RAINBOW HAMILTON CYCLES IN RANDOM REGULAR GRAPHS

SVANTE JANSON AND NICHOLAS WORMALD

To Alan Frieze on the occasion of his 60th birthday

ABSTRACT. A rainbow subgraph of an edge-coloured graph has all edges of distinct colours. A random d-regular graph with d even, and having edges coloured randomly with $d/2$ of each of n colours, has a rainbow Hamilton cycle with probability tending to 1 as $n \to \infty$, for fixed $d \geq 8$.

1. INTRODUCTION

An edge-coloured graph is a *rainbow* if no colour appears more than once. We will study rainbow Hamilton cycles in edge-coloured graphs with n vertices where the number of colours available is also n; thus a rainbow Hamilton cycle uses each of the colours exactly once.

We consider in this paper only (random) regular graphs and (random) colourings where each colour occurs the same number of times. If each colour occurs on q edges, we thus have qn edges, and hence the vertex degrees are $2q$. We use the standard notation $G(n, d)$ for a uniformly chosen random d-regular graph on n given (labelled) vertices. We will only consider the case $d = 2q$ even (and hence there is no parity restriction on n), because otherwise it impossible to have n colours occurring equally often. Having sampled a random graph $G(n, 2q)$, we then randomly colour its qn edges by n colours ($1, \dots, n$, say) with q edges of each colour, again choosing uniformly among all possibilities. We denote the resulting randomly coloured random graph by $G_c(n, 2q)$.

Our main result is the following on randomly coloured random regular graphs. (For some related results on the random graph $G(n, m)$, see Cooper and Frieze [3].) We say that an event holds *with high probability* (**whp**), if it holds with probability tending to 1 as $n \to \infty$. (All unspecified limits in this paper are for $n \to \infty$.)

Theorem 1.1. *Consider the randomly coloured random $2q$-regular graph $G_c(n, 2q)$, with n colours and q edges of each colour. Then,* ***whp****, there exists a rainbow Hamilton cycle if $q \geq 4$, and not if $q \leq 3$.*

Date: August 1, 2005; revised September 19, 2006.

2000 *Mathematics Subject Classification.* 05C80 (05C15, 05C45, 60C05).

The second author acknowledges the support of the Canadian Research Chairs Program and NSERC.

Recall that it was shown by Robinson and Wormald [16, 17] that $G(n,d)$ **whp** contains a Hamilton cycle as soon as $d \geq 3$. In our setting, when $d = 2q$ has to be even, we thus **whp** have Hamilton cycles, ignoring the colouring, when $d \geq 4$, but rainbow Hamilton cycles only when $d \geq 8$. It is nevertheless remarkable that **whp** some Hamilton cycle manages to pick up an edge of each colour in a random 8-regular graph, when there are only four edges of each colour to choose from.

Remark 1.2. In a similar direction, Robinson and Wormald [18] showed that a random 3-regular graph with $o(\sqrt{n})$ randomly specified edges **whp** has a Hamilton cycle passing through all the specified edges (and, moreover, in randomly prespecified directions). It has further been shown by Kim and Wormald [14] that a random $2q$-regular graph **whp** has an edge-decomposition into q Hamilton cycles, provided $q \geq 2$.

It is natural to ask whether, similarly, a randomly coloured $2q$-regular graph with n colours and q edges of each colour, as above, **whp** has an edge-decomposition into q rainbow Hamilton cycles. By computing the expected number of such decompositions (similarly to the proof of Lemma 3.2 below), it is easily seen that this is **whp** false when $q \leq 4$. We leave the case $q \geq 5$ (when the expected number tends to infinity) as an open problem.

The proof of Theorem 1.1 is based on the small subgraph conditioning method introduced by Robinson and Wormald [16, 17], and further developed in [12], [15], [19] and [13, Chapter 9]. However, for this problem we have to consider the colourings of the small subgraphs too, see Section 3.

Acknowledgements. This problem was suggested by Alan Frieze during the Conference on Random Structures and Algorithms at Emory University, Atlanta, 1995; he was originally intended as a coauthor but later declined this. Consequently we are pleased to be able finally to dedicate this work to him, marking 10 years since its beginnings, in which he was involved, and 60 since his.

We also acknowledge the assistance of the Maple algebraic manipulation package for the variance calculations in Section 6. Although the proof we found can be verified by hand, Maple was instrumental in finding that proof.

2. Multigraphs, a bipartite graph, and traffic rules

As usual in the study of random regular graphs (with small degree), it is convenient to extend the study to multigraphs. Recall that a convenient way (at least for theoretical purposes) to generate a random regular graph is the so-called *configuration model* or *pairing model*, see e.g. [1] or [19]: We start with nd points partitioned into n cells of d points each. We then take a random pairing of the points into $nd/2$ pairs (assuming nd to be even). Collapsing each cell to a vertex and regarding each pair as an edge, we obtain a random d-regular multigraph that may contain loops and multiple edges; we denote this random multigraph by $G^*(n,d)$. (The points themselves are called *half-edges*.) It is well-known, and easily seen, that if we condition $G^*(n,d)$ on being a simple graph (no loops nor multiple edges), then we obtain the uniformly distributed random regular graph $G(n,d)$. Moreover, it is well-known that for fixed d, the probability $\mathbb{P}(G^*(n,d)$ is simple) tends to a non-zero limit as $n \to \infty$; hence, every property that $G^*(n,d)$ has **whp**, is **whp** enjoyed by $G(n,d)$ too. In particular, we may (and will) prove Theorem 1.1 by proving the following extension of it; we define $G^*_c(n,2q)$ by analogy with $G_c(n,2q)$, by choosing uniformly at random a colouring of the edges with n colours with q edges of each colour.

