

# Graph Imperfection II

Stefanie Gerke<sup>1</sup>

*Mathematical Institute, University of Oxford, 24–29 St. Giles,  
Oxford OX1 3LB, United Kingdom*  
E-mail: [gerke@maths.ox.ac.uk](mailto:gerke@maths.ox.ac.uk)

and

Colin McDiarmid

*Department of Statistics, University of Oxford, 1 South Parks Road,  
Oxford OX1 3TG, United Kingdom*  
E-mail: [cmcd@stats.ox.ac.uk](mailto:cmcd@stats.ox.ac.uk)

Received December 17, 1999; published online June 29, 2001

The imperfection ratio is a graph invariant which indicates how good a lower bound the weighted clique number gives on the weighted chromatic number, in the limit as weights get large. Its introduction was motivated by investigations of the radio channel assignment problem, where one has to assign channels to transmitters and the demands for channels at some transmitters are large. In this paper we show that the imperfection ratio behaves multiplicatively under taking the lexicographic product, which permits us to identify its possible values, investigate its extremal behaviour and its behaviour on random graphs, explore three upper bounds, and show that it is NP-hard to determine. © 2001 Academic Press

*Key Words:* imperfection ratio; perfect graphs; random graphs; radio channel assignment; cochromatic number.

## 1. INTRODUCTION

We are interested in a graph invariant called the imperfection ratio. This concept arose in investigations of the radio channel assignment problem, where one has to assign radio channels or frequencies to transmitters and some of the demands at transmitters are large; see [12, 22]. The *imperfection ratio* of a graph  $G = (V, E)$  may be defined as

$$\text{imp}(G) = \max \left\{ \frac{\chi_f(G, \mathbf{x})}{\omega(G, \mathbf{x})} : \mathbf{0} \neq \mathbf{x} \in \mathbb{N}^V \right\}. \quad (1)$$

<sup>1</sup> The research was partially supported by the EPSRC under Grant 97004215 and the UK Radiocommunications Agency.

Here the maximum is over all integral weight vectors  $\mathbf{x}$ , where a *weight vector* is a non-zero non-negative vector indexed by the nodes. Also  $\chi_f(G, \mathbf{x})$  denotes the *weighted fractional chromatic number*, that is the value of the following linear program with a variable  $y_S$  for each stable set  $S$  of  $G$ :  $\min \sum_S y_S$  subject to  $\sum_{S \ni v} y_S \geq x_v$  for each node  $v \in V(G)$ , and  $y_S \geq 0$  for each stable set  $S$  of  $G$ . Furthermore,  $\omega(G, \mathbf{x})$  denotes the *weighted clique number*, which is the maximum of  $\sum_{v \in K} x_v$  over all cliques  $K$  in  $G$ .

In [12] several alternative descriptions are given, for example

$$\text{imp}(G) = \min\{t: QSTAB(G) \subseteq t STAB(G)\} \quad (2)$$

$$= \max\{\chi_f(G, \mathbf{x}): \mathbf{x} \text{ is a vertex of } QSTAB(G)\} \quad (3)$$

$$= \max\{\mathbf{x} \cdot \mathbf{y}: \mathbf{x} \in QSTAB(G), \mathbf{y} \in QSTAB(\bar{G})\}. \quad (4)$$

Here the *stable set polytope*  $STAB(G) \subseteq [0, 1]^V$  is the convex hull of the incidence vectors of the stable (or independent) sets in  $G$ ; the *fractional node-packing polytope*  $QSTAB(G) \subseteq [0, 1]^V$  is the set of non-negative real vectors  $\mathbf{x} = (x_v: v \in V)$  such that  $\sum_{v \in K} x_v \leq 1$  for every clique  $K$  in  $G$ ; and we denote by  $tP$  the scaled set  $\{t\mathbf{x}: \mathbf{x} \in P\}$ . Recently, the imperfection ratio was also characterised in terms of graph entropy [25]; see also [22].

It was noted in [12] that  $\text{imp}(G) = 1$  if and only if  $G$  is perfect, and that  $\text{imp}(G) = \text{imp}(\bar{G})$  for any graph  $G$  and its complement  $\bar{G}$ . Another fact we will use is that the imperfection ratio of a triangle-free graph  $G$  equals  $\chi_f(G)/2$  where  $\chi_f(G) = \chi_f(G, \mathbf{1})$  is the *fractional chromatic number*. In this paper we investigate the imperfection ratio further.

In Section 2 we show that the lexicographic product  $G[H]$  of two graphs  $G$  and  $H$  satisfies the equation  $\text{imp}(G[H]) = \text{imp}(G) \text{imp}(H)$ . This extends the well known result that  $G[H]$  is perfect if and only if both  $G$  and  $H$  are perfect [17]. We use this equation to identify the possible values of  $\text{imp}(G)$ : we find that there is a graph  $G$  with  $\text{imp}(G) = r$  if and only if  $r$  is rational and at least 1.

In Section 3 we investigate the extremal behaviour of the imperfection ratio and its behaviour on random graphs. We prove that the maximum value of  $\text{imp}(G)$  over all graphs  $G$  on  $n$  nodes is  $O(n \log \log n / \log^2 n)$ . (We denote by  $\log$  the binary logarithm and by  $\ln$  the natural logarithm.) The maximum value of  $\text{imp}(G)$  over all graphs  $G$  with maximal degree  $d$  is  $o(d)$ , and the corresponding maximum over all triangle-free graphs on  $n$  nodes is  $\Theta(\sqrt{n/\log n})$ . On the other hand the imperfection ratio of the random graph  $G_{n, 1/2}$  is usually close to  $n/(4 \log^2 n)$ . We obtain corresponding results for sparse random graphs and for random regular graphs.

In Section 4 we consider three upper bounds on the imperfection ratio. The first bound asserts that  $\text{imp}(G) \leq b/c$  if every induced subgraph of  $G$

contains a node the neighbourhood of which can be covered  $c$  times by  $b$  cliques. This bound yields for example the result that  $\text{imp}(G) \leq 6$  for any disk graph  $G$ . The second bound is the *fractional perfect covering number*  $\text{pc}_f(G)$ , which is the minimum value of  $p/q$  over all  $p$  and  $q$  such that the nodes of  $G$  can be covered  $q$  times by  $p$  induced perfect subgraphs. We see that  $\text{imp}(G) \leq \text{pc}_f(G)$ . Clearly,  $\text{imp}(G) = \text{pc}_f(G)$  for perfect graphs  $G$ , and the same holds for minimal imperfect graphs. On the other hand the ratio  $\text{pc}_f(G)/\text{imp}(G)$  can grow like  $n^\delta$  for  $n$  node graphs  $G$ . The third bound is in terms of the *cochromatic number*  $z(G)$ , which is the least number of stable sets or cliques to cover the nodes of  $G$ . For any non-trivial graph  $G$  we have  $\text{pc}_f(G) \leq z(G)/2$  and hence  $\text{imp}(G) \leq z(G)/2$ . On the other hand we show that  $\text{imp}(G) \geq z(G)^2/(4n)$  if  $G$  has  $n$  nodes.

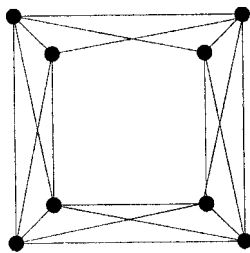
In Section 5 we prove that it is NP-hard to determine  $\text{imp}(G)$ . If we could determine  $\text{imp}(G)$  then we could tell whether  $G$  is perfect, since  $\text{imp}(G) = 1$  if and only if  $G$  is perfect: but it is not known whether this is hard to do. We show that it is NP-hard to determine the fractional chromatic number of triangle-free graphs, which implies the result since for these graphs  $\text{imp}(G) = \chi_f(G)/2$ , as we noted earlier.

## 2. LEXICOGRAPHIC PRODUCT

Let  $G[H]$  denote the *lexicographic product* of the graphs  $G$  and  $H$ , that is the graph  $G[H]$  with node set  $V(G) \times V(H)$ , and two nodes  $(u_1, v_1)$  and  $(u_2, v_2)$  are adjacent if either  $\{u_1, u_2\} \in E(G)$  or  $u_1 = u_2$  and  $\{v_1, v_2\} \in E(H)$ . For an example see Fig. 1.

**THEOREM 2.1.** *For any two graphs  $G$  and  $H$ ,  $\text{imp}(G[H]) = \text{imp}(G) \text{imp}(H)$ .*

Since  $\text{imp}(G) = 1$  if and only if  $G$  is perfect, this theorem extends the well known result that  $G[H]$  is perfect if and only if  $G$  and  $H$  are perfect [17]. It also yields the next result, which we prove before it.



