Cleaning random d-regular graphs with brushes and Brooms

Pawel Pralat

Department of Mathematics and Statistics, Dalhousie University
(joint work with Noga Alon and Nick Wormald)

Jagiellonian University, April 2009

The web graph

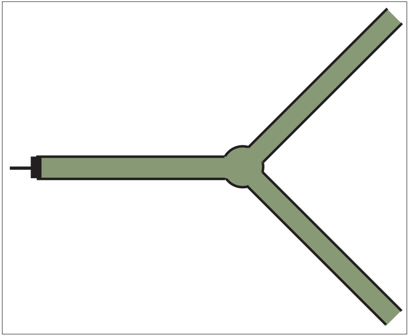
nodes: *web pages* edges: *hyperlinks*

Outline

Outline

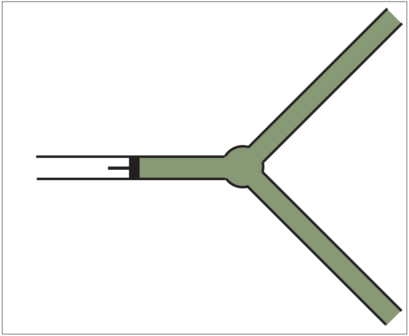
Consider a network of pipes with a biological contaminant (e.g. water pipes that have algae or zebra mussels contamination) that

- is mobile so that contamination can spread from one area to another,
- grows back over time if removed.



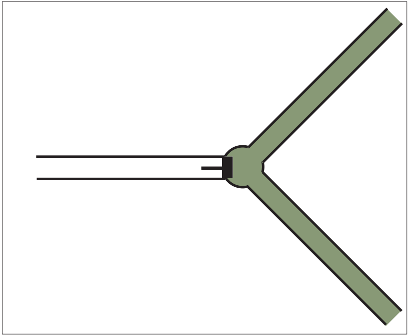
Consider a network of pipes with a biological contaminant (e.g. water pipes that have algae or zebra mussels contamination) that

- is mobile so that contamination can spread from one area to another,
- grows back over time if removed.



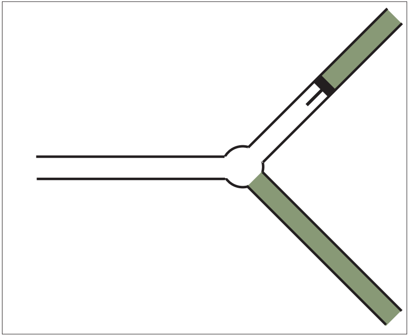
Consider a network of pipes with a biological contaminant (e.g. water pipes that have algae or zebra mussels contamination) that

- is mobile so that contamination can spread from one area to another,
- grows back over time if removed.



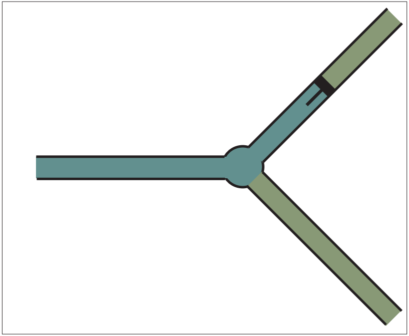
Consider a network of pipes with a biological contaminant (e.g. water pipes that have algae or zebra mussels contamination) that

- is mobile so that contamination can spread from one area to another,
- grows back over time if removed.



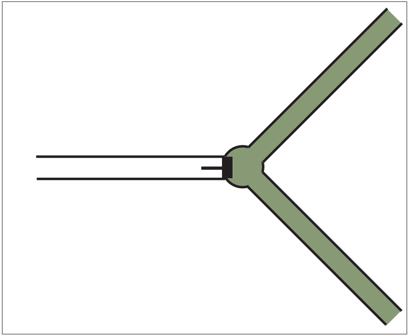
Consider a network of pipes with a biological contaminant (e.g. water pipes that have algae or zebra mussels contamination) that

- is mobile so that contamination can spread from one area to another,
- grows back over time if removed.



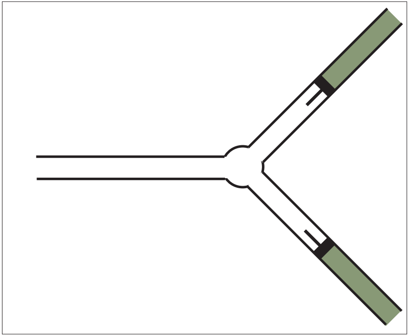
Consider a network of pipes with a biological contaminant (e.g. water pipes that have algae or zebra mussels contamination) that

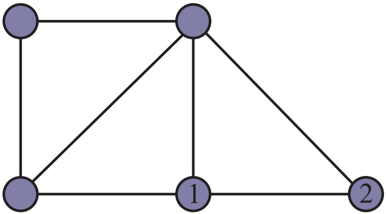
- is mobile so that contamination can spread from one area to another,
- grows back over time if removed.



Consider a network of pipes with a biological contaminant (e.g. water pipes that have algae or zebra mussels contamination) that

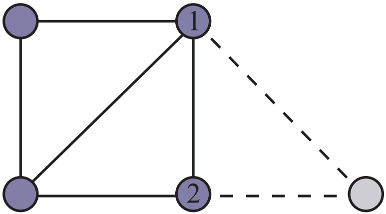
- is mobile so that contamination can spread from one area to another,
- grows back over time if removed.





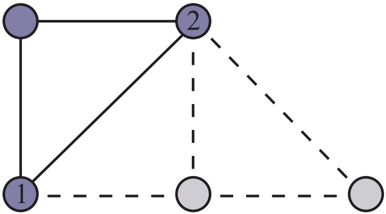
Definition (Cleaning algorithm)

- 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.
- Brushes cannot traverse a clean edge.



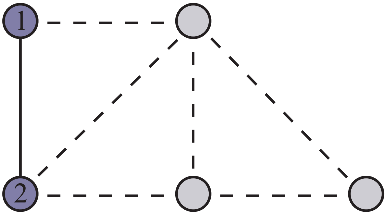
Definition (Cleaning algorithm)

- 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.
- Brushes cannot traverse a clean edge.



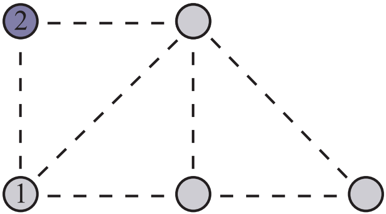
Definition (Cleaning algorithm)

- 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.
- Brushes cannot traverse a clean edge.



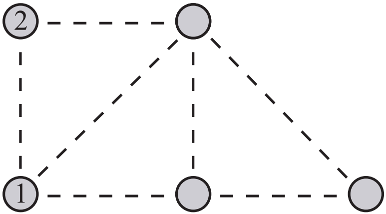
Definition (Cleaning algorithm)

- 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.
- Brushes cannot traverse a clean edge.



Definition (Cleaning algorithm)

- 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.
- Brushes cannot traverse a clean edge.



Definition (Cleaning algorithm)

- 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.
- Brushes cannot traverse a clean edge.