Theorem 2.1. *Theorem 1.1 holds for the randomly coloured random regular multigraph $G^*_c(n,2q)$ too.*

We find it useful to introduce an associated bipartite graph. (This is really a multigraph too, since it may have multiple edges.) Given the randomly coloured multigraph $G^*_c(n,2q)$, add a new vertex on each edge. We give each new vertex the colour of the edge it bisects, leaving the original vertices uncoloured. Finally, we combine the q vertices of each colour into a single coloured vertex of that colour. This gives us a $2q$-regular bipartite (multi)graph with $n+n$ vertices; the original n vertices form one side of the bipartition, and the n coloured vertices the other. Moreover, each coloured vertex comes with a pairing of the edges (or half-edges) attached to it; this pairing shows which pairs of edges correspond to edges in the multigraph. We may think of the coloured vertices as having $2q$ attached half-edges arranged in a circle, with each half-edge matched to the opposite one. There is then a one-to-one correspondence between walks in the multigraph and walks in the bipartite graph (of twice the length, and beginning and ending at uncoloured vertices) that pass 'straight ahead' between matched half-edges at each coloured vertex. In particular, rainbow Hamilton cycles in the multigraph correspond to Hamilton cycles in the bipartite graph that obey this traffic rule.

The colours are no longer important in the bipartite graph, but it will be convenient to refer to the two sets of vertices in the bipartition as 'plain' and 'coloured'.

Conversely, we may start with the bipartite graph, with given traffic rules, and obtain the original multigraph by combining the edges two by two at the coloured vertices. Note that choosing the bipartite (multi)graph at random using the configuration model (in its bipartite version, and with traffic rules as above given in each coloured cell) gives back the random coloured multigraph $G^*_c(n,2q)$ with the right distribution.

We let $B^*(n,n;2q)$ denote this random bipartite multigraph with traffic rules as above, that is, with $n+n$ vertices of degree $2q$ and with a random pairing of the half-edges at each vertex in the coloured part of the bipartition. (Although viewing it as a multigraph when referring to cycles etc., all computations are done with the equivalent configuration model.) To prove Theorems 1.1 and 2.1, it is enough to prove the following.

Theorem 2.2. *The random bipartite multigraph $B^*(n,n;2q)$ **whp** has a rainbow Hamilton cycle obeying the traffic rules if $q \geq 4$, and not if $q \leq 3$.*

Remark 2.3. One may study random (regular) graphs with other traffic rules at the vertices. In general, we may equip each vertex of degree d with a (possibly directed, and possibly random) *connection graph* with d vertices representing the incident edges; the edges in the connection graph show the allowed connections between incoming and outgoing edges. In our case, the connection graph is the complete graph (no restrictions) for one side of the bipartition, and a matching with $d/2$ edges for the other side. We do not know of any general study, but a few examples of this type have appeared in the literature:

Garmo [8, 9] studied random railways; these are regular (typically cubic) graphs where the vertices (representing switches) have connection graphs that are stars. In [10], this was extended to graphs where a random subset of the vertices have a star as connection graph and the rest the complete graph.

Gamburd [7] studied long cycles in random oriented cubic graphs; here the connection graph is a directed 3-cycle at each vertex.

3. Small subgraphs

The small subgraph conditioning method introduced by Robinson and Wormald [16, 17] has been successfully applied to several problems, in particular in the theory of random regular graphs, see e.g. [13, Chapter 9], [19] and [11]. (For applications to random hypergraphs, see [5, 4].)

As often pointed out by Alan Frieze, see [4, 5, 6], the method can be regarded as an *analysis of variance.* The main idea is that we consider some random variable, Y say, that counts occurrences of some structure, and let a parameter $n \to \infty$. Typically, it is easy to prove that the expectation $\mathbb{E}\, Y$ tends to infinity, but we want to show that $\mathbb{P}(Y > 0) \to 1$. If the variance $\operatorname{Var}(Y)$ is $o(\mathbb{E}\, Y)^2$, then the second moment method (i.e. Chebyshev's inequality) immediately shows the desired result. The small subgraph conditioning method applies to cases where the variance $\operatorname{Var}(Y)$ is of the same order as $(\mathbb{E}\, Y)^2$, by showing that the variance can be explained, up to a factor $1-o(1)$, by the interaction between the numbers of some small subgraphs and the random variable Y. The desired conclusion $Y > 0$ **whp** then follows by conditioning on the numbers of these small subgraphs and using Chebyshev's inequality on the conditioned variables. For details, see [13, Theorem 9.12–Remark 9.18] and [19, Theorem 4.1]. We state the results there in the following form (an immediate consequence of [19, Corollary 4.2]). We use $[x]_m := x(x-1)\cdots(x-m+1)$ to denote falling factorials.

Theorem 3.1. *Let $\lambda_i > 0$ and $\delta_i \geq -1$ be real numbers for $i = 1, 2, \dots$ and suppose that for each n there are random variables $X_i = X_i(n)$, $i = 1, 2, \dots$ and $Y = Y(n)$, all defined on the same probability space $\mathcal{G} = \mathcal{G}_n$ such that*

X_i is nonnegative integer valued, Y is nonnegative and $\mathbb{E}\, Y > 0$ (for n sufficiently large). Suppose furthermore that

(i) *For each $k \geq 1$, the variables $X_1, \dots, X_k$ are asymptotically independent Poisson random variables with $\mathbb{E}\, X_i \to \lambda_i$,*

(ii) *if $\mu_i = \lambda_i(1 + \delta_i)$, then*

$$\frac{\mathbb{E}(Y[X_1]_{m_1} \cdots [X_k]_{m_k})}{\mathbb{E}\, Y} \to \prod_{i=1}^{k} \mu_i^{m_i} \tag{3.1}$$

for every finite sequence $m_1, \dots, m_k$ of nonnegative integers,

(iii) *$\sum_i \lambda_i \, \delta_i^{\,2} < \infty$,*

(iv) *$\mathbb{E}\, Y^2/(\mathbb{E}\, Y)^2 \leq \exp(\sum_i \lambda_i \, \delta_i^{\,2}) + o(1)$ as $n \to \infty$.*

*Then, if $\mathcal{E}$ is the event $\bigwedge_{\delta_i = -1} \{X_i = 0\}$, $\mathbb{P}(Y > 0 \mid \mathcal{E}) \to 1$. In particular, if $\delta > -1$ for every i, then $Y > 0$ **whp**.*

We will actually use Theorem 3.1 with a doubly indexed sequence X_{ij}; obviously, this is just a matter of notation.

