Uniform generation of d-factors in dense host graphs

Pu Gao[*] [†]
Max-Planck-Institut für Informatik
janegao@mpi-inf.mpg.de

Abstract

Given a constant integer $d \geq 1$ and a host graph H that is sufficiently dense, we lower bound the number of d-factors H contains. When the complement of H is sufficiently sparse, we provide an algorithm that uniformly generates the d-factors of H and we justify the efficiency of the algorithm.

1 Introduction

Enumeration and uniform generation of d-regular graphs on n vertices have been studied for decades. The earliest result on enumerating d-regular graphs for bounded integers d was obtained by Bender and Canfield [2]. The boundedness of d was relaxed by Bollobás [1], and then further by McKay [6], and McKay and Wormald [8]. The currently best result on enumerating sparse d-regular graphs is due to McKay and Wormald [8] for $d = o(n^{1/2})$. On the other hand, enumerating dense d-regular graphs (where $\min\{d, n-d\} \geq cn/\ln n$ for some $c > 0$) was achieved by McKay and Wormald [9].

The technique used to enumerate sparse d-regular graphs leads to algorithms that generate uniformly random d-regular graphs. For $d = O(n^{1/3})$, McKay and Wormald [7] used switchings which repeatedly switch off loops and double edges. These switching operations were a modification and refinement of those first introduced and applied by McKay in [6].

A natural generalisation of the above problems is to enumerate and uniformly generate d-factors of a given graph H_n on n vertices instead of the complete graph K_n. We call H_n the host graph and we let $\overline{H}_n$ denote the complement of H_n. A recent result [3] proves that every bridgeless cubic graph contains exponentially many perfect matchings. A more accurate asymptotic enumeration result was given by McKay when $\overline{H}_n$ is sufficiently sparse. To be specific, given the degree sequence of H_n, McKay [5, Theorem 2.3 (a)] estimated the asymptotic number of d-factors when $d = o(n^{1/3})$ and the maximum degree of $\overline{H}_n$ is $o(\sqrt{n/d})$. We haven't seen any result on uniform generation of d-factors so far. Jerrum and Sinclair [4] analysed an algorithm which generates near-uniform perfect and near-perfect matchings. The

[*]Research supported by the Humboldt Foundation
[†]Current affiliation: University of Toronto, Canada

algorithm is based on designing a Markov chain which was shown to mix rapidly and has uniform stationary distribution. The purpose of this paper is to provide a simple algorithm that uniformly generates d-factors of a given host graph H_n for bounded d, and justify its efficiency when the number of edges in $\overline{H}_n$ is $O(n)$ and the maximum degree of $\overline{H}_n$ is bounded by a certain constant times n. When $\overline{H}_n$ is denser, we also give asymptotic bounds on the number of d-factors of H_n. We do not attempt to achieve tight bounds but rather make an effort to weaken the conditions on H_n. For instance, our bounds require no information on the degree sequence of H_n and allow the maximum degree of $\overline{H}_n$ to be linear in n, whereas the number of edges in $\overline{H}_n$ can be proportional to n^2.

2 Main results

Let $(H_n)_{n\geq 1}$ be a sequence of graphs on n vertices. Let $e(H_n)$ denote the number of edges in H_n. Colour all edges in H_n blue and all edges in $\overline{H}_n$ red. Let K_n denote the complete graph on the same vertex set of H_n, i.e. $K_n = H_n \cup \overline{H}_n$. For any red edge $x = uv$, let $d_r(x) = d_r(u) + d_r(v) - 2$, where $d_r(u)$ denotes the number of vertices adjacent to u by a red edge. Let $\Delta_r(H_n) = \max_{x\in R} d(x)$. For simplicity, we use $\Delta_r(n)$ instead of $\Delta_r(H_n)$ when the context is clear.

All asymptotics in this paper refer to $n \to \infty$. For two functions $f(n)$ and $g(n)$, we say $f(n)$ is asymptotically at least $g(n)$ if there exists a sequence of positive reals $(a_n)_{n\geq 1}$ such that $a_n \to 0$ as $n \to \infty$ and $f(n) \geq (1 - a_n)g(n)$ for all n.

