Almost All Cubic Graphs are Hamiltonian

R. W. Robinson*
Computer Science Department
University of Georgia
Athens, GA 30602, U.S.A.

N. C. Wormald†
Department of Mathematics
University of Melbourne
Parkville, VIC 3052, Australia

Abstract

In a previous paper the authors showed that at least 98.4% of large labelled cubic graphs are hamiltonian. In the present paper, this is improved to 100% in the limit by asymptotic analysis of the variance of the number of Hamilton cycles with respect to populations of cubic graphs with fixed numbers of short odd cycles.

*Research supported by the National Science Foundation under contract MCS-8302282.
†Research supported by the Australian Research Council

1 Introduction

It was first proved by Read [4] that the number M of labelled cubic graphs on $2n$ vertices is asymptotically

$$M \sim \frac{(6n)!}{e^2(3n)!3^{2n}2^{5n}},$$

where, throughout this paper, asymptotics are for $n \to \infty$, unless otherwise specified. Turn this set of graphs into a probability space $\Omega = \Omega_n$ with the uniform distribution. The authors showed in an earlier paper [5, Theorem 2.4] that the expectation and variance of the number H of Hamilton cycles of a graph $G \in \Omega$ are given asymptotically by

$$\mathrm{Exp} H \sim \frac{e\sqrt{\pi}}{2\sqrt{n}} \left(\frac{4}{3}\right)^n, \tag{1.1}$$

$$\mathrm{Var} H \sim \left(\frac{3}{e} - 1\right) (\mathrm{Exp} H)^2. \tag{1.2}$$

From this, it can be deduced using the standard second moment method that asymptotically at least $2 - 3e^{-1}$ of all cubic graphs are hamiltonian. This bound was improved in [5] to $2 - 3e^{-13/12}$, using computations in the space of all triangle-free cubic graphs. The aim of that exercise was to get as close as possible to showing the following result, whose truth has been suspected for some time (see also Frieze [3]), and is finally demonstrated in the present paper.

Theorem. *If H is the number of Hamilton cycles in a cubic graph chosen uniformly at random from all labelled cubic graphs on $2n$ vertices, then* $\lim_{n\to\infty} Pr(H > 0) = 1$.

For $G \in \Omega$ let $X_i(G)$ denote the number of cycles of length i in G. The following is now well known. It was first proved in [6] and is found in Bollobás [2] and in [7].

Lemma 1. *For any fixed k the variables X_i, $3 \le i \le k$, are asymptotically independent Poisson random variables, with*

$$\mathrm{Exp} X_i \sim \frac{2^{i-1}}{i}. \tag{1.3}$$

The approach of the present paper is to divide the cubic graphs in Ω into groups according to the values of the variables X_i, then refine the second moment method for these groups. It turns out that the variable X_i for even i has no effect asymptotically in our argument, so the groups studied are characterized by the sequence $c_1, \ldots, c_b$ for fixed b, where $c_i = X_{2i+1}$ for $i = 1, \ldots, b$. It will be shown that by taking b to be large, the variance of the group means can be made arbitrarily close (in proportion) to the total variance $\mathrm{Var}H$, and thus the mean of the group variances can be made arbitrarily small compared to $\mathrm{Var}H$. As a consequence of this, after showing that the group means are sufficiently well behaved, it is deduced that $\Pr(H = 0)$ is bounded above asymptotically by $\epsilon(b)$, where

$$\lim_{b\to\infty} \epsilon(b) = 0.$$

An alternative way of bounding the group variances is to compute them directly using the techniques in [5] combined with those in the proof of Lemma 2 of the present paper, but this would appear to be rather more complicated.

Note that since X_i plays no role when i is even, this gives a broader perspective to the improvement obtained in [5] by considering triangle-free graphs. In addition it provides an intuitive basis for the fact also proved in [5] that $VarH = o((ExpH)^2)$ among labelled bipartite cubic graphs, from which it follows at once that almost all are hamiltonian.

