Pegging Graphs Yields a Small Diameter

STEFANIE GERKE[1], ANGELIKA STEGER[2]

and NICHOLAS WORMALD[3†]

[1]Mathematics Department, Royal Holloway College, University of London, Egham, TW20 0EX, UK
(e-mail: `stefanie.gerke@rhul.ac.uk`)

[2]Institute for Theoretical Computer Science, ETH Zurich, CH-8092 Zurich, Switzerland
(e-mail: `steger@inf.ethz.ch`)

[3]Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, N2L 3G1, Canada
(e-mail: `nwormald@uwaterloo.ca`)

We consider the following process for generating large random cubic graphs. Starting with
a given graph, repeatedly add edges that join the midpoints of two randomly chosen edges.
We show that the growing graph asymptotically almost surely has logarithmic diameter.
This process is motivated by a particular type of peer-to-peer network. Our method extends
to similar processes that generate regular graphs of higher degree.

1. The problem

Given a graph $G = (V, E)$ on $V = \{1, \ldots, n\}$ and two distinct edges $e = ab, f = cd \in E$, the
edges e and f are *pegged* if they are subdivided and a new edge is introduced between the
two new vertices. That is, two vertices $e' = n + 1$ and $f' = n + 2$ are added to V, the edges
e and f are deleted, and the edges ae', be', cf', df' and $e'f'$ are added. Note that if any
two edges of a 3-regular graph are pegged, then another 3-regular graph results. In this
paper we consider the following (random) pegging process $\mathcal{P}(G_0)$. Start with any fixed
graph G_0 with at least two edges. At each step $t \geqslant 1$, choose a pair of edges uniformly at
random from the set of all pairs of distinct edges of G_{t-1}, and peg this pair of edges to
obtain G_t. Let n_t denote the number of vertices of G_t, so that $n_t = n_0 + 2t$. We show that
asymptotically almost surely (a.a.s.), that is, with probability tending to one as t tends to
infinity, the diameter of G_t is at most $D \log t$, where D is an absolute constant.

The motivation for this process goes back to a peer-to-peer network introduced by Bourassa and Holt [2], called the SWAN network. The basic feature of this network is that it remains 4-regular, with any new node being added by inserting it at the midpoints of two existing edges. The reverse operation was used to delete nodes. Cooper, Dyer and Greenhill [3] showed under some mild simplifying assumptions that, restricted to the times when the network has n nodes, the stationary distribution is uniform on the set of 4-regular networks on n nodes. It then follows from results on uniformly random regular graphs that, as $n \to \infty$, these networks a.a.s. are 4-connected and have diameter logarithmic in the number of nodes [3].

Gao and the third author [5] recently introduced processes with the aim of approximating a rapidly growing SWAN peer-to-peer network. Processes were described there that produce random d-regular graphs for any $d \geqslant 3$; in the case of the 4-regular graph it merely consists of repeatedly adding nodes in the manner described above. For the case $d = 3$, this is the process $\mathcal{P}(G_0)$ described above. The short cycle distribution of the graphs produced by this process was studied in [5], and in [4] it was shown that they are a.a.s. d-connected. Probably the main outstanding unstudied parameter of these graphs, from the communication network point of view, was the diameter.

For most of this paper, we study the process $\mathcal{P}(G_0)$, which produces random cubic graphs if G_0 is cubic. In the final section we point out how our results extend to the general pegging processes for graphs of higher degree. Recall that G_t is the graph after t pegging steps in $\mathcal{P}(G_0)$, and $n_t = |V(G_t)|$.

Theorem 1.1. *Let G_0 be a fixed connected graph. Then the diameter of G_t in the pegging process $\mathcal{P}(G_0)$ is a.a.s. $O(\log t)$.*

Notes. **(1)** The constant implicit in the $O()$ notation is independent of G_0; that is, '$O(\log t)$' may be replaced by 'at most $D \log t$ for some absolute constant D'. The assumption that G_0 is connected is not necessary, as it was shown by Gao [4] that, starting with any G_0, the graph G_t is a.a.s. connected.

(2) It is immediate from the proof that the conclusion of the theorem holds even when G_0 is not fixed, provided only that it is connected and has at most Ct edges for some constant C.

