Cleaning random d-regular graphs with Brooms

Paweł Prałat *

Abstract

A model for *cleaning* a graph with brushes was recently introduced. Let $\alpha = (v_1, v_2, \dots, v_n)$ be a permutation of the vertices of G; for each vertex v_i let $N^+(v_i) = \{j : v_j v_i \in E \text{ and } j > i\}$ and $N^-(v_i) = \{j : v_j v_i \in E \text{ and } j < i\}$; finally let $b_\alpha(G) = \sum_{i=1}^n \max\{|N^+(v_i)| - |N^-(v_i)|, 0\}$. The *Broom* number is given by $B(G) = \max_\alpha b_\alpha(G)$.

We consider the Broom number of d-regular graphs, focusing on the asymptotic number for random d-regular graphs. Various lower and upper bounds are proposed. To get an asymptotically almost sure lower bound we use a degree-greedy algorithm to clean a random d-regular graph on n vertices (with dn even) and analyze it using the differential equations method (for fixed d). We further show that for any d-regular graph on n vertices there is a cleaning sequence such at least $n(d+1)/4$ brushes are needed to clean a graph using this sequence. For an asymptotically almost sure upper bound, the pairing model is used to show that at most $n(d + 2\sqrt{d \ln 2})/4$ brushes can be used when a random d-regular graph is cleaned. This implies that for fixed large d, the Broom number of a random d-regular graph on n vertices is asymptotically almost surely $\frac{n}{4}(d + \Theta(\sqrt{d}))$.

1 Introduction

The cleaning model, introduced in [13], is a combination of the chip-firing game and edge-searching on a simple finite graph. Initially, every edge and vertex of a graph is *dirty* and a fixed number of brushes start on a set of vertices. At each step, a vertex v and all its incident edges which are dirty may be *cleaned* if there are at least as many brushes on v as there are incident dirty edges. When a vertex is cleaned, every incident dirty edge is traversed (i.e. cleaned) by one and only one brush, and brushes cannot traverse a clean edge. See Figure 1 for an example of this cleaning process. The initial configuration has only 2 brushes. The solid edges are dirty and the dotted edges are clean. The circle indicates which vertex is cleaned next.

In [2, 9, 10, 13, 15, 17], the focus was on determining the minimum number of brushes required. To this end, a different but equivalent formulation of the problem was introduced. Let $\alpha = (v_1, v_2, \dots, v_n)$ be a permutation of the vertices of G; for each vertex v_i

*Department of Mathematics, West Virginia University, Morgantown, WV 26506-6310, USA.

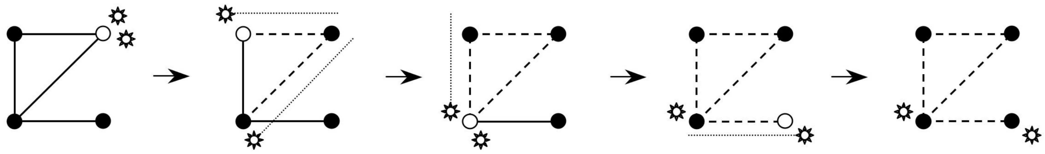


Figure 1: An example of the cleaning process.

let $(v_i) = \{j : v_j v_i \in E \text{ and } j > i\}$ and $N^-(v_i) = \{j : v_j v_i \in E \text{ and } j < i\}$; finally let

$$b_\alpha(G) = \sum_{i=1}^{n} \max\{|(v_i)| - |N^-(v_i)|, 0\}.$$

The *brush* number is given by $b(G) = \min_\alpha b_\alpha(G)$.

In this paper, we concentrate on the *Broom number*, $B(G) = \max_\alpha b_\alpha(G)$, considered in [14]. This is the maximum number of brushes needed to clean the graph where every brush has to clean at least one edge. We are interested in the asymptotic possible number of brushes that can be used to clean d-regular subject to this constraint, and mainly random d-regular (finite, simple) graphs.

In Section 2 we introduce the formal definitions for the cleaning process and also include a description of the pairing model which is used in the results on random regular graphs, instead of working directly in the uniform probability space.

We first observe that if $d = 2$, then the Broom number of a random d-regular graph on n vertices is asymptotically almost surely $n - (1/4 + o(1)) \log n$; see Section 3. Unfortunately, this is the only case where our upper and a lower bounds match; for other values of d we provide bounds only.

In Section 4 we describe some general upper bounds for the Broom number following from both expansion properties of random d-regular graph and direct calculations based on the pairing model. In particular, we show that for random d-regular graphs on n vertices, the maximum number of brushes needed is, asymptotically almost surely, at most $\frac{n}{4}(d + 2\sqrt{d \ln 2})$.

In Section 5 we show that for d-regular graphs on n vertices, there is a cleaning sequence such that $n(d+1)/4$ brushes are needed to clean the graph using this sequence if d is odd, and $\frac{n}{4}(d + 1 - \frac{1}{d+1})$ brushes are needed if d is even. These bounds are tight. In order to improve this for a random case, an asymptotically almost sure bound on the Broom number is obtained by considering a degree-greedy algorithm to clean the graph and then using the differential equation method, studied in [21], to find the asymptotic possible number of brushes that can be used. We also consider the case of large d, and show that the Broom number in this case is roughly $nd/4$.

Note that the paper uses similar approaches that were used in [2], where the problem of finding the minimum number of brushes needed to clean random d-regular graph was considered. Some necessary modifications are straightforward, however, sometimes non-trivial adjustments and new ideas are needed. For example, an upper bound requires a new approach; we use the fact that a large value of the Broom number implies that there is at least one large induced subgraph that is sparse comparing to the expected number of edges. This is, asymptotically almost surely, not the case for a random graph.

2 Definitions

The following cleaning algorithm and terminology were recently introduced in [13].