**FIG. 1.**  $C_4[K_2]$ .

**THEOREM 2.2.** *There exists a graph  $G$  with  $\text{imp}(G) = r$  if and only if  $r$  is rational and at least 1.*

*Proof.* That  $r$  is rational and at least 1 if there is a graph  $G$  with  $\text{imp}(G) = r$  follows immediately from the definition (1). Recall that for positive integers  $a$  and  $b$  with  $2b \leq a$ , the Kneser graph  $K_{a:b}$  has a node for each  $b$ -element subset of  $\{1, \dots, a\}$ , and two nodes are adjacent when the corresponding subsets are disjoint. If  $a$  and  $b$  are integers with  $0 < 2b < a < 3b$ , then the Kneser graph  $K_{a:b}$  has no triangles, and so

$$\text{imp}(K_{a:b}) = \frac{1}{2} \chi_f(K_{a:b}) = \frac{a}{2b}$$

—see, for example, [24]. But any rational number greater than 1 is a product of rationals in the open interval  $(1, 3/2)$ , and so Theorem 2.1 yields the result. ■

We break the proof of Theorem 2.1 into an upper bound part (Lemma 2.2) and a lower bound part (Lemma 2.3). Before we can do this we need some more notation and one preliminary lemma from [12].

We denote by  $G[H^v: v \in V]$  the graph which is constructed by replacing each node  $v$  of  $G = (V, E)$  by the graph  $H^v$ . Here, replacing a node  $v$  of  $G$  by  $H^v$  means that all the nodes of  $H^v$  are joined to the nodes adjacent to  $v$  in  $G$  and  $v$  is deleted. The lexicographic product of  $G$  and  $H$  is therefore the graph where each node of  $G$  is replaced by a copy  $H^v$  of  $H$ . If we are given a weight vector  $\mathbf{x}$  of  $G$ , we abuse notation by using  $\mathbf{x}$  also for its restriction to the node set of an induced subgraph  $H$  of  $G$ . In the same vein we say that a node of  $G[H^v: v \in V]$  is an element of  $V(H^v)$ , if this node belongs to the copy of  $H^v$  in  $G[H^v: v \in V]$ .

**LEMMA 2.1** [12]. *For any weight vector  $\mathbf{x}$  of  $G[H^v: v \in V]$ , we have  $\chi_f(G[H^v: v \in V], \mathbf{x}) = \chi_f(G, \hat{\mathbf{x}})$  where  $\hat{\mathbf{x}}$  is the weight vector of  $G$  with  $\hat{x}_v = \chi_f(H^v, \mathbf{x})$  for all  $v \in V$ .*

**LEMMA 2.2.**  $\text{imp}(G[H^v: v \in V]) \leq \text{imp}(G) \max\{\text{imp}(H^v): v \in V\}$ .

Note that it follows from this lemma that  $G[H^v: v \in V]$  is perfect if and only if  $G$  and each graph  $H^v$  are perfect.

*Proof.* Let  $\mathbf{x}$  be a weight vector for  $G[H^v: v \in V]$ . Define the weight vector  $\hat{\mathbf{x}}$  for  $G$  by setting  $\hat{x}_v = \chi_f(H^v, \mathbf{x})$  for each  $v \in V$ . By Lemma 2.1 and the definition of the imperfection ratio,

$$\chi_f(G[H^v: v \in V], \mathbf{x}) = \chi_f(G, \hat{\mathbf{x}}) \leq \text{imp}(G) \omega(G, \hat{\mathbf{x}}). \quad (5)$$

Let  $K$  be a maximal weight clique for  $(G, \hat{\mathbf{x}})$ , and let  $p = \max\{\text{imp}(H^v): v \in V\}$ . Then

$$\omega(G, \hat{\mathbf{x}}) = \sum_{v \in K} \hat{x}_v = \sum_{v \in K} \chi_f(H^v, \mathbf{x}) \leq p \sum_{v \in K} \omega(H^v, \mathbf{x}).$$

Now the union of maximal weight cliques of  $(H^v, \mathbf{x})$  over all  $v \in K$  forms a clique in  $G[H^v: v \in V]$ , and so

$$p \omega(G[H^v: v \in V], \mathbf{x}) \geq p \sum_{v \in K} \omega(H^v, \mathbf{x}) \geq \omega(G, \hat{\mathbf{x}}).$$

By (5) and the last inequality,

$$\chi_f(G[H^v: v \in V], \mathbf{x}) \leq \text{imp}(G) p \omega(G[H^v: v \in V], \mathbf{x}).$$

Therefore for any weight vector  $\mathbf{x}$ ,

$$\frac{\chi_f(G[H^v: v \in V], \mathbf{x})}{\omega(G[H^v: v \in V], \mathbf{x})} \leq \text{imp}(G) \max\{\text{imp}(H^v): v \in V\},$$

which yields the required inequality.  $\blacksquare$

LEMMA 2.3.  $\text{imp}(G[H^v: v \in V]) \geq \text{imp}(G) \min\{\text{imp}(H^v): v \in V\}$ .

*Proof.* Let  $\mathbf{x} \in \text{QSTAB}(G)$  satisfy  $\text{imp}(G) = \chi_f(G, \mathbf{x})$ , and for each  $v \in V$ , let  $\mathbf{y}^v \in \text{QSTAB}(H^v)$  satisfy  $\text{imp}(H^v) = \chi_f(H^v, \mathbf{y}^v)$ . Note that such vectors always exist by (3). We abbreviate  $\min\{\text{imp}(H^v): v \in V\}$  by  $p$ . Denote by  $\mathbf{z}$  the weight vector of  $G[H^v: v \in V]$  with

$$z_u = x_v y_u^v \quad \text{for each } u \in V(H^v).$$

Since  $\omega(H^v, \mathbf{y}^v) = 1$ , we have  $\omega(H^v, \mathbf{z}) = x_v$ . Thus  $\text{imp}(H^v) = \hat{x}_v/x_v$  where  $\hat{x}_v = x_v \chi_f(H^v, \mathbf{y}^v) = \chi_f(H^v, \mathbf{z})$ . Therefore

$$p x_v \leq \text{imp}(H^v) x_v = \hat{x}_v.$$

This together with Lemma 2.1 implies that

$$\chi_f(G[H^v: v \in V], \mathbf{z}) = \chi_f(G, \hat{\mathbf{x}}) \geq \chi_f(G, p\mathbf{x}) = \chi_f(G, \mathbf{x}) p.$$

In addition,  $\omega(G[H^v: v \in V], \mathbf{z}) = \omega(G, \mathbf{x}) = 1$ . Hence

$$\begin{aligned} \text{imp}(G[H^v: v \in V]) &\geq \chi_f(G[H^v: v \in V], \mathbf{z}) \geq \chi_f(G, \mathbf{x}) p \\ &= \text{imp}(G) \min\{\text{imp}(H^v): v \in V\} \end{aligned}$$

as required.  $\blacksquare$

The above two lemmas establish Theorem 2.1. A result which can be proved in a similar way to them is the following proposition.

**PROPOSITION 2.1** [11]. *The imperfection ratio of  $G[H^v: v \in V]$  depends only on the graph  $G=(V, E)$  and the values  $\text{imp}(H^v)$ , i.e.,  $\text{imp}(G[H^v: v \in V]) = \text{imp}(G[\tilde{H}^v: v \in V])$  if  $\text{imp}(H^v) = \text{imp}(\tilde{H}^v)$  for each  $v \in V$ .*

### 3. EXTREMAL RESULTS AND RANDOM GRAPHS

In this section we show that there exists a constant  $c$  such that for every graph  $G$  on  $n \geq 3$  nodes

$$\text{imp}(G) \leq c \frac{n(\log \log n)}{\log^2 n};$$

see Theorem 3.1. We see that the maximum value of  $\text{imp}(G)$  over all graphs on  $n$  nodes with maximal degree  $d$  is  $o(d)$ , and the corresponding maximum over all triangle-free graphs is  $\Theta(\sqrt{n/\log n})$ .

Recall that the random graph  $G_{n,p}$  has  $n$  nodes  $1, \dots, n$  and the  $\binom{n}{2}$  possible edges appear independently with probability  $p$ . We see in Theorem 3.3 that a.s.  $\text{imp}(G_{n,1/2}) \sim n/(4 \log^2 n)$ , which shows that the upper bound on  $\text{imp}(G)$  given above is at most a factor  $\log \log n$  too large. We also investigate the imperfection ratio for sparse random graphs, and for random regular graphs. In addition we see that  $\text{imp}(G_{n,p})$  is concentrated around its mean value.

Before we consider the results just mentioned, we give an upper bound on the imperfection ratio in terms of the logarithm of the clique number and the *binary imperfection ratio*  $\text{imp}_b(G)$ , that is the maximum of  $\chi_f(H)/\omega(H)$  over all induced subgraphs  $H$  of  $G$ , or equivalently

$$\text{imp}_b(G) = \max \left\{ \frac{\chi_f(G, \mathbf{x})}{\omega(G, \mathbf{x})} : \mathbf{0} \neq \mathbf{x} \in \{0, 1\}^V \right\}.$$

Clearly,  $\text{imp}_b(G) \leq \text{imp}(G)$  for any graph  $G$ .