Let F_n denote a d-regular graph chosen uniformly at random from all d-factors of K_n, and let $X_n = X_n(H_n)$ denote the number of red edges contained in F_n. In this paper, we assume d is constant. Our main result is the following theorem.

Theorem 2.1 *Let H_n be a graph on n vertices and let $t_n = \binom{n}{2} - e(H_n)$. Assume $a := \limsup_{n\to\infty} \Delta_r(n)/n < 1$. Then provided t_n/n^2 is sufficiently small,*

$$\mathbf{P}(X_n = 0) \geq \exp\left(\frac{-8d^2 t_n}{(1-a)n} + o(1)\right).$$

A special case of Theorem 2.1 is $t_n = o(n)$ and d is constant, in which the probability $\mathbf{P}(X_n) = 1 + o(1)$. Thus, removing $o(n)$ arbitrary edges from a complete graph does not affect asymptotically the number of d-factors it contains for any constant d.

Since the total number of d-factors of K_n is asymptotically

$$\frac{(dn)!}{2^{dn/2}(dn/2)!(d!)^n} \exp\left(-\frac{d^2-1}{4}\right), \tag{2.1}$$

as proved in [2], we immediately obtain a lower bound of the number of d-factors in H_n.

Corollary 2.2 *Let H_n be a graph on n vertices and let $t_n = \binom{n}{2} - e(H_n)$. Assume $a := \limsup_{n\to\infty} \Delta_r(n)/n < 1$. Then provided t_n/n^2 is sufficiently small, the number of d-factors of H_n is asymptotically at least*

$$\frac{(dn)!}{2^{dn/2}(dn/2)!(d!)^n} \exp\left(-\frac{d^2-1}{4} - \frac{8d^2 t_n}{(1-a)n}\right).$$

The assumption $\limsup_{n\to\infty} \Delta_r(n)/n < 1$ in Theorem 2.1 and Corollary 2.2 can not be weakened without imposing other assumptions of H_n. This can be easily seen by considering the following counterexample. Let H_{4n+2} be a graph composed of the union of two copies of K_{2n+1}. Clearly, $\limsup_{n\to\infty} \Delta_r(H_{4n+2})/(4n+2) = 1$ but H_{4n+2} contains no perfect matchings.

A special case of Corollary 2.2 is when H_n is regular.

Corollary 2.3 *There exists a constant $1/2 < \alpha < 1$, such that for any k-regular graph H_n on n vertices with $k \geq \alpha n$, the number of d-factors that are contained in H_n is asymptotically at least*

$$\frac{(dn)!}{2^{dn/2}(dn/2)!(d!)^n} \exp\left(-\frac{d^2-1}{4} - 4d^2 n \cdot \frac{n-k-1}{2k-n+4}\right).$$

The following is another direct corollary of Theorem 2.1.

Corollary 2.4 *Let $d \geq 1$ and $M > 0$ be bounded. Let H_n be a graph on n vertices with at least $\binom{n}{2} - Mn$ edges. Assume $a := \limsup_{n\to\infty} \Delta_r(n)/n < 1$. Then*

$$\liminf_{n\to\infty} \mathbf{P}(X_n = 0) \geq \exp(-8d^2M/(1-a)) > 0.$$

Consider the following algorithm, which obviously outputs a uniformly random d-factor of an input graph H. The efficiency of the algorithm, when H is sufficiently dense, is guaranteed by Corollary 2.4.

Algorithm: *The d-Factor Generator*

Input: A graph H on the vertex set $[n]$.

Output: A uniformly random d-factor of H.

1. Uniformly at random choose a d-factor F_n from K_n, the complete graph on the vertex set $[n]$.
2. Output F_n if all edges in F_n are in H. Otherwise, go to step 1.

The first step of the d-Factor Generator can be performed efficiently when $d = O(n^{1/3})$ (with expected running time $O(nd^3)$) using the algorithm in [7]. Corollary 2.4 implies that the d-Factor Generator algorithm runs efficiently when the input graph H is sufficiently dense. Given H as the input of the d-Factor Generator, let $T(H)$ denote the number of times that step 1 of the algorithm is called. The following corollary follows directly from Corollary 2.4.