Formally, at each step t, $\omega_t(v)$ denotes the number of brushes at vertex v ($\omega_t : V \to \mathbb{N} \cup \{0\}$) and D_t denotes the set of dirty vertices. An edge $uv \in E$ is dirty if and only if both u and v are dirty: $\{u, v\} \subseteq D_t$. Finally, let $D_t(v)$ denote the number of dirty edges incident to v at step t:

$$D_t(v) = \begin{cases} |N(v) \cap D_t| & \text{if } \ v \in D_t \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2.1 *The* ***cleaning process*** $\mathfrak{P}(G, \omega_0) = \{(\omega_t, D_t)\}_{t=0}^{T}$ *of an undirected graph* $G = (V, E)$ *with an* ***initial configuration of brushes*** ω_0 *is as follows:*

(0) *Initially, all vertices are dirty:* $D_0 = V$*; set* $t := 0$

(1) *Let* α_{t+1} *be any vertex in* D_t *such that* $\omega_t(\alpha_{t+1}) \geq D_t(\alpha_{t+1})$*. If no such vertex exists, then stop the process (*$T = t$*), return the* ***cleaning sequence*** $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_T)$*, the* ***final set of dirty vertices*** D_T*, and the* ***final configuration of brushes*** ω_T

(2) *Clean* α_{t+1} *and all dirty incident edges by traversing a brush from* α_{t+1} *to each dirty neighbour. More precisely,* $D_{t+1} = D_t \setminus \{\alpha_{t+1}\}$*,* $\omega_{t+1}(\alpha_{t+1}) = \omega_t(\alpha_{t+1}) - D_t(\alpha_{t+1})$*, and for every* $v \in N(\alpha_{t+1}) \cap D_t$*,* $\omega_{t+1}(v) = \omega_t(v) + 1$*, the other values of* ω_{t+1} *remain the same as in* ω_t*.*

(3) $t := t + 1$ *and go back to* **(1)**

Note that for a graph G and initial configuration ω_0, the cleaning process can return different cleaning sequences and final configurations of brushes. It was shown (see [13, Theorem 2.1]), however, that the final set of dirty vertices is determined by G and ω_0. Thus, the following definition is natural.

Definition 2.2 *A graph* $G = (V, E)$ ***can be cleaned*** *by the initial configuration of brushes* ω_0 *if the cleaning process* $\mathfrak{P}(G, \omega_0)$ *returns an empty final set of dirty vertices (*$D_T = \emptyset$*).*

The brush number, $b(G)$*, is the minimum number of brushes needed to clean* G*, that is,*

$$b(G) = \min_{\omega_0 : V \to \mathbb{N} \cup \{0\}} \Big\{ \sum_{v \in V} \omega_0(v) : G \text{ can be cleaned by } \omega_0 \Big\}.$$

Similarly, $b_\alpha(G)$ *is defined as the minimum number of brushes needed to clean* G *using the cleaning sequence* α*.*

It is clear that for every cleaning sequence α, $b_\alpha(G) \geq b(G)$ and $b(G) = \min_\alpha b_\alpha(G)$. (The last relation can be used as an alternative definition of $b(G)$.) In this paper we focus on the worst-case scenario, that is, we try to determine the worst cleaning sequence which uses as many brushes as possible (see [14] for more). This, of course, gives an upper bound for any cleaning sequence.

Definition 2.3 *Let the Broom number, $B(G)$, of a given graph $G = (V, E)$ be the maximum number of brushes needed to clean G using the cleaning sequence α, that is,*

$$B(G) = \max_\alpha b_\alpha(G).$$

For example, we may clean $C_8 = (v_1, v_2, \ldots, v_8)$ using only two brushes, using cleaning sequence $\gamma = (v_1, v_2, \ldots, v_8)$. In fact, $b(C_8) = b_\gamma(C_8) = 2$. However, we could also clean C_8 using eight brushes and cleaning sequence $\alpha = (v_1, v_3, v_5, v_7, v_2, v_4, v_6, v_8)$. That is, $b_\alpha(C_8) = 8$. Clearly the maximum number of brushes one could use to clean any graph G is $|E(G)|$: each brush cleans exactly one edge. Consequently, $B(C_8) = b_\alpha(C_8) = 8$.

In general, it is difficult to find $b(G)$. (In [9], the cleaning process was translated into the BALANCED VERTEX-ORDERING problem, which was known from [4] to be $\mathcal{NP}$-complete. The complexity of computing $B(G)$ is still unknown.) However, $b_\alpha(G)$ can be easily computed. For this, it seems better not to choose the function ω_0 in advance, but to run the cleaning process in the order α, and compute the initial number of brushes needed to clean a vertex. We can adjust ω_0 along the way

$$\omega_0(\alpha_{t+1}) = \max\{2D_t(\alpha_{t+1}) - \deg(\alpha_{t+1}), 0\}, \tag{1}$$

for $t = 0, 1, \ldots, |V| - 1$, since that is the number of brushes we have to add over and above what we get for free. Alternatively, (1) can be rewritten as

$$\omega_0(\alpha_{t+1}) = \max\{|_\alpha(\alpha_{t+1})| - |N^-_\alpha(\alpha_{t+1})|, 0\}, \tag{2}$$

where $N^-_\alpha(\alpha_{t+1})$ denotes the left-neighbours of α_{t+1} in the cleaning sequence α: the set of vertices cleaned before α_{t+1}; $N^+_\alpha(\alpha_{t+1})$ is defined similarly. This will be a useful representation in later sections.

The model presented in this paper is one where the edges are continually recontaminated, say by algae, so that cleaning is regarded as an on-going process. Ideally, the final configuration of the brushes, after all the edges have been cleaned, should be a viable starting configuration to clean the graph again. We know that this is possible, even with the least number of brushes. The following theorem has been proven in [13] (Theorem 2.3) although the statement presented here is a little bit stronger focusing on the cleaning sequence that can be used.

Theorem 2.4 ([13]) ***The Reversibility Theorem***
Given the initial configuration ω_0, suppose G can be cleaned using cleaning sequence $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and yielding final configuration ω_n, $n = |V(G)|$. Then, given initial configuration $\tau_0 = \omega_n$, G can be cleaned using cleaning sequence $\overline{\alpha} = (\alpha_n, \alpha_{n-1}, \ldots, \alpha_1)$ and yielding the final configuration $\tau_n = \omega_0$. Moreover, $b_\alpha(G) = b_{\overline{\alpha}}(G)$.

When a graph G is cleaned using the cleaning process described in Definition 2.1, each edge of G is traversed exactly once and by exactly one brush. Note that no brush may return to a vertex it has already visited, motivating the following definition.

Definition 2.5 *The **brush path** of a brush b is the path formed by the set of edges cleaned by b.*

Our main results refer to the probability space of random d-regular graphs with uniform probability distribution. This space is denoted $\mathcal{G}_{n,d}$, and asymptotics (such as 'asymptotically almost surely', which we abbreviate to a.a.s.) are for $n \to \infty$ with $d \geq 2$ fixed, and n even if d is odd.