**LEMMA 3.1.** *For any graph  $G$ ,  $\text{imp}(G) \leq (2 \lceil \log \omega(G) \rceil) \text{imp}_b(G)$ .*

*Proof.* Denote  $\lceil \log \omega(G) \rceil$  by  $l$ . First we claim that for each point  $\mathbf{x}$  of  $QSTAB(G)$  there exist positive numbers  $\alpha_i$  and vectors  $\mathbf{x}^i \in \{0, j_i\}^V \cap QSTAB(G)$ ,  $j_i > 0$ ,  $i = 1, \dots, l$ , such that  $\mathbf{x} \leq \sum_{i=1}^l \alpha_i \mathbf{x}^i$  and  $\sum_{i=1}^l \alpha_i \leq 2l$ .

Assume for now that the claim is true. By (3) there exists a vertex  $\mathbf{x}$  of  $QSTAB(G)$  such that  $\text{imp}(G) = \chi_f(G, \mathbf{x})$ . Hence

$$\begin{aligned} \text{imp}(G) = \chi_f(G, \mathbf{x}) &\leq \chi_f\left(G, \sum_{i=1}^l \alpha_i \mathbf{x}^i\right) \leq \sum_{i=1}^l \alpha_i \chi_f(G, \mathbf{x}^i) \\ &\leq \sum_{i=1}^l \alpha_i \text{imp}_b(G) \leq (2\lceil \log \omega(G) \rceil) \text{imp}_b(G). \end{aligned}$$

*Proof of the claim.* For  $i = 1, 2, \dots, l-1$ , let

$$A_i = \{v \in V : 2^{-i} < x_v \leq 2^{-(i-1)}\}$$

and let

$$A_l = V \setminus \bigcup_{i=1}^{l-1} A_i.$$

Then  $0 \leq \mathbf{x}^i := 2^{-i} \mathbf{1}_{A_i} \leq \mathbf{x}$  for  $i = 1, 2, \dots, l-1$  and hence  $\mathbf{x}^i \in QSTAB(G)$ . In addition, if  $v \in A_i$  then  $x_v \leq 2x_v^i$ . If  $v \in A_l$  then  $x_v \leq 2^{-(l-1)} \leq 2/\omega(G)$ . Set  $\mathbf{x}^l = 1/\omega(G) \mathbf{1}_{A_l}$ . Observe that  $\mathbf{x}^l \in QSTAB(G)$  and if  $v \in A_l$  then  $x_v \leq 2x_v^l$ . Hence  $\mathbf{x} \leq \sum_{i=1}^l 2\mathbf{x}^i$ , and each  $\mathbf{x}^i$  belongs to  $QSTAB(G) \cap \{0, j_i\}^V$  with  $j_i = 2^{-i}$  for  $i = 1, \dots, l-1$  and  $j_l = 1/\omega(G)$ . ■

**THEOREM 3.1.** *There exists a constant  $c$  such that for all graphs  $G$  with  $n \geq 3$  nodes,*

$$\text{imp}(G) \leq c \frac{n(\log \log n)}{\log^2 n}.$$

*Proof.* In [8] Erdős showed that there exists a constant  $c_0$  such that for any graph  $G$  on  $n \geq 2$  nodes,  $\chi(G)/\omega(G) \leq c_0 n/\log^2 n$ . When  $n = 2$  the bound is  $2c_0$ , and  $c_0 \geq 1/2$ . We take  $c = 10c_0 \geq 5$ . We shall prove by induction on  $n$  that for any graph  $G$  on  $n \geq 3$  nodes and any integral weight vector  $\mathbf{x}$  for  $G$

$$\frac{\chi(G, \mathbf{x})}{\omega(G, \mathbf{x})} \leq c \frac{n(\log \log n)}{\log^2 n}. \tag{6}$$

To begin, we note that every graph  $G$  on 3 nodes is perfect, so by [17] we always have  $\chi(G, \mathbf{x})/\omega(G, \mathbf{x}) = 1$ , and hence (6) holds for  $n = 3$ .

Now let  $n \geq 4$ , and suppose that (6) holds for all graphs  $G$  on at most  $n-1$  nodes and all corresponding integral weight vectors  $\mathbf{x}$ . Let  $G = (V, E)$  be a graph of order  $n$ , and let  $\mathbf{x}$  be an integral weight vector for  $G$ . Let  $x_{\min}$  denote  $\min\{x_v : v \in V\}$  and let  $x_{\max}$  denote  $\max\{x_v : v \in V\}$ . We may assume

that  $x_{\min} \neq 0$  since otherwise the result follows directly from the induction hypothesis (as  $n \log \log n / \log^2 n$  is an increasing function). We consider two cases, depending on the ratio of  $x_{\max}$  to  $x_{\min}$ .

*Case 1.*  $x_{\max}/x_{\min} \leq \log^2 n$ . If  $\omega(G) \geq \log^4 n$  then  $\omega(G, \mathbf{x}) \geq x_{\min} \log^4 n$ , and we obtain

$$\frac{\chi(G, \mathbf{x})}{\omega(G, \mathbf{x})} \leq \frac{x_{\max} n}{x_{\min} \log^4 n} \leq \frac{n}{\log^2 n} \leq c \frac{n \log \log n}{\log^2 n}.$$

If  $\omega(G) < \log^4 n$  then by Lemma 3.1 and Erdős' result mentioned above, we have

$$\begin{aligned} \text{imp}(G) &\leq 2 \lceil \log \omega(G) \rceil \text{imp}_b(G) \leq (8 \log \log n + 2) \frac{c_0 n}{\log^2 n} \\ &\leq \frac{10c_0 n \log \log n}{\log^2 n}. \end{aligned}$$

*Case 2.*  $x_{\max}/x_{\min} > \log^2 n$ . Let  $v$  be a vertex with  $x_v = x_{\min}$ , and denote by  $G - v$  the graph obtained from  $G$  by deleting  $v$ . It follows that

$$\frac{\chi(G, \mathbf{x})}{\omega(G, \mathbf{x})} \leq \frac{x_{\min} + \chi(G - v, \mathbf{x})}{\omega(G, \mathbf{x})} \leq \frac{x_{\min}}{x_{\max}} + \frac{\chi(G - v, \mathbf{x})}{\omega(G - v, \mathbf{x})}.$$

Now the function  $f(n) = n/\log^2 n$  is concave so  $f(n-1) \leq f(n) - f'(n)$ , and  $f'(n) = (\log n - 2 \ln 2)/\log^3 n$ . Hence by the induction hypothesis

$$\begin{aligned} \frac{\chi(G, \mathbf{x})}{\omega(G, \mathbf{x})} &\leq \frac{1}{\log^2 n} + c \log \log n \frac{(n-1)}{\log^2 (n-1)} \\ &\leq \frac{1}{\log^2 n} + c \log \log n \left( \frac{n}{\log^2 n} - \frac{\log n - 2 \ln 2}{\log^3 n} \right) \\ &\leq c \frac{n \log \log n}{\log^2 n} \end{aligned}$$

since  $n \geq 4$ . ■

**THEOREM 3.2.** *For each  $\varepsilon > 0$ , there exists a constant  $d_0$  such that, for each  $d \geq d_0$  and each graph  $G$  with maximal degree  $\Delta(G) \leq d$ ,*

$$\text{imp}(G) \leq \varepsilon d.$$

*Proof.* For graphs  $H$  with uniformly bounded clique number  $\omega(H)$ , we have [15]

$$\chi(H) = O(d(\log \log d)/\log d),$$

where  $d = \Delta(H)$ . Let  $\varepsilon > 0$ . Then there exists  $d_0$  such that for all  $d \geq d_0$  and all graphs  $H$  with  $\omega(H) \leq \frac{2}{\varepsilon}$  and  $\Delta(H) \leq d$  we have  $\chi(H) \leq \frac{\varepsilon}{2}(d-1)$ .