We note that our usage of a.a.s. in conjunction with other asymptotic notation conforms to the conventions in [9]. With respect to a.a.s., if the parameter tending to infinity is not t, it will be specified. The base of log is always e if not otherwise specified.

Before starting the formal proof, we give some intuition as to why the result should be true. We will use the result of Bollobás and Chung [1], implying that the union of a random perfect matching and a Hamilton cycle on the same set of n vertices a.a.s. gives a (multi-) graph with diameter $O(\log n)$. To apply this for Theorem 1.1 we will first let the process continue until some time $t_0 < t$ so that the graph G_{t_0} is large, then define a cycle C that is closely related to G_{t_0}, then show that some of the subsequent pegging steps essentially create a random matching M of some (though not all) of the vertices of C, arguing that the resulting graph has small diameter via the above-mentioned result. There

will still remain the problem that some of the pegging steps insert extra new vertices in the edges of $C \cup M$, which may increase the diameter. We will show that such an increase is not significant by showing that the extra vertices do not increase the lengths of paths in $C \cup M$ by more than $O(\log t)$, and also that they all lie 'close' to vertices of $C \cup M$.

2. The proof

We start our proof of Theorem 1.1 with a result that will help to treat the last-mentioned issue above.

Lemma 2.1. *Let G_0 be a connected graph on n vertices and let G_t be the graph obtained after t steps in the pegging process $\mathcal{P}(G_0)$. Then with probability at least $1 - O(1/n^2)$, for all $t \leqslant \frac{1}{50}n$ every vertex in $V(G_t) \setminus V(G_0)$ has distance at most $4 \log_2 n$ to some vertex in $V(G_0)$.*

Proof. Define a cover of $E(G_t)$ by subsets S_i^t, $i = 1, \ldots, |E(G_0)|$, recursively as follows. For $t = 0$ each edge of G_0 is placed in a unique set S_i^0. If a pegging step at time t involves two edges $e = ab$ and $f = cd$ and two new vertices e' and f' are created, then ae' and be' are placed in the set that e belonged to, cf' and df' are placed in the set f belonged to, and the edge $e'f'$ is placed in the set with the smaller index of the two sets that e and f belonged to. (If e and f are in the same set S_i^t then the edge $e'f'$ is simply added to this set.) The remaining sets stay the same, that is, $S_i^t = S_i^{t+1}$ for all sets S_i^t that do not contain any pegged edge. Note that the subsets S_i^t induce connected subgraphs of G_t and that each such subgraph contains at least one vertex of $V(G_0)$. We only need to show that for $t^* = \lceil n/50 \rceil$, with probability $1 - O(1/n^2)$ every set $S_i^{t^*}$ has size at most $4 \log_2 n$.

Fix the index i. Note that if we choose at most one edge from S_i^{t-1} at time t then $|S_i^t| \leqslant |S_i^{t-1}| + 2$, and if we choose both edges from S_i^t then $|S_i^t| \leqslant |S_i^{t-1}| + 3$. In the tth pegging step, conditional upon the history of the process, the probability that the first edge is in S_i^{t-1} is at most $|S_i^{t-1}|/(n + 3t - 4)$ since G_0 is connected and hence G_{t-1} has at least $n + 3t - 4$ edges to choose from. For the second edge this probability is at most $|S_i^{t-1}|/(n + 3t - 5)$ (as we have to choose a different edge).

We divide each pegging step into two time steps, one for each choice of the edges to be pegged. The event that at pegging step t^* a set S_i has size bigger than $D = \lfloor 4 \log_2 n \rfloor$ is contained in the union of the events $E(t_1, \ldots, t_{D/2})$ that the first $D/2$ times an edge of S_i is pegged are $t_1, \ldots, t_{D/2}$, and $t_{D/2} \leqslant 2t^*$. From the above we know that each of the events $E(t_1, \ldots, t_{D/2})$ has probability at most $(D/(n - 2))^{D/2}$. The number of choices of the times t_i is at most $\binom{2\lceil n/50 \rceil}{D/2}$. Hence, for sufficiently large n the probability that the size of S_i increases to at least $D + 1$ is at most

$$\binom{2\lceil n/50 \rceil}{D/2} \left(\frac{D}{n-2} \right)^{D/2} \leqslant \left(\frac{4e\lceil n/50 \rceil}{D} \right)^{D/2} \left(\frac{D}{n-2} \right)^{D/2} \leqslant \left(\frac{1}{4} \right)^{D/2} \leqslant 4n^{-4}.$$