We can get two different final configurations but the final set of dirty vertices is determined.

Theorem (Messinger, Nowakowski, Pralat)

Given a graph G and the initial configuration of brushes $\omega_0 : V \to \mathbb{N} \cup \{0\}$, the cleaning algorithm returns a unique final set of dirty vertices.

We can get two different final configurations but the final set of dirty vertices is determined.

Theorem (Messinger, Nowakowski, Pralat)

Given a graph G and the initial configuration of brushes $\omega_0 : V \to \mathbb{N} \cup \{0\}$, the cleaning algorithm returns a unique final set of dirty vertices.

Thus, the following definition is natural.

Definition (brush number)

A graph $G = (V, E)$ **can be cleaned** by the initial configuration of brushes ω_0 if the cleaning process returns an empty final set of dirty vertices.

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

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

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

Thus, the following definition is natural.

Definition (brush number)

A graph $G = (V, E)$ **can be cleaned** by the initial configuration of brushes ω_0 if the cleaning process returns an empty final set of dirty vertices.
Let the brush number, $b(G)$, be the minimum number of brushes needed to clean G, that is,

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

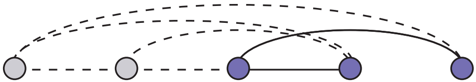
Similarly, $b_\alpha(G)$ is defined as the minimum number of brushes needed to clean G using the cleaning sequence α.

In general, it is difficult to find $b(G)$...

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

The problem is $\mathcal{NP}$-complete and remains $\mathcal{NP}$-complete for bipartite graphs of maximum degree 6, planar graphs of maximum degree 4, and 5-regular graphs.

...but $b_\alpha(G)$ can be easily computed.

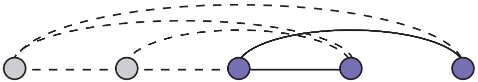


In general, it is difficult to find $b(G)$...

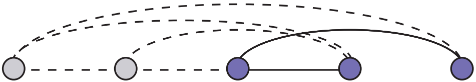
Theorem (Gaspers, Messinger, Nowakowski, Pralat)

The problem is $\mathcal{NP}$-complete and remains $\mathcal{NP}$-complete for bipartite graphs of maximum degree 6, planar graphs of maximum degree 4, and 5-regular graphs.

...but $b_\alpha(G)$ can be easily computed.



...but $b_\alpha(G)$ can be easily computed.

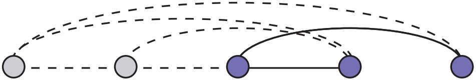


Let $D(v)$ be the number of dirty neighbours of v at the time when v is cleaned.

- The number of brushes arriving at a vertex before it is cleaned equals $\deg(v) - D(v)$
- The total number of brushes needed is D(v)

$$\omega_0(v) = \max\{2D(v) - \deg(v), 0\}, \quad \text{for } v \in V.$$

...but $b_\alpha(G)$ can be easily computed.

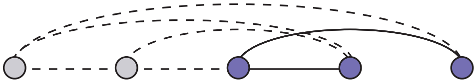


Let $D(v)$ be the number of dirty neighbours of v at the time when v is cleaned.

- The number of brushes arriving at a vertex before it is cleaned equals $\deg(v) - D(v)$
- The total number of brushes needed is D(v)

$$\omega_0(v) = \max\{2D(v) - \deg(v), 0\}, \quad \text{for } v \in V.$$

...but $b_\alpha(G)$ can be easily computed.

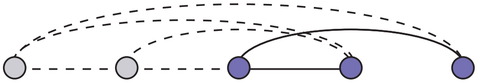


Let $D(v)$ be the number of dirty neighbours of v at the time when v is cleaned.

- The number of brushes arriving at a vertex before it is cleaned equals $\deg(v) - D(v)$
- The total number of brushes needed is D(v)

$$\omega_0(v) = \max\{2D(v) - \deg(v), 0\}, \quad \text{for } v \in V.$$

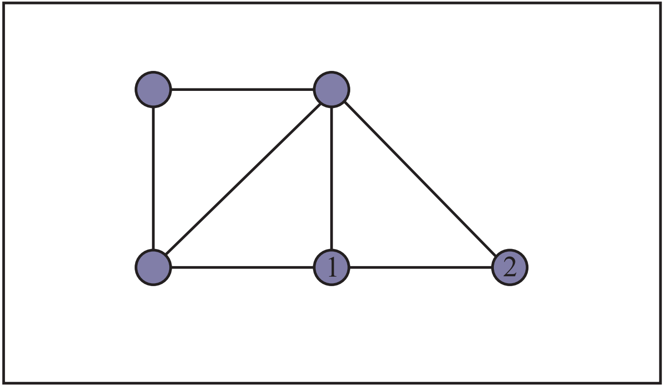
...but $b_\alpha(G)$ can be easily computed.



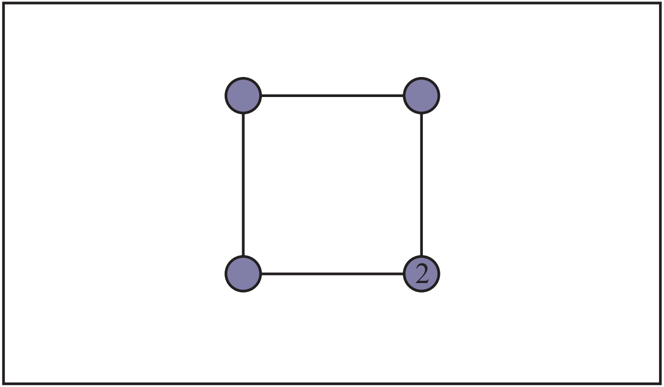
Let $D(v)$ be the number of dirty neighbours of v at the time when v is cleaned.

- The number of brushes arriving at a vertex before it is cleaned equals $\deg(v) - D(v)$
- The total number of brushes needed is D(v)

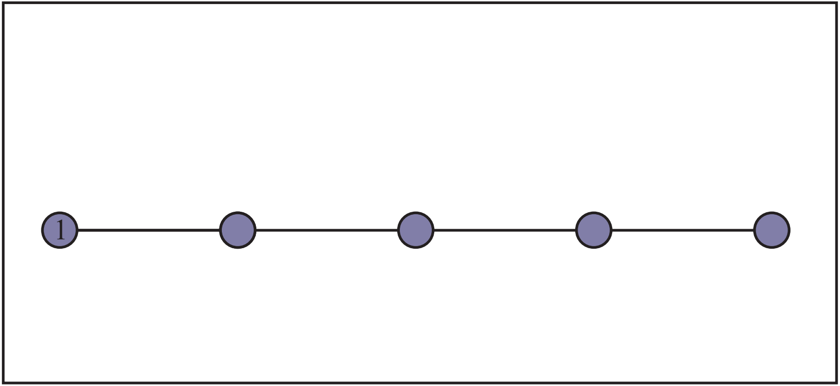
$$\omega_0(v) = \max\{2D(v) - \deg(v), 0\}, \quad \text{for } v \in V.$$