In many applications of the small subgraph conditioning method, the variables X_i are the numbers of cycles of different lengths. This has perhaps misled some into the belief that the short cycles are expected to provide the answer in all cases. But they play the central role for most problems only because they are the only possible 'unusual' small subgraphs. The subgraphs of fixed size in a random d-regular graph are very well behaved. Near a random vertex, such a graph looks locally like a tree. But even that statement can be misleading when we consider what comes shortly. The thing to focus on is that, because **whp** no two short cycles are near each other, the number of small subgraphs of any particular type, even disconnected, is determined by the numbers of short cycles.

In our case, it will turn out that conditioning on the numbers of small cycles in $G_c^*(n, 2q)$ does not explain all of the variance of the number of rainbow Hamilton cycles; we have to consider also colourings. Note that for each fixed length i, there are only a few cycles of length i (the expected number is $O(1)$), and **whp** they are all rainbow, so we would expect no explanation of variance to be caused by the numbers of intrinsically differently coloured short cycles. However, we may consider, for example, the number of paths of length i where the first and last edges have the same colour. The expected number is $\Theta(1)$. These are intrinsically different from short rainbow paths, and it turns out that these structures too will be significant in the analysis of variance.

Perhaps surprisingly, there is even more to consider. It turns out that the existence of two short paths, each joining a blue edge to a red edge, somehow affects the expected number of rainbow Hamilton cycles (and therefore these structures contribute to the variance), even when the paths occur in distant parts of the graph. However, this should not be surprising given the discussion above about small subgraphs. Colours are clearly relevant in our present problem, so we should consider coloured subgraphs. Typical small subgraphs, not necessarily connected, are, for our present problem, forests with distinctly coloured edges. The numbers of small forests in which some of the edges are coloured the same thus qualify as special small subgraphs. It was for this reason that the method was called small subgraph conditioning in [19], rather than short cycle conditioning. This is indeed the first application of the method in which some of the small subgraphs involved are disconnected.

To describe the general situation precisely, we work with the random bipartite multigraph $B^*(n, n; 2q)$ defined in Section 2, and let Y be the number of Hamilton cycles in the multigraph that obey the traffic rules. Recall that Y equals the number of rainbow Hamilton cycles in $G^*_c(n, 2q)$. Further, for each $i \ge 1$ and j with $0 \le j \le i$, we let X_{ij} be the number of cycles of length $2i$ in $B^*(n, n; 2q)$ that violate the traffic rules at exactly j coloured vertices. (Thus, $Y = X_{n0}$, but we are mainly interested in X_{ij} for small i.) Note that X_{i0} equals the number of rainbow i-cycles in $G^*_c(n, 2q)$, and thus it **whp** equals the number of i-cycles in $G^*_c(n, 2q)$, while X_{i1} **whp** equals the number of paths of length $i+1$ where the first and last edges have the same colour. (This holds only **whp**, since for X_{i1} the endpoints of the path may coincide with each other or with some interior point.) We may similarly interpret X_{ij} for $j \ge 2$, at least **whp**, as the number of certain collections of j paths, generalising the example mentioned above, but we leave the details to the reader.

We state three lemmas that will be proven in the following sections.

Lemma 3.2. *Suppose that $d = 2q \ge 4$. Then*

$$\mathbb{E}(Y) = \Theta\left(\frac{(d-1)(d-2)^{d-2}}{d^{d-2}}\right)^n.$$

Hence, as $n \to \infty$, $\mathbb{E}(Y) \to 0$ for $d \le 6$ but $\mathbb{E}(Y) \to \infty$ for $d \ge 8$.

Lemma 3.3. *Conditions* (i) *and* (ii) *in Theorem 3.1 are satisfied for the variables $(X_{ij})_{ij}$ and*

$$\lambda_{ij} = \frac{1}{2i}\binom{i}{j}(d-1)^i(d-2)^j,$$

$$\delta_{ij} = \begin{cases} (-1)^{i+j}\frac{2^j}{(d-1)^i(d-2)^j}, & j > 0, \\ -\frac{2}{(d-1)^i}, & j = 0 \text{ and } i \text{ odd}, \\ 0, & j = 0 \text{ and } i \text{ even}. \end{cases}$$

Lemma 3.4. *Suppose that $d > 4$. Then*

$$\mathbb{E}\, Y^2/(\mathbb{E}\, Y)^2 \to \left(\frac{d}{d-4}\right)^{1/2}.$$

Proof of Theorem 2.2. First note that if $d \le 6$, then $\mathbb{E}(Y) \to 0$ by Lemma 3.2, and thus $\mathbb{P}(Y > 0) \to 0$, i.e. $Y = 0$ **whp**. In other words, there is then **whp** no rainbow Hamilton cycle in $G^*_c(n, 2q)$.

For the remainder of the proof, assume that $d = 2q \geq 8$. By Lemma 3.2, $\mathbb{E}(Y) \to \infty$. We want to show that $Y > 0$ **whp**. We employ Theorem 3.1 with X_{ij} as defined above, and λ_{ij} and δ_{ij} as given in Lemma 3.3. Note that $\delta_{ij} > -1$ for all i and j, so it remains only to show that the assumptions (i)–(iv) in Theorem 3.1 hold. For (i) and (ii), this is Lemma 3.3.