Hence, as $|E(G_0)| \leqslant n^2$ by the union bound, with probability at least $1 - 4/n^2$ none of the sets S_i, $i = 1, \ldots, |E(G_0)|$, has more than D elements when $t = t^*$. □

By means of the next lemma we will associate a graph G_{t_0} with a 2-edge-coloured cycle C, according to the proof outline given earlier.

Lemma 2.2. *Let $G = (V, E)$ be a connected graph. We denote by $G' = (V, E')$ the graph that is obtained from G by replacing each edge by a double edge and colouring one of these edges blue and the other red. Then there exists a closed walk in G' such that the edges along the walk are alternately coloured red and blue and each edge in E' is used exactly once.*

Proof. We prove the lemma by induction on the number m of edges of the graph. If $m = 0$ then the only connected graph consists of a single vertex and the statement is true. If $e \in E$ then by induction $(G - e)'$ has one such walk W' if it is connected, and otherwise it has two such walks, W_1 and W_2, in the two components that e connects. In the first case, we may obtain a new walk by traversing W' starting from an endvertex of e, then following it by the two versions of e, using the one of appropriate colour first. In the second case, the desired walk is easily obtained by traversing W_1, then one version of e of appropriate colour, then W_2, then the other version of e. $\qquad\square$

The next few lemmas show that the diameter of $\mathcal{P}(G_0)$ is small if G_0 is a Hamilton cycle. To start with we define a slightly simpler process. Given a graph G and a set S of its edges, the *random insertion process* applied to G with respect to S consists of repetitions of the following step. Choose an edge uv in S uniformly at random, delete uv from G and S, and add two new edges uw and wv to both G and S, where w is a new vertex of G. The *length* of the process is the number of repetitions of this step. The final set S is a subset of the edge set of the final G, which is a subdivision of the input version of G. Denote the diameter of a graph G by $\mathrm{diam}(G)$.

Lemma 2.3. *Given constants $c > 0$ and $d \geqslant 2$, there exists a constant D for which the following holds. Let G be a connected graph with maximum vertex degree at most d, and with n vertices. Obtain G' from G by the random insertion process of length at most cn with respect to some set S of edges of G with $|S| \geqslant n$. Then a.a.s. $\mathrm{diam}(G') < D\,\mathrm{diam}(G)$ as $n \to \infty$.*

Proof. We first consider the case that $c \leqslant 1/10$. Let P be any path in G of length $k \leqslant \mathrm{diam}(G)$, consider an integer D (to be determined later), and set $j = (D - 1)\,\mathrm{diam}(G)$. The event that P expands to length at least $D\,\mathrm{diam}(G)$ during the insertions is contained in the union of the events $E(t_1, \ldots, t_j)$ that the first j insertion steps occur in the path at times $t_1, \ldots, t_j$. Note that at time t_j the path has length at most $D\,\mathrm{diam}(G)$ as $t_1, \ldots, t_j$ are the first insertions into the path. Thus the probability of inserting the next vertex into P at any of these times, is at most $D\,\mathrm{diam}(G)/n$, since S contains at least n edges. It follows that each of these events has probability at most $(D\,\mathrm{diam}(G)/n)^j$. The number of choices of the times t_i is at most $\binom{cn}{j}$. Hence, the probability that the length of P does increase to at least $D\,\mathrm{diam}(G)$ is at most (since $c \leqslant 1/10$ and we may assume $D \geqslant 2$)

$$\binom{cn}{j}\left(\frac{D\,\mathrm{diam}(G)}{n}\right)^j \leqslant \left(\frac{ecn}{j}\right)^j\left(\frac{2j}{10cn}\right)^j = \left(\frac{e}{5}\right)^j = \left(\frac{e}{5}\right)^{(D-1)\,\mathrm{diam}(G)}.$$

Since successive neighbourhoods of a vertex grow at rate at most $d - 1$, $\mathrm{diam}(G)$ is necessarily at least $(\log n)/(\log d)$ for n sufficiently large. As d is fixed, the above bound

can be made $o(n^{-2}d^{-\mathrm{diam}(G)})$ by taking D a sufficiently large constant (depending on d, but not on c). The number of possible paths P of length k is at most $nd^k \leqslant nd^{\mathrm{diam}(G)}$ since $k \leqslant \mathrm{diam}(G)$, and we are done by the union bound.