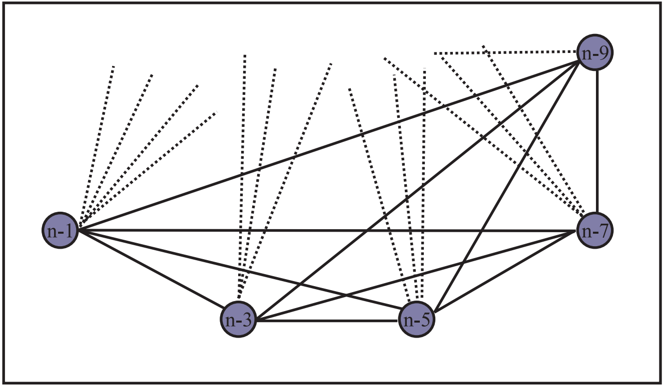
$b(G) = 3$



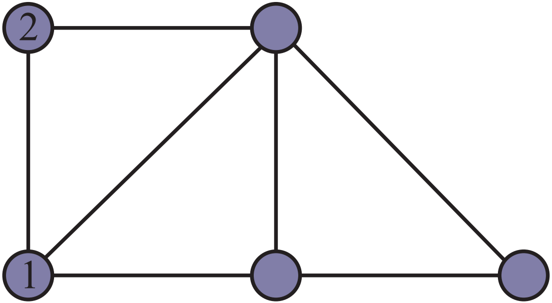
$b(C_n) = 2, n \geq 3$

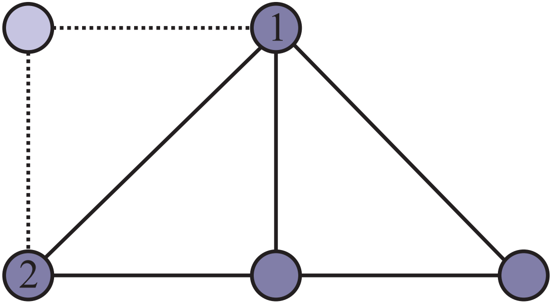


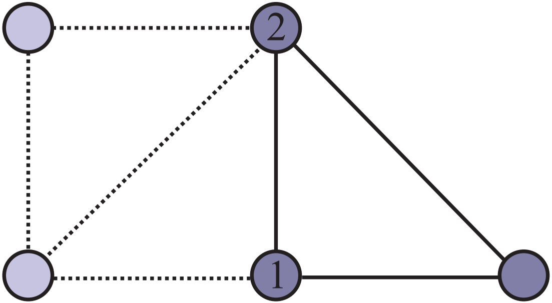
$b(P_n) = 1$, $n \geq 2$

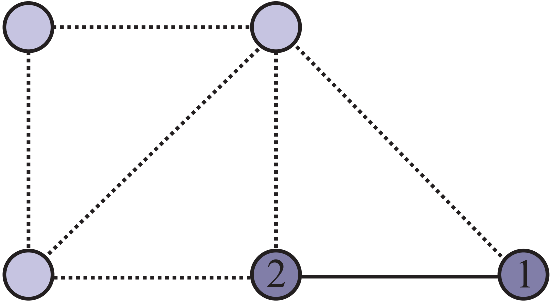


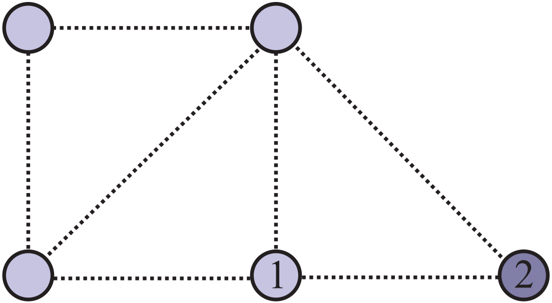
$$b(K_n) = \begin{cases} \frac{n^2}{4} & \text{if n is even} \\ \frac{n^2-1}{4} & \text{if n is odd.} \end{cases}$$

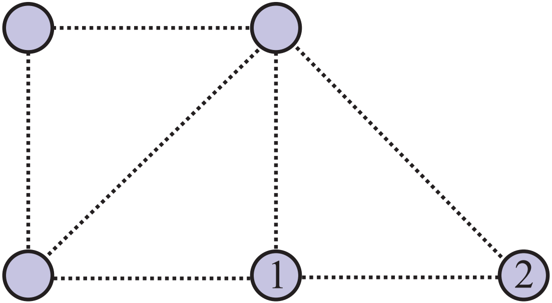

2
1

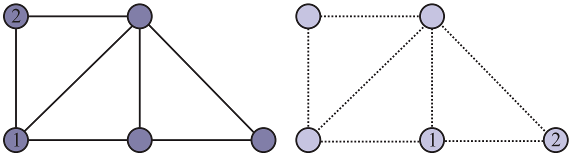

1
2


2
1


2
1


1
2


1
2



Theorem (Messinger, Nowakowski, Pralat)

The Reversibility Theorem

Given the initial configuration ω_0, suppose G can be cleaned yielding final configuration ω_n, $n = |V(G)|$. Then, given the initial configuration $\tau_0 = \omega_n$, G can be cleaned yielding the final configuration $\tau_n = \omega_0$.

The cleaning process is a combination of both

- the edge-searching problem:
  - modeling sequential computation,
  - assuring privacy when using bugged channels,
  - VLSI circuit design,
  - security in the web graph.
- the chip firing game:
  - the Tutte polynomial and group theory,
  - algebraic potential theory (social science).

There is also a relationship between the Cleaning problem and the Balanced Vertex-Ordering problem (this has consequences for both problems).

The cleaning process is a combination of both

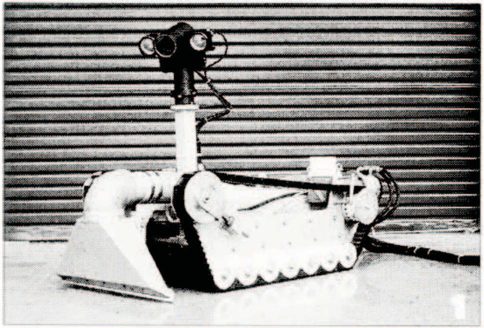
- the edge-searching problem:
  - modeling sequential computation,
  - assuring privacy when using bugged channels,
  - VLSI circuit design,
  - security in the web graph.
- the chip firing game:
  - the Tutte polynomial and group theory,
  - algebraic potential theory (social science).

There is also a relationship between the Cleaning problem and the Balanced Vertex-Ordering problem (this has consequences for both problems).

The cleaning process is a combination of both

- the edge-searching problem:
  - modeling sequential computation,
  - assuring privacy when using bugged channels,
  - VLSI circuit design,
  - security in the web graph.
- the chip firing game:
  - the Tutte polynomial and group theory,
  - algebraic potential theory (social science).

There is also a relationship between the Cleaning problem and the Balanced Vertex-Ordering problem (this has consequences for both problems).



