ smaller than the general (not-necessarily-bipartite) case. (The constant factor 32 in Theorem A.1 below arises since graphs in $\mathcal{B}(n,d,d)$ have $2n$ nodes.)

**Theorem A.1.** *Let $\mathcal{B}(n,d,d)$ be the set of $d$-regular bipartite graphs with $2n$ nodes and a given bipartition. Then the mixing time of the bipartite switch chain on $\mathcal{B}(n,d,d)$ satisfies*

$$\tau(\varepsilon) \le 32d^{17}n^6\big(2dn\log(2dn) + \log(\varepsilon^{-1})\big).$$

*Proof.* Temporarily, write $N = 2n$, for ease of comparison with [10, 11]. The flow can be defined in exactly the same way as in [10], though a shortcut edge will never be needed. (Every circuit decomposes into 1-circuits.) Hence there will be at most 3 defect edges, two labelled $-1$ and one labelled 2, all incident with the "start vertex" $x_0$ of the 1-circuit. At most two switches are needed to transform any encoding into a graph, and [10, Lemma 4] becomes $|\mathcal{L}(Z)| \le 2d^4N^3|\mathcal{B}(n,d,d)|$. (This is smaller by a factor of $d^2N^2$ than in the non-bipartite case, essentially because we save one (-1)-switch in the worst case, which costs $d^2N^2$.)

In the bipartite case, we save a factor of $d^4$ compared with [11, Lemma 1] (which is a correction of [10, Lemma 5]). This is because there can be at most 10 bad pairs in the yellow-green colouring, not 14. Recall that each bad pair contributes a factor of $d$. (There are at most 3 bad pairs from each defect edge, plus at most one additional bad pair from wrapping around at $x_0$. Alternatively, we no longer have a shortcut edge, which in [10, 11] was responsible for up to 4 bad pairs, so 14 goes down to 10.)

Combining these effects, we save a factor of $d^6N^2$ compared to the mixing time from [11, Theorem 1]. Note that we have $\ell(f) \le dN/2$ and, as there are $dn$ edges in any element of $\mathcal{B}(n,d,d)$,

$$1/Q(e) \le 4\binom{nd}{2}\,|\mathcal{B}(n,d,d)| \le \tfrac{1}{2}d^2N^2\,|\mathcal{B}(n,d,d)|,$$

which saves an additional factor of $\frac{1}{2}$ compared with [10, 11]. $\square$

Let $\Delta = \max\{d_{\max}, r_{\max}\}$. By adapting the analysis from [19], we can prove a bound in the irregular case which is a factor of $\Delta^4 M^2$ smaller than in the general (not-necessarily-bipartite) case.

**Theorem A.2.** *Let $\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})$ be the set of all bipartite graphs with a given node bipartition, degrees $\boldsymbol{d}$ on the left and degrees $\boldsymbol{r}$ on the right. Suppose that all degrees are at least 1 and that $3 \le d_{\max}, r_{\max} \le \frac{1}{3}\sqrt{M}$. Then the mixing time of the bipartite switch chain satisfies*

$$\tau(\varepsilon) \le \Delta^{10} M^7 \big(\tfrac{1}{2} M \log(M) + \log(\varepsilon^{-1})\big).$$

*Proof.* We have $1/Q(e) \le M^2\, |\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})|$ and $\ell(f) \le M/2$, as in [19, Theorem 1.1]. Arguing as above, the number of bad pairs is at most 10, not 14, saving a factor of $\Delta^4$. The main thing is the critical lemma [19, Lemma 2.5], where we give an upper bound on the number of encodings. We claim that

$$|\mathcal{L}^*(Z)| \le 2M^4 |\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})|$$

so long as $\Delta \le \frac{1}{3}\sqrt{M}$, say.

In [19, Lemma 2.5], we only performed a $(-1, 2)$-switching if we had four defect edges. This was to ensure that we definitely had a $(-1)$-defect edge incident with a 2-defect edge: but when there is no shortcut edge, this is already guaranteed when we have three defect edges. Therefore, letting

$$a = 2\Delta^2 M, \qquad b = 2\Delta^2, \qquad c = \tfrac{9}{8} M^2$$

be the upper bounds that we proved in [19, Lemma 2.5] on the various ratios, we obtain

$$|\mathcal{L}^*(Z)| \le (1 + b + c + bc + c^2 + ac)\, |\mathcal{B}(\boldsymbol{d}, \boldsymbol{r})|.$$

(The saving here is replacing $bc^2$ by $ac$, and in omitting the terms involving $b^2$ or $abc$ or $ac^2$, which were needed in [19] to deal with the shortcut edge.) Using the bounds $3 \le \Delta$ and $\Delta^2 \le M/9$, we see that

$$|\mathcal{L}^*(Z)| \le 2M^4.$$

This is a factor of $M^2$ smaller than the corresponding bound given in [19, Lemma 2.5], again because (in the worst case) we save a $(-1)$-switch, which gives a ratio of $\frac{9}{8}M^2$. Combining this with the earlier saving of $\Delta^4$, we obtain the stated bound. □

The following corollary is most relevant to sampling uniform hypergraphs with given degrees. It follows directly from Theorem A.2 by considering a regular sequence $\boldsymbol{r}$.

**Corollary A.3.** *Let $\mathcal{B}(\boldsymbol{d}, r)$ be the set of all bipartite graphs with a given node bipartition, degrees $\boldsymbol{d} = (d_1, \ldots, d_n)$ on one side and with $m$ nodes of degree $r$ on the other. Let $M = rm = \sum_{j=1}^{n} d_j$. Suppose that all degrees are at least 1 and that $3 \le d_{\max}, r \le \frac{1}{3}\sqrt{M}$. Then the mixing time of the bipartite switch chain satisfies*

$$\tau(\varepsilon) \le \Delta^{10} M^7 \big(\tfrac{1}{2} M \log(M) + \log(\varepsilon^{-1})\big)$$

*where $\Delta = \max\{d_{\max},\, r\}$.*