Let  $d \geq d_0$  and let  $G$  be a graph with  $\Delta(G) \leq d$ . Let  $\mathbf{x} \in QSTAB(G)$  satisfy  $\chi_f(G, \mathbf{x}) = \text{imp}(G)$ . Let  $\mathbf{y}$  denote the vector indexed by the nodes  $v$  of  $G$ , with  $y_v = x_v$  if  $x_v \geq \frac{\varepsilon}{2}$  and  $y_v = 0$  otherwise. The subgraph  $H$  of  $G$  induced by the nodes  $v$  with  $y_v > 0$  has clique number  $\omega(H) \leq 2/\varepsilon$ , and hence

$$\chi_f(G, \mathbf{y}) \leq \chi_f(H) \leq \chi(H) \leq \frac{\varepsilon}{2}(d-1).$$

Also

$$\chi_f(G, \mathbf{x} - \mathbf{y}) \leq \chi_f\left(G, \frac{\varepsilon}{2} \mathbf{1}\right) = \frac{\varepsilon}{2} \chi_f(G) \leq \frac{\varepsilon}{2}(d+1).$$

Hence

$$\text{imp}(G) = \chi_f(G, \mathbf{x}) \leq \chi_f(G, \mathbf{y}) + \chi_f(G, \mathbf{x} - \mathbf{y}) \leq \varepsilon d,$$

as required. ■

It is known [16] that for a triangle-free graph  $G$  on  $n$  nodes, we have  $\chi(G) \leq (1 + o(1)) 2 \sqrt{2} \sqrt{n/\log n}$  and hence that  $\text{imp}(G) \leq (1 + o(1)) \sqrt{2} \sqrt{n/\log n}$ . There is also a matching lower bound, that is there are triangle-free graphs  $G$  on  $n$  nodes with  $\text{imp}(G) \geq c \sqrt{n/\log n}$  where  $c > 0$  is a constant. This follows from the fact that there exists for  $n$  sufficiently large a triangle-free graph  $G$  on  $n$  nodes with  $\alpha(G) \leq 9 \sqrt{n \log n}$ ; see [16]. Thus the maximum value of  $\text{imp}(G)$  over all triangle-free graphs  $G$  with  $n$  nodes is  $\Theta(\sqrt{n/\log n})$ .

We now consider random graphs. As remarked earlier, the next result shows that a.s.  $\text{imp}(G_{n, 1/2}) \sim n/(4 \log^2 n)$ .

**THEOREM 3.3.** *For any (fixed)  $0 < p < 1$ , and any  $\eta > 0$  a.s.*

$$\frac{n}{4 \log_{1/p} n \log_{1/q} n} \leq \text{imp}(G_{n, p}) \leq (1 + \eta) \frac{n}{4 \log_{1/p} n \log_{1/q} n}, \tag{7}$$

where  $q = 1 - p$ .

*Proof.* First we consider the lower bound. If  $G$  is a graph on  $n$  nodes which satisfies

$$\alpha(G) \leq 2 \log_{1/q} n \quad (8)$$

and

$$\omega(G) \leq 2 \log_{1/p} n, \quad (9)$$

then

$$\text{imp}(G) \geq \frac{n}{\alpha(G) \omega(G)} \geq \frac{n}{4 \log_{1/p} n \log_{1/q} n}.$$

A simple first moment argument shows that the random graph  $G_{n,p}$  a.s. satisfies the conditions (8) and (9) (see for example [5]), and so the lower bound in (7) follows.

Now we consider the upper bound. Let  $\eta > 0$ , and let  $\varepsilon > 0$  satisfy  $(1 + \varepsilon)/(1 - \varepsilon) + \varepsilon \leq 1 + \eta$ . Let  $l = l(n) = n/\log n$ . Note that  $l(\log \log l)/\log^2 l = o(n/\log^2 n)$ . Thus by Theorem 3.1 we have that for any graph  $G_l$  on at most  $l$  nodes

$$\text{imp}(G_l) \leq \varepsilon \frac{n}{4 \log_{1/p} n \log_{1/q} n} \quad (10)$$

for all sufficiently large  $n$ . For the remainder of this proof we assume that  $n$  is large enough such that (10) is fulfilled.

Consider the following two conditions on a graph  $G$  with  $n$  nodes:

$$\chi(G) \leq (1 + \varepsilon) \frac{n}{2 \log_{1/q} n}, \quad (11)$$

and for all induced subgraphs  $H$  of  $G$  on at least  $l$  nodes

$$\omega(H) \geq (1 - \varepsilon) 2 \log_{1/p} n. \quad (12)$$

The random graph  $G_{n,p}$  a.s. satisfies these conditions: Condition (11) is essentially a celebrated theorem of Bollobás' [6] and Condition (12) (for stable sets rather than cliques) is a natural step towards proving that theorem—see also [2, 21].

Suppose that  $G$  is a graph on  $n$  nodes satisfying (11) and (12). We show that

$$\text{imp}(G) \leq (1 + \eta) \frac{n}{4 \log_{1/p} n \log_{1/q} n}, \quad (13)$$

which proves the upper bound in (7).

Let  $\mathbf{x}$  be any integer weighting for  $G$  with  $\text{imp}(G) = \chi_f(G, \mathbf{x})/\omega(G, \mathbf{x})$ . Let  $V_i(\mathbf{x}) = \{v \in V(G) : x_v \geq i\}$ , and let  $G_i$  be the subgraph of  $G$  induced by  $V_i(\mathbf{x})$  for  $i = 0, 1, \dots$ , so that  $V_0(\mathbf{x}) = V$ ,  $G_0 = G$ . Let  $k = \max\{i : |V_i(\mathbf{x})| \geq l\}$ . If  $k = 0$  then (13) follows from (10), and thus we may assume that  $k \geq 1$ . Let  $y_v = \min\{x_v, k\}$  and  $z_v = x_v - y_v$  for each  $v \in V$ , so that  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ . Now by (11)

$$\chi(G, \mathbf{y}) \leq \chi(G, k\mathbf{1}) \leq \frac{k(1 + \varepsilon)n}{2 \log_{1/q} n},$$

and by (12)

$$\omega(G, \mathbf{y}) \geq k\omega(G_k) \geq k(1 - \varepsilon) 2 \log_{1/p} n;$$

and so

$$\frac{\chi(G, \mathbf{y})}{\omega(G, \mathbf{y})} \leq \frac{1 + \varepsilon}{1 - \varepsilon} \frac{n}{4 \log_{1/p} n \log_{1/q} n}.$$

Also, since  $|V_{k+1}(\mathbf{x})| < l$ , by (10)

$$\frac{\chi_f(G, \mathbf{z})}{\omega(G, \mathbf{z})} \leq \text{imp}(G_{k+1}) \leq \varepsilon \frac{n}{4 \log_{1/p} n \log_{1/q} n}.$$

But

$$\text{imp}(G) = \frac{\chi_f(G, \mathbf{x})}{\omega(G, \mathbf{x})} \leq \frac{\chi(G, \mathbf{y})}{\omega(G, \mathbf{y})} + \frac{\chi_f(G, \mathbf{z})}{\omega(G, \mathbf{z})},$$

and the required inequality (13) follows from the last two inequalities. ■

Since  $\log p \log(1 - p) \leq 1$  for all  $0 < p < 1$ , we have  $\log^2 n \leq \log_{1/p} n \log_{1/q} n$  for all  $n$  and all  $0 < p < 1$ , and hence the best choice is  $p = 1/2$  if one is interested in a large imperfection ratio.

It is known that (11) and (12) are not only a.s. satisfied but that the probability that either of these conditions fails is  $o(1/n)$  (and indeed is far smaller) [6, 21]. This, together with the fact that  $\text{imp}(G) \leq |V(G)|$  for any graph  $G$ , yields that the expected value  $\mathbf{E}(\text{imp}(G_{n,p}))$  of the imperfection ratio of  $G_{n,p}$  satisfies

$$\mathbf{E}(\text{imp}(G_{n,p})) = (1 + o(1)) \frac{n}{4 \log_{1/p} n \log_{1/q} n}.$$

The next theorem states that the imperfection ratio is concentrated around its expected value.

**THEOREM 3.4.**  $P(|\text{imp}(G_{n,p}) - \mathbf{E}(\text{imp}(G_{n,p}))| \geq t) \leq 2e^{-2t^2/n}$  for  $t \geq 0$ .

*Proof.* Let  $G$  be a graph with at least two nodes, let  $v$  be a node in  $G$ , and let  $G - v$  denote the graph which is obtained from  $G$  by deleting  $v$ . Let  $\mathbf{x} \in QSTAB(G)$  satisfy  $\text{imp}(G) = \chi_f(G, \mathbf{x})$ . Then

$$\text{imp}(G) = \chi_f(G, \mathbf{x}) \leq \chi_f(G - v, \mathbf{x}) + x_v \leq \text{imp}(G - v) + 1,$$

and so

$$\text{imp}(G - v) \leq \text{imp}(G) \leq \text{imp}(G - v) + 1.$$

Now let  $G$  and  $G'$  be two graphs which differ only in edges incident to a single node  $v$ . Then by the above

$$\text{imp}(G - v) \leq \text{imp}(G), \text{imp}(G') \leq \text{imp}(G - v) + 1,$$

and hence

$$|\text{imp}(G) - \text{imp}(G')| \leq 1.$$

The result now follows from Lemma 3.3 of [20], which says that if the graph function  $f$  satisfies  $|f(G) - f(G')| \leq 1$  whenever  $G'$  can be obtained from  $G$  by changing edges incident with a single node, then the corresponding random variable  $Y = f(G_{n,p})$  satisfies

$$P(|Y - \mathbf{E}(Y)| \geq t) \leq 2e^{-2t^2/n}$$

for  $t \geq 0$ . ■

We next determine the behaviour of the imperfection ratio also for sparse random graphs, at least if the average degree is not too small.