S.R. Kotler, E.C. Mallen, K.M. Tammus, Robotic Removal of Zebra Mussel Accumulations in a Nuclear Power Plant Screenhouse, *Proceedings of The Fifth International Zebra Mussel and Other Aquatic Nuisance Organisms Conference*, Toronto, Canada, February 1995

Our results refer to the probability space of random d-regular graphs with uniform probability distribution. This space is denoted $\mathcal{G}_{n,d}$.

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.

For example, random 4-regular graph is connected a.a.s.; that is,

$$\lim_{n\to\infty} \frac{\text{\# of connected 4-regular graphs on } n \text{ vertices}}{\text{\# of 4-regular graphs on } n \text{ vertices}} = 1\,.$$

Our results refer to the probability space of random d-regular graphs with uniform probability distribution. This space is denoted $\mathcal{G}_{n,d}$.

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.

For example, random 4-regular graph is connected a.a.s.; that is,

$$\lim_{n\to\infty} \frac{\text{\# of connected 4-regular graphs on } n \text{ vertices}}{\text{\# of 4-regular graphs on } n \text{ vertices}} = 1\,.$$

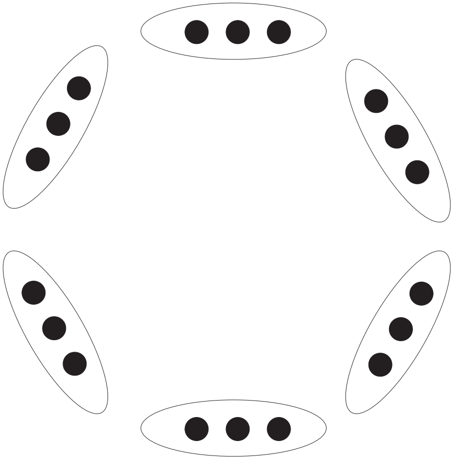
Our results refer to the probability space of random d-regular graphs with uniform probability distribution. This space is denoted $\mathcal{G}_{n,d}$.

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.

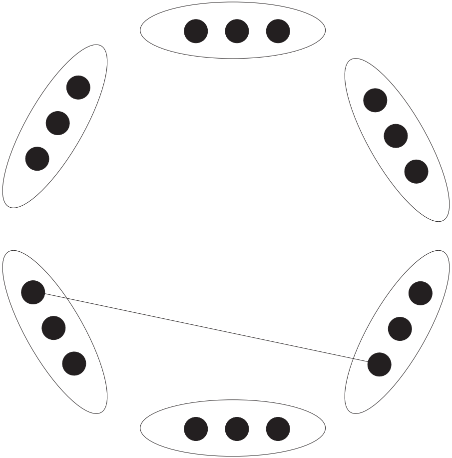
For example, random 4-regular graph is connected a.a.s.; that is,

$$\lim_{n\to\infty} \frac{\text{\# of connected 4-regular graphs on } n \text{ vertices}}{\text{\# of 4-regular graphs on } n \text{ vertices}} = 1\,.$$

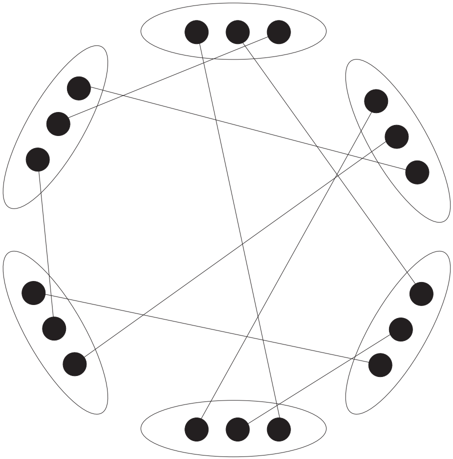
Pairing model: $n = 6$, $d = 3$



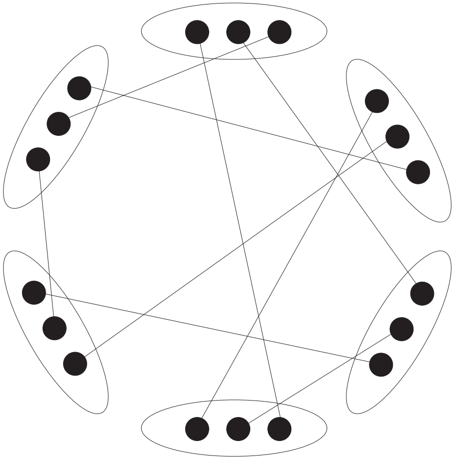
Pairing model: $n = 6$, $d = 3$

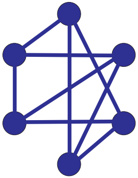


Pairing model: $n = 6$, $d = 3$

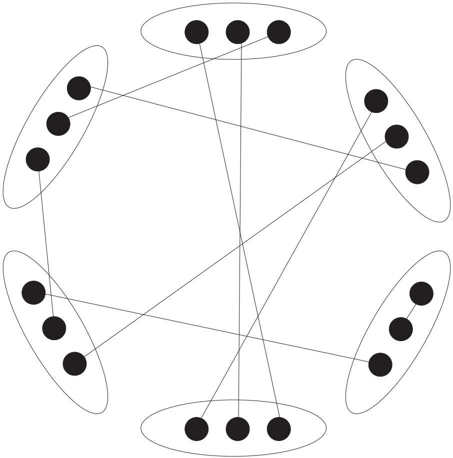


Pairing model: $n = 6$, $d = 3$

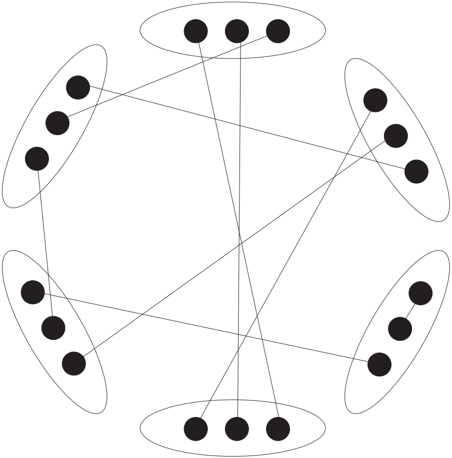


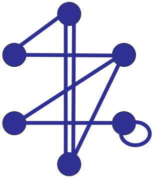


Pairing model: $n = 6$, $d = 3$



Pairing model: $n = 6$, $d = 3$





Pairing model:

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, a random pairing generates a simple graph with probability asymptotic to $e^{(1-d^2)/4}$ depending on d.

Therefore, any event holding a.a.s. over the probability space of random pairings also holds a.a.s. over the corresponding space $\mathcal{G}_{n,d}$.

Pairing model:

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, a random pairing generates a simple graph with probability asymptotic to $e^{(1-d^2)/4}$ depending on d.

Therefore, any event holding a.a.s. over the probability space of random pairings also holds a.a.s. over the corresponding space $\mathcal{G}_{n,d}$.

Pairing model:

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, a random pairing generates a simple graph with probability asymptotic to $e^{(1-d^2)/4}$ depending on d.