Using techniques related to those of the present paper it can be shown that almost all k-regular graphs are hamiltonian, for any fixed $k \geq 3$. This and similar topics will be pursued in a subsequent publication.

2 Proof of the Theorem

Define

$$\lambda_i = \frac{4^i}{2i+1}, \qquad \mu_i = 1 - 4^{-i}.$$

Note that for fixed i, Lemma 1 implies $\mathrm{Exp}X_{2i+1} \sim \lambda_i$.

Define $S(y,b)$ to be the event that $X_{2i+1} < \lambda_i + y\lambda_i^{2/3}$ for $1 \leq i \leq b$. We will show

$$\Pr(\bar{S}(y,b)) = O(e^{-y/4}) + o(1) \tag{2.1}$$

where $\bar{S}(y,b)$ denotes the complement of $S(y,b)$. We will bound the probability of non-hamiltonicity by way of the obvious inequality

$$\Pr(H=0) \leq \Pr(\{H=0\} \wedge S(y,b)) + \Pr(\bar{S}(y,b)). \tag{2.2}$$

Throughout the paper, unless y and b are explicitly required to vary with n, the asymptotics are for y and b fixed as n goes to ∞. In (2.1) the constant implicit in $O()$ is uniform over all b and y, whereas the constant function implicit in $o()$ may depend on both y and b.

Define the group mean

$$E_{c_1,\ldots,c_b} = \mathrm{Exp}(H|X_3 = c_1, \ldots, X_{2b+1} = c_b),$$

and the group variance

$$V_{c_1,\ldots,c_b} = \mathrm{Exp}(H^2|X_3 = c_1, X_5 = c_2, \ldots, X_{2b+1} = c_b) - E^2_{c_1,\ldots,c_b}.$$

One of our main steps in the proof is to establish the following.

Lemma 2. *For fixed $b \geq 3$,*

$$E_{c_1,\ldots,c_b} \sim \mathrm{Exp}H \prod_{i=1}^{b} \mu_i^{c_i} e^{\lambda_i(1-\mu_i)}.$$

The calculations in the proof of Lemma 2 can also be applied for short cycles of even length, but it is then seen that these have an asymptotically insignificant effect on the expected number of Hamilton cycles in G.

Lemma 2 will help us to prove the inequality

$$\mathrm{Exp}V_{X_3,\ldots,X_{2b+1}} \geq (1+o(1))(\mathrm{Exp}H)^2 \Pr(\{H=0\} \wedge S(y,b)) e^{-O(y)}, \tag{2.3}$$

where the constant implicit in $O()$ is independent of y and b.

We have

$$\begin{aligned}
\mathrm{Var}H &= \mathrm{Exp}V_{X_3,\ldots,X_{2b+1}} + \mathrm{Var}E_{X_3,\ldots,X_{2b+1}} \\
&= \mathrm{Exp}V_{X_3,\ldots,X_{2b+1}} + \mathrm{Exp}E^2_{X_3,X_5,\ldots,X_{2b+1}} - (\mathrm{Exp}H)^2,
\end{aligned}$$

and we will show that

$$\mathrm{Exp}E^2_{X_3,X_5,\ldots,X_{2b+1}} \geq \frac{3}{e}(1 - O(e^{-b}) + o(1))(\mathrm{Exp}H)^2 \tag{2.4}$$

where the constant implicit in $O()$ is uniform over all $b \geq 1$. Hence (1.2) implies that

$$\mathrm{Exp} V_{X_3,\ldots,X_{2b+1}} = (O(e^{-b}) + o(1))(\mathrm{Exp} H)^2, \tag{2.5}$$

where the constant implicit in $O()$ is uniform over all $b \geq 1$.