Now consider the case $c > 1/10$. Observe that if we apply the insertion process for t steps to obtain the graph G', then $|S'| = |S| + t \geqslant |V(G')| = n + t$. That is, regardless of the value t, the new graph again satisfies the assumption of the lemma. Thus, we can chain together a bounded number of such processes, with c replaced by $1/10$, to create the original process. As d does not change during these processes, the diameter goes up by at most a factor $D^{\lceil 10c \rceil}$ overall, which implies the lemma by redefining D. $\qquad\square$

Our next aim is to use this lemma in order to show that it suffices to add $c\ell$ edges to a Hamilton cycle on ℓ vertices in order to reduce the diameter to $O(\log \ell)$.

Lemma 2.4. *Let C_ℓ be a cycle on ℓ vertices and let $0 < c < 1/2$ be a constant. Let M be a random matching of $V(C_\ell)$ of size $s = s(\ell) \geqslant c\ell$. Then a.a.s. as $\ell \to \infty$ the graph obtained by adding M to C_ℓ has diameter at most $c_0 \log \ell$ for some constant c_0 depending only upon c.*

Proof. Let G_n be the union of a Hamilton cycle of even length n and a random perfect matching of its vertices. A result of Bollobás and Chung [1] gives, with room to spare, that G_n a.a.s has diameter at most $2 \log n$.

Now set $n = 2s$ and take G_n as above. Then let S be the set of edges of the cycle, and apply the random insertion process for $\ell - 2s$ steps. It is easy to check that this produces graphs in which the matched vertices form a uniformly random subset of all the vertices. The lemma consequently follows from Lemma 2.3. $\qquad\square$

With respect to any pegging process under consideration, an edge of G_t is *fresh* if it is also an edge of G_0. This means that it is an edge of G_0 that has not been pegged in the first t pegging steps.

Lemma 2.5. *Let G_0 be any graph with at least ℓ edges and let $0 < c < 1/8$ be a constant. Consider the pegging process $\mathcal{P}(G_0)$ up to $t = \lfloor c\ell \rfloor$ pegging steps. Then a.a.s. as $\ell \to \infty$, at least $t/6$ of these pegging steps involve two fresh edges being pegged.*

Proof. Let X_i denote the indicator variable for the event that the ith edge joins two fresh edges. Clearly, we have

$$\mathbb{P}[X_i = 1 \mid X_1, \ldots, X_{i-1}] \geqslant \frac{\binom{\ell - 2(i-1)}{2}}{\binom{\ell + 3(i-1)}{2}} \geqslant \frac{2}{7} \qquad \text{for all } 1 \leqslant i \leqslant \ell/8,$$

as after $i - 1$ steps the graph contains $\ell + 3(i - 1)$ edges and at least $\ell - 2(i - 1)$ of the edges are still fresh.

Thus $Y_i := \frac{2}{7} - X_i$ forms a supermartingale difference sequence, and for $Y = \sum_{i=1}^{t} Y_i$, since $|Y_i| \leqslant 1$, the one-sided version of Azuma's inequality for supermartingales ([8,

Lemma 4.2] for example) gives

$$\mathbb{P}[Y \geqslant \varepsilon t] \leqslant e^{-\varepsilon^2 t/2},$$

and hence

$$\mathbb{P}\left[\sum_{i=1}^{t} X_i \leqslant 2t/7 - \varepsilon t\right] \leqslant e^{-\varepsilon^2 t/2},$$

from which the claim of the lemma follows immediately. $\qquad\square$

With these preliminaries in hand, we proceed to proving our main result. Several times we will use without comment the fact that the graph G_t in a pegging process $\mathcal{P}(G_0)$ has $|V(G_t)| = |V(G_0)| + 2t$ vertices and $|E(G_t)| = |E(G_0)| + 3t > 3t$ edges.