Corollary 2.5 *Let d, M and $0 < a < 1$ be constants. Assume H_n is an input graph with n vertices and at least $\binom{n}{2} - Mn$ edges and $\Delta_r(H_n) \leq an$. Then the d-Factor Generator algorithm outputs a uniformly random d-factor of H and the expectation of $T(H_n)$ is uniformly bounded, i.e., $\mathbf{E}T(H_n)$ is bounded above by some absolute constant that depends only on d, M and a.*

On the other hand, we have the following theorem, implying that $\mathbf{E}T(H_n)$ is unbounded once $c_n \to \infty$ as $n \to \infty$, where $c_n = (\binom{n}{2} - e(H_n))/n$.

Theorem 2.6 *Let $c_n = (\binom{n}{2} - e(H_n))/n$. Assume $c_n \to \infty$ as $n \to \infty$. Then $\lim_{n\to\infty} \mathbf{P}(X_n = 0) = 0$.*

We prove Theorems 2.1 and 2.6 in the next section.

3 Proofs

In order to prove Theorem 2.1, we introduce a switching operation called the *r-switching*. Let F_n denote a d-factor of K_n.
The r-switching: Given F_n containing at least one red edge, choose a red edge $x \in F_n$; label its end vertices as u and v and choose a blue edge $y \in F_n$ which is not incident with x; label its end vertices as u' and v'. Replace x and y by uu' and vv'. The r-switching is applicable if and only if both uu' and vv' are blue and are not contained in F_n.

The inverse r-switching: Choose a blue edge in F_n and label its end vertices as u and u'; choose another blue edge in F_n that is not incident with uu' and label its end vertices as v and v'. Replace these two edges by uv and $u'v'$. The inverse r-switching is applicable if and only if both edges uv and $u'v'$ are not in F_n and uv is red and $u'v'$ is blue.

See an example of the r-switching and its inverse in Figure 1, where the dashed line denotes a red edge and a solid line denotes a blue edge.

Let $\mathcal{R}(\ell)$ be the set of all d-factors F_n containing exactly ℓ red edges. For every $\ell \geq 1$, an r-switching converts an $F_n \in \mathcal{R}(\ell)$ into an $F'_n \in \mathcal{R}(\ell - 1)$. Conversely, an inverse r-switching converts an $F'_n \in \mathcal{R}(\ell - 1)$ into an $F_n \in R(\ell)$. Let $N(F_n)$ denote the number of r-switchings applicable on F_n. For $F'_n \in \mathcal{R}(\ell - 1)$, let $N'(F'_n)$ denote the number of inverse r-switchings applicable on F'_n.

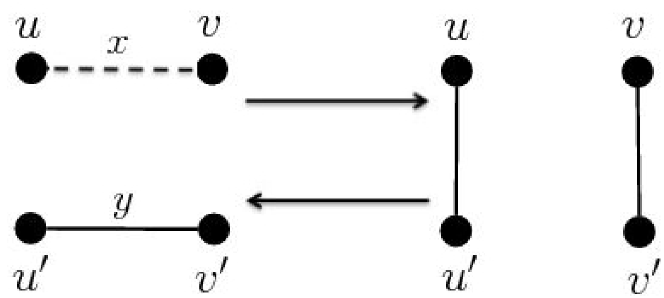

Figure 1: *r-switching and its inverse*

In the rest of the paper, we always let $t_n = \binom{n}{2} - e(H_n)$ and let $c_n = t_n/n$. We first prove the following lemma.