From (2.1) – (2.3) it follows that for $y \geq 1$ and $b \geq 1$,

$$\Pr(H = 0) = O(e^{-y/4}) + e^{O(y)} \mathrm{Exp} V_{X_3,\ldots,X_{2b+1}} / (\mathrm{Exp} H)^2 + o(1)$$

where the implied constants in each $O()$ are independent of y and b. Hence by (2.5),

$$\Pr(H = 0) = O(e^{-y/4} + e^{O(y)-b}) + o(1),$$

where each $O()$ is independent of y and b but $o(1)$ is not. On choosing, say, $b = y^2$, we obtain

$$\limsup \Pr(H = 0) = O(e^{-y/4}).$$

Since y can be chosen arbitrarily large, the theorem follows. Another way to argue this last step is to let $y \to \infty$ sufficiently slowly with n, so that the $O(e^{-y/4})$ term dominates the $o(1)$ term. This can be done with y fixed in all parts of the proof except for this last step.

To complete the proof, it is required to establish Lemma 2 as well as (2.1), (2.3) and (2.4).

Proof of Lemma 2. We first show that

$$\mathrm{Exp}(H X_m) \sim \nu_m \mathrm{Exp} H \tag{2.6}$$

where

$$\nu_m = \begin{cases} 2^{m-1}/m & \text{if } m \text{ is even} \\ \lambda_i \mu_i & \text{if } m \text{ is odd, } m = 2i+1. \end{cases}$$

This requires only some modifications of the derivation of (1.1) as given in [5]. For that, we counted the cubic graphs which possess a prescribed Hamilton cycle (the number is asymptotic to $(2n)!/(n!2^n e)$ by the result of Bender and Canfield [1, Theorem 1] or by an elementary argument using inclusion-exclusion), multiplied by the number of possible Hamilton cycles $((2n)!/4n)$, and divided by M. The present aim is to count cubic graphs G with a given Hamilton cycle D once for every m-cycle C that they contain, and again divide by M. It is clear that there must be s edges of C which are not contained in D for some s in the range $1 \leq s \leq m/2$, since these edges must be mutually non-adjacent and not all edges of C can lie in D.

We claim that the number of choices for C, given s, is asymptotic to $(4n)^s/2m$. This claim is most easily established by regarding C to be cyclically oriented and to have a special vertex distinguished, called the *initial* vertex. This has the effect of multiplying the numbers in the enumeration by $2m$, and induces a labelling $1, 2, \ldots, m$ of the edges of C around C in the direction of orientation, beginning with the initial vertex. We classify C according to labels of the edges of C which are not contained in D. By Lemma 2.1 of [5], there are exactly

$$\frac{m}{s}\binom{m-s-1}{s-1}$$

possible choices of s non-adjacent edges from an m-cycle, and hence this is the number of choices for the labels of the edges of C not in D. So suppose that these labels have been chosen. We now choose the cycle C by choosing the vertices $v_1, \ldots, v_m$ on D which are in C one by one around C, beginning with the initial vertex of C. Firstly, assume that the edge immediately following the initial vertex is in D. Then the number of choices for the vertex v_i will be either

(a) at most $2n$, if $i = 1$, or if v_i is a vertex immediately following an edge of C not in D except for the last such edge; in total this happens s times, or

(b) at most 2, if v_i immediately follows any of the vertices in (a); so this happens s times also, or

(c) at most 1, otherwise, since v_i is then the next vertex around D in a predetermined direction.

In case (a) there are always at least $2n - m^2/2$ choices which place v_i so that its distance from $v_1, \ldots, v_{i-1}$ is at least $m - i + 1$. This guarantees exactly 2 and 1 choices in the cases (b) and (c), respectively, for the following vertices being determined until the next occurrence of case (a) or the final (forced) move back to the initial vertex of C. Thus the number of choices for $v_1, \ldots, v_m$ is bounded below by $(4n - m^2)^s$ and above by $(4n)^s$, which are asymptotically equal for fixed m. On the other hand, if the edge immediately following the initial vertex is not in D, the same asymptotic result is obtained. This finishes the proof of the claim.