Instead of working directly in the uniform probability space of random regular graphs on n vertices $\mathcal{G}_{n,d}$, we use the *pairing model* of random regular graphs, first introduced by Bollobás [5], which is described next. Suppose that dn is even, as in the case of random regular graphs, and consider dn points partitioned into n labeled buckets $v_1, v_2, \ldots, v_n$ of d points each. A *pairing* of these points is a perfect matching into $dn/2$ pairs. Given a pairing P, we may construct a multigraph $G(P)$, with loops allowed, as follows: the vertices are the buckets $v_1, v_2, \ldots, v_n$, and a pair $\{x, y\}$ in P corresponds to an edge $v_i v_j$ in $G(P)$ if x and y are contained in the buckets v_i and v_j, respectively. It is an easy fact that the probability of a random pairing corresponding to a given simple graph G is independent of the graph, hence the restriction of the probability space of random pairings to simple graphs is precisely $\mathcal{G}_{n,d}$. Moreover, it is well known that a random pairing generates a simple graph with probability asymptotic to $e^{(1-d^2)/4}$ depending on d, so that any event holding a.a.s. over the probability space of random pairings also holds a.a.s. over the corresponding space $\mathcal{G}_{n,d}$. For this reason, asymptotic results over random pairings suffice for our purposes. One of the advantages of using this model is that the pairs may be chosen sequentially so that the next pair is chosen uniformly at random over the remaining (unchosen) points. For more information on this model, see [20].

3 2-regular graphs

By definition, G can be decomposed into $b_\alpha(G)$ brush paths. Since no brush can stay at its initial vertex in the minimal brush configuration, these paths have at least one edge. Thus, the maximum number of paths into which a graph G can be decomposed (that is, the number of edges) yields an upper bound for $B(G)$. The following simple property has been observed in [14].

Proposition 3.1 *For any graph $G = (V, E)$, $B(G) \leq |E|$.*

It is clear that this upper bound can be obtained if $G = (V = X \cup Y, E)$ is bipartite; clean every vertex in X (in any order) and then clean vertices in Y (again, the order is not important). The Broom number is smaller than this trivial bound otherwise (again, see [14] for more). Thus,

$$\begin{aligned} B(C_{2k}) &= |E(C_{2k})| &= 2k \\ B(C_{2k-1}) &= |E(C_{2k-1})| - 1 &= 2k-2, \end{aligned} \tag{3}$$

for $k \geq 2$.

Let $Y = Y_n$ be the total number of cycles in a random 2-regular graph on n vertices and let $Z = Z_n$ be the number of odd cycles. From (3) if follows that we need $n - Z_n \geq n - Y_n$ brushes in order to clean a 2-regular graph. We know that the random 2-regular graph is a.a.s. disconnected; by simple calculations one can show that the probability of having a Hamiltonian cycle is asymptotic to $\frac{1}{2}e^{3/4}\sqrt{\pi}n^{-1/2}$ (see, for example, [20]).

We also know that the total number of cycles Y_n is sharply concentrated near $(1/2)\log n$. It is not difficult to see this by generating the random graph sequentially using the pairing model. The probability of forming a cycle in step i is exactly $1/(2n - 2i + 1)$, so the expected number of cycles is $(1/2)\log n + O(1)$. The variance can be calculated in a similar way. So we get that a.a.s. the Broom number for a random 2-regular graph is at least $n - (1/2 + o(1))\log n$.

In order to estimate Z_n, one can show that until near the end of the process, the probability that the next cycle is odd is close to 1/2. So the Azuma-Hoeffding inequality shows that the number of odd cycles is close to half the total.

Theorem 3.2 *For $G \in \mathcal{G}_{n,2}$, a.a.s.*

$$B(G) = n - (1/4 + o(1))\log n\,.$$

4 Upper bounds

Before we move to proving a general upper bound, we study the following useful property of a cleaning sequence that yields $B(G)$.

Lemma 4.1 *For any graph $G = (V, E)$, there is a cleaning sequence α yielding $B(G)$ which is sorted with respect to $|N^+(\alpha_i)| - |N^-(\alpha_i)|$, that is, $|N^+(\alpha_i)| - |N^-(\alpha_i)| \geq |N^+(\alpha_{i+1})| - |N^-(\alpha_{i+1})|$ for $1 \leq i \leq |V| - 1$.*

Proof: Let α be any cleaning sequence that yields $B(G)$. Consider any pair of consecutive vertices α_i, α_{i+1} and suppose that $|N^+(\alpha_i)| - |N^-(\alpha_i)| < |N^+(\alpha_{i+1})| - |N^-(\alpha_{i+1})|$. It is clear that if $\alpha_i\alpha_{i+1} \notin E$, then we can change the order of cleaning of these two vertices (keeping the rest as before) and $b_\alpha(G)$ is not affected. Since $|N^+(\alpha_i)| - |N^-(\alpha_i)| < |N^+(\alpha_{i+1})| - |N^-(\alpha_{i+1})|$, swapping α_i and α_{i+1} does not decrease the brush number if $\alpha_i\alpha_{i+1} \in E$ (in fact, it does not change $b_\alpha(G)$ since α yields $B(G)$). In order to get a cleaning sequence we claim to exist, one can use bubble-sort. ■

To show the result, one can use the expansion properties of random d-regular graphs that follow from their eigenvalues. The adjacency matrix $A = A(G)$ of a given d-regular graph G with n vertices, is an $n \times n$ real and symmetric matrix. Thus, the matrix A has n real eigenvalues which we denote by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. It is known that certain properties of a d-regular graph are reflected in its spectrum but, since we focus on expansion properties, we are particularly interested in the following quantity: $\lambda = \lambda(G) = \max(|\lambda_2|, |\lambda_n|)$. In words, λ is the largest absolute value of an eigenvalue other than $\lambda_1 = d$ (for more details, see the general survey [11] about expanders, or [3], Chapter 9).

The value of λ for random d-regular graphs has been studied extensively. A major result due to Friedman [8] is the following:

Lemma 4.2 ([8]) *For every fixed $\varepsilon > 0$ and for $G \in \mathcal{G}_{n,d}$,*

$$\mathbb{P}(\lambda(G) \le 2\sqrt{d-1} + \varepsilon) = 1 - o(1)\,.$$

The number of edges $|E(S,T)|$ between sets S and T is expected to be close to the expected number of edges between S and T in a random graph of edge density d/n, namely $d|S||T|/n$. A small λ (or large spectral gap) implies that this deviation is small. The following useful bound is essentially proved in [1] (see also [3]):

Lemma 4.3 (Expander Mixing Lemma) *Let G be a d-regular graph with n vertices and set $\lambda = \lambda(G)$. Then for all $S, T \subseteq V$*

$$\left| |E(S,T)| - \frac{d|S||T|}{n} \right| \le \lambda \sqrt{|S||T|}\,.$$

(Note that $S \cap T$ does not have to be empty; in general, $|E(S,T)|$ is defined to be the number of edges between $S \setminus T$ to T plus twice the number of edges that contain only vertices of $S \cap T$.)