Proof of Theorem 1.1. Let G_0 be a connected graph. We may assume that $t \geqslant 300$ is large enough such that $t > |E(G_0)|$. We select an integer t_0 satisfying $\frac{1}{50}t \geqslant t - t_0 \geqslant \frac{1}{60}t$, and define $t' = t - t_0$. We will prove the result by just analysing the graph process starting from G_{t_0}. In other words, we let the process run for the first t_0 steps, take whatever graph we might have at that time, and then argue that the remaining $t' = t - t_0$ steps a.a.s. reduce the diameter to $O(\log t)$, where the convergence in the 'a.a.s.' and the bound implicit in $O()$ are uniform over all the graphs possible for G_{t_0}. Note that by Lemma 2.1, it is sufficient to show that a.a.s. each pair of vertices in G_{t_0} has distance $O(\log t)$ after t' further pegging steps. (Recall that by the above assumptions we have $t' \leqslant \frac{1}{50}t \leqslant \frac{1}{50}2t_0 \leqslant \frac{1}{50}|V(G_{t_0})|$ and Lemma 2.1 can thus be applied.) This is what we will show in the remainder of the proof.

Let W be the walk in G_{t_0} obtained by Lemma 2.2. Consider the undirected cycle C obtained from W by splitting each vertex of G_{t_0} into several parts, such that the edges of W appear in the same order around C. We may define a process $\mathcal{P}(C)$, beginning with C, that 'shadows' the process $\mathcal{P}(G_{t_0})$ as follows. To begin with, each edge of G_{t_0} corresponds to the red version of that edge in C. Whenever two edges are pegged in $\mathcal{P}(G_{t_0})$, the corresponding edges in $\mathcal{P}(C)$ are pegged, and the new edges formed in the pegging step are put into correspondence in the obvious way.

Since $2t \leqslant 3t_0 < |E(G_{t_0})| \leqslant 4t$ and $\frac{1}{60}t \leqslant t' \leqslant \frac{1}{50}t$, Lemma 2.5 implies that a.a.s. for $t \to \infty$ at least $\lfloor t'/6 \rfloor$ of the first t' pegging steps in $\mathcal{P}(G_{t_0})$ involve two fresh edges. Let C^* denote the graph produced from C by performing just the set $\mathcal{S}$ of the first $\lfloor t'/6 \rfloor$ of these selected pegging steps. This corresponds to a matching of the fresh edges chosen uniformly at random. We may apply Lemma 2.4 to the graph which is obtained by contracting the red edges of C into vertices. Then a.a.s. the resulting graph has diameter at most $c_0 \log |E(G_{t_0})|$ for an appropriately chosen constant c_0. Because every path in C^* has at most 2 edges in a row that are either red or are red edges subdivided by a pegging step, a.a.s. the diameter of C^* is at most $3c_0 \log(|E(G_{t_0})|) + 2 < 4c_0 \log t$.

Note that each vertex in $V(G_{t_0})$ can be identified with vertices in $V(C)$ and that each edge x of C^* corresponds to a path $F(x)$ in G_t, the extra vertices in the path being those vertices inserted during the pegging steps not included in $\mathcal{S}$. In this way, every path P in C^* corresponds to a path $F(P)$ in G_t. So, to get the required bound on distances between

vertices in G_{t_0}, it suffices to show that a.a.s. for every path P of length at most $4c_0 \log t$ in C^*, the length of $F(P)$ is at most $(4c_0 + c_1) \log t$ for some constant c_1.

To show this, we first note that each path of length $\ell \leqslant 4c_0 \log t$ in C^* consists of some number, k, of 'diagonal' edges in C^* created when red edges are pegged, where $0 \leqslant k \leqslant \ell/2 \leqslant 2c_0 \log t$, together with $k+1$ sub-paths $P_1, \ldots, P_{k+1}$ of C (though actually the end edges of each of these sub-paths are subdivided in C^*). For each k under consideration, we consider all possible families S of such sub-paths of C that have combined length $\ell - k$, with each sub-path starting and ending with a red edge.