Therefore, any event holding a.a.s. over the probability space of random pairings also holds a.a.s. over the corresponding space $\mathcal{G}_{n,d}$.

Outline

2-regular graphs

Let Y_n be the total number of cycles in a random 2-regular graph on n vertices. Since exactly two brushes are needed to clean one cycle, we need $2Y_n$ brushes in order to clean a 2-regular graph.

It can be shown that the total number of cycles Y_n is sharply concentrated near $(1/2) \log n$.

Theorem (Alon, Pralat, Wormald)

Let G be a random 2-regular graph on n vertices. Then, a.a.s.

$$b(G) = (1 + o(1)) \log n \, .$$

2-regular graphs

Let Y_n be the total number of cycles in a random 2-regular graph on n vertices. Since exactly two brushes are needed to clean one cycle, we need $2Y_n$ brushes in order to clean a 2-regular graph.

It can be shown that the total number of cycles Y_n is sharply concentrated near $(1/2)\log n$.

Theorem (Alon, Pralat, Wormald)

Let G be a random 2-regular graph on n vertices. Then, a.a.s.

$$b(G) = (1 + o(1)) \log n \, .$$

2-regular graphs

Let Y_n be the total number of cycles in a random 2-regular graph on n vertices. Since exactly two brushes are needed to clean one cycle, we need $2Y_n$ brushes in order to clean a 2-regular graph.

It can be shown that the total number of cycles Y_n is sharply concentrated near $(1/2)\log n$.

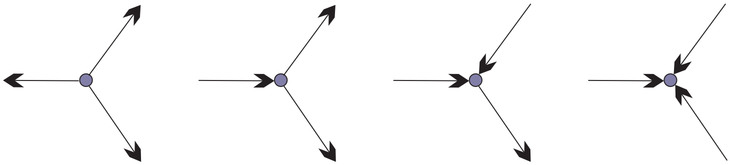
Theorem (Alon, Pralat, Wormald)

Let G be a random 2-regular graph on n vertices. Then, a.a.s.

$$b(G) = (1 + o(1)) \log n \, .$$

3-regular graphs

The first vertex cleaned must start three brush paths, the last one terminates three brush paths, and all other vertices must start or finish at least one brush path, so the number of brush paths is at least $n/2+2$.



It is known that a random 3-regular graph a.a.s. has a Hamilton cycle. The edges not in a Hamilton cycle must form a perfect matching. Such a graph can be cleaned by starting with three brushes at one vertex, and moving along the Hamilton cycle with one brush, introducing one new brush for each edge of the perfect matching.

Theorem (Alon, Pralat, Wormald)

Let G be a random 3-regular graph on n vertices. Then, a.a.s.

$$b(G) = n/2 + 2 .$$

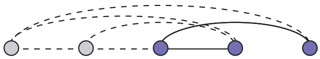
It is known that a random 3-regular graph a.a.s. has a Hamilton cycle. The edges not in a Hamilton cycle must form a perfect matching. Such a graph can be cleaned by starting with three brushes at one vertex, and moving along the Hamilton cycle with one brush, introducing one new brush for each edge of the perfect matching.

Theorem (Alon, Pralat, Wormald)

Let G be a random 3-regular graph on n vertices. Then, a.a.s.

$$b(G) = n/2 + 2\,.$$

Outline

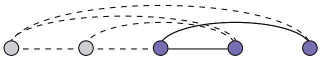


$$b(G) \geq \max_j \min_{S \subseteq V, |S|=j} |E(S, V \setminus S)| .$$

(The proof is simply to observe that the minimum is a lower bound on the number of edges going from the first j vertices cleaned to elsewhere in the graph.)

Suppose that $x = x(n)$ and $y = y(n)$ are chosen so 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 tending to zero with n.

Then this, together with the first moment principle, gives that the brush number is a.a.s. at least yn.

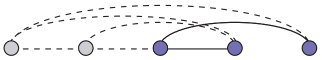


$$b(G) \geq \max_{j} \min_{S \subseteq V, |S|=j} |E(S, V \setminus S)| .$$

(The proof is simply to observe that the minimum is a lower bound on the number of edges going from the first j vertices cleaned to elsewhere in the graph.)

Suppose that $x = x(n)$ and $y = y(n)$ are chosen so 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 tending to zero with n.

Then this, together with the first moment principle, gives that the brush number is a.a.s. at least yn.



$$b(G) \geq \max_{j} \min_{S \subseteq V, |S|=j} |E(S, V \setminus S)| .$$

(The proof is simply to observe that the minimum is a lower bound on the number of edges going from the first j vertices cleaned to elsewhere in the graph.)

Suppose that $x = x(n)$ and $y = y(n)$ are chosen so 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 tending to zero with n.

Then this, together with the first moment principle, gives that the brush number is a.a.s. at least yn.

In order to find an optimal values of x and y we use the pairing model. It is clear that

$$
\begin{aligned}
S(x,y) &= \binom{n}{xn}\binom{xdn}{yn}M(xdn-yn)\binom{(1-x)dn}{yn}(yn)! \\
&\quad \times M((1-x)dn-yn)/M(dn)
\end{aligned}
$$

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

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

After simplification and using Stirling's formula we get that $S(x,y) < O(n^{-1})e^{f(x,y,d)}$ where

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

Thus, if $f(x,y,d) = 0$, then $S(x,y)$ tends to zero with n and the brush number is at least yn a.a.s.

After simplification and using Stirling's formula we get that $S(x,y) < O(n^{-1})e^{f(x,y,d)}$ where

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

Thus, if $f(x,y,d)=0$, then $S(x,y)$ tends to zero with n and the brush number is at least yn a.a.s.

Not surprisingly, the strongest bound is obtained for $x = 1/2$, in which case $f(x, y, d)$ becomes

$$(d-1)\ln(1/2) + (d/2)\ln d - y\ln y - (d/2-y)\ln(d/2-y)$$
$$= -\frac{d}{4}((1+z)\ln(1+z) + (1-z)\ln(1-z)) + \ln 2$$

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

It is straightforward to see that this function is decreasing in z for $z \geq 0$. Let l_d/n denote the value of y for which it first reaches 0.

Since the Taylor expansion of $(1+z)\ln(1+z) + (1-z)\ln(1-z)$ is $z^2 + z^4/6 + \ldots$, $l_d/n \geq (d/4)(1 - 2\sqrt{\ln 2}/\sqrt{d})$.

Not surprisingly, the strongest bound is obtained for $x = 1/2$, in which case $f(x, y, d)$ becomes

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

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

It is straightforward to see that this function is decreasing in z for $z \geq 0$. Let l_d/n denote the value of y for which it first reaches 0.

Since the Taylor expansion of $(1+z)\ln(1+z) + (1-z)\ln(1-z)$ is $z^2 + z^4/6 + \ldots$, $l_d/n \geq (d/4)(1 - 2\sqrt{\ln 2}/\sqrt{d})$.

Not surprisingly, the strongest bound is obtained for $x = 1/2$, in which case $f(x, y, d)$ becomes

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

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

It is straightforward to see that this function is decreasing in z for $z \geq 0$. Let l_d/n denote the value of y for which it first reaches 0.

