Obtaining Online Ecological Colourings by Generalizing First-Fit

Matthew Johnson[1*], Viresh Patel[2**], Daniël Paulusma[1***], and Théophile Trunck[3]

[1]School of Engineering and Computing Sciences, Durham University,
Science Laboratories, South Road, Durham DH1 3LE, U.K.
{matthew.johnson2,daniel.paulusma}@dur.ac.uk
[2]School of Mathematics, Univeristy of Birmingham,
Edgbaston, Birmingham, B15 2TT, U.K.
viresh.s.patel@gmail.com
[3]École Normale Supérieure de Lyon,
46 Allée d'Italie, 69364 Lyon Cedex 07, France.
theophile.trunck@ens-lyon.fr

Abstract. A colouring of a graph is ecological if every pair of vertices that have the same set of colours in their neighbourhood are coloured alike. We consider the following problem: if a graph G and an ecological colouring c of G are given, can further vertices added to G, one at a time, be coloured so that at each stage the current graph is ecologically coloured? If the answer is yes, then we say that the pair (G, c) is ecologically online extendible. By generalizing the well-known First-Fit algorithm, we are able to characterize when (G, c) is ecologically online extendible, and to show that deciding whether (G, c) is ecologically extendible can be done in polynomial time. We also describe when the extension is possible using only colours from a given finite set C.

For the case where c is a colouring of G in which each vertex is coloured distinctly, we give a simple characterization of when (G, c) is ecologically online extendible using only the colours of c, and we also show that (G, c) is always online extendible using the colours of c plus one extra colour.

We also study (off-line) ecological H-colourings (an H-colouring of G is a homomorphism from G to H). We study the problem of deciding whether G has an ecological H-colouring for some fixed H and give a characterization of its computational complexity in terms of the structure of H.

1 Introduction

One of the goals of social network theory is to determine patterns of relationships amongst actors in a society. Social networks can be represented by graphs, where vertices of the graph represent individuals and edges represent relationships amongst them. One way to study patterns of relationships in such networks is to assign *labels* in such a way that those who are assigned the same label have similar sorts of relationships within

* Supported by EPSRC Grant EP/E048374/1
** Supported by EPSRC Grant EP/F064551/1
*** Supported by EPSRC Grant EP/G043434/1

the network; see, for example, Hummon and Carley [10]. Several graph-theoretic concepts such as ecological colourings [2], role assignments [7] and perfect colourings [3], have been introduced to facilitate the study of social networks in this way.

This paper focuses on ecological colourings. The term "ecological" is derived from certain models of population *ecology* in which individuals are assumed to be determined by their environment. For example, in biology, features of a species' morphology are usually defined in relation to the way such a species interacts with other species. Also, some network theories of attitude formation assume that one's attitude is predicted by the combination of the attitudes of surrounding individuals [4, 6].

We introduce some basic notation and terminology. Throughout the paper, all graphs are undirected and without loops or multiple edges unless otherwise stated. We denote the vertex and edge sets of a graph G by V_G and E_G respectively (but often omit subscripts). An edge between u and v is denoted (u, v). The *neighbourhood* of a vertex u in G is denoted $N_G(u) = \{v \mid (u, v) \in E_G\}$. For a subset $S \subseteq V_G$ and a function c on V_G (for example, a colouring of the vertices), we use the short-hand notation $c(S)$ for the set $\{c(u) \mid u \in S\}$. The *colourhood* of a vertex v in a graph G with colouring c is defined to be $c(N_G(v))$. For a set C, we write $A + x$ to denote $A \cup \{x\}$ for some subset $A \subseteq C$ and $x \in C$.

Ecological colourings were introduced by Borgatti and Everett in [2] to analyse power in experimental exchange networks. Formally, an *ecological colouring* of a graph $G = (V_G, E_G)$ is a vertex mapping $c : V_G \to \{1, \ldots\}$ such that any two vertices $u, v \in V_G$ with the same *colourhood*, that is, with $c(N_G(u)) = c(N_G(v))$, have the same colour $c(u) = c(v)$. Note that such a colouring does not have to be *proper*, that is, two adjacent vertices may receive the same colour. This reflects that two individuals that play the same role in their environment might be related to each other. See Fig. 1 for an example of a proper ecological colouring.

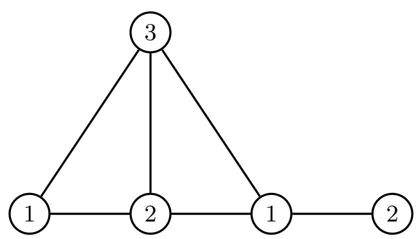

Fig. 1. A proper ecological colouring that is also an ecological K_3-colouring.