When C has been chosen, the number of ways of filling in the rest of G, avoiding edges already present, is asymptotic to $(2n-2s)!/((n-s)!2^{n-s}e)$, for the same reasons as the factor $(2n)!/(n!2^n e)$ in the calculation of Exp$H$. Comparing with the calculation for Exp$H$ and summing over all possible s-sets of excluded edges of the m-cycle and all possible values of s, we have (with square brackets denoting the extraction of coefficients)

$$
\begin{aligned}
\frac{\mathrm{Exp}(HX_m)}{\mathrm{Exp}H} &\sim \sum_{1\le s\le m/2} \frac{2^{s-1}}{s}\binom{m-s-1}{s-1},\\
\frac{\mathrm{Exp}(HX_m)}{2^{m-1}\mathrm{Exp}H} &\sim \sum_{1\le s\le m/2} \frac{2^{s-m}}{m-s}\binom{m-s}{s}\\
&= -\frac{2^{-m}}{m} + \sum_{0\le s\le m/2} \frac{2^{s-m}}{m-s}\binom{m-s}{s}\\
&= -\frac{2^{-m}}{m} + [x^m]\sum_{0\le s\le m/2} \frac{1}{m-s}\left(\frac{x(1+x)}{2}\right)^{m-s}\\
&= -\frac{2^{-m}}{m} + [x^m]\sum_{j=1}^{\infty} \frac{1}{j}\left(\frac{x(1+x)}{2}\right)^{j}\\
&= -\frac{2^{-m}}{m} + [x^m]\Big(-\ln(1-x)-\ln(1+x/2)\Big)\\
&= \frac{-2^{-m}+1+(-1)^m 2^{-m}}{m}.
\end{aligned}
$$

This gives (2.6).

Define the uniform probability space $\Phi = \Phi_n$ of all labelled cubic graphs on $2n$ vertices with a distinguished Hamilton cycle. Note that the number of elements of Φ is MExp$H$. We use the subscript Φ on Pr and Exp to distinguish references to this space from those to Ω, which remain unsubscripted. Extend the definition of X_i in the obvious way to Φ. Since $\mathrm{Exp}_\Phi X_m = \mathrm{Exp}(HX_m)/\mathrm{Exp}H$, it follows from (2.6) that

$$\mathrm{Exp}_\Phi X_m \sim \nu_m. \qquad (2.7)$$

Next, repeat the above method of calculation for the joint factorial moment to show that

$$\mathrm{Exp}_\Phi([X_3]_{i_3}\cdots[X_{2b+1}]_{i_{2b+1}}) \sim \prod_{m=3}^{2b+1} \nu_m^{i_m}, \tag{2.8}$$

where $[x]_i$ denotes $x(x-1)\cdots(x-i+1)$. For this, one needs to count cubic graphs G with a given Hamilton cycle D once for every ordered set of i_3 3-cycles, ..., i_{2b+1} $(2b+1)$-cycles that they contain, and divide by M. Classify the j'th cycle in the ordered set according to the number s_j of edges of D which it does not contain. The resulting asymptotic expression for $\mathrm{Exp}(H[X_3]_{i_3}\cdots[X_{2b+1}]_{i_{2b+1}})/\mathrm{Exp}H$ factorises into summations over the separate variables s_j, yielding (2.8) as a result.