Lemma 3.1 *For all ℓ such that $n - 2\ell/d - 2d - \Delta_r(n) > 0$,*

$$2\ell(dn - 2\ell - 2d^2 - d\Delta_r(n)) \leq \quad N(F_n) \quad \leq 2d\ell n, \tag{3.1}$$

$$0 \leq \quad N'(F'_n) \quad \leq 2d^2 c_n n. \tag{3.2}$$

Proof. Given $F_n \in \mathcal{R}(\ell)$, the number of ways to choose x and label its end vertices is 2ℓ. There are at most $dn/2$ choices for y and given each choice of y there are two ways to label its end vertices. So the upper bound of $N(F_n)$ follows immediately. The lower bound is obtained by a little more careful estimation of the choices of y. Clearly, given any choice of x and its end-vertex labelling, the number of ways to choose the edge y and label its end vertices so that y is blue and not incident with x and that $uu' \notin F_n$, $vv' \notin F_n$ is at least $2 \cdot dn/2 - 2\ell - 2d^2$. On the other hand, given any choice of x and labelling of its end vertices, the number of ways to choose y and label its end vertices so that either uu' or vv' is red is at most $d\Delta_r(n)$. Therefore, the total number of r-switchings applicable on F_n is at least $2\ell(dn - 2\ell - 2d^2 - d\Delta_r(n))$.

Given $F'_n \in \mathcal{R}(\ell - 1)$, we choose uu' and vv' in the following way. First choose a red edge $x \notin F'_n$. Label its end vertices as u and v. Then choose a neighbour of u in F'_n and label it as u'. Choose a neighbour of v in F'_n and label it as v'. The inverse switching is applicable with such choice of uu' and vv' if and only if $u' \neq v'$ and all edges uu', vv' and $u'v'$ are blue. The upper bound of $N'(F'_n)$ follows immediately by noting that there are at most $c_n n$ ways to choose x and two ways to label its end vertices, whereas given the labelling of u and v, there are at most d ways to choose u' and d ways to choose v'. ∎

Next we prove the following lemma.

Lemma 3.2 *Assume $a := \limsup_{n\to\infty} \Delta_r(n)/n < 1$ and c_n/n is sufficiently small. Let $b = (1-a)/8$. Then $\mathbf{P}(X_n \geq bn) = o(1)$.*

Proof. There are at most $c_n n$ red edges, and thus at most $\binom{c_n n}{bn}$ ways to choose bn red edges for F_n. The number of ways to choose the remaining $dn/2 - bn$ edges to form a d-factor of K_n is at most

$$\frac{(dn-2bn)!}{2^{dn/2-bn}(dn/2-bn)!} = \exp\left((d/2-b)n\ln\left(\frac{dn-2bn}{e}\right) + O(1)\right) = \exp\left((d/2-b)n\ln n + O(n)\right).$$

Hence, the number of d-factors of K_n that contain at least bn red edges is at most

$$\binom{c_n n}{bn}\exp\left((d/2-b)n\ln n + O(n)\right). \tag{3.3}$$

We first show that

$$\binom{c_n n}{bn} = \exp(bn\ln c_n + O(n)). \tag{3.4}$$

This is clearly true if $c_n = O(1)$. Assume $c_n \to \infty$. Then using the Stirling's formula, we obtain

$$\begin{aligned}\binom{c_n n}{bn} &= \exp(n(c_n\ln c_n - b\ln b - (c_n - b)\ln(c_n - b)) + O(n)) \\ &= \exp(c_n n(\ln c_n - \ln(c_n(1 - b/c_n))) + bn\ln c_n + O(n)) \\ &= \exp(c_n n(\ln c_n - (\ln c_n - b/c_n + O(c_n^{-2}))) + bn\ln c_n + O(n)) \\ &= \exp(bn\ln c_n + O(n)).\end{aligned}$$

This completes the proof of (3.4). Combining with (3.3), the number of d-factors of K_n that contain at least bn red edges is at most

$$\exp\left((d/2-b)n\ln n + bn\ln c_n + O(n)\right). \tag{3.5}$$

By (2.1), the total number of d-factors of K_n is asymptotically

$$\frac{(dn)!}{2^{dn/2}(dn/2)!(d!)^n}\exp\left(-\frac{d^2-1}{4}-\frac{d^3}{12n}\right)=\exp\left((dn/2)\ln n+O(n)\right). \tag{3.6}$$