One of the appealing features of ecological colourings is a result of Crescenzi et al. [5]. In order to state the result precisely, we need to introduce some terminology. A *twin-free* graph (also known as a neighbourhood-distinct graph) is a graph in which no two vertices have the same neighbourhood (including empty neighbourhoods). A graph G that is not twin-free can be made twin-free as follows: whenever we find a pair of vertices u and v for which $N_G(u) = N_G(v)$, we delete one of them until no such pair remains. It is easy to check that the resulting graph is independent of the order in which vertices are deleted and is twin-free; it is called the *neighbourhood graph* of G and is denoted by G_N. The main result of Crescenzi et al. [5] states that an ecological colouring of a graph G using exactly k colours can be found in polynomial time for each $1 \leq k \leq |V_{G_N}|$ and does not exist for $k \geq |V_{G_N}| + 1$.

Our motivation for studying online ecological colourings. In static optimization problems, one is often faced with the challenge of determining efficient algorithms that solve a particular problem optimally for any given instance of the problem. In the area of *dynamic* optimization the situation is more complicated: here, one often lacks knowledge of the complete instance of the problem.

This paper studies ecological colourings for dynamic networks. Gyárfás and Lehel [8] introduced the concept of *online colouring* to tackle dynamical storage allocations. An *online colouring algorithm* irrevocably colours the vertices of a graph one by one, as they are revealed, where determination of the colour of a new vertex can only depend on the coloured subgraph induced by the revealed vertices. See [11] for a survey on online colouring.

Perhaps the most well-known online colouring algorithm is FIRST-FIT. Starting from the empty graph, this algorithm assigns each new vertex the least colour from $\{1, 2, \ldots\}$ that does not appear in its neighbourhood. It is easy to check that an ecological colouring is obtained at each stage and hence FIRST-FIT is an example of an *online ecological colouring algorithm*. Note, however, that it may use an unbounded number of colours. If we wish to use at most k colours when we start from the empty graph, then we can alter FIRST-FIT so that each new vertex v is assigned, if possible, the least colour in $\{1, 2, \ldots, k-1\}$ not in the colourhood of v, or else v is coloured k. We call the modified algorithm k-FIRST-FIT. It gives a colouring that is ecological but not necessarily proper (consider, for example, 1-FIRST-FIT which assigns all vertices the same colour).

A natural situation to consider is when we are given a nonempty *start graph* $G_0 = G$, the vertices of which are coloured by an ecological colouring c. At each stage i, a new vertex v_i is added to G_{i-1} (the graph from the previous stage) together with (zero or more) edges between v_i and the vertices of G_{i-1}, to give the graph G_i. Knowledge of G_i is the *only* information we have at stage i. Our task is to colour the new vertex v_i at each stage i, without changing the colours of the existing vertices, to give an ecological colouring of G_i. If there exists an online colouring algorithm that accomplishes this task, we say that the pair (G, c) is *(ecologically) online extendible*. If there is an algorithm that uses only colours from a finite set C, we say that (G, c) is online extendible with C (of course, we assume throughout that $C \supseteq c(V)$). Motivated by our observation that colourings obtained by FIRST-FIT and k-FIRST-FIT are ecological, we examine which pairs (G, c) are online extendible.

Our motivation for studying ecological H-colourings. In order to analyse the salient features of a large network G, it is often desirable to compress G into a smaller network H in such a way that important aspects of G are maintained in H. Extracting relevant information about G becomes much easier using H. This idea of compression is encapsulated by the notion of *graph homomorphisms*, which are generalizations of graph colourings. Let G and H be two graphs. An *H-colouring* or *homo-*

morphism from G to H is a function $f : V_G \to V_H$ such that for all $(u, v) \in E_G$ we have $(f(u), f(v)) \in E_H$. An *ecological* H-colouring of G is a homomorphism $f : V_G \to V_H$ such that, for all pairs of vertices $u, v \in V_G$, we have

$$f(N_G(u)) = f(N_G(v)) \implies f(u) = f(v).$$

See Fig. 1 for an example of an ecological K_3-colouring, where K_3 denotes the complete graph on $\{1, 2, 3\}$. The ECOLOGICAL H-COLOURING problem asks if a graph G has an ecological H-Colouring. Classifying the computational complexity of this problem is our second main goal in this paper. This research was motivated by Crescenzi et al. [5] who posed the computational complexity of a variant of ECOLOGICAL H-COLOURING as an open problem. This variant is that of testing whether a graph has a surjective ecological H-colouring, that is, an ecological H-colouring f with $f(V_G) = V_H$. We call this problem the SURJECTIVE ECOLOGICAL H-COLOURING problem. We will show that our results on ECOLOGICAL H-COLOURING provide a partial complexity classification of SURJECTIVE ECOLOGICAL H-COLOURING.