From (2.8) by inclusion-exclusion, we deduce that $X_3,\ldots,X_{2b+1}$ are asymptotically independent Poisson random variables in Φ with expectations given by (2.7). (See Bollobás [2, p.23] for a statement of the appropriate principle.) Thus, if we now restrict our attention to odd cycle lengths,

$$\Pr_\Phi(X_3=c_1,\ldots,X_{2b+1}=c_b) \sim \prod_{i=1}^{b} (\lambda_i\mu_i)^{c_i} e^{-\lambda_i\mu_i}/c_i!.$$

Note that

$$\Pr_\Phi(X_3=c_1,\ldots,X_{2b+1}=c_b) = \frac{\sum_h h\Pr(H=h, X_3=c_1,\ldots,X_{2b+1}=c_b)}{\mathrm{Exp}H},$$

$$E_{c_1,\ldots,c_b} = \frac{\sum_h h\Pr(H=h, X_3=c_1,\ldots,X_{2b+1}=c_b)}{\Pr(X_3=c_1,\ldots,X_{2b+1}=c_b)},$$

and, from Lemma 1,

$$\Pr(X_3=c_1,\ldots,X_{2b+1}=c_b) \sim \prod_{i=1}^{b} e^{-\lambda_i}\lambda_i^{c_i}/c_i!. \tag{2.9}$$

The last four relations combine to give Lemma 2. ∎

To show that $\Pr(S(x,b))$ tends to 1 as x increases, and that certain other events occur almost surely, we use the following loose bound.

Lemma 3. *Let $\eta_1,\eta_2,\ldots$ be given. Suppose that $\eta_1>0$ and that for some $c>1$, $\eta_{i+1}/\eta_i>c$ for all $i\ge 1$. Then uniformly over $x\ge 1$,*

$$R(x) = \sum_{i=1}^{\infty}\sum_{t=\eta_i(1+y_i)}^{\infty} \frac{\eta_i^t}{t!e^{\eta_i}} = O(e^{-c_0 x})$$

where $y_i = x\eta_i^{-1/3}$ *and* $c_0 = \min\{\eta_1^{1/3}, \eta_1^{2/3}\}/4$.

Proof. Apply Stirling's formula to the first term in the summation over t, giving $(\eta_i(1+y_i))^{-1/2}e^{\eta_i(y_i-(1+y_i)\log(1+y_i))}$, and bound the rest by a geometric series with common ratio $1/(1+y_i)$. This yields

$$R(x) = O(1)\sum_{i=1}^{\infty}\frac{\sqrt{1+y_i}}{y_i\sqrt{\eta_i}}e^{\eta_i(y_i-(1+y_i)\log(1+y_i))}.$$

It is readily verified that $(1+y_i)\log(1+y_i)) - y_i \geq \min\{y_i, y_i^2\}/4$ for all $y_i \geq 0$. Hence

$$R(x) = O(1)\sum_{i=1}^{\infty} x^{-1/2}\eta_i^{-1/6}e^{-x\min\{\eta_i^{1/3},\eta_i^{2/3}\}/4}.$$

The bound on the term for $i=1$ dominates, which gives the lemma. ∎

Proof of (2.1). This follows immediately from Lemma 1 and Lemma 3 with $\eta_i = \lambda_i$, $x = y$. ∎

Proof of (2.4). For any real x and non-negative integer b we have

$$\begin{aligned}\mathrm{Exp}E^2_{X_3,X_5,\ldots,X_{2b+1}} &= E_S + \Pr(\bar{S}(x,b))\mathrm{Exp}(E^2_{X_3,\ldots,X_{2b+1}}|\bar{S}(x,b)) \\ &\geq E_S\end{aligned}$$

where

$$\begin{aligned}E_S &= \sum_{c_1,\ldots,c_b\in S(x,b)} E^2_{c_1,\ldots,c_b}\Pr(X_3=c_1, X_5=c_2,\ldots,X_{2b+1}=c_b) \\ &\sim (\mathrm{Exp}H)^2\prod_{i=1}^{b}(1-Z_i)e^{\lambda_i(1-\mu_i)^2}\end{aligned}$$

by (2.9) and Lemma 2, with

$$Z_i = \sum_{t=\lambda_i+x\lambda_i^{2/3}}^{\infty}\frac{\lambda_i^t\mu_i^{2t}}{t!e^{\lambda_i\mu_i^2}}.$$