For (iii) and (iv) we split the sum into two parts.

$$\begin{aligned}
\sum_{i,j>0} \lambda_{ij}\delta_{ij}^2 &= \sum_{i,j\geq 1} \frac{1}{2i}\binom{i}{j}\frac{4^j}{(d-1)^i(d-2)^j} \\
&= \sum_{i=1}^{\infty} \frac{1}{2i}(d-1)^{-i}\Big(\Big(1+\frac{4}{d-2}\Big)^i - 1\Big) \\
&= \frac{1}{2}\sum_{i=1}^{\infty} \frac{1}{i}\Big(\Big(\frac{d+2}{(d-1)(d-2)}\Big)^i - \frac{1}{(d-1)^i}\Big) \\
&= -\frac{1}{2}\ln\Big(1-\frac{d+2}{(d-1)(d-2)}\Big) + \frac{1}{2}\ln\Big(1-\frac{1}{d-1}\Big) \\
&= -\frac{1}{2}\ln\frac{d^2-4d}{(d-1)(d-2)} + \frac{1}{2}\ln\frac{d-2}{d-1} \\
&= \frac{1}{2}\ln\frac{(d-2)^2}{d(d-4)}
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i} \lambda_{i0}\delta_{i0}^2 &= \sum_{i \text{ odd}} \frac{1}{2i}\frac{4}{(d-1)^i} = -\ln\Big(1-\frac{1}{d-1}\Big) + \ln\Big(1+\frac{1}{d-1}\Big) \\
&= -\ln\frac{d-2}{d-1} + \ln\frac{d}{d-1} = \ln\frac{d}{d-2}
\end{aligned}$$

Consequently,

$$\sum_{i,j} \lambda_{ij}\delta_{ij}^2 = \frac{1}{2}\ln\frac{(d-2)^2}{d(d-4)} + \ln\frac{d}{d-2} = \frac{1}{2}\ln\frac{d}{d-4}.$$

This proves (iii), and together with Lemma 3.4 also (iv). □

4. Expectation

Proof of Lemma 3.2. There are $n!^2/2n$ ways to arrange the $2n$ vertices in a cycle, with plain and coloured vertices alternating, and for each such arrangement $d(d-1)$ ways to choose the half-edges at each plain vertex and d ways to choose the half-edges at each coloured vertex (obeying the traffic rules). For each such choice, the probability that the selected $4n$ half-edges are connected to each other in the specified order equals $((d-2)n)!/(dn)!$. Consequently, using Stirling's formula,

$$\mathbb{E}\,Y = \frac{d^{2n}(d-1)^n n!^2((d-2)n)!}{2n\,(dn)!} = \Theta\Big(\frac{d^2(d-1)(d-2)^{d-2}}{d^d}\Big)^n = \Theta\big(f(d)\big)^n,$$

where $f(d) := (d-1)(1-2/d)^{d-2}$. We have $f(4) = 3/4 < 1$, $f(6) = 80/81 < 1$, $f(8) = 5103/4096 > 1$, and $f(d) > (d-1)e^{-2} > 1$ for $d > 8$. □

5. Short cycles

Proof of Lemma 3.3. We use arguments that have become standard for similar problems for random regular graphs, see e.g. [13, Section 9.4] or [19, Section 4.2]; we will thus omit some details.

For (i), we use the method of moments. It suffices to show that

$$\mathbb{E}\prod_{ij}[X_{ij}]_{m_{ij}} \to \prod_{ij}\lambda_{ij}^{m_{ij}}$$

for every finite set of non-negative integers $\{m_{ij}\}$. For convenience, we will only treat the expectation of a single X_{ij}; as in all similar problems, the argument extends immediately to (mixed) higher factorial moments.

To calculate $\mathbb{E}\,X_{ij}$, we count the appropriate oriented cycles with a designated initial vertex, which we require to be plain; this counts each cycle $2i$ times. The vertices in the cycle may now be chosen in $[n]_i^2 \sim n^{2i}$ ways. Consider first the case $j = 0$, i.e. cycles obeying the traffic rules everywhere. For each choice of vertices there are, as in Section 4, $d(d-1)$ ways to choose the half-edges at each of the i plain vertices and d ways to choose the half-edges at each of the i coloured vertex. Finally, the probability of pairing the $4i$ chosen half-edges into $2i$ edges is $1/[dn]_{2i} \sim (dn)^{-2i}$. Hence,

$$2i\,\mathbb{E}\,X_{i0} = \frac{d^{2i}(d-1)^i[n]_i^2}{[dn]_{2i}} \to (d-1)^i.$$

For $j > 0$ we argue similarly. The traffic rules are to be violated at precisely j coloured vertices. These may be chosen in $\binom{i}{j}$ ways, and at each of them there is additional factor of $d-2$ for the choice of the out-going half-edge. Hence we obtain, for all i and j,

$$2i\,\mathbb{E}\,X_{ij} = \binom{i}{j}\frac{d^{2i}(d-1)^i(d-2)^j[n]_i^2}{[dn]_{2i}} \to \binom{i}{j}(d-1)^i(d-2)^j,$$

or $\mathbb{E}\,X_{ij} \to \lambda_{ij}$.

For (ii), we first observe that the left hand side of (3.1), by symmetry, remains the same if we fix two half-edges at each vertex, always choosing two opposite half-edges at the coloured vertices, and then replace Y by the indicator that the chosen half-edges comprise a rainbow Hamilton cycle. Denoting this event by $\mathcal{H}_1$, we thus want to show

$$\mathbb{E}\Big(\prod_{ij}[X_{ij}]_{m_{ij}}\,\Big|\,\mathcal{H}_1\Big) \to \prod_{ij}\mu_{ij}^{m_{ij}}. \tag{5.1}$$

Condition on $\mathcal{H}_1$, and let H_1 be the (unique) rainbow Hamilton cycle that uses the chosen half-edges. It is easily seen that the remainder of the graph $G_c^*(n,2q)$ can be regarded as the random multigraph $G_c^*(n,2q-2)$, and that this is independent of H_1. Hence, the left hand side of (5.1) equals the expectation in the union of a random rainbow Hamilton cycle H_1 and an independent $G^*_c(n, 2q-2)$ on the same vertex set.