**THEOREM 3.5.** *Suppose  $p = p(n)$  satisfies  $p = o(n^{-2/3})$  and  $np \rightarrow \infty$  as  $n \rightarrow \infty$ . Then for every  $\varepsilon > 0$  a.s.*

$$(1 - \varepsilon) \frac{np}{4 \ln np} \leq \text{imp}(G_{n,p}) \leq (1 + \varepsilon) \frac{np}{4 \ln np}.$$

*Proof.* If  $p = o(1)$  and  $np \rightarrow \infty$  as  $n \rightarrow \infty$ , then for any  $\varepsilon > 0$ , a.s.

$$\chi(G_{n,p}) \leq (1 + \varepsilon) \frac{np}{2 \ln np}; \tag{14}$$

see [18]. But  $\text{imp}(G) \leq \chi(G)/2$  for any graph  $G$  with at least one edge, by Proposition 3.4 of [12], and the required upper bound on the imperfection ratio follows.

Let  $\delta > 0$  satisfy  $(1 - \delta)/(1 + \delta) \geq (1 - \varepsilon)$ , and let  $G$  be any graph on  $n$  nodes satisfying

$$\alpha(G) \leq (1 + \delta) \frac{2 \ln np}{p}. \tag{15}$$

Then for any induced subgraph  $H$  of  $G$  on at least  $n - \delta n$  nodes

$$\chi_f(H) \geq \frac{n - \delta n}{\alpha(G)} \geq \frac{1 - \delta}{1 + \delta} \frac{np}{2 \ln np} \geq (1 - \varepsilon) \frac{np}{2 \ln np}. \tag{16}$$

By [9], the random graph  $G_{n,p}$  a.s. satisfies (15). Also, the expected number of triangles in  $G_{n,p}$  is less than  $(np)^3/6$ , and hence the probability that the number of triangles in  $G_{n,p}$  is greater than  $\delta n$  is at most  $n^2 p^3/6\delta = o(1)$ . Hence there a.s. is a triangle-free induced subgraph  $H$  with at least  $n - \delta n$  nodes. Since  $\text{imp}(G) \geq \text{imp}(H) = \chi_f(H)/2$  the result follows by (16). ■

There is a similar result for random  $r$ -regular graphs  $G_{n,r}$ , that is graphs taken uniformly at random from the set of all  $r$ -regular graphs on the  $n$  nodes  $\{1, 2, \dots, n\}$  (where  $rn$  is even). The limit in the following theorem refers to  $n \rightarrow \infty$  with  $n$  restricted to even integers if  $r$  is odd.

**THEOREM 3.6.** *There exists  $\varepsilon = \varepsilon(r) > 0$  for integers  $r \geq 2$  with  $\varepsilon(r) \rightarrow 0$  as  $r \rightarrow \infty$ , such that for each fixed  $r \geq 2$  a.s.*

$$\frac{r}{4 \ln r} \leq \text{imp}(G_{n,r}) \leq (1 + \varepsilon) \frac{r}{4 \ln r}.$$

*Proof.* We may argue much as in the proof of Theorem 3.5. To do this, the upper bound (14) has to be replaced by the statement that for each  $r \geq 2$  there exist  $\varepsilon = \varepsilon(r) > 0$  with  $\varepsilon(r) \rightarrow 0$  as  $r \rightarrow \infty$  such that

$$\chi(G_{n,r}) \leq (1 + \varepsilon) \frac{r}{2 \ln r},$$

which is proved in [10].

For the lower bound we need the results that for each fixed  $r \geq 2$  there exists  $\varepsilon = \varepsilon(r) > 0$  such that a.s.

$$\alpha(G_{n,r}) \leq (1 - \varepsilon) \frac{2 \ln r}{r} n,$$

which can be found in [4] or [5, p.277, Corollary 28], and that the expected number of triangles in  $G_{n,r} \rightarrow (r-1)^3/6$  as  $n \rightarrow \infty$ ; see [3] or [5, p.53, Corollary 19]. A simple calculation similar to (16) now yields the lower bound. ■

#### 4. UPPER BOUNDS ON THE IMPERFECTION RATIO

In this section we consider three upper bounds on the imperfection ratio. The first bound is motivated by the well known fact that one can colour a graph  $G$  with  $t+1$  colours if every induced subgraph of  $G$  contains a node of degree at most  $t$ ; see, for example, [14]. It follows that one can colour a graph  $G$  with  $t(\omega(G) - 1) + 1$  colours if every induced subgraph of  $G$  contains a node the neighbourhood of which can be covered by  $t$  cliques. (Graphs for which this is true with  $t=1$  are known as chordal or triangulated graphs: it is well known that such graphs are perfect.) This method also yields a bound on the chromatic number of a *disk graph*  $G$ , that is a graph the nodes of which can be represented by open disks in the plane such that two disks intersect if and only if the represented nodes are adjacent. It is not hard to verify that the neighbourhood of a node which is represented by a smallest size disk can be covered by 6 cliques, and clearly any induced subgraph of a disk graph is a disk graph. Hence  $\chi(G) \leq 6\omega(G) - 5$  for every disk graph  $G$ . The following proposition yields that  $\text{imp}(G) \leq 6$  for these graphs.

**PROPOSITION 4.1.** *For each graph  $G$  and  $t \geq 1$ , if each induced subgraph of  $G$  contains a node the neighbourhood of which can be covered at least  $p/t$  times by a family of  $p$  cliques (for some  $p \geq 1$ ), then  $\text{imp}(G) \leq t$ .*

*Proof.* Let  $G$  have  $n$  nodes. We can order the nodes of  $G$  in such a way that, for each  $i=2, \dots, n$ , the nodes of  $\{v_1, \dots, v_{i-1}\}$  which are adjacent to  $v_i$  can be covered  $q_i$  times by a family of  $p_i$  cliques, where  $p_i/q_i \leq t$ . We claim that, for any integral weight vector  $\mathbf{x}$ , we need at most  $\omega(G, \mathbf{x}) t$  colours to colour each node  $v$  with  $x_v$  colours, if we greedily colour the nodes of  $G$  in the order above. To see this, let  $w(i)$  denote the sum of the values  $x_{v_j}$  such that  $j < i$  and  $v_j$  is adjacent to  $v_i$ . Observe that for each  $i=2, \dots, n$  we have  $p_i(\omega(G, \mathbf{x}) - x_{v_i}) \geq q_i w(i)$ , and so  $w(i) \leq t(\omega(G, \mathbf{x}) - x_{v_i})$ . When we come to colour  $v_i$ , at most  $w(i) \leq \omega(G, \mathbf{x}) t - x_{v_i}$  colours are already used for the neighbours of  $v_i$ . Thus we can colour  $v_i$  with  $x_{v_i}$  colours using at most  $\omega(G, \mathbf{x}) t$  colours all together. Therefore  $\chi(G, \mathbf{x}) \leq \omega(G, \mathbf{x}) t$ , as required. ■

The last result does not always yield good bounds. For the complete bipartite graph  $K_{n/2, n/2}$  on  $n$  nodes (with  $n$  even), the smallest  $t$  which fulfils

the conditions in Proposition 4.1 equals  $n/2$ , and hence the best bound which can be derived from this proposition for a graph on  $n$  nodes can be  $n/2$  times bigger than the imperfection ratio.

A *unit disk graph* is a disk graph which has a representation with unit diameter disks. Since any induced subgraph of a unit disk graph is a unit disk graph, and since the neighbours of the left-most bottom node can be covered by 3 cliques, Proposition 4.1 shows that the imperfection ratio of such graphs is at most 3. However, this bound can be improved. In [12] it is shown by a method involving coverings by perfect graphs that  $\text{imp}(G) \leq 2.155$  for any unit disk graph  $G$ . This leads to the second general bound on the imperfection ratio we discuss in this section.

The *fractional perfect covering number* is defined by

$$\text{pc}_f(G) := \inf \left\{ \frac{p}{q} : \begin{array}{l} \text{each node of } G \text{ can be covered } q \text{ times by } p \text{ perfect} \\ \text{induced subgraphs} \end{array} \right\}.$$

The infimum in the definition of  $\text{pc}_f(G)$  can be replaced by minimum, since it is not hard to verify that  $\text{pc}_f(G)$  is the value of the following linear program with a variable  $y_P$  for every induced perfect subgraph  $P$  of  $G$ :  $\min \sum_P y_P$  subject to  $\sum_{P \ni v} y_P \geq 1$  for each node  $v$  of  $G$  and  $y_P \geq 0$  for each induced perfect subgraph  $P$ .