Let us introduce one more useful definition before we move to the next theorem. At a given time-step of the process, we define the *dirty degree* of v to be the degree of v in a graph induced by the set of dirty vertices (at that point). Now we are ready to prove an upper bound of the Broom number.

Theorem 4.4 *Let $G \in \mathcal{G}_{n,d}$, where $d \ge 3$. Then, for every fixed $\varepsilon > 0$ a.a.s.*

$$B(G) \le \frac{n}{4}\left(d + 4\sqrt{d-1} + \varepsilon\right)(1 + o(1)).$$

Proof: Suppose that d is even and let α be a cleaning sequence that yields $B(G)$. By Lemma 4.1, we can assume that α is sorted with respect to $|N^+(\alpha_i)| - |N^-(\alpha_i)|$; we clean vertices of dirty degree d first (up to time t_d), then we move to cleaning vertices of dirty degree $d-1$ (up to time t_{d-1}), and so on. Therefore,

$$B(G) = dt_d + (d-2)(t_{d-1} - t_d) + \cdots + 2(t_{d/2+1} - t_{d/2+2}) = 2 \sum_{i=d/2+1}^{d} t_i. \tag{4}$$

Moreover, without loss of generality, we can assume that $t_{d/2+1} \le n/2$, that is, at least $n/2$ vertices are cleaned 'for free' (if α does not have this property, then $\overline{\alpha}$ (α reversed) has it; see Theorem 2.4).

Now, consider a subgraph induced by the set $X = \{\alpha_1, \alpha_2, \ldots, \alpha_{t_{d/2+1}}\}$. The number of edges in $G[X]$ is

$$\begin{aligned} |E(G[X])| &= \sum_{j=1}^{t_{d/2+1}} N^-(\alpha_j) \\ &= (t_{d-1} - t_d) + 2(t_{d-2} - t_{d-1}) + \cdots + \left(\frac{d}{2} - 1\right)(t_{d/2+1} - t_{d/2+2}) \\ &= \frac{d}{2} t_{d/2+1} - \sum_{i=d/2+1}^{d} t_i. \end{aligned}$$

On the other hand, Lemma 4.3 implies that

$$|E(G[X])| \geq \frac{d\ (t_{d/2+1})^2}{2n} - \frac{1}{2}\ \lambda\ t_{d/2+1}, \tag{5}$$

so we get that

$$\sum_{i=d/2+1}^{d} t_i \leq \frac{d}{2} t_{d/2+1} \left(1 - \frac{t_{d/2+1}}{n}\right) + \frac{1}{2}\ \lambda\ t_{d/2+1} \leq \frac{dn}{8} + \frac{\lambda n}{4}.$$

For d even, the theorem holds by (4) and Lemma 4.2.

Exactly the same argument can be used for d odd. This time,

$$B(G) = 2 \sum_{i=(d+3)/2}^{d} t_i + t_{(d+1)/2},$$

we can assume that $t_{(d+1)/2} \leq n/2$, and

$$|E(G[X])| = \frac{d}{2} t_{(d+1)/2} - \left(\sum_{i=(d+3)/2}^{d} t_i + t_{(d+1)/2} \right).$$

■

Note that this result can be improved a little, namely, (5) can be replaced by a stronger statement that follows from Lemma 4.5 below.

Suppose that x ($0 \leq x \leq 1/2$) and y ($w(x) = w \leq y \leq xd$) are real numbers such that the expected number $S(x, y)$ of sets S of xn vertices in $G \in \mathcal{G}_{n,d}$ with yn edges to the complement $V(G) \setminus S$ is $o(n^{-2})$. Then a.a.s. no set S, $|S| = xn \leq n/2$ has at least wn edges to the complement, by the first moment principle. In order to find optimal value of w for a given x we use the pairing model. It is clear that

$$S(x, y) = \frac{\binom{n}{xn}\binom{xdn}{yn} M(xdn - yn) \binom{(1-x)dn}{yn} (yn)! M((1-x)dn - yn)}{M(dn)}$$

where $M(i)$ is the number of perfect matchings on i vertices, that is,

$$M(i) = \frac{i!}{(i/2)! 2^{i/2}}.$$

To see this, we consider the pairing model discussed before. We fix xn vertices (xdn points) to form set S (term $\binom{n}{xn}$), and yn points in S that correspond to yn edges to the complement of S (term $\binom{xdn}{yn}$). These edges are incident to yn points in $V(G) \setminus S$ (term $\binom{(1-x)dn}{yn}$). After fixing points in both S and $V(G) \setminus S$, we need to connect them in all possible ways (term $(yn)!$). Finally, we need to take a perfect matching of remaining points in S (term $M(xdn - yn)$) and a perfect matching of remaining points in $V(G) \setminus S$ (term $M((1-x)dn - yn)$) to consider all possible configurations satisfying our assumption.

After simplification we get

$$S(x, y) = \frac{n!(xdn)!((1-x)dn)!(dn/2)!2^{yn}}{(xn)!((1-x)n)!(yn)!((xd-y)n/2)!(((1-x)d-y)n/2)!(dn)!}.$$

Using Stirling's formula ($n! \sim \sqrt{2\pi n}(n/e)^n$) and taking the exponential part we obtain

$$\begin{aligned} S(x,y) &\leq e^{o(n)} \frac{x^{x(d-1)n}(1-x)^{(1-x)(d-1)n} d^{dn/2}}{y^{yn}(xd-y)^{(xd-y)n/2}((1-x)d-y)^{((1-x)d-y)n/2}} \\ &= e^{f(x,y,d)n+o(n)}, \end{aligned} \tag{6}$$

where

$$\begin{aligned} f(x,y,d) &= x(d-1)\ln x + (1-x)(d-1)\ln(1-x) + 0.5d\ln d - y\ln y \\ &\quad -0.5(xd-y)\ln(xd-y) - 0.5((1-x)d-y)\ln((1-x)d-y). \end{aligned}$$

Thus, if $f(x,y,d) < 0$, then $S(x,y)$ is exponentially small (n large). Since we expect $dx(1-x)n$ edges between S and $V(G) \setminus S$, we put $y = dx(1-x) + dz/4$. Not surprisingly, function $f(x, dx(1-x) + dz/4, d)$ is increasing in x so we focus on $x = 1/2$, in which case $f(x,y,d)$ becomes

$$\begin{aligned} &(d-1)\ln(1/2) + (d/2)\ln d - y\ln y - (d/2-y)\ln(d/2-y) \\ &\qquad = -\frac{d}{4}((1+z)\ln(1+z) + (1-z)\ln(1-z)) + \ln 2 \end{aligned}$$

where $y = (d/4)(1+z)$.