For the same reasons as in (i), we will only consider a single expectation $\mathbb{E}(X_{ij} \mid \mathcal{H}_1)$. Consider first the case $j = 0$. We may, as for (i), choose the vertices of the $2i$-cycle in $n^{2i}(1+o(1))$ ways. We then decide whether the $2i$ edges are in the Hamilton cycle H_1 or in $G^*_c(n, 2q-2)$; we denote the choices by $\alpha_s \in \{1, 2\}$ for $s = 1, \dots, 2i$. At a plain vertex where the incoming edge is to have type $\alpha \in \{1, 2\}$ and the outgoing edge type β, there is for each possible incoming half-edge $a_{\alpha\beta}$ choices of the outgoing, where the numbers $a_{\alpha\beta}$ are conveniently collected in the matrix

$$A := (a_{\alpha\beta}) = \begin{pmatrix} 1 & d-2 \\ 2 & d-3 \end{pmatrix}.$$

At the coloured vertices there is only one choice for the outgoing edge, and it has to have the same type as the incoming; we encode this as $b_{\alpha\beta}$ with $B := (b_{\alpha\beta}) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, the identity matrix. Note that we have not counted the number of incoming half-edges; this is because these numbers cancel when we take into account the probability of making the connections; the probability of connecting a half-edge of either type to *some* incoming half-edge of the same type at a given vertex (in the opposite part of the bipartition) is $(1+o(1))n^{-1}$. Moreover, the probability that all $2i$ connections are made is $(1+o(1))n^{-2i}$, except in the case when all $\alpha_s = 1$, which is impossible for $n > i$. Consequently,

$$\begin{aligned} 2i\,\mathbb{E}(X_{i0} \mid \mathcal{H}_1) &\to \sum_{\alpha_1,\dots,\alpha_{2i}=1}^{2} a_{\alpha_1\alpha_2} b_{\alpha_2\alpha_3} \cdots b_{\alpha_{2i}\alpha_1} - a_{11}^i b_{11}^i \\ &= \operatorname{Tr}(AB)^i - 1 = \operatorname{Tr}(A^i) - 1 = (d-1)^i + (-1)^i - 1, \end{aligned}$$

since A has the eigenvalues $d-1$ and -1. We have thus shown (5.1) for this case, with

$$\mu_{i0} = \frac{(d-1)^i + (-1)^i - 1}{2i} = \lambda_{i0}(1 + \delta_{i0})$$

as required.

For $j > 0$ we argue similarly. Now we have to choose j coloured vertices where the traffic rule is violated, and for these vertices the matrix B is replaced by

$$\overline{B} := A - B = \begin{pmatrix} 0 & d-2 \\ 2 & d-4 \end{pmatrix}.$$

Luckily, A and $\overline{B} = A - I$ commute, so all $\binom{i}{j}$ choices of the violating vertices give the same result, and thus, for $j > 0$,

$$\begin{aligned} 2i\,\mathbb{E}(X_{ij} \mid \mathcal{H}_1) &\to \binom{i}{j} \operatorname{Tr}(A^i \overline{B}^j) = \binom{i}{j} \operatorname{Tr}\big(A^i (A-I)^j\big) \\ &= \binom{i}{j} \Big((d-1)^i (d-2)^j + (-1)^i (-2)^j\Big), \end{aligned}$$

which equals $2i\lambda_{ij}(1+\delta_{ij})$ as required in this case too. □

6. Variance

In this section we prove Lemma 3.4, thus completing the proof of Theorem 1.1.

Proof of Lemma 3.4. We compute $\mathbb{E}\,Y^2$ by calculating the probability that a given ordered pair of Hamilton cycles (H_1, H_2) are contained in the pairing corresponding to $B^*(n,n;2q)$, and summing over all possible ordered pairs (H_1, H_2).

The first part is similar to the treatment of H_1 in Section 5. Let k denote the number of coloured vertices at which the same two half-edges are used by both H_1 and H_2, and let j denote the number of plain vertices of this type. As befits their name, we will sometimes view the half-edges as the two halves of each edge. Note that, if $H_1 \neq H_2$, the edges shared by the two cycles induce a collection of paths of length at least 1 around H_1. We will call the set of edges of each such path a *string*. Each string ends at two different vertices; each of these vertices has exactly one half-edge (the one in the string) used by both H_1 and H_2. Each of these end vertices must be plain, because of the traffic rules. The strings may also have interior vertices; these are the vertices where the same two half-edges are used by both H_1 and H_2. Each string has exactly one more coloured interior vertex than plain ones. Consequently, there are $k-j$ strings, and hence precisely $2(k-j)$ plain vertices where exactly one half-edge is used for both H_1 and H_2.

The same strings also occur in H_2, though in general in a different order, and moreover their orientations with respect to H_2 are independent of their orientations with respect to H_1. Together, H_1 and H_2 determine the pairs of the pairing containing two of the half-edges at k coloured vertices and four at all other coloured vertices, so $4n-2k$ pairs in all (as each pair contains just one coloured vertex). Hence, defining $\mathcal{H}_i$ as the event that H_i occurs,

$$\mathbb{P}(\mathcal{H}_2 \mid \mathcal{H}_1) = \frac{\big((d-4)n+2k\big)!}{\big((d-2)n\big)!},$$

and we may write

$$\frac{\mathbb{E}\,Y^2}{(\mathbb{E}\,Y)^2} = \frac{1+\sum_{k,j<n} N(k,j)\,\mathbb{P}(\mathcal{H}_2 \mid \mathcal{H}_1)}{\mathbb{E}\,Y}$$

where $N(k,j)$ is the number of different H_2 (or, to be precise, sets of pairs corresponding to H_2) overlapping any given cycle H_1 with particular values of k and j. The term 1 accounts for the case $H_2 = H_1$, when $k=j=n$.

Since $\mathbb{E}\,Y$ was given precisely in Section 4, all that is left to evaluate is $N(k,j)$. Note that the cardinality of the set S_1 of plain vertices with three half-edges contained in $H_1 \cup H_2$ is $2k-2j$, and for the set S_2 of plain vertices with four half-edges, it is $n-2k+j$.

By elementary counting, the number of ways to place $k-j$ strings referred to above onto H_1 is

$$\frac{n}{k}\binom{k}{j}\binom{n-k-1}{k-j-1}.$$

Here the first factor converts the problem from a cyclical one to that of counting linear arrangements of the n coloured vertices (as unlabelled objects), with k of them designated as special (members of strings), in which the first coloured vertex of the linear arrangement is one of the k special ones, and the special ones form exactly $k-j$ strings (with at least one non-special one between different strings). The second factor is for deciding the relative positions of the k coloured vertices in a sequence of $k-j$ strings and the third is for deciding the positions of the vertices not in strings.