It is known [17] that, given any integral weight vector  $\mathbf{x}$  for a perfect graph  $H$ , each node can be covered  $x_v$  times by  $\omega(H, \mathbf{x})$  stable sets. It follows that, if each node of  $G$  can be covered  $q$  times by  $p$  perfect induced subgraphs, and  $\mathbf{x}$  is an integral weight vector then each node  $v$  can be covered  $x_v q$  times with  $\omega(G, \mathbf{x})$   $p$  stable sets. Hence  $\chi_f(G, \mathbf{x}) \leq \omega(G, \mathbf{x}) p/q$ , which implies that

$$\text{imp}(G) \leq \text{pc}_f(G). \tag{17}$$

Clearly  $\text{imp}(G) = \text{pc}_f(G)$  for any perfect graph  $G$ . It follows from results in [12] that  $\text{imp}(G) = \text{pc}_f(G)$  for minimal imperfect graphs, too. The discussion on unit disk graphs in [12] which was mentioned above in fact establishes the given upper bound on  $\text{pc}_f(G)$  and hence on  $\text{imp}(G)$ .

The third bound we consider in this section involves the *cochromatic number*  $z(G)$ , that is the least number of stable sets and cliques needed to cover the nodes of  $G$ ; see, for example, [14]. Since  $G$  is perfect if  $z(G) \leq 2$ , it follows that  $\text{pc}_f(G) \leq z(G)/2$  provided  $G$  contains at least one edge and is not the complete graph (i.e.,  $z(G) \geq 2$ ), and so  $\text{imp}(G) \leq z(G)/2$  for such graphs. The following proposition gives an upper bound on  $z(G)$  (and thus on  $\text{pc}_f(G)$ ) in terms of  $\text{imp}(G)$ .

**PROPOSITION 4.2.** *For each graph  $G$  on  $n$  nodes which is neither complete nor the complement of a complete graph,*

$$\text{imp}(G) \geq \frac{z(G)^2}{4n} \geq \frac{\text{pc}_f(G)^2}{n}.$$

*Proof.* For  $t \geq 1$ , let  $f_t(n) = \max\{z(H) : |V(H)| = n, \text{imp}(H) \leq t\}$ . We show by induction on  $n$  that  $f_t(n) \leq 2\sqrt{tn}$  for all positive integers  $n$ . Clearly  $f_t(n) \leq n \leq 2\sqrt{tn}$  for  $n = 1, 2, 3, 4$ . Now let  $n \geq 5$ , and assume that  $f_t(k) \leq 2\sqrt{tk}$  for all  $1 \leq k \leq n-1$ . Let  $H$  be a graph on  $n$  nodes with  $\text{imp}(H) \leq t$ . Then

$$\frac{n}{\alpha(H)\omega(H)} \leq \text{imp}(H) \leq t,$$

and so  $\alpha(H)\omega(H) \geq n/t$ . Therefore  $\max\{\alpha(G), \omega(G)\} \geq \lceil \sqrt{n/t} \rceil$ . Hence

$$\begin{aligned} f_t(n) &\leq 1 + f_t(n - \lceil \sqrt{n/t} \rceil) \leq 1 + 2\sqrt{t} \sqrt{n - \lceil \sqrt{n/t} \rceil} \\ &\leq 1 + 2\sqrt{tn} \sqrt{1 - \frac{1}{\sqrt{tn}}} \\ &\leq 1 + 2\sqrt{tn} \left(1 - \frac{1}{2\sqrt{tn}}\right) \quad \text{since } \sqrt{1-x} \leq 1 - x/2 \text{ for } 0 \leq x \leq 1 \\ &= 2\sqrt{tn}. \quad \blacksquare \end{aligned}$$

This proposition yields  $z(G) \leq 2\sqrt{n}\sqrt{\text{imp}(G)}$  which extends the result [7] that for a perfect graph  $G$ ,  $z(G) \leq 3/2 + \sqrt{2n+9/4}$ , apart from a factor of about  $\sqrt{2}$ . Let us remark that there are perfect graphs  $G$  with  $z(G) \geq \sqrt{n}$ . For example the interval graph  $H$  on  $\{1, 2, \dots, k^2\}$  with intervals  $\{1, \dots, k\}, \{2, \dots, k+1\}, \dots, \{k^2-k+1, \dots, k^2\}$  is perfect, and

$$z(H) \geq \frac{k^2 - k + 1}{k} = k - \frac{k-1}{k},$$

and so  $z(H) = k$ .

There can also be a gap between  $\text{pc}_f(G)$  and  $\text{imp}(G)$ . Consider, for example, the Petersen graph  $P$  shown in Fig. 2. The maximum number of nodes in an induced bipartite subgraph is 7, and since there are no triangles, this is also the maximum number of nodes in an induced perfect subgraph. Hence  $\text{pc}_f(P) \geq |V(P)|/7 = 10/7$ . (In fact  $\text{pc}_f(P) = 10/7$  since  $P$  is node-transitive, see [11].) But  $\text{imp}(P) = \chi_f(P)/2 = 5/4$  since the Petersen

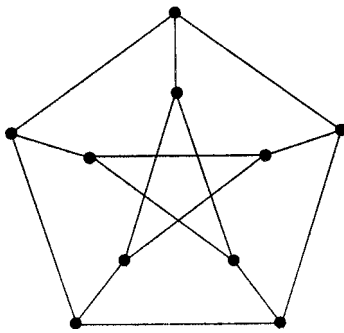


FIG. 2. The Petersen graph.

graph is triangle-free. The next proposition allows us to increase the gap between  $pc_f(G)$  and  $imp(G)$ .

PROPOSITION 4.3. *For any two graphs  $G$  and  $H$ , the lexicographic product  $G[H]$  satisfies  $pc_f(G[H]) = pc_f(G) pc_f(H)$ .*

*Proof.* Let  $pc_f(G) = a/b$ , and let  $P_1, P_2, \dots, P_a$  be a family of perfect induced subgraphs of  $G$  which cover each node of  $G$   $b$  times. Let  $pc_f(H) = c/d$ , and let  $Q_1, Q_2, \dots, Q_c$  be a family of perfect induced subgraphs of  $H$  which cover each node of  $H$   $d$  times. Then the graphs  $P_i[Q_j]$ ,  $i = 1, \dots, a$ ;  $j = 1, \dots, c$  are perfect and cover every node  $bd$  times. Hence  $pc_f(G[H]) \leq ac/bd = pc_f(G) pc_f(H)$ .

For the reverse inequality, let  $pc_f(G[H]) = e/f$ , and consider a covering of  $G[H] = G[H^v: v \in V(G)]$  with a family  $P_1, \dots, P_e$  of perfect induced subgraphs of  $G[H]$  such that each node is covered  $f$  times. This covering gives rise to a covering of  $G$  by perfect induced subgraphs of  $G$ : for  $i = 1, \dots, e$ , let  $P'_i$  be the subgraph of  $G$  induced by the nodes  $v$  of  $G$  such that  $P_i$  contains at least one node from the copy  $H^v$  of  $H$  corresponding to  $v$ . Since every node of  $H^v$  is covered  $f$  times by the graphs  $P_1, \dots, P_e$ , the node  $v$  is covered at least  $fpc_f(H)$  times by  $P'_1, \dots, P'_e$ . Hence  $pc_f(G) \leq e/(fpc_f(H))$ , which is equivalent to  $pc_f(G[H]) \geq pc_f(G) pc_f(H)$ . ■

There is a corresponding result for the maximum number of nodes in an induced perfect subgraph, that is if we let  $pn(G)$  be the maximum number of nodes in a perfect induced subgraph of  $G$ , then  $pn(G[H]) = pn(G) pn(H)$ . Each of these extends the result that  $G[H]$  is perfect if and only if  $G$  and  $H$  are perfect (as did Theorem 2.1).

PROPOSITION 4.4. *For every positive integer  $i$ , there is a graph  $G$  on  $n = 10^i$  nodes with  $pc_f(G) = n^\delta imp(G)$ , where  $\delta = \ln(8/7)/\ln 10 \approx 0.058$ .*

*Proof.* Let  $P$  denote the Petersen graph (see Fig. 2), let  $G^1 = P$ , and let  $G^{i+1} = G^i[P]$  for  $i = 1, 2, \dots$ . Then  $|V(G^i)| = 10^i$ ;  $\text{pc}_f(G^i) = (10/7)^i$  by Proposition 4.3; and  $\text{imp}(G^i) = (5/4)^i$  by Theorem 2.1. Hence

$$\ln \left( \frac{\text{pc}_f(G^i)}{\text{imp}(G^i)} \right) = i \ln \left( \frac{8}{7} \right) = \frac{\ln |V(G^i)|}{\ln 10} \ln \left( \frac{8}{7} \right) = \delta \ln(|V(G^i)|),$$

which gives the required result. ■

The last result shows that the ratio  $\text{pc}_f(G)/\text{imp}(G)$  can grow like  $n^\delta$  for a graph on  $n$  nodes. In fact the ratio usually grows with  $n$ , but like  $\log n$  as also does the ratio  $z(G)/\text{imp}(G)$ . To prove this we need one preliminary lemma. For any graphs  $G$  and  $H$ , let  $\text{ex}_H(G)$  be the maximum number of nodes in an induced subgraph  $G'$  of  $G$  such that  $G'$  has no induced subgraph isomorphic to  $H$ . (Here  $\text{ex}$  stands for excluding.)