Since the Taylor expansion of $(1+z)\ln(1+z) + (1-z)\ln(1-z)$ is $z^2 + z^4/6 + \ldots$, $l_d/n \geq (d/4)(1 - 2\sqrt{\ln 2}/\sqrt{d})$.

Theorem (Alon, Pralat, Wormald)

Let G be a random d-regular graph on n vertices. Then, a.a.s.

$$b(G) \geq \frac{dn}{4}\left(1 - \frac{2\sqrt{\ln 2}}{\sqrt{d}}\right).$$

Outline

We study an algorithm that cleans random vertices of minimum degree in the graph induced by a set of dirty vertices.

In the kth phase a mixture of vertices of degree $d-k$ and $d-k-1$ are cleaned. There are two possible endings.

1. the vertices of degree $d-k$ are becoming so common that the vertices of degree $d-k-1$ start to explode (in which case we move to the next phase),
2. the vertices of degree $d-k+1$ are getting so rare that those of degree $d-k$ disappear (in which case the process goes “backwards”).

With various initial conditions, either one could occur.

This degree-greedy algorithm can be analyzed using the differential equations method.

We study an algorithm that cleans random vertices of minimum degree in the graph induced by a set of dirty vertices.

In the kth phase a mixture of vertices of degree $d - k$ and $d - k - 1$ are cleaned. There are two possible endings.

1. the vertices of degree $d - k$ are becoming so common that the vertices of degree $d - k - 1$ start to explode (in which case we move to the next phase),
2. the vertices of degree $d - k + 1$ are getting so rare that those of degree $d - k$ disappear (in which case the process goes "backwards").

With various initial conditions, either one could occur.

This degree-greedy algorithm can be analyzed using the differential equations method.

We study an algorithm that cleans random vertices of minimum degree in the graph induced by a set of dirty vertices.

In the kth phase a mixture of vertices of degree $d - k$ and $d - k - 1$ are cleaned. There are two possible endings.

1. the vertices of degree $d - k$ are becoming so common that the vertices of degree $d - k - 1$ start to explode (in which case we move to the next phase),
2. the vertices of degree $d - k + 1$ are getting so rare that those of degree $d - k$ disappear (in which case the process goes "backwards").

With various initial conditions, either one could occur.

This degree-greedy algorithm can be analyzed using the differential equations method.

We study an algorithm that cleans random vertices of minimum degree in the graph induced by a set of dirty vertices.

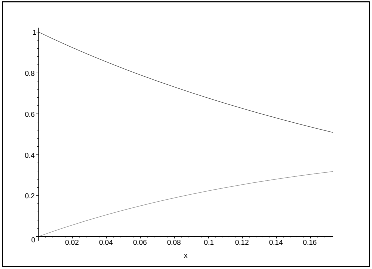
In the kth phase a mixture of vertices of degree $d-k$ and $d-k-1$ are cleaned. There are two possible endings.

1. the vertices of degree $d-k$ are becoming so common that the vertices of degree $d-k-1$ start to explode (in which case we move to the next phase),
2. the vertices of degree $d-k+1$ are getting so rare that those of degree $d-k$ disappear (in which case the process goes “backwards”).

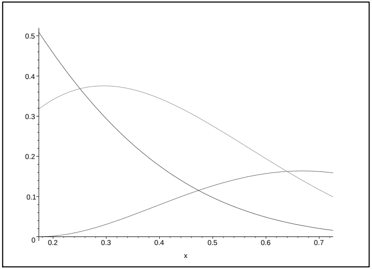
With various initial conditions, either one could occur.

This degree-greedy algorithm can be analyzed using the differential equations method.

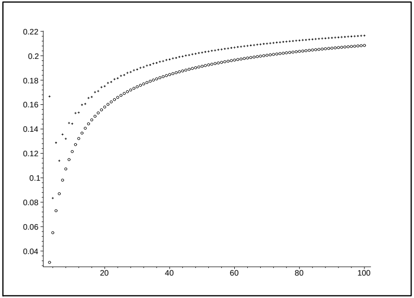
Cleaning a random 5-regular graph



(a) phase 1
(we clean vertices
of degree 4 and 3)

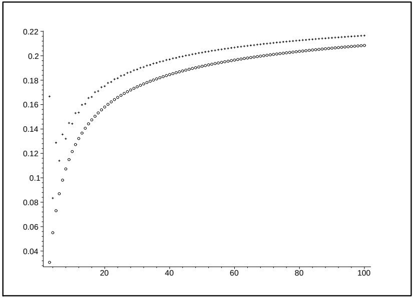


(b) phase 2
(we clean vertices
of degree 3 and 2)



A graph of u_d/dn and l_d/dn versus d (from 3 to 100).

Does $\lim_{d\to\infty} b(G)/dn$ exist?



A graph of u_d/dn and l_d/dn versus d (from 3 to 100).

Does $\lim_{d\to\infty} b(G)/dn$ exist?

The eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ of a graph are the eigenvalues of its adjacency matrix.

The value of $\lambda = \max(|\lambda_2|, |\lambda_n|)$ for a random d-regular graphs has been studied extensively. It is known that for every $\varepsilon > 0$ and $G \in \mathcal{G}_{n,d}$,

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

The eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ of a graph are the eigenvalues of its adjacency matrix.

The value of $\lambda = \max(|\lambda_2|, |\lambda_n|)$ for a random d-regular graphs has been studied extensively. It is known that for every $\varepsilon > 0$ and $G \in \mathcal{G}_{n,d}$,

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

Lemma (Expander Mixing Lemma; Alon, Chung, 1988)

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

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

(Note that $S \cap T$ does not have to be empty; $|E(S,T)|$ 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 $S \cap T$.)

Theorem (Alon, Pralat, Wormald)

Let G be a random d-regular graph on n vertices. Then, a.a.s.

$$\frac{dn}{4}\left(1-\frac{2\sqrt{\ln 2}}{\sqrt{d}}\right) \le b(G) \le \frac{dn}{4}\left(1+\frac{O(1)}{\sqrt{d}}\right).$$

Moreover, $\lim_{d\to\infty} u_d/dn = 1/4$.

Note that the differential equation method gives a better upper bound.

Theorem (Alon, Pralat, Wormald)

Let G be a random d-regular graph on n vertices. Then, a.a.s.

$$\frac{dn}{4}\left(1-\frac{2\sqrt{\ln 2}}{\sqrt{d}}\right) \le b(G) \le \frac{dn}{4}\left(1+\frac{O(1)}{\sqrt{d}}\right).$$

Moreover, $\lim_{d\to\infty} u_d/dn = 1/4$.

Note that the differential equation method gives a better upper bound.

Theorem (Alon, Pralat, Wormald)

Let G be a d-regular graph on n vertices. Then,

$$b(G) \le \begin{cases} \frac{n}{4}\left(d+1-\frac{1}{d+1}\right) & \text{if } d \text{ is even;} \\ \frac{n}{4}(d+1) & \text{if } d \text{ is odd.} \end{cases}$$

This holds for **any** d-regular graph. Proof is nonconstructive (does not reveal how to construct an initial configuration of brushes).