It is straightforward to see that this function is decreasing in z for $z \geq 0$. Let $\overline{z} = \overline{z}(x)$ denote the value of z for which it first reaches 0. Then any function of the form $w(x) = dx(1-x) + d\overline{z}/4 + \varepsilon$ can be used to get the result we aim for. To get the strongest asymptotically almost sure upper bound for the Broom number (for a fixed d), one should solve it numerically. To obtain a result useful for all values of d, it is straightforward to show (since the Taylor expansion of $(1+z)\ln(1+z) + (1-z)\ln(1-z)$ is $z^2 + z^4/6 + \ldots$) that $\overline{z} < 2\sqrt{\ln 2}/\sqrt{d}$.

Lemma 4.5 *Let $G \in \mathcal{G}_{n,d}$, where $d \geq 3$. Then, for every sufficiently small but fixed $\varepsilon > 0$ a.a.s. for every $S \subseteq V$*

$$|E(S, V \setminus S)| \leq \frac{d|S||V \setminus S|}{n} + \frac{dn(\overline{z}+\varepsilon)}{4} \leq \frac{d|S||V \setminus S|}{n} + \frac{\sqrt{d\ln 2}\; n}{2},$$

where $\overline{z}$ is the solution of $d((1+z)\ln(1+z) + (1-z)\ln(1-z)) = 4\ln 2$.

This result has the following implication giving a nontrivial upper bound for $d \geq 3$.

Theorem 4.6 *Let $G \in \mathcal{G}_{n,d}$, where $d \geq 3$. Then, for every sufficiently small but fixed $\varepsilon > 0$ a.a.s.*

$$B(G) \leq \frac{dn}{4}(1 + \overline{z} + \varepsilon) \leq \frac{dn}{4}\left(1 + \frac{2\sqrt{\ln 2}}{\sqrt{d}}\right),$$

where $\overline{z}$ is the solution of $d((1+z)\ln(1+z) + (1-z)\ln(1-z)) = 4\ln 2$.

This gives us the following asymptotically almost sure upper bounds u_d for the Broom number of random d-regular graph: $u_3 = 1.41n$, $u_4 = 1.78n$, $u_5 = 2.14n$, and $u_6 = 2.48n$. (In this paper, whenever we quote numerical values for computed constants such as u_d/n and l_d/n, we use a few decimal places rounded down for lower bounds and up for upper bounds.) In Figure 5, the values of u_d/dn have been presented for all d-values up to 100; we have also listed the first 30 and a few more values for higher d in Table 1 (see Section 5.5).

5 Lower bounds

5.1 A general lower bound

An argument that provides an upper bound for the brush number of a general graph (see [2]) can be easily modified to obtain a lower bound for the Broom number. This has been done in [14] to obtain the following result.

Theorem 5.1

$$B(G) \geq \frac{|E|}{2} + \frac{|V|}{4} - \frac{1}{4} \sum_{v \in V(G), \deg(v) \text{ is even}} \frac{1}{\deg(v)+1}$$

for any graph $G = (V, E)$.

Note that the bound is tight when G is a union of cliques. From this we get immediately the following corollary presented in [14].

Corollary 5.2 *Let* $G = (V, E)$ *be a* d*-regular graph on* n *vertices. If* d *is even, then*

$$B(G) \geq \frac{n}{4} \left(d + 1 - \frac{1}{d+1} \right),$$

and if d *is odd, then*

$$B(G) \geq \frac{n}{4}(d+1).$$

Both bounds are tight for every n and d satisfying $(d+1)|n$, as shown by a disjoint union of complete graphs K_{d+1}.

The bound in Theorem 5.1 holds for every d-regular graph, and for a random d-regular graph G one can slightly improve the result as follows. Denote by $maxcut(G)$ the maximum value of a cut in $G = (V, E)$. Let $V = A \cup B$ be a partition of V with $|E(A, B)| = maxcut(G)$. Define a permutation $\pi = (v_1, v_2, \ldots, v_n)$ of V by putting the vertices of A first in an arbitrary order, followed by the vertices of B in an arbitrary order. Now, let us clean graph G using permutation π. It is clear that after cleaning the vertices of A exactly one brush is sent from every edge between A and B. Moreover, these brushes cannot be reused (at least, not at this point). Thus

$$B(G) \geq b_\pi(G) \geq maxcut(G).$$

The above bound can be used to show that for a random d regular graph (for a fixed $d \geq 3$), a.a.s. $B(G) \geq nd/4 + cn\sqrt{d}$, for some absolute constant $c > 0$. Indeed, such a random graph has a.a.s. only $O(1)$ triangles, and one can (after a trivial alternation of such G) apply the result of Shearer [18] who showed that a triangle-free graph $G = (V, E)$ with degree sequence $(d_1, d_2, \ldots, d_n)$ has a cut of size at least $|E|/2 + c\sum_{i=1}^{n} \sqrt{d_i}$ for some (explicit) constant $c > 0$. We get the following result.

Theorem 5.3 *Let* $G \in \mathcal{G}_{n,d}$*, where* $d \geq 3$*. Then, a.a.s.,*

$$B(G) \geq \frac{n}{4} \left(d + \Omega(\sqrt{d}) \right).$$

5.2 Degree-greedy algorithm

The differential equations method (described in [21]) is used here to find a lower bound on the number of brushes needed to clean a graph using a degree-greedy algorithm. We describe the approach used, state some general results, and apply them to the special cases of $d = 4$ before discussing higher values of d.

5.2.1 The general setting

In this subsection, we assume $d \geq 3$ is fixed with dn even. In order to get an asymptotically almost sure lower bound on the Broom number, we study an algorithm that cleans random vertices of maximum dirty degree. This algorithm is called *degree-greedy* because the vertex being cleaned is chosen from those with the highest dirty degree.