For convenience, we will count oriented versions of H_2 and divide by 2 at the end. The number of ways to choose the half-edges being used by H_2 at each plain vertex is $(d-2)^{|S_1|}((d-2)(d-3))^{|S_2|}$, and $(d-2)^{n-k}$ for the half-edges at the set S_3 of coloured vertices not used by H_1 (obeying the traffic rules). These choices determine the direction that H_2 passes through each vertex except for those in S_1, which are determined by the direction it passes through each string. These directions can be chosen in 2^{k-j} ways, giving

$$2^{k-j}(d-2)^{2n-k-j}(d-3)^{n-2k+j}$$

ways to make these choices.

Now that the order of half-edges used by H_2 is determined at each vertex, it remains to choose the remaining $2(n-k)$ pairs of the pairing. These pairs must connect the strings and the vertices in $S_2 \cup S_3$ into a Hamilton cycle in a bipartite fashion (regarding each string as a vertex) and connecting the 'out' half-edge at a vertex to the 'in' one at the next. The number of such choices of pairs is

$$(n-k)!(n-k-1)!.$$

Multiplying all the displayed factors together and dividing by 2 to un-orient H_2 gives $N(k,j)$. Combining with the earlier equations then produces

$$\frac{\mathbb{E}\,Y^2}{(\mathbb{E}\,Y)^2} = \frac{1}{\mathbb{E}\,Y} + \sum_{k<n}\sum_{j<k} f(n,d,k,j) \tag{6.1}$$

where

$$f(n,d,k,j) =$$
$$\frac{n^2(k-1)!\,(n-k)!^{\,3}2^{k-j}(d-2)^{2n-k-j}(d-3)^{n-2k+j}\big((d-4)n+2k\big)!\,(dn)!}{(n-k)^2(k-j)!\,(k-j-1)!\,j!\,(n-2k+j)!\,\big((d-2)n\big)!^{\,2}d^{2n}(d-1)^n n!^2}$$

As usual, upon applying Stirling's formula we find that the powers of n/e cancel and we are left with

$$f(n,d,k,j) = f_0(n,d,k,j)G^n\Big(1+O\Big(\frac{1}{j+1}+\frac{1}{k-j}+\frac{1}{n-k}\Big)\Big) \tag{6.2}$$

where

$$f_0(n,d,k,j) = \frac{\sqrt{(dn-4n+2k)nd}}{2\pi(d-2)\sqrt{k(n-k)j(n-2k+j)}}$$

and (with $\kappa = k/n$, $\gamma = j/n$)

$$\begin{aligned} G &= \frac{2^{\kappa-\gamma} d^{d-2} (d-3)^{1+\gamma-2\kappa} (d-4+2\kappa)^{d-4+2\kappa} \kappa^\kappa (1-\kappa)^{3-3\kappa}}{(d-1)(d-2)^{2d-6+\kappa+\gamma} \gamma^\gamma (\kappa-\gamma)^{2\kappa-2\gamma} (1-2\kappa+\gamma)^{1-2\kappa+\gamma}} \\ &= F(\alpha,\delta) := \frac{2^\alpha (t+2)^t (t-1)^{\delta-\alpha} (t-2\delta)^{t-2\delta} (1-\delta)^{1-\delta} \delta^{3\delta}}{(t+1) t^{2t-2\delta-\alpha} (1-\delta-\alpha)^{1-\delta-\alpha} \alpha^{2\alpha} (\delta-\alpha)^{\delta-\alpha}} \end{aligned} \tag{6.3}$$

where $\delta = 1-\kappa$, $\alpha = \kappa-\gamma$ and $t = d-2$.

We seek the maximum value of F in the triangle

$$T = \big\{(\delta,\alpha) : 0 \le \alpha \le \delta,\ \delta+\alpha \le 1\big\}.$$

The partial derivatives of $\ln F$ are

$$\frac{\partial(\ln F)}{\partial\delta} = \ln \frac{t^2(t-1)\delta^3(1-\delta-\alpha)}{(t-2\delta)^2(1-\delta)(\delta-\alpha)},$$

$$\frac{\partial(\ln F)}{\partial\alpha} = \ln \frac{2t(\delta-\alpha)(1-\delta-\alpha)}{(t-1)\alpha^2}. \tag{6.4}$$

Setting these equal to zero gives necessary conditions for a stationary point of F:

$$t^2(t-1)\delta^3(1-\delta-\alpha) = (t-2\delta)^2(1-\delta)(\delta-\alpha) \tag{6.5}$$

and

$$(t+1)\alpha^2 - 2t\alpha + 2t\delta(1-\delta) = 0. \tag{6.6}$$

Solving the first equation for α and substituting this into the second shows that the value of δ at a stationary point in the interior of T must be a root of

$$g(\delta) = (1-\delta)\big((t-2\delta)^2 - t^2(t-1)\delta^2\big)^2 - 2t^3\delta(1-2\delta)^2(t-2\delta)^2, \tag{6.7}$$

which is quintic in δ. Now g factorises as

$$g(\delta) = (t+2)(\delta_0-\delta)h(\delta) \tag{6.8}$$

where

$$\begin{aligned} h(\delta) &= t^5\delta^4 - 2\delta(2\delta^3+\delta^2-2\delta+1)t^4 + (9\delta^4 - 12\delta^3 + 6\delta^2 + 1)t^3 \\ &\quad + 2\delta(\delta-1)(3\delta^2+2\delta+3)t^2 - 4\delta^2(\delta+3)(\delta-1)t + 8\delta^3(\delta-1), \end{aligned}$$

and $\delta_0 = t/(t+2)$ which, as we will show, determines the maximum value of F in T.