For a random d-regular graph G one can improve the result confirming the conjecture that $b(G) \le dn/4$ a.a.s. (proof uses martingales, cleaning along the Hamilton cycle).

Theorem (Alon, Pralat, Wormald)

Let G be a d-regular graph on n vertices. Then,

$$b(G) \leq \begin{cases} \frac{n}{4}\left(d+1-\frac{1}{d+1}\right) & \text{if } d \text{ is even;} \\ \frac{n}{4}(d+1) & \text{if } d \text{ is odd.} \end{cases}$$

This holds for **any** d-regular graph. Proof is nonconstructive (does not reveal how to construct an initial configuration of brushes).

For a random d-regular graph G one can improve the result confirming the conjecture that $b(G) \leq dn/4$ a.a.s. (proof uses martingales, cleaning along the Hamilton cycle).

Theorem (Alon, Pralat, Wormald)

Let G be a d-regular graph on n vertices. Then,

$$b(G) \leq \begin{cases} \frac{n}{4}\left(d+1-\frac{1}{d+1}\right) & \text{if } d \text{ is even;} \\ \frac{n}{4}(d+1) & \text{if } d \text{ is odd.} \end{cases}$$

This holds for **any** d-regular graph. Proof is nonconstructive (does not reveal how to construct an initial configuration of brushes).

For a random d-regular graph G one can improve the result confirming the conjecture that $b(G) \leq dn/4$ a.a.s. (proof uses martingales, cleaning along the Hamilton cycle).

Theorem (Alon, Pralat, Wormald)

Let G be a d-regular graph on n vertices. Then,

$$b(G) \leq \begin{cases} \frac{n}{4}\left(d+1-\frac{1}{d+1}\right) & \text{if } d \text{ is even;} \\ \frac{n}{4}(d+1) & \text{if } d \text{ is odd.} \end{cases}$$

Theorem (Alon, Pralat, Wormald)

Let G be a random d-regular graph on n vertices. Then, a.a.s.

$$b(G) \leq \begin{cases} \frac{n}{4}\left(d-1-\frac{1}{d-1}\right)(1+o(1)) & \text{if } d \text{ is even;} \\ \frac{n}{4}(d-1)(1+o(1)) & \text{if } d \text{ is odd.} \end{cases}$$

Nonconstructive proof:

Let G be **any** d-regular graph. Let π be a random permutation of the vertices of G taken with uniform distribution. We clean G according to this permutation.

We have to assign to vertex v exactly

$$X(v) = \max\{0, 2N^+(v) - \deg(v)\}$$

brushes in the initial configuration, where $N^+(v)$ is the number of neighbors of v that follow it in the permutation.

The random variable $N^+(v)$ attains each of the values $0, 1, \ldots, d$ with probability $1/(d+1)$.

Nonconstructive proof:

Let G be **any** d-regular graph. Let π be a random permutation of the vertices of G taken with uniform distribution. We clean G according to this permutation.

We have to assign to vertex v exactly

$$X(v) = \max\{0, 2N^+(v) - \deg(v)\}$$

brushes in the initial configuration, where $N^+(v)$ is the number of neighbors of v that follow it in the permutation.

The random variable $N^+(v)$ attains each of the values $0, 1, \ldots, d$ with probability $1/(d+1)$.

Nonconstructive proof:

Let G be **any** d-regular graph. Let π be a random permutation of the vertices of G taken with uniform distribution. We clean G according to this permutation.

We have to assign to vertex v exactly

$$X(v) = \max\{0, 2N^+(v) - \deg(v)\}$$

brushes in the initial configuration, where $N^+(v)$ is the number of neighbors of v that follow it in the permutation.

The random variable $N^+(v)$ attains each of the values $0, 1, \ldots, d$ with probability $1/(d+1)$.

Therefore the expected value of $X(v)$, for even d, is

$$\frac{d+(d-2)+\cdots+2}{d+1} = \frac{d+1}{4} - \frac{1}{4(d+1)},$$

and for odd d it is

$$\frac{d+(d-2)+\cdots+1}{d+1} = \frac{d+1}{4}.$$

Thus, by linearity of expectation,

$$\begin{aligned} \mathbb{E}b_\pi(G) &= \mathbb{E}\left(\sum_{v\in V} X(v)\right) = \sum_{v\in V} \mathbb{E}X(v) \\ &= \begin{cases} \frac{n}{4}\left(d+1-\frac{1}{d+1}\right) & \text{if } d \text{ is even;} \\ \frac{n}{4}(d+1) & \text{if } d \text{ is odd,} \end{cases} \end{aligned}$$

which means that there is a permutation π_0 such that $b(G) \le b_{\pi_0}(G) \le \mathbb{E}b_\pi(G)$.

Therefore the expected value of $X(v)$, for even d, is

$$\frac{d + (d-2) + \cdots + 2}{d+1} = \frac{d+1}{4} - \frac{1}{4(d+1)},$$

and for odd d it is

$$\frac{d + (d-2) + \cdots + 1}{d+1} = \frac{d+1}{4}.$$

Thus, by linearity of expectation,

$$\begin{aligned} \mathbb{E} b_\pi(G) &= \mathbb{E}\left(\sum_{v \in V} X(v)\right) = \sum_{v \in V} \mathbb{E} X(v) \\ &= \begin{cases} \frac{n}{4}\left(d+1-\frac{1}{d+1}\right) & \text{if } d \text{ is even;} \\ \frac{n}{4}(d+1) & \text{if } d \text{ is odd,} \end{cases} \end{aligned}$$

which means that there is a permutation π_0 such that $b(G) \le b_{\pi_0}(G) \le \mathbb{E} b_\pi(G)$.

Outline

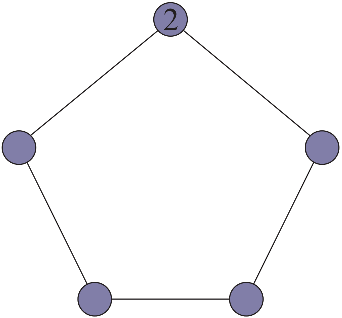
Other research directions in graph cleaning:

- Parallel cleaning,
- Cleaning with Brooms,
- Cleaning binomial random graphs,
- Generalized cleaning (for example, send at most k brushes),
- Combinatorial game,
- Cleaning the web graph.

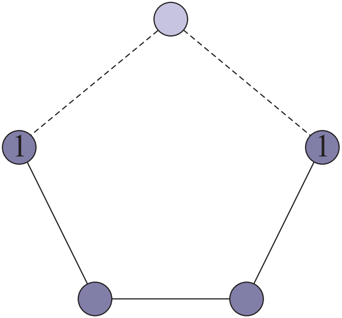
Other research directions in graph cleaning:

- **Parallel cleaning**,
- **Cleaning with Brooms**,
- Cleaning binomial random graphs,
- Generalized cleaning (for example, send at most k brushes),
- Combinatorial game,
- Cleaning the web graph.