Taking the ratio of (3.5) and (3.6) yields

$$\mathbf{P}(X_n\geq bn)\leq \exp(-bn\ln n+bn\ln c_n+O(n))=o(1),$$

provided c_n/n is sufficiently small. ∎

Proof of Theorem 2.1. Recall that $a=\limsup_{n\to\infty}\Delta_r(n)/n<1$. Let $b=(1-a)/8$. Then there exists $n_0>0$ such that for all $n\geq n_0$,

$$\frac{\Delta_r(n)}{n}<a+\frac{1-a}{2}=1-4b,\quad \text{and } bn\geq 2d^2.$$

Therefore, for all $n\geq n_0$ and for all $\ell\leq bn$,

$$dn-2\ell-2d^2-d\Delta_r(n)\geq n(d-2b-d(1-4b))-2d^2\geq bn.$$

By Lemma 3.1, for all $\ell\leq bn$,

$$2\ell bn|\mathcal{R}(\ell)|\leq 2d^2c_nn|\mathcal{R}(\ell-1)|.$$

So,

$$\frac{|\mathcal{R}(\ell)|}{|\mathcal{R}(\ell-1)|}\leq\frac{2d^2c_nn}{2\ell bn}\leq\frac{d^2c_n}{b\ell}.$$

It follows that for any $n\geq n_0$ and for any $\ell\leq bn$,

$$\frac{|\mathcal{R}(\ell)|}{|\mathcal{R}(0)|}\leq\frac{(d^2c_n/b)^\ell}{\ell!}.$$

Since t_n/n^2 is sufficiently small, so is c_n/n. Then by Lemma 3.2, for all $n\geq n_0$,

$$\frac{1}{\mathbf{P}(X_n=0)}=(1+o(1))\frac{\sum_{\ell=0}^{bn}|\mathcal{R}(\ell)|}{|\mathcal{R}(0)|}\leq(1+o(1))\sum_{\ell=0}^{bn}\frac{(d^2c_n/b)^\ell}{\ell!}\leq(1+o(1))\exp(d^2c_n/b),$$

which implies that

$$\liminf_{n\to\infty}\mathbf{P}(X_n=0)\geq\exp(-d^2c_n/b)=\exp(-8d^2t_n/(1-a)n).\ ∎$$

Next, we prove Theorem 2.6. We define a few other switching operations.

The r_0-switching: Given $F_n\in\mathcal{R}(0)$, choose two edges x and y in F_n, and label their end vertices as u, v and u', v' respectively. Replace x and y by uu' and vv'. This switching is applicable if and only if both uu' and vv are not in F_n and at least one of these two edges is red.

The r_1-switching: Given $F_n \in \mathcal{R}(1)$, let $x \in F_n$ be the red edge and choose another blue edge y in F_n, and label their end vertices as u, u' and v, v' respectively. Replace x and y by uv and $u'v'$. This switching is applicable if and only if both uv and $u'v'$ are not in F_n and both edges are blue.

The r_2-switching: Given $F_n \in \mathcal{R}(2)$, let x and y be the two red edges in F_n, and label their end vertices as u, u' and v, v' respectively. Replace x and y by uv and $u'v'$. This switching is applicable if and only if both uv and $u'v'$ are not in F_n and both edges are blue.

Examples of the r_i-switchings are given in Figures 2–4. In Figures 3 and 4, the notation (b), or (r), means the edge under discussion needs to be blue, or red, respectively.