The second derivative of h with respect to δ is

$$\begin{aligned} &12t^5\delta^2 + (-48\delta^2 - 12\delta + 8)t^4 + 12(9\delta^2 - 6\delta + 1)t^3 \\ &+ (72\delta^2 - 12\delta + 4)t^2 + (-48\delta^2 - 48\delta + 24)t + 96\delta^2 - 48\delta. \end{aligned}$$

This can be rearranged as

$$12(t-6)t^4\delta^2+(24\delta^2-12\delta+8)t^4+12(9\delta^2-6\delta+1)t^3+12\delta(t^2-4)$$
$$+12\delta^2(t^2-6t)+4(15\delta^2-6\delta+1)t^2+24(\delta^2-2\delta+1)t+96\delta^2,$$

in which each collected term is clearly nonnegative for $t \geq 6$ and all $\delta \geq 0$. Thus h is a convex function of δ for each t.

From (6.7) we compute firstly $g(1/2) = (t-1)^2(t-2)^4/32 > 0$, secondly

$$g(1/\sqrt{t}) < t^2 - 2t^{5/2}(1-2/\sqrt{t})^2(t-1)^2 < 0$$

since $t \geq 6$ and $25(1-2/\sqrt{6})^2 > 0.8$, and thirdly, since $t \geq 6$ implies $t^2-7t+8>0$ and thus $1-1/t < 2(1-2/t)^2$,

$$g(1/t) < (1-1/t)\big((t-2/t)^4 - t^2(t-2/t)^2\big) < 0.$$

Note that h has the same sign as g for $0 \leq \delta \leq 1/2$ as $\delta_0 > 1/2$. Thus $h(1/2)$ is positive and $h(1/\sqrt{t})$ and $h(1/t)$ are negative. So by the convexity of h, taking into consideration that $h(0) > 0$, the only zeros of h for $0 < \delta < 1$ lie in $0 < \delta < 1/t$ or $1/\sqrt{t} < \delta < 1/2$. We show separately that these two subsets of T can hold no stationary points of F.

Case 1: $0 < \delta < 1/t$
Substituting $\alpha = \beta\delta$ into $\ln F$, and taking the second derivative with respect to δ, we obtain

$$\frac{\partial^2 \ln F(\beta\delta,\delta)}{\partial\delta^2} = \frac{2t-\beta t-4t\delta+2t\delta^2-2\beta t\delta+2\beta t\delta^2+2\beta\delta}{(1-\delta)(1-\delta-\beta\delta)\delta(t-2\delta)}$$

Since $\alpha \leq \delta$ in T, we have $\beta \leq 1$ and so the factors in the denominator are all positive. The numerator is at least $2t-t-4t\delta-2t\delta > 0$ as $\delta < 1/t \leq 1/6$. Hence, $\ln F$ can have no local maximum in T for such δ.

Case 2: $1/\sqrt{t} < \delta < 1/2$
For such δ, from (6.6) we obtain $2t\alpha > 2t\delta(1-\delta)$, i.e. $\alpha > \delta(1-\delta)$. So, using (6.5), at a stationary point

$$\frac{(1-\delta)^2}{\delta^2} = \frac{1-\delta-\delta(1-\delta)}{\delta-\delta(1-\delta)} < \frac{1-\delta-\alpha}{\delta-\alpha} = \frac{(t-2\delta)^2(1-\delta)}{t^2(t-1)\delta^3}$$

and so, since $\delta^2 > 1/t$,

$$\frac{1-\delta}{\delta} < \frac{(t-2\delta)^2}{t(t-1)}.$$

But this fails at $\delta = 1/2$, and the derivative of the left hand side with respect to δ is less than -4 for $\delta < 1/2$, whilst that of the right is easily greater than -4. So the inequality fails, and there is no such stationary point.

We conclude that $\delta = \delta_0$ determines the unique local maximum in the interior of T. The boundary of T must also be investigated. Considering (6.4), there is no local maximum at a boundary point with $0 < \delta < 1$, since $\partial(\ln F)/\partial\alpha$ tends to ∞ as α tends to 0 from the right, and to $-\infty$ as α tends to $\min(\delta, 1-\delta)$ from the left. A similar argument applies to eliminate $(\alpha, \delta) = (0, 1)$ from consideration, since moving along the boundary where $\alpha = 1 - \delta$, F is a smooth function times $\alpha^{-\alpha}$. This leaves only the point $\alpha = \delta = 0$, which is indeed a local maximum, with value $F(0, 0) = (1 + 2/t)^t/(t + 1) < 1$ for $t \geq 6$.

To deduce that δ_0 determines the unique global maximum in T, we only need to observe that the corresponding value of α is $\alpha_0 = 2\delta_0/(t + 1)$, and that $F(\alpha_0, \delta_0) = 1$.

The rest of the argument is totally standard for such variance calculations, as in [6, Section 4] for example, so we omit the justifications. The point (α_0, δ_0) corresponds to $\kappa = \kappa_0 = 2/d$, $\gamma = \kappa_0/(d-1)$. Putting $\kappa = \kappa_0 + \hat{\kappa}/\sqrt{n}$ and $\gamma = \gamma_0 + \hat{\gamma}/\sqrt{n}$, and expanding $\ln(G^n)$ (G defined in (6.3)) about $\hat{\kappa} = \hat{\gamma} = 0$, we find up to quadratic terms in $\hat{\kappa}$ and $\hat{\gamma}$

$$G^n \approx e^{c_1\hat{\kappa}^2 + c_2\hat{\kappa}\hat{\gamma} + c_3\hat{\gamma}^2}$$

where

$$\begin{aligned} c_1 &= -\frac{d(d^3 - 3d^2 + 4d + 4)}{4(d-2)^2(d-3)} \\ c_2 &= \frac{d(d-1)^2}{(d-2)(d-3)} \\ c_3 &= -\frac{d(d-1)^2}{4(d-3)}. \end{aligned}$$

Here c_3 is clearly negative, and the determinant $D = 4c_1c_3 - c_2^2$ of the Hessian of the quadratic form is positive, as we expect since the expansion is at a local maximum. The routine argument now gives from (6.2)

$$\begin{aligned} \sum_{k,j<n} f(n, d, k, j) &\sim f_0(n, d, n\kappa_0, n\gamma_0) 2\pi n/\sqrt{D} \\ &= \sqrt{\frac{d}{d-4}}. \end{aligned}$$

Recalling (6.1) and that $\mathbb{E}\, Y \to \infty$ now establishes Lemma 3.4. $\square$

7. Rainbow matchings

In this section we briefly consider the analoguous problem of the existence of a rainbow perfect matching in a randomly coloured random regular graph. We will omit the details of the calculations.