Our results and paper organisation. In Section 2, we characterize when a pair (G, c) is online extendible. We extend the characterization to the case where the extension must use only colours from a give finite set C. In Section 3, we investigate the computational complexity of deciding whether online extensions can be obtained and show that this can be done in polynomial time if the set of colours is either unspecified or fixed. We then focus on the case where each vertex of a k-vertex graph G is coloured distinctly by c. In Section 4, we show that such a pair (G, c) is always online extendible with $k + 1$ colours, and give a polynomial-time online colouring algorithm for achieving this. We show that this result is tight by giving a simple characterization of exactly which (G, c) are not ecologically online extendible with k colours. This characterization can be verified in polynomial time. In Section 5, we classify the computational complexity of the ECOLOGICAL H-COLOURING problem. We show that if H is bipartite or contains a loop then ECOLOGICAL H-COLOURING is polynomial-time solvable, and is NP-complete otherwise. Section 6 contains conclusions and open problems. Amongst others, we will discuss the complexity status of the SURJECTIVE ECOLOGICAL H-COLOURING problem and indicate which cases are still open.

2 Online Ecological Colouring

2.1 Examples

We first give an example to demonstrate that not all pairs (G, c) are online extendible. Consider the ecologically coloured graph in Fig. 2.(i). Suppose that a further vertex is added as shown in Fig. 2.(ii). Its colourhood is $\{1, 3, 4\}$ so it must be coloured 2 to keep the colouring ecological (since there is already a vertex with that colourhood). Finally suppose that a further vertex is added as shown in Fig. 2.(iii). Its colourhood is $\{2, 3, 4\}$ so it must be coloured 1. But now the two vertices of degree 2 have the same colourhood but are not coloured alike so the colouring is not ecological.

We also give an example of a pair (G, c) that is online extendible but for which we cannot use FIRST-FIT or k-FIRST-FIT. Let G be the path $v_1v_2v_3v_4$ on four vertices

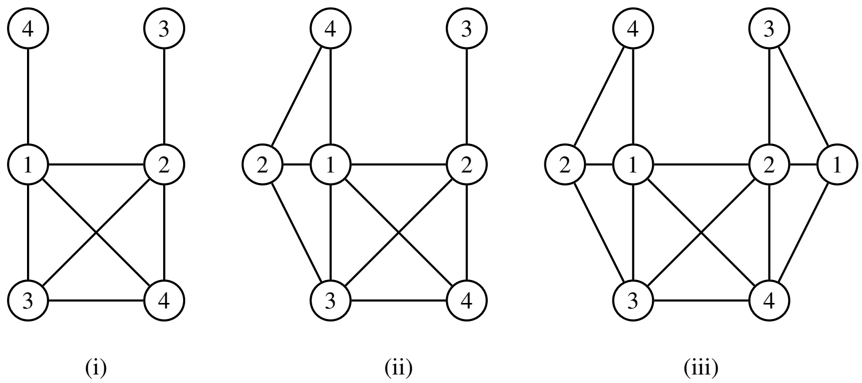

Fig. 2. A pair (G, c) that is not online extendible.

coloured $abcd$. We will show in Theorem 3 that (G, c) is online extendible (even if we are forced to use only colours from $\{a, b, c, d\}$). However, FIRST-FIT or k-FIRST FIT (arbitrary k) cannot be used with any ordering of $\{a, b, c, d\}$. To see this, add a new vertex adjacent to v_1 and v_3. Any correct online colouring algorithm must colour it b. So if the algorithm is FIRST-FIT, b is before d in the ordering of the colours. Next add a new vertex adjacent to v_3. If this vertex is not coloured d then the colouring will not be ecological, but FIRST-FIT will not use d as b (or possibly a) is preferred.

2.2 Rules

Let us now describe our general approach for obtaining online ecological colourings when they exist. As before, let G be a graph with an ecological colouring c and let C be a set of colours where $C \supseteq c(V)$. What we would like to do is to write down a set of rules: for each subset $A \subseteq C$, a colour x should be specified such that whenever a vertex is added and its colourhood is exactly A, we will colour it x. We would like to construct a fixed set of rules that, when applied, always yields an ecological colouring. There is no reason a priori that a general online colouring algorithm should be based on a set of rules. We show in Theorem 1 though that every online *ecological* colouring algorithm can be simply assumed to follow a set of rules. However, finding such a set turns out to be nontrivial. We start by making the following definitions.