LEMMA 4.1. *For any graph  $H$  with  $h$  nodes, a.s.  $\text{ex}_H(G_{n, 1/2}) \leq c \ln n$ , where we may take  $c = h^2 2^{\binom{h}{2}}$ .*

*Proof.* Fix a graph  $H$  with  $h$  nodes. For each  $n$ , in the complete graph  $K_n$  on  $n$  nodes there is a packing of  $t_n$  edge-disjoint copies of  $K_h$  such that  $t_n \sim \binom{n}{2} / \binom{h}{2}$ , and so  $t_n \geq n^2/h^2$  for  $n$  sufficiently large [2, 23]. Now let  $c = h^2 2^{\binom{h}{2}}$ , and let  $z = z(n) = \lceil c \ln n \rceil$ . We denote by  $H \subset G$  the event that  $H$  is (isomorphic to) an induced subgraph of  $G$ . Then

$$P(H \not\subset G_{h, 1/2}) \leq 1 - 2^{-\binom{h}{2}},$$

and hence

$$P(H \not\subset G_{z, 1/2}) \leq (1 - 2^{-\binom{h}{2}})^{t_z} \leq (e^{-2^{-\binom{h}{2}}})^{t_z}.$$

It follows that

$$\begin{aligned} P(\text{ex}_H(G_{n, 1/2}) \geq z) &\leq \binom{n}{z} P(H \not\subset G_{z, 1/2}) \leq \left( \frac{ne}{z} \right)^z e^{-t_z 2^{-\binom{h}{2}}} \\ &\leq \left( \frac{ne}{z} e^{-(z/h^2) 2^{-\binom{h}{2}}} \right)^z \quad \text{for sufficiently large } n \\ &\leq \left( \frac{ne}{z} e^{-\ln n} \right)^z \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ . ■

Recall that the fractional perfect graph covering number  $pc_f(G)$  and the cochromatic number  $z(G)$  satisfy  $pc_f(G) \leq z(G)/2$  for every graph  $G$  with at least an edge which is not complete.

PROPOSITION 4.5. *For any  $\varepsilon > 0$ , the random graph  $G_{n, 1/2}$  a.s. satisfies*

$$(1 - \varepsilon) \log n \leq \frac{z(G_{n, 1/2})}{2 \text{imp}(G_{n, 1/2})} \leq (1 + \varepsilon) \log n,$$

and there is a constant  $c > 0$  such that a.s.

$$c \log n \leq \frac{pc_f(G_{n, 1/2})}{\text{imp}(G_{n, 1/2})}.$$

*Proof.* We have seen in the proof of Theorem 3.3 that  $G_{n, p}$  almost surely satisfies (8), (9) and (11). Hence

$$\frac{n}{2 \log n} \leq z(G_{n, 1/2}) \leq (1 + \varepsilon) \frac{n}{2 \log n} \quad \text{a.s.}$$

since  $z(G) \leq \chi(G)$  always. Also by Lemma 4.1 a.s.  $ex_{C_5}(G_{n, 1/2}) \leq c_1 \ln n$  for a constant  $c_1 \leq 5^2 2^{\binom{5}{2}} = 25600$ , and hence a.s.  $pc_f(G_{n, 1/2}) \geq n/(c_1 \ln n)$ . The result now follows from Theorem 3.3. ■

## 5. HARDNESS RESULT

In this final section we prove the following theorem.

THEOREM 5.1. *It is NP-hard to determine the fractional chromatic number of a triangle-free graph.*

Since  $\text{imp}(G) = \chi_f(G)/2$  for any triangle-free graph  $G$  [12], we obtain:

COROLLARY 5.1. *It is NP-hard to determine the imperfection ratio of a graph.*

Before we prove the theorem we need some preliminaries. First we consider what happens to the fractional chromatic number when two graphs overlap in a clique.

LEMMA 5.1. *Let the graph  $G$  be formed from two graphs  $H^1$  and  $H^2$  which overlap in a clique. Then*

$$\chi_f(G, \mathbf{x}) = \max\{\chi_f(H^1, \mathbf{x}), \chi_f(H^2, \mathbf{x})\}$$

for any weight vector  $\mathbf{x}$ .

*Proof.* Without loss of generality we may assume that the weight vector  $\mathbf{x}$  is integral because otherwise we could consider the vector  $N\mathbf{x}$  where  $N$  is chosen such that  $N\mathbf{x}$  is integral. Let  $\chi(G, \mathbf{x})$  denote the chromatic number of the replicated graph  $G_{\mathbf{x}}$ , where each node  $v$  is replaced by a clique with  $x_v$  nodes. Clearly  $\chi(G) = \max\{\chi(H^1), \chi(H^2)\}$ , and similarly we have  $\chi(G, \mathbf{x}) = \max\{\chi(H^1, \mathbf{x}), \chi(H^2, \mathbf{x})\}$ . But  $\chi_f(G, \mathbf{x}) = \lim_{t \rightarrow \infty} \chi(G, t\mathbf{x})/t$  (see, for example, [24]), and the lemma follows. ■

Next we introduce the notion of a *tight* node. Denote the set of all stable sets of a graph  $G$  by  $\mathcal{S}_G$ . We call a non-negative vector  $\mathbf{y} = (y_S : S \in \mathcal{S}_G)$  a *fractional stable set cover* for  $G$  if  $\sum_{S \ni v} y_S \geq 1$  for all  $v \in V(G)$ . The *cost* of  $\mathbf{y}$  is  $\sum_{S \in \mathcal{S}_G} y_S$ . A fractional stable set cover is called *optimal* if its cost is minimal, that is if it equals  $\chi_f(G) (= \chi_f(G, \mathbf{1}))$ . A node  $v$  of a graph  $G$  is called *tight* if every optimal fractional stable set cover  $\mathbf{y}$  satisfies  $\sum_{S \ni v} y_S = 1$ . The next lemma shows that every graph  $G$  has a tight node.

LEMMA 5.2. *Let  $U$  be a minimal set of nodes in  $G$  such that the corresponding induced subgraph  $H$  satisfies  $\chi_f(H) = \chi_f(G)$ . Then each node in  $U$  is tight.*

*Proof.* We may assume that  $G$  has at least one edge. Let  $v \in U$ . Assume for a contradiction that there exists an optimal fractional stable set cover  $\mathbf{y}$  for  $G$  with  $\sum_{S \ni v} y_S = 1 + \delta > 1$ . Then there is an optimal fractional stable set cover  $\tilde{\mathbf{y}}$  for  $H$  with  $\sum_{S \ni v} \tilde{y}_S = 1 + \delta$ . But by the minimality of  $U$  there is a fractional stable set cover  $\mathbf{z}$  for  $H - v$  with cost  $< \chi_f(G)$ . Then  $1/(1 + \delta) \tilde{\mathbf{y}} + \delta/(1 + \delta) \mathbf{z}$  is a stable set cover for  $H$  with cost  $< \chi_f(G)$ , a contradiction. ■

Tight nodes  $v$  can be characterised as follows. Let  $G_v$  be the graph obtained from  $G$  by adding a node  $\bar{v}$  with exactly the same neighbours as  $v$  (so  $v$  and  $\bar{v}$  are not adjacent). Then the node  $v$  is tight if for every optimal fractional stable set cover  $\mathbf{y}$  of  $G_v$ , and every stable set  $S$  with  $y_S > 0$ , either both  $v, \bar{v} \in S$  or neither do. For a node  $v \in V(G)$ , consider the graph  $G \cdot v$  which is obtained by adding two nodes  $\bar{v}$  and  $\tilde{v}$  to  $V(G)$  and connecting  $\bar{v}$  to all nodes adjacent to  $v$  in  $G$  and to  $\tilde{v}$ ; see Fig. 3.

*Observation 5.1.* If  $v$  is a tight node of a graph  $G$  with at least one edge, then  $\chi_f(G) = \chi_f(G \cdot v)$ ; and for every optimal stable set cover  $\mathbf{y}$  of  $G \cdot v$ , and every stable set  $S$  with  $y_S > 0$ , at most one of the nodes  $v, \tilde{v}$  is in  $S$ .

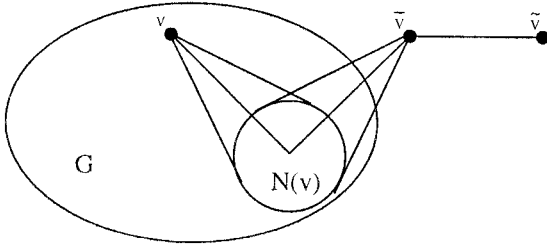


FIG. 3. The graph  $G \cdot v$ .