We start with a random d-regular graph $G = (V, E)$ on n vertices. Initially, all vertices are dirty: $D_0 = V$. In every step t of the cleaning process, we clean a random vertex α_t, chosen uniformly at random from those vertices with the highest dirty degree in the induced subgraph $G[D_{t-1}]$, where $D_t = D_{t-1} \setminus \{\alpha_t\}$. In the first step, d brushes are needed to clean a random vertex α_1. The induced subgraph $G[D_1]$ now has d vertices of dirty degree $d-1$ and $n-d-1$ vertices of dirty degree d. In the second step, another d brushes are needed to clean a random vertex α_2 of dirty degree d. We keep cleaning vertices of dirty degree d until there are no more vertices of dirty degree d left and the first phase ends. In general, in the kth phase ($1 \leq k \leq d+1$) vertices of dirty degree $d-k+1$ are cleaned.

For $0 \leq i \leq d$, let $Y_i = Y_i(t)$ denote the number of vertices of degree i in $G[D_t]$ (that is, dirty degree i). (Note that $Y_0(t) = n - t - \sum_{i=1}^{d} Y_i(t)$ so $Y_0(t)$ does not need to be calculated, but it is useful in the discussion.) Let $S(t) = \sum_{l=1}^{d} lY_l(t)$ and for any statement A, let δ_A denote the Kronecker delta function

$$\delta_A = \begin{cases} 1 & \text{if } A \text{ is true,} \\ 0 & \text{otherwise.} \end{cases}$$

It is not difficult to see that

$$\begin{aligned} \mathbb{E}(Y_i(t) - Y_i(t-1) \mid G[D_{t-1}] \ \wedge \ \deg_{G[D_{t-1}]}(\alpha_t) = r) & \\ = f_{i,r}((t-1)/n, Y_1(t-1)/n, Y_2(t-1)/n, \ldots, Y_d(t-1)/n) & \\ = -\delta_{i=r} - r\frac{iY_i(t-1)}{S(t-1)} + r\frac{(i+1)Y_{i+1}(t-1)}{S(t-1)}\delta_{i+1\leq d} & \qquad (7) \end{aligned}$$

for $r \in [d]$, $0 \leq i \leq r$. Indeed, α_t has dirty degree r, hence the term $-\delta_{i=r}$. When a pair of points in the pairing model is exposed, the probability that the other point is in a bucket of dirty degree i (that is, the bucket contains i unchosen points) is asymptotic to $iY_i(t-1)/S(t-1)$. Thus $riY_i(t-1)/S(t-1)$ stands for the expected number of the r buckets found adjacent to α_t which have dirty degree i. This contributes negatively to the expected change in Y_i, whilst buckets of dirty degree $i+1$ which are reached contribute positively (of course, only if this type of vertices (buckets) exist in a graph; thus $\delta_{i+1\leq d}$). This explains (7).

From (7) it follows that the following system of differential equations should be considered when vertices of dirty degree r are cleaned

$$\frac{dy_i}{dx} = f_{i,r}(x, \mathbf{y}), \quad i = 0, 1, \ldots, r.$$

At this point we may formally define the interval $[x_{k-1}, x_k]$ to be phase k, where the termination point x_k is defined as the infimum of those $x > x_k$ for which $y_{d-k+1}(x) = 0$. Using final values $y_i(x_k)$ in phase k as initial values for phase $k+1$ we can repeat the argument inductively moving from phase to phase starting from phase 1 with obvious initial conditions $y_d(0) = 1$ and $y_i(0) = 0$ for $0 \le i \le d-1$.

The conclusion is that, for the degree-greedy algorithm we consider, with variables Y_i defined as above, we have that a.a.s.

$$Y_i(t) = ny_i(t/n) + o(n)$$

for $1 \le i \le d-k+1$ for phases $k = 1, 2, \ldots, d+1$. We omit all details, pointing the reader to the general survey [21] about the differential equations method. Let us also mention that we are interested in the first $\lceil d/2 \rceil$ phases only; the rest of the graph is cleaned 'for free'. For the same reason, we do not have to control the number of vertices of dirty degree i, $0 \le i \le \lfloor d/2 \rfloor$.

In the kth phase ($1 \le k \le \lceil d/2 \rceil$) vertices of dirty degree $d-k+1$ are cleaned. Since $d-2k+2$ brushes are needed to clean a vertex (see (1)), we need

$$l_d^k = (1+o(1))n(d-2k+2)(x_k - x_{k-1})$$

brushes in phase k. Thus, the total number of brushes needed to clean a graph using the degree-greedy algorithm is a.a.s. equal to

$$l_d = \sum_{k=1}^{\lceil d/2 \rceil} l_d^k = (1+o(1))n \sum_{k=1}^{\lceil d/2 \rceil} (d-2k+2)(x_k - x_{k-1}).$$

5.3 The end of the first phase

Before we move to a specific values of d, let us discuss the behaviour of vertices of dirty degree d during the first phase that can be studied in general. During this phase we clean vertices of dirty degree d. In order to control the number of vertices of dirty degree d, we have to consider the following differential equation.

$$z_d'(x) = -1 - \frac{d \cdot z(x)}{1-2x}$$

with the initial condition $z_d(0) = 1$. The solution is

$$z_d(x) = -\frac{1-2x}{d-2} + \frac{d-1}{d-2}(1-2x)^{d/2}, \tag{8}$$

and thus the first phase finishes at time

$$t_1 = \frac{n}{2}\left(1 - \left(\frac{1}{d-1}\right)^{\frac{2}{d-2}}\right) \tag{9}$$

(the second root of the equation $z_d(x) = 0$ is $1/2$).

This initial result suggests that we should focus on differential equations of the following form

$$z'(x) = \frac{a \cdot z(x)}{q-2x} + \sum_{i=1}^{s} b_i(q-2x)^{c_i}.$$

The general solution to this differential equation is

$$\begin{aligned} z(x) &= C(q-2x)^{-a/2} - \sum_{i=1}^{s} \frac{b_i(q-2x)^{c_i+1}}{a+2(c_i+1)} \delta_{a+2(c_i+1)\neq 0} \\ & -\frac{b}{2}(q-2x)^{-a/2}\ln(q-2x)\delta_{\exists i, a+2(c_i+1)=0}. \end{aligned} \tag{10}$$

5.4 4-regular graphs

For 4-regular graphs, to estimate the Broom number one has to carefully analyze phases 1 and 2. During the first phase, we need four brushes to clean vertices of dirty degree 4. From (8) and (9), we get that

$$z_4(x) = -\frac{1}{2}(1-2x) + \frac{3}{2}(1-2x)^2$$

and the first phase ends at time $t_1 = \frac{n}{3}$. To study the number of vertices of dirty degree 3 during the first phase, we consider the following differential equation

$$z_3'(x) = \frac{-3z_3(x)}{1-2x} + \frac{4z_4(x)}{1-2x} = \frac{-3z_3(x)}{1-2x} - 2 + 6(1-2x)$$

with the initial condition $z_3(0) = 0$. The solution (see (10) and Figure 3 (a)) is

$$z_3(x) = (1-2x)\left(8(1-2x)^{1/2} - 2 - 6(1-2x)\right)$$

so a.a.s. $Y_3(t_1) = (\frac{8\sqrt{3}}{9} - \frac{4}{3})n(1+o(1)) \approx 0.2063n$.