We first give an upper bound on the probability of the event that (i) all diagonal edges between paths P_i and P_{i+1} indeed 'materialize' (by pegging the corresponding red edges on P_i and P_{i+1}) and that additionally (ii) the paths $P_1, \ldots, P_{k+1}$, together with these diagonals, are extended in length by more than $c_1 \log t$ during the formation of G_t. Later we will apply the union bound over all families S and all relevant ℓ and k.

Let $j = \lceil D \log t \rceil$ where $D = D(c_0) \geqslant 4c_0$ is a constant chosen sufficiently large to satisfy (2.1) below in all cases. Let $E_S(u_1, \ldots, u_k; t_1, \ldots, t_j)$ be the event that, in $\mathcal{P}(C)$, at time u_i the diagonal edge connecting P_i and P_{i+1} is created, and at times $t_1, \ldots, t_j$ at least one of the edges in the subpaths or diagonals is pegged. Note that, since each edge of G_{t_0} was replaced by a blue and a red edge, at each time t_i up to four vertices may be inserted into $P_1, \ldots, P_{k+1}$ or the red diagonals that are already present (or, more precisely, into the corresponding subdivided paths). Hence, with j 'insertion' steps, the increase in length of the paths is at most $4j$.

The next part of the argument is similar to the proof of Lemma 2.3. At any time the number of edges in the subdivision of $P_1, \ldots, P_{k+1}$ and the present diagonals is bounded by $\ell + 4j \leqslant 5j$ by choice of D. Let m_0 denote $|E(G_{t_0})|$. There are at least m_0 edges present in each step of $\mathcal{P}(G_{t_0})$, so the probability that the two paths P_i and P_{i+1} are joined correctly at time u_i is at most $\binom{m_0}{2}^{-1}$. The probability that the first edge pegged at time t_i lies in the required set of paths and diagonals is at most $5j/m_0$, and for the second edge it is at most $5j/(m_0 - 1)$. At least one of these two events must happen. Thus, the probability of any one event $E_S(u_1, \ldots, u_k; t_1, \ldots, t_j)$ is at most

$$\binom{m_0}{2}^{-k} \left(\frac{10j}{m_0 - 1} \right)^j.$$

On the other hand, for fixed k the number of such events $E_S(u_1, \ldots, u_k; t_1, \ldots, t_j)$ to consider is at most (recalling $t' = t - t_0$)

$$m_0^{k+1} 2^{k+1} \binom{\ell}{k} (t')^k \binom{t'}{j} \leqslant m_0^{k+1} 2^{\ell+k+1} (t')^k \left(\frac{et'}{j} \right)^j,$$

where the first factor chooses the leading red edges of the $k+1$ segments, the next decides which direction along the cycle the segments are laid out, the next (binomial) factor bounds the number of compositions of $\ell - k$ which determine the numbers of red edges in each of $P_1, \ldots, P_{k+1}$, the next factor gives a bound on the choices of the times u_i, and the final factor chooses the times t_j.

Thus the probability that at least one of these events holds is at most the product of the two quantities above, which, using $m_0 - 1 \geqslant 3t_0$ (by the comment before this proof), $\ell \leqslant j$ and $t' = t - t_0 \leqslant t/50$, is at most

$$2m_0 \left(\frac{t'}{t_0} \right)^{k+j} \left(\frac{4}{3} \right)^k \left(\frac{40e}{3} \right)^j < 2m_0 \left(\frac{40t'}{t_0} \right)^{k+j} \leqslant 2m_0 \left(\frac{40}{49} \right)^{k+j} < 1/t \qquad (2.1)$$

as $m_0 = O(t_0)$, $j \geqslant D \log t$, and we may choose D sufficiently large. It follows that a.a.s. none of the events $E_S(u_1, \ldots, u_k; t_1, \ldots, t_j)$ hold, for any of the $O(\log^2 t)$ choices for ℓ and k. When none of these events hold, it follows that for every path P of length at most $4c_0 \log t$ in C^*, the length of $F(P)$ is at most $4c_0 \log t + \lceil D \log t \rceil$ as required. □