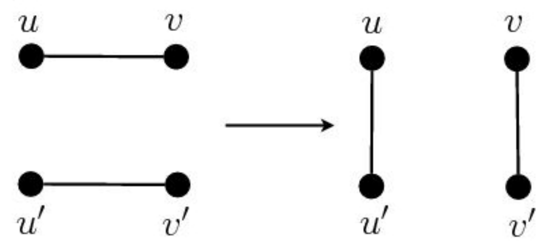

Figure 2: r_0-*switching*

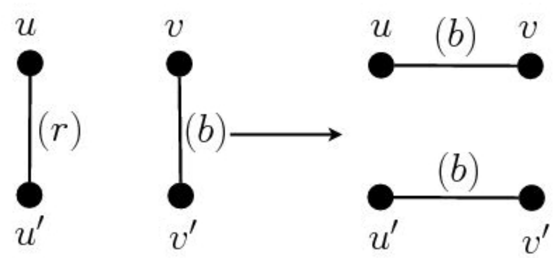

Figure 3: r_1-*switching*

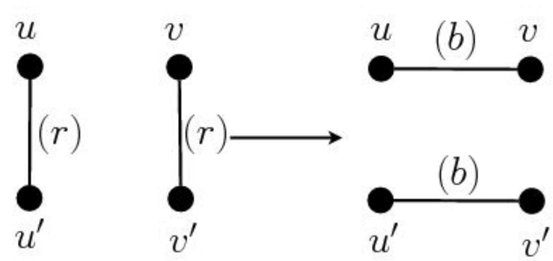

Figure 4: r_2-*switching*

Let $N_i(H_i)$ denote the number of r_i-switchings applicable on $H_i \in \mathcal{R}(i)$ for $i = 0, 1, 2$. Clearly,

$$\sum_{H_0 \in \mathcal{R}(0)} N_0(H_0) = \sum_{H_1 \in \mathcal{R}(1)} N_1(H_1) + \sum_{H_2 \in \mathcal{R}(2)} N_2(H_2). \tag{3.7}$$

Proof of Theorem 2.6. We first show the following claim. For any $H_i \in \mathcal{R}(i)$, $i = 0, 1, 2$,

$$N_0(H_0) \geq 2dc_n n, \quad N_1(H_1) \leq 4dn, \quad N_2(H_2) \leq 8. \tag{3.8}$$

The upper bound of $N_i(H_i)$ for $i = 1, 2$ follows trivially by considering the number of ways to choose the red and blue edges and label their end vertices respectively. For any $H_0 \in \mathcal{R}(0)$, let $z \notin H_0$ be a red edge. Label its end vertices as u and u'. There are d neighbours of u in H_0. Choose one of them and label it as v. Given the choice of v, there are at least $\max\{1, d-1\}$ neighbours of u' in H_0 that are distinct from v. Choose one of them and label it as v'. Then an r_0-switching can be applied to the edges uv and $u'v'$. There are $c_n n$ ways to choose the edge z and two ways to label its end vertices. Therefore, the number of r_0-switchings applicable on H_0 such that uu' is red is at least $2dc_n n$. Hence, $N_0(H_0) \geq 2dc_n n$. Note that the inequality holds also because we omit the count of r_0-switchings such that uu' is blue and vv' is red. This completes the proof of the claim (3.8).

By (3.7) and (3.8), for any $n \geq 2$,

$$2dc_n n|\mathcal{R}(0)| \leq 4dn|\mathcal{R}(1)| + 8|\mathcal{R}(2)| \leq 4dn(|\mathcal{R}(1) + \mathcal{R}(2)|),$$

which yields

$$\mathbf{P}(X_n = 0) \leq \frac{|\mathcal{R}(0)|}{|\mathcal{R}(1)| + |\mathcal{R}(2)|} \leq \frac{4dn}{2dc_n n} = 2c_n^{-1}.$$

Since $c_n \to \infty$ as $n \to \infty$, $\lim_{n\to\infty} \mathbf{P}(X_n = 0) = 0$. ∎

References

[1] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combin.* **1** (1980), 311–316.

[2] E.A. Bender and E.R. Canfield, The asymptotic number of labeled graphs with given degree sequences, *J. Combinatorial Theory Ser. A* **24** (1978), 296–307.

[3] L. Esperet, F. Kardos, A. King, D. Kral and S. Norine, Exponentially many perfect matchings in cubic graphs, preprint.

[4] M. Jerrum and A. Sinclair, Approximating the permanent, *SIAM J. Comput.* **18**, pp. 1149–1178.

[5] B. D. McKay, Subgraphs of random graphs with specified degrees, *Proceedings of the International Congress of Mathematicians*, Volume **IV**, 2489-2501, Hindustan Book Agency, New Delhi, 2010.

[6] B.D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums, *Ars Combinatoria* **19A** (1985), 15–25.

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

[8] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(\sqrt{n})$, *Combinatorica* **11** (1991), 369-382.

[9] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, *European journal of combinatorics*, **11**, 1990, 565–580.