During the second phase, we need two brushes to clean each vertex of dirty degree 3. Now, the differential equation we need to consider is

$$z_3'(x) = -1 - \frac{3z_3(x)}{10/9 - 2x}$$

with the initial condition $z_3\left(\frac{1}{3}\right) = \frac{8\sqrt{3}}{9} - \frac{4}{3}$ (note that the number of points in the pairing model at time t is $\frac{10n}{3} - 6t$). The solution (see (10) and Figure 3 (b)) is

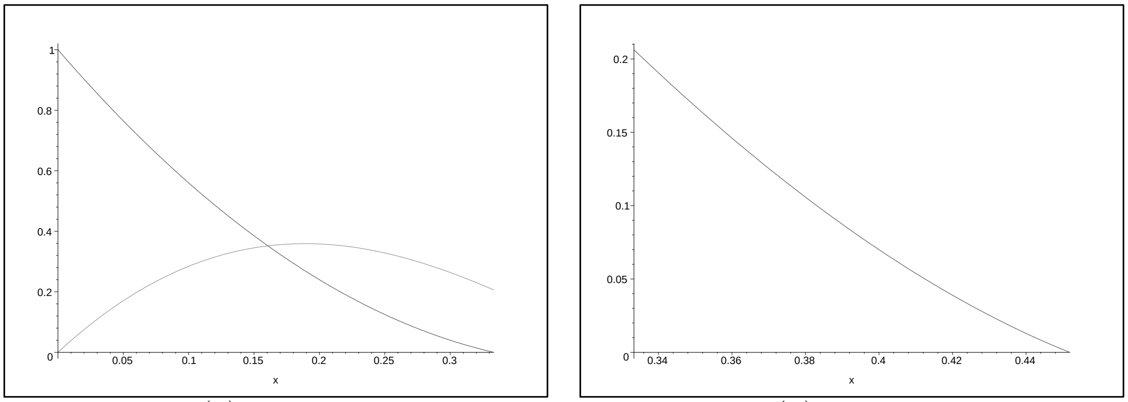
$$z_3(x) = \left(\frac{10}{9} - 2x\right)\left(\left(3\sqrt{3}-3\right)\left(\frac{10}{9} - 2x\right)^{1/2} - 1\right)$$

so the second phase ends at time $t_2 = (\frac{1}{2} - \frac{\sqrt{3}}{36})n \approx 0.4519n$. Therefore, we get an asymptotically almost sure lower bound of

$$l_4 = (1+o(1))(4t_1 + 2(t_2 - t_1)) = (1+o(1))n\left(\frac{5}{3} - \frac{\sqrt{3}}{18}\right) \approx 1.57044n.$$

On the other hand, it is true that a.a.s. a random 4-regular graph can be decomposed into two edge-disjoint Hamilton cycles [12], and hence four paths.

The solutions to the relevant differential equations for $d = 4$ are shown in Figure 3.



(a) phase 1 (b) phase 2

Figure 2: Solution to the differential equations: 4-regular graph.

d	l_d/n	u_d/n	d	l_d/n	u_d/n	d	l_d/n	u_d/n	d	l_d/n	u_d/n
3	1.25	1.41	11	3.78	4.11	19	6.13	6.55	27	8.41	8.90
4	1.57	1.78	12	4.07	4.42	20	6.41	6.84	28	8.69	9.19
5	1.92	2.14	13	4.38	4.73	21	6.71	7.14	29	8.97	9.48
6	2.22	2.48	14	4.66	5.04	22	6.98	7.44	30	9.25	9.77
7	2.56	2.82	15	4.97	5.34	23	7.28	7.73	31	9.54	10.06
8	2.85	3.15	16	5.25	5.65	24	7.55	8.02	32	9.81	10.34
9	3.18	3.47	17	5.55	5.95	25	7.85	8.32	99	28.00	28.89
10	3.47	3.79	18	5.83	6.25	26	8.12	8.61	100	28.26	29.16

Table 1: Approximate upper and lower bounds on the Broom number.

5.5 d-regular graphs of higher order

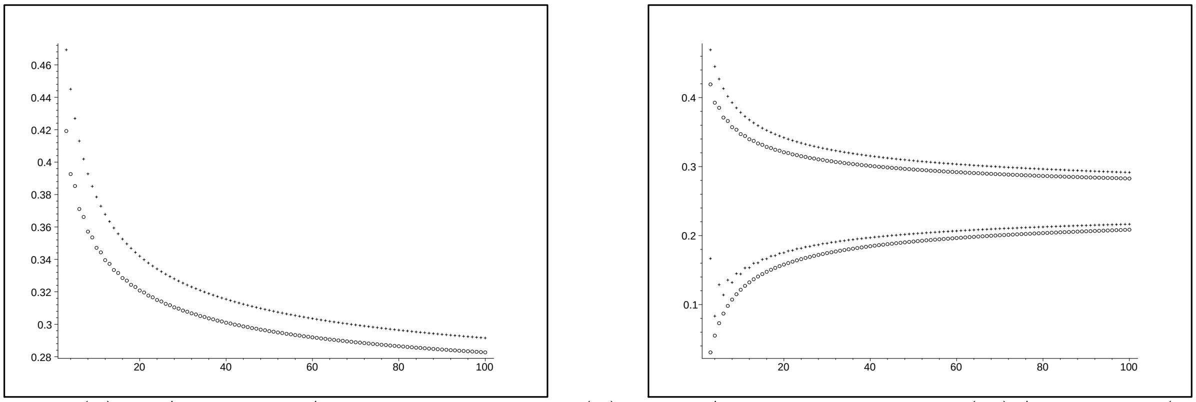
In Figure 5, the values of l_d/dn and u_d/dn have been presented for all d-values up to 100, although we have only listed the first 30 and a few more values for higher d in Table 1. The values of a lower and an upper bounds of $b(G)/dn$ have been presented as well (see [2] for more). The computations presented in the paper were performed by using Maple$^{\text{TM}}$ [16]. The worksheets can be found at the following address: "`http://www.math.wvu.edu/~pralat/`".