Since Z_i is strictly smaller than the summation over t in Lemma 3 with $\eta_i = \lambda_i\mu_i^2$, that lemma gives

$$\sum_{i=1}^{b} Z_i = O(e^{-x/5}).$$

As $\sum_{i\geq 0} 4^{-i}/(2i+1) = \log 3$, one also has

$$\prod_{i=1}^{b} e^{\lambda_i(1-\mu_i)^2} \;=\; \frac{3}{e}(1 - O(4^{-b}/b)).$$

Thus, taking $x \geq 5b$, we obtain (2.4). ∎

Proof of (2.3). Let Ind_S denote the indicator function for an event S. Let S denote $S(y,b)$, and note that for $G \in S$,

$$\begin{aligned} \prod_{i=1}^{b} \mu_i^{X_{2i+1}} e^{\lambda_i(1-\mu_i)} &\geq \prod_{i=1}^{b} e^{1/(2i+1)}(1-4^{-i})^{\lambda_i + y\lambda_i^{2/3}} \\ &\geq e^{-O(y)} \end{aligned}$$

by routine calculations using $\log \mu_i \geq -2^{-2i} - 2^{-4i}$.

The expected group variance can be written as

$$\begin{aligned} &\mathrm{Exp} V_{X_3,\ldots,X_{2b+1}} \\ &\quad= \mathrm{Exp}\Big(\mathrm{Exp}\Big((H - E_{X_3,X_5,\ldots,X_{2b+1}})^2 | X_3,\ldots,X_{2b+1}\Big)\Big) \\ &\quad\geq \mathrm{Exp}\Big(\mathrm{Exp}\Big((H - E_{X_3,X_5,\ldots,X_{2b+1}})^2 \mathrm{Ind}_{\{H=0\}\wedge S} | X_3,\ldots,X_{2b+1}\Big)\Big) \\ &\quad= \mathrm{Exp}\Big(\mathrm{Exp}\Big(E^2_{X_3,X_5,\ldots,X_{2b+1}} \mathrm{Ind}_{\{H=0\}\wedge S} | X_3,\ldots,X_{2b+1}\Big)\Big) \\ &\quad\sim (\mathrm{Exp} H)^2 \mathrm{Exp}\Big(\mathrm{Exp}\Big(\mathrm{Ind}_{\{H=0\}\wedge S} \prod_{i=1}^{b} \mu_i^{2X_{2i+1}} e^{2\lambda_i(1-\mu_i)} | X_3,\ldots,X_{2b+1}\Big)\Big) \\ &\quad\geq (\mathrm{Exp} H)^2 \mathrm{Exp}\Big(\mathrm{Exp}\Big(\mathrm{Ind}_{\{H=0\}\wedge S} e^{-O(y)} | X_3,\ldots,X_{2b+1}\Big)\Big) \\ &\quad\geq (\mathrm{Exp} H)^2 \Pr(\{H=0\} \wedge S) e^{-O(y)}, \end{aligned}$$

where the third-last step uses Lemma 2, and the second-last uses the bound derived above. ∎

References

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

[2] B. Bollobás, *Random graphs*, Academic Press, London, 1985.

[3] A. M. Frieze, Finding Hamilton cycles in sparse random graphs, *J. Combin. Theory Ser. B* **44** (1988), 230-250.

[4] R. C. Read, *Some Enumeration Problems in Graph Theory.* Doctoral thesis, London University, 1958.

[5] R. W. Robinson and N. C. Wormald, Existence of long cycles in random cubic graphs, in *Enumeration and Design* (D. M. Jackson and S. A. Vanstone, eds.), Academic Press, Toronto (1984), 251-270.

[6] N. C. Wormald, *Some Problems in the Enumeration of Labelled Graphs.* Doctoral thesis, Newcastle University, 1978.

[7] N. C. Wormald, The asymptotic distribution of short cycles in random regular graphs, *J. Combin. Theory Ser. B* **31** (1981), 168-182.