The model is now slightly different. We consider a random regular graph $G(2n, d)$ with an even number of vertices (d may now be arbitrary), and colour randomly the nd edges with n colours, d edges of each colour. We then ask whether there exists a rainbow matching consisting of n disjoint edges of different colours.

We can translate this to a random bipartite (multi)graph as above; now the bipartite graph has $2n$ plain vertices of degree d and n coloured vertices of degree $2d$. Let Z be the number of rainbow perfect matchings; in the bipartite version, Z is the number of decompositions of the graph into n vertex-disjoint paths of length 2, with 2 plain and 1 coloured vertex each.

Calculations as above yield

$$\mathbb{E}\, Z = d^{3n} \frac{(2n)!\,\big((2d-2)n\big)!}{(2dn)!} = \Theta\left(n^{1/2} \Big(\frac{(d-1)^{2d-2}}{d^{2d-3}} \Big)^n \right)$$

and it is easily checked that, as $n \to \infty$, $\mathbb{E}\, Z \to 0$ for $d \leq 6$, while $\mathbb{E}\, Z \to \infty$ for $d \geq 7$. In particular, for $d \leq 6$ there is **whp** no rainbow perfect matching.

Furthermore, for $d \geq 7$, an argument similar to the one in [2] (and much simpler than the proof of Lemma 3.4 above, since we only need to maximize over one variable) yields

$$\frac{\mathbb{E}(Z^2)}{(\mathbb{E}\, Z)^2} \to \frac{d-1}{\sqrt{d(d-3)}}.$$

Finally, defining X_{ij} as before, Theorem 3.1 applies when $d \geq 7$ with

$$\lambda_{ij} = \frac{1}{2i} \binom{i}{j} 2^j (d-1)^{i+j},$$
$$\delta_{ij} = \frac{(-1)^{i+j}}{(d-1)^{i+j}}.$$

Hence there exists a rainbow perfect matching **whp** when $d \geq 7$.

By analogy with the open problem in Remark 1.2 one might further ask whether there exists a decomposition into d rainbow perfect matchings. Computing the expected number of such decompositions reveals that when $d \leq 11$, **whp** no such decomposition exists. This compares with the corresponding result for uncoloured graphs, that $G(2n, d)$ **whp** has a decomposition into d perfect matchings as soon as $d \geq 3$ [12, 15].

References

[1] B. Bollobás, *Random Graphs.* Academic Press, New York, 1985; Second ed. Cambridge University Press, Cambridge, 2001.

[2] B. Bollobás & B.D. McKay, The number of matchings in random regular graphs and bipartite graphs. *J. Combin. Theory Ser. B* **41** (1986), no. 1, 80–91.

[3] C. Cooper & A. Frieze, Multi-coloured Hamilton cycles in randomly coloured random graphs. *Combin. Probab. Comput.* **11** (2002), 129–134.

[4] C. Cooper, A. Frieze, M. Molloy & B. Reed, Perfect matchings in random r-regular, s-uniform hypergraphs. *Combin. Probab. Comput.* **5** (1996), 1–15.

[5] A. Frieze & S. Janson Perfect matchings in random s-uniform hypergraphs. *Random Strucures Algorithms* **7** (1995), 41–57.

[6] A. Frieze, M. Jerrum, M. Molloy, R. Robinson & N. Wormald, Generating and counting Hamilton cycles in random regular graphs, *J. Algorithms* **21** (1996), 176–198.

[7] A. Gamburd, Poisson-Dirichlet distribution for random Belyi surfaces. Preprint, 2005. arXiv:math.PR/0501283

[8] H. Garmo, Random railways modeled as random 3-regular graphs. *Random Strucures Algorithms* **9** (1996), no. 1-2, 113–136.

[9] H. Garmo, Asymptotic properties of the connectivity number of random railways. *Adv. Appl. Probab.* **31** (1999), no. 3, 720–741.

[10] H. Garmo, S. Janson & M. Karoński, On generalized random railways. *Combin. Probab. Comput.* **13** (2004), no. 1, 31–35.

[11] C. Greenhill, S. Janson, J.H. Kim & N.C. Wormald, Permutation pseudographs and contiguity. *Combin. Probab. Comput.* **11** (2002), no. 3, 273–298.

[12] S. Janson, Random regular graphs: asymptotic distributions and contiguity. *Combin. Probab. Comput.* **4** (1995), 369–405.

[13] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[14] J.H. Kim & N.C. Wormald, Random matchings which induce Hamilton cycles, and hamiltonian decompositions of random regular graphs, *J. Combin. Theory Ser. B* **81** (2001), 20–44.

[15] M. Molloy, H. Robalewska, R.W. Robinson & N.C. Wormald, 1-factorisations of random regular graphs. *Random Strucures Algorithms* **10** (1997), 305–321.

[16] R.W. Robinson & N.C. Wormald, Almost all cubic graphs are hamiltonian, *Random Strucures Algorithms* **3** (1992), 117–125.

[17] R.W. Robinson & N.C. Wormald, Almost all regular graphs are hamiltonian, *Random Strucures Algorithms* **5** (1994), 363–374.

[18] R.W. Robinson & N.C. Wormald, Hamilton cycles containing randomly selected edges in random regular graphs, *Random Strucures Algorithms* **19** (2001), 128–147.

[19] N.C. Wormald, Models of random regular graphs. *Surveys in Combinatorics 1999*, eds. J.D. Lamb & D.A. Preece, LMS Lecture Note Series **267**, Cambridge University Press, Cambridge, 1999, 239–298.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`

Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, Canada N2L 3G1

E-mail address: `nwormald@uwaterloo.ca`

URL: `http://www.math.uwaterloo.ca/~nwormald/`