Let  $G[E \leftarrow H \cdot v]$  denote the graph obtained by replacing each edge  $\{u, w\}$  of  $G$  by a copy of  $H \cdot v$  and identifying  $u$  with  $v$  and identifying  $w$  with  $\tilde{v}$ . Observe that if the graph  $H$  is triangle-free, then  $G[E \leftarrow H \cdot v]$  must be triangle free.

LEMMA 5.3. *Let  $H$  be a graph with at least one edge and let  $\chi_f(H) = c$ . Then for all graphs  $G$*

$$\chi_f(G[E \leftarrow H \cdot v]) \leq c \text{ for all } v \in V(H) \Leftrightarrow \chi_f(G) \leq c.$$

*Proof.* Let  $v$  be a tight node of  $H$ . Denote  $G[E \leftarrow H \cdot v]$  by  $\tilde{G}$ . Suppose that  $\chi_f(\tilde{G}) \leq c$ . Let  $\mathbf{y}$  be an optimal fractional stable set cover of  $\tilde{G}$ . For each stable set  $S$  in  $G$ , let  $x_S$  be the sum of the values  $y_{\tilde{S}}$  over the stable sets  $\tilde{S}$  of  $\tilde{G}$  such that  $V(G) \cap \tilde{S} = S$ , i.e.,

$$x_S = \sum_{\substack{\tilde{S} \in \mathcal{S}_{\tilde{G}} \\ V(G) \cap \tilde{S} = S}} y_{\tilde{S}}.$$

Since  $v$  is a tight node of  $H$ , Observation 5.1 shows that the set  $V(G) \cap \tilde{S}$  is a stable set in  $G$  for each stable set  $\tilde{S}$  of  $\tilde{G}$  with  $y_{\tilde{S}} > 0$ . Thus we have

$$\sum_{S \ni u} x_S = \sum_{\tilde{S} \ni u} y_{\tilde{S}} \geq 1 \quad \text{for all } u \in V(G).$$

Also  $\sum_{S \in \mathcal{S}_G} x_S = \sum_{\tilde{S} \in \mathcal{S}_{\tilde{G}}} y_{\tilde{S}} = c$  and therefore  $\chi_f(G) \leq c$ .

Suppose now that  $\chi_f(G) \leq c$ . Let  $v$  be any node of  $V(H)$  and let  $H: v$  be the graph obtained by adding the edge  $\{v, \tilde{v}\}$  to  $H \cdot v$ . Observe that  $\chi_f(H: v) = \chi_f(H \cdot v) = c$  by Observation 5.1. Hence by Lemma 5.1,

$$\chi_f(\tilde{G}) \leq \chi_f(G[E \leftarrow H: v]) = \max\{\chi_f(G), \chi_f(H: v)\} = c,$$

and thus  $\chi_f(\tilde{G}) \leq c$ . ■

*Proof (of Theorem 5.1).* We want to reduce the problem of approximating the fractional chromatic number of a graph up to a constant, to the problem of finding the fractional chromatic number for a triangle-free graph. The former problem is known to be NP-hard [19]. (Indeed it is shown that it is hard to approximate the fractional chromatic number of a graph  $G$  up to a factor  $|V(G)|^\delta$  for some  $\delta > 0$ .)

Assume we have a subroutine to determine the fractional chromatic number of a triangle-free graph. Let  $G$  be any graph of order  $n$ . Following, for example, [1], one may construct a triangle-free graph  $H$  of order  $N$  such that  $N \geq n^3$ ,  $N = O(n^3)$ , and  $\alpha(H) \leq n^2$ , and thus  $\chi_f(H) \geq n$ . By removing nodes of  $H$  one by one, we may construct a sequence of triangle-free graphs  $H = H_N, \dots, H_1$ , with  $\chi_f(H_i) \leq \chi_f(H_{i+1}) \leq \chi_f(H_i) + 1$ ,  $i = 1, \dots, N-1$ ,  $\chi_f(H_1) = 1$  and  $\chi_f(H_N) \geq n$ .

For all  $i = 1, \dots, N$ , we can determine  $\chi_f(H_i)$ , and for each node  $v$  of  $H_i$  we can calculate  $\chi_f(G[E \leftarrow H_i \cdot v])$  by means of the subroutine, since these graphs are triangle-free as we observed earlier. By Lemma 5.3 we can find a number  $c$  such that  $c \leq \chi_f(G) \leq c + 1$ , and thus approximate  $\chi_f(G)$  within 1.

The subroutine is called  $O(nN) = O(n^4)$  times and each triangle-free graph involved contains  $O(n^5)$  nodes. Thus the reduction can be performed in polynomial time. ■

Finally, let us note that Lemma 5.1 which we needed above yields the following (unsurprising?) result.

**PROPOSITION 5.1.** *Let the graph  $G$  be formed from two graphs  $H^1$  and  $H^2$  which overlap in a clique. Then  $\text{imp}(G) = \max\{\text{imp}(H^1), \text{imp}(H^2)\}$ .*

*Proof.* Since each  $H^i$  is an induced subgraph of  $G$ ,  $\text{imp}(G) \geq \max\{\text{imp}(H^1), \text{imp}(H^2)\}$ . For the reverse inequality we use Lemma 5.1: for every weight vector  $\mathbf{x}$

$$\begin{aligned} \frac{\chi_f(G, \mathbf{x})}{\omega(G, \mathbf{x})} &= \frac{\max\{\chi_f(H^1, \mathbf{x}), \chi_f(H^2, \mathbf{x})\}}{\omega(G, \mathbf{x})} \\ &\leq \max\left\{\frac{\chi_f(H^1, \mathbf{x})}{\omega(H^1, \mathbf{x})}, \frac{\chi_f(H^2, \mathbf{x})}{\omega(H^2, \mathbf{x})}\right\} \\ &\leq \max\{\text{imp}(H^1), \text{imp}(H^2)\}, \end{aligned}$$

which completes the proof. ■

## ACKNOWLEDGMENT

We thank Stephan Brandt for pointing out some references and the referees for many helpful comments.

## REFERENCES

1. N. Alon, Explicit Ramsey graphs and orthonormal labelings, *Electr. J. Combinatorics* **1** (1994), 12.
2. N. Alon and J. Spencer, "The Probabilistic Method," Wiley, New York, 1992.
3. B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combin.* **1** (1980), 311–316.
4. B. Bollobás, The independence ratio of regular graphs, *Proc. Amer. Math. Soc.* **83** (1981), 433–436.
5. B. Bollobás, "Random Graphs," Academic Press, San Diego, 1985.
6. B. Bollobás, The chromatic number of random graphs, *Combinatorica* **8** (1988), 49–55.
7. Y. Caro, On the covering number of combinatorial structures, *Ars Combinatoria* **29** (1990), 111–124.
8. P. Erdős, Some remarks on chromatic graphs, *Colloq. Math.* **16** (1967), 253–256.
9. A. M. Frieze, On the independence numbers of random graphs, *Discrete Math.* **81** (1990), 171–175.
10. A. M. Frieze and T. Łuczak, On the independence numbers of random regular graphs, *J. Combin. Theory Ser. B* **54** (1992), 123–132.
11. S. Gerke, "Channel Assignment Problems," DPhil thesis, University of Oxford, 2000.
12. S. Gerke and C. McDiarmid, Graph imperfection, *J. Combin. Theory Ser. B* **83** (2001), 58–78.
13. M. Grötschel, L. Lovász, and A. Schrijver, "Geometric Algorithms and Combinatorial Optimization," Springer-Verlag, 1988.
14. T. R. Jensen and B. Toft, "Graph Colouring Problems," Wiley, New York, 1995.
15. A. Johansson, The choice number of sparse graphs, *J. Combin. Theory Ser. B*, in press.
16. J. H. Kim, The Ramsey number  $R(3, t)$  has order of magnitude  $t^2/\log t$ , *Random Structures Algorithms* **7** (1995), 173–207.
17. L. Lovász, Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2** (1972), 253–267.
18. T. Łuczak, The chromatic number of random graphs, *Combinatorica* **11** (1991), 45–54.
19. C. Lund and M. Yannakakis, On the hardness of approximating minimization problems, *J. Assoc. Comput. Mach.* **41** (1994), 960–981.
20. C. McDiarmid, Concentration, in "Probabilistic Methods for Algorithmic Discrete Mathematics" (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, Eds.), pp. 195–248, Springer-Verlag, Berlin, 1998.
21. C. McDiarmid, On the chromatic number of random graphs, *Random Structures Algorithms* **1** (1990), 435–442.
22. C. McDiarmid, Graph imperfection and channel assignment, in "Perfect Graphs" (J. L. Ramirez Alfonsin and B. Reed, Eds.), Wiley and Sons, Chichester, UK, in press.
23. V. Rödl, On a packing and covering problem, *European J. Combinatorics* **5** (1985), 69–78.
24. E. R. Scheinerman and D. H. Ullman, "Fractional Graph Theory," Wiley, New York, 1997.
25. G. Simonyi, Imperfection ratio and graph entropy, manuscript, 1999.