3. Final remarks

Pegging processes generating regular graphs of any fixed degree $d \geqslant 3$ were defined in [5]. The proof of Theorem 1.1 is easily modified for other pegging processes. Here we outline the ideas and necessary modifications only for 4-regular pegging processes. Pegging in this case, which we call *4-pegging*, is accomplished by picking two edges $e = ab, f = cd \in E$, adding a new vertex e' and new edges ae', be', ce', de', and deleting the edges e and f. Note that in order to avoid multiple edges it is required that the edges e and f are non-adjacent. In order to show that also in a 4-pegging process the diameter is $O(\log t)$ after t steps, we only have to change the above proof at a few places. As only non-adjacent edges may be pegged, we have to modify Lemma 2.4 so that the matching is edge-disjoint from the cycle. (Here we may use [7, Corollary 4.21] for a random perfect matching plus a cycle when multiple edges are excluded.) Also within the proof of Lemma 2.1, the fact that the second edge has to be non-adjacent to first decreases the denominator of the probability from $n - 1$ to $n - 2d$, which is asymptotically negligible, since we may assume that $d \leqslant \log n$ if G_0 is fixed (if this is not the case we run the process for n steps and call the new graph G_0). Thus, the diameter of the graph in the 3-pegging process in which adjacent edges may not be chosen is again $O(\log n)$ a.a.s. It is then routine to modify the proof so that it also remains valid if we forbid pegging of any edge joining two new vertices in any pegging step. Finally, this process may be used to simulate the 4-pegging process by counting the new edge as having zero length.

We claim that our method extends without any trouble to the pegging processes of higher degree defined in [5]. Hence, in each case the diameter of G_t is $O(\log t)$.

Figure 1 shows the results of simulations involving thousands of random 4-regular graphs with n vertices for each of 220 values of n from 10 to 8000, for each of three models: the uniform model; the 4-regular pegging process beginning with $G_0 = K_5$ (the complete graph); the 4-regular pegging process beginning with two disjoint copies of K_5. In each case, let $f_d(n)$ denote the probability that the random graph on n vertices has diameter d. Each continuous line shows a graph of $f_d(n)$ for a fixed value of $d \geqslant 3$, plotted against $x = \log_3(n \ln n)$, with the points for distinct n joined in a piecewise linear fashion. In each case, the values of $f_d(n)$ drop close to 0 somewhat before x reaches $d + 1$. In the third model, the plot for $d = 4$ has a very low peak and that for $d = 5$ has an unusual shape (due to the fact that the first 4-pegging of two K_5s often creates a graph of diameter
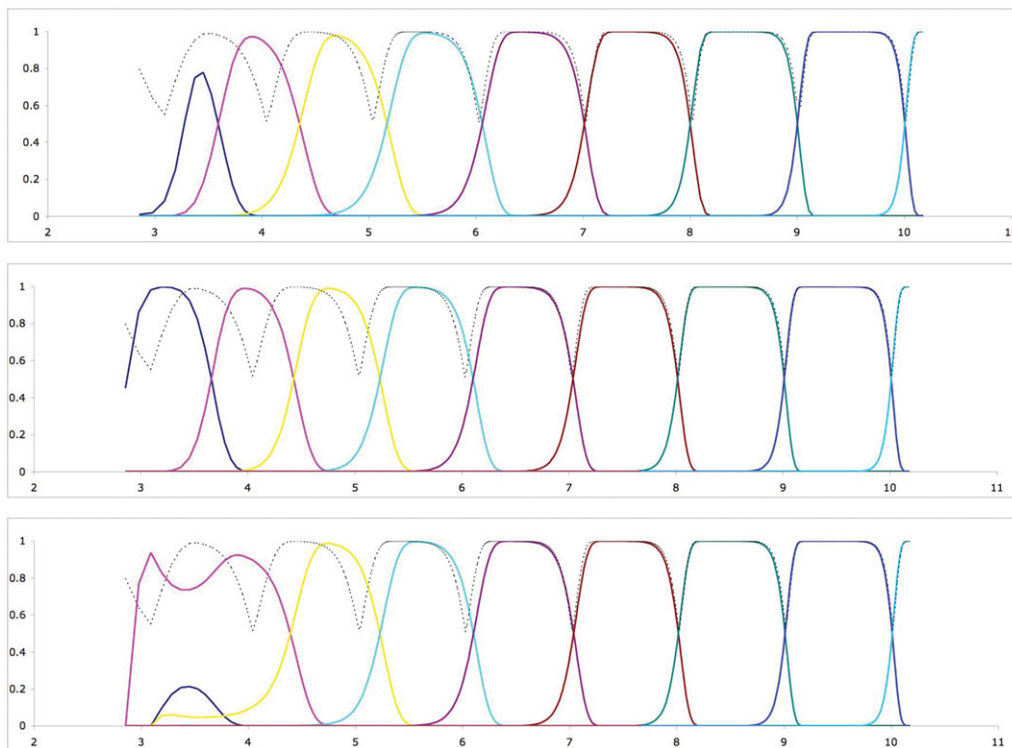
Figure 1. (Colour online) Diameters of random 4-regular graphs by simulation. Upper: uniform. Middle: 4-pegging, starting with K_5. Lower: 4-pegging, starting with $2K_5$.