For each value of $d \geq 3$ there is a gap between l_d/dn and u_d/dn but the gap becomes smaller for higher values of d which follows from the following theorem.

Theorem 5.4 *The Broom number of a random d-regular graph is asymptotically almost surely* $\frac{n}{4}(d + \Theta(\sqrt{d}))$. *Moreover,* $\lim_{d\to\infty} l_d/dn = 1/4$, *that is, for large d, the degree-greedy algorithm a.a.s. achieves the optimal number of brushes up to a lower order term.*

Proof: The first part of the theorem follows from Theorem 5.1 (see also Theorem 5.3) and Theorem 4.4 (see also Theorem 4.6).

It remains to estimate the performance of the degree-greedy algorithm. Let $d \geq 3$ be an integer, and let $G \in \mathcal{G}_{n,d}$, as before. It follows from Lemmas 4.2 and 4.3 that a.a.s. for all



(a) u_d/dn and l_d/dn versus d (b) lower/upper bounds of $b(G)/dn$ and $B(G)/dn$

Figure 3: A graph of lower/upper bound of $b(G)/dn$ and $B(G)/dn$ versus d.

$m \in \{0, 1, \ldots, n-1\}$ and all sets $X \subseteq V$ with $|X| = m$,

$$|E(G[V \setminus X])| \geq \frac{(n-m)^2 d}{2n} - \frac{1}{2} 2\sqrt{d}(n-m)\,,$$

since the number of edges inside $G[V \setminus X]$ is $|E(V \setminus X, V \setminus X)|/2$. So the average degree of $G[V \setminus X]$ (and thus the maximum degree as well) is at least

$$\xi_m = \max\left\{\frac{(n-m)d}{n} - 2\sqrt{d}, 0\right\}.$$

Thus, using (1) we get that a.a.s. the number of brushes used by the degree-greedy algorithm is at least

$$\sum_{m=0}^{n-1} \max\{2\xi_m - d, 0\} \geq \frac{dn}{4} - O(\sqrt{d}n)\,.$$

It follows, by Theorem 4.6, that for large d the greedy algorithm achieves, a.a.s., essentially the optimum number of brushes. This completes the proof of the theorem. ∎

6 Possible improvement

We conclude the paper with an open problem that can be posed for any $d \geq 3$. It follows from Theorem 5.4 that the degree greedy algorithm we study a.a.s. achieves the optimal number of brushes up to a lower order term. However, for a small values of d, one can try to improve the algorithm and use more brushes by starting with a larger independent set in the first phase; the cleaning process can be continued using the degree-greedy approach as before. In our case, the number of vertices cleaned during the first phase is $(1-(d-1)^{-2/(d-2)})n/2$ (see (9)) but other algorithms for finding a larger independent set are known (see, for example, [19, 6, 7]).

7 Acknowledgements

We would like to thank the anonymous referee who helped improve the paper.

References

[1] N. Alon and F.R.K. Chung, Explicit construction of linear sized tolerant networks, *Discrete Math.*, **72** (1988), 15–19.

[2] N. Alon, P. Prałat, and N.C. Wormald, Cleaning d-regular graphs with brushes, *SIAM Journal on Discrete Mathematics* **23** (2008), 233–250.

[3] N. Alon and J.H. Spencer, The Probabilistic Method, Wiley, 1992 (Third Edition, 2008).

[4] T. Biedl, T. Chan, Y. Ganjali, M. Hajiaghayo, and D. Wood, Balanced vertex–orderings of graphs, *Discrete Applied Mathematics* **148**(1) (2005) 27–48.

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

[6] W. Duckworth and M. Zito, Uncover low degree vertices and minimise the mess: independent sets in random regular graphs, *Proccedings of the 32nd International Symposium, Mathematical Foundations of Computer Science*, Lecture Notes in Computer Science **4708**, Springer, 2007, 56–66.

[7] D. Fernholz, Independent set heuristics for random 3-regular graphs, submitted, 39pp.

[8] J. Friedman, A proof of Alon's second eigenvalue conjecture, *Memoirs of the A.M.S.*, accepted, 118pp.

[9] S. Gaspers, M.E. Messinger, R. Nowakowski, and P. Prałat, Clean the graph before you draw it!, *Information Processing Letters* **109** (2009), 463–467.

[10] S. Gaspers, M.E. Messinger, R. Nowakowski, and P. Prałat, Parallel Cleaning of a Network with Brushes, *Discrete Applied Mathematics* **158** (2010), 467–478.

[11] S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their applications, *Bull. Amer. Math. Soc. (N.S.)* **43** (2006), no. 4, 439–561.

[12] J.H. Kim and 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.

[13] M.E. Messinger, R.J. Nowakowski, and P. Prałat, Cleaning a Network with Brushes, *Theoretical Computer Science* **399** (2008), 191–205.

[14] M.E. Messinger, R.J. Nowakowski, and P. Prałat, Cleaning with Brooms, *Graphs and Combinatorics*, accepted, 16pp.

[15] M.E. Messinger, R.J. Nowakowski, P. Prałat, and N.C. Wormald, Cleaning random d-regular graphs with brushes using a degree-greedy algorithm, *Proceedings of the 4th Workshop on Combinatorial and Algorithmic Aspects of Networking*, Lecture Notes in Computer Science **4852**, Springer, 2007, 13–26

[16] M.B. Monagan, K.O. Geddes, K.M. Heal, G. Labahn, S.M. Vorkoetter, J. McCarron, and P. DeMarco, Maple 10 Programming Guide, Maplesoft, Waterloo ON, Canada, 2005.

[17] P. Prałat, Cleaning random graphs with brushes, *Australasian Journal of Combinatorics* **43** (2009), 237–251.

[18] J. Shearer, A note on bipartite subgraphs of triangle-free graphs, *Random Structures Algorithms* **3** (1992), 223–226.

[19] N.C. Wormald, Differential equations for random processes and random graphs, *Annals of Applied Probability* **5** (1995), 1217–1235.

[20] N.C. Wormald, Models of random regular graphs, Surveys in Combinatorics, 1999, J.D. Lamb and D.A. Preece, eds. London Mathematical Society Lecture Note Series, vol 276, pp. 239–298, Cambridge University Press, Cambridge, 1999.

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