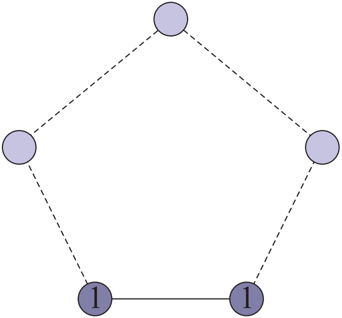
Parallel cleaning



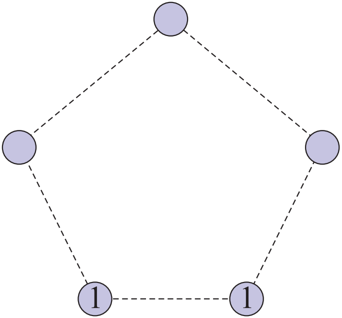
Parallel cleaning



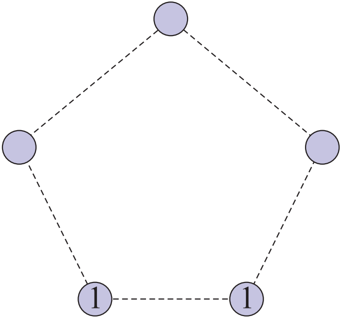
Parallel cleaning



Parallel cleaning



Parallel cleaning



The process is not reversible! We wish to determine the minimum number of brushes, $cpb(G)$, needed to ensure a graph G can be parallel cleaned *continually*.

Parallel cleaning

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any tree T, $cpb(T) = b(T) = d_o(T)/2$.

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any complete bipartite graph $K_{m,n}$,

$$cpb(K_{m,n}) = b(K_{m,n}) = \lceil mn/2 \rceil.$$

Conjecture

$cpb(G) = b(G)$ for any bipartite graph G.

- true if $|V(G)| \leq 11$
- there is one graph on 12 vertcies for which $cpb(G) \neq b(G)$

Parallel cleaning

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any tree T, $cpb(T) = b(T) = d_o(T)/2$.

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any complete bipartite graph $K_{m,n}$,

$$cpb(K_{m,n}) = b(K_{m,n}) = \lceil mn/2 \rceil.$$

Conjecture

$cpb(G) = b(G)$ for any bipartite graph G.

- true if $|V(G)| \leq 11$
- there is one graph on 12 vertcies for which $cpb(G) \neq b(G)$

Parallel cleaning

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any tree T, $cpb(T) = b(T) = d_o(T)/2$.

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any complete bipartite graph $K_{m,n}$,

$$cpb(K_{m,n}) = b(K_{m,n}) = \lceil mn/2 \rceil.$$

Conjecture

$cpb(G) = b(G)$ for any bipartite graph G.

- true if $|V(G)| \leq 11$
- there is one graph on 12 vertcies for which $cpb(G) \neq b(G)$

Parallel cleaning

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any tree T, $cpb(T) = b(T) = d_o(T)/2$.

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any complete bipartite graph $K_{m,n}$,

$$cpb(K_{m,n}) = b(K_{m,n}) = \lceil mn/2 \rceil.$$

Conjecture

$cpb(G) = b(G)$ for any bipartite graph G.

- true if $|V(G)| \leq 11$
- there is one graph on 12 vertcies for which $cpb(G) \neq b(G)$

Parallel cleaning

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any tree T, $cpb(T) = b(T) = d_o(T)/2$.

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any complete bipartite graph $K_{m,n}$,

$$cpb(K_{m,n}) = b(K_{m,n}) = \lceil mn/2 \rceil.$$

Conjecture

$cpb(G) = b(G)$ for any bipartite graph G.

- true if $|V(G)| \leq 11$
- there is one graph on 12 vertcies for which $cpb(G) \neq b(G)$

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any complete graph K_n

$$5/16n^2 + O(n) \leq cpb(K_n) \leq 4/9n^2 + O(n).$$

Conjecture

$$\lim_{n \to \infty} b(K_n)/cpb(K_n) = (1/4)/(4/9) = 9/16.$$

Theorem (Gaspers, Messinger, Nowakowski, Pralat)

For any complete graph K_n

$$5/16n^2 + O(n) \leq cpb(K_n) \leq 4/9n^2 + O(n).$$

Conjecture

$$\lim_{n\to\infty} b(K_n)/cpb(K_n) = (1/4)/(4/9) = 9/16.$$

Cleaning with Brooms

The *brush* number: $b(G) = \min_\alpha b_\alpha(G)$.

The *Broom* number: $B(G) = \max_\alpha b_\alpha(G)$.

Theorem (Pralat)

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

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

Cleaning with Brooms

The *brush* number: $b(G) = \min_\alpha b_\alpha(G)$.

The *Broom* number: $B(G) = \max_\alpha b_\alpha(G)$.

Theorem (Pralat)

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

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

Cleaning with Brooms

The *brush* number: $b(G) = \min_\alpha b_\alpha(G)$.

The *Broom* number: $B(G) = \max_\alpha b_\alpha(G)$.

Theorem (Pralat)

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

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

Cleaning with Brooms – upper bound

Theorem (Pralat)

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

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

where $\bar{z}$ is the solution of

$$d((1+z)\ln(1+z) + (1-z)\ln(1-z)) = 4\ln 2.$$

Cleaning with Brooms – lower bound

Theorem (Pralat (non-constructive proof))

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

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

Cleaning with Brooms – lower bound

Theorem (Pralat (cleaning along the Hamilton cycle))

Let $G \in \mathcal{G}_{n,d}$, where $d \geq 3$. Then, a.a.s., if d is even

$$B(G) \geq \frac{n}{4}\left(d + 3 + \frac{3}{d-1} - 2^{-d+4}\binom{d-2}{d/2-1}\right)(1+o(1))$$

and if d is odd then

$$B(G) \geq \frac{n}{4}\left(d + 3 + \frac{4}{d-1} - 2^{-d+3}\frac{d}{d-1}\binom{d-1}{(d-1)/2}\right)(1+o(1)).$$

Degree-greedy algorithm yields the best lower bound (numerical).

Cleaning with Brooms – lower bound

Theorem (Pralat (cleaning along the Hamilton cycle))

Let $G \in \mathcal{G}_{n,d}$, where $d \geq 3$. Then, a.a.s., if d is even

$$B(G) \geq \frac{n}{4}\left(d+3+\frac{3}{d-1}-2^{-d+4}\binom{d-2}{d/2-1}\right)(1+o(1))$$

and if d is odd then

$$B(G) \geq \frac{n}{4}\left(d+3+\frac{4}{d-1}-2^{-d+3}\frac{d}{d-1}\binom{d-1}{(d-1)/2}\right)(1+o(1)).$$

Degree-greedy algorithm yields the best lower bound (numerical).

Cleaning with Brooms – lower bound

The *brush* number: $b(G) = \min_\alpha b_\alpha(G)$.

The *Broom* number: $B(G) = \max_\alpha b_\alpha(G)$.