5 on 11 vertices). The horizontal scale is specially chosen to obtain a repeating pattern, in view of the formula for the diameter of a uniformly random 4-regular graph in [6], where it is shown that the diameter is 2-point concentrated, and 1-point concentrated for almost all n. (The dashed line in each diagram shows the theoretical asymptotic formula for $\max_{d \geqslant 3} f_d(n)$ from [6].) The data suggest something much more precise than the $O(\log n)$ bound on diameter that we proved in this paper. From these results, we do not hesitate to make the following conjecture, which says, loosely speaking, that the diameter of the graph G_t in the pegging process is 2-point concentrated, and for almost all values of t is it is 1-point concentrated. The form of x given in this conjecture is such that $x = d$ is the point at which the typical diameter jumps from d to $d + 1$ for the uniform model, as determined by the formula proved in [6].

Conjecture 1. *There exists a function $s(n) \to 0$ such that the following holds. Let $r \geqslant 3$, and let G_P be a the random graph with n vertices in the pegging process beginning with any fixed r-regular graph G_0. Let $x := \log_{r-1}((2(r-2)/r)n \ln n)$. Then a.a.s. either the diameter of G_P is equal to $\lceil x \rceil$, or $|x - k| < s(n)$ for some integer k and the diameter is k or $k + 1$.*

The results of [6] show that the corresponding statement for the uniform model is true provided only that $s(n) \log n \to \infty$.

References

[1] Bollobás, B. and Chung, F. R. K. (1988) The diameter of a cycle plus a random matching. *SIAM J. Discrete Math.* **1** 328–333.

[2] Bourassa, V. and Holt, F. (2003) SWAN: Small-world wide area networks. In *Proc. International Conference on Advances in Infrastructures: SSGRR 2003w* (L'Aquila, Italy, 2003), paper #64.

[3] Cooper, C., Dyer, M. and Greenhill, C. (2005) Sampling regular graphs and a peer-to-peer network. In *Proc. Sixteenth Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 980–988.

[4] Gao, P. Connectivity of random regular graphs generated by the pegging algorithm. *J. Graph Theory*, to appear.

[5] Gao, P. and Wormald, N. C. (2009) Short cycle distributions in random regular graphs recursively generated by pegging. *Random Struct. Alg.* **34** 54–86.

[6] Sanwalani, V. and Wormald, N. The diameter of random regular graphs. In preparation.

[7] Wormald, N. C. (1999) Models of random regular graphs. In *Surveys in Combinatorics, 1999*, Vol. 267 of *London Mathematical Society Lecture Notes* (J. D. Lamb and D. A. Preece, eds), Cambridge University Press, Cambridge, pp. 239–298.

[8] Wormald, N. C. (1999) The differential equation method for random graph processes and greedy algorithms. In *Lectures on Approximation and Randomized Algorithms* (M. Karoński and H. J. Prömel, eds), PWN, Warsaw, pp. 73–155.

[9] Wormald, N. C. (2004) Random graphs and asymptotics. Section 8.2 in *Handbook of Graph Theory* (J. L. Gross and J. Yellen, eds), CRC, Boca Raton, pp. 817–836.