(i) A *rule* is a pair containing a set of colours A and a colour x and is denoted $A \to x$.
(ii) A rule $A \to x$ *represents* a vertex v in G if v has colourhood A and $c(v) = x$.
(iii) The set of rules that represent each vertex of G is said to be *induced* by (G, c) and is denoted $R_{(G,c)}$.
(iv) A set of rules R is *valid* for (G, c) if $R \supseteq R_{(G,c)}$ and for each pair of distinct rules $A \to x$, $B \to y$ in R, we have $A \neq B$.
(v) A set of rules R is *full* on a set of colours C if for every $A \subseteq C$, there is exactly one colour $x \in C$ such that $A \to x$ is in R.

We will sometimes refer to a set of rules R *on* C meaning that for each rule $A \to x$ of R, we have $A \subseteq C$ and $x \in C$.

Notice that a full set of rules R constitutes an online colouring algorithm: if v is a newly revealed vertex with colourhood A and $A \to x$ is the unique rule for A in R, then v is coloured x by R. Notice also that the k-FIRST-FIT algorithm can be written down as the full set of rules

$$R^k_{FF} = \{A \to \min\{y \geq 1 \mid y \notin A\} \mid A \subset \{1, \ldots, k\}\} \cup \{\{1, \ldots, k\} \to k\},$$

that is, the k-FIRST-FIT algorithm assigns colours to new vertices purely as a function of their colourhoods. In this way, the notion of rules generalises FIRST-FIT.

While a full set of rules R gives an online colouring algorithm, it does not guarantee that each colouring will be ecological. For this, we must impose conditions on R. The following observation, which follows trivially from the definitions, shows that having a valid set of rules for a coloured graph ensures that it is ecologically coloured. We state the observation formally so that we can refer to it later.

Observation 1 *Let $G = (V, E)$ be a graph with colouring c. Let R be a valid set of rules for (G, c). Then c is an ecological colouring of G.*

Proof. Suppose c is not ecological. Then there are two vertices coloured x and y, $x \neq y$, which both have colourhood $A \subseteq c(V)$. Then the set of rules induced by (G, c) contains two rules $A \to x$ and $A \to y$. Since R is valid for (G, c), it must contain the rules induced by (G, c), but this contradicts that R must contain at most one rule for each set A. □

Note that if we have a valid and full set of rules R on C for (G, c) and further vertices are added and coloured according to the rules, R might not necessarily remain valid for the new graph, that is, R might not be a superset of the induced rules for the new graph. Let us see what might happen. Suppose that a new vertex u is added such that its colourhood is B and that, according to a rule $B \to y$ in R, it is coloured y. Now consider a neighbour v of u. Suppose that it had been coloured x at some previous stage according to a rule $A \to x$ in R. But now the colour y has been added to its colourhood. So R is valid for the altered graph only if it contains the rule $A + y \to x$.

This motivates the following definition. Let R be a set of rules. We say that R is a *good* set of rules if for all sets A, B and all colours x, y the following holds:

if $(A \to x) \in R$ and $(B \to y) \in R$ and $x \in B$ then $(A + y \to x) \in R$.

It is an easy exercise to check that the rules R^k_{FF} for k-FIRST-FIT are good.

2.3 Conditions for obtaining online extensions

We are now able to characterize when a pair (G, c) is online extendible.

Theorem 1. *Let $G = (V, E)$ be a graph with ecological colouring c. Then (G, c) is online extendible if and only if there exists a set of rules R such that* *(i)* R *is valid for* (G, c)*, and*

(ii) R *is good.*

Furthermore (G, c) *is online extendible with a finite set* C *if, in addition to (i) and (ii),*

(iii) R *is full on* C'*, where* C' *is a set of colours satisfying* $C \supseteq C' \supseteq c(V)$*.*

The purpose of C' in the statement of Theorem 1 is to account for the possibility that some of the colours of C may, under all circumstances, not be required.

Proof. The theorem describes two results corresponding to the fact that, when seeking online extensions, we may or may not specify the set of colours C to be used. The only difference is the extra condition (iii) in the case where the set is given. The two results can be proved together if we take care to note that, throughout, a set C may or may not have been specified.

($\Longrightarrow$) We assume that (G, c) is online extendible, and show that this implies (i) and (ii), and, also, that if the extension can be made using colours from a given finite set C, then (iii) is satisfied.

Assuming that (G, c) is online extendible, there exists, by definition, an algorithm α that can be used to obtain an ecological colouring of any graph constructed by adding vertices to G. We shall show that, by carefully choosing how to add vertices to G and colouring them with α, we obtain a graph which induces a set of rules that is valid for (G, c), good and, if C is given, full on some set $C' \subseteq C$.

First, we describe one way in which we add vertices. If a graph contains a vertex u coloured x with colourhood A, then the set of rules induced by the graph includes $A \to x$. To *protect* that rule means to add another vertex v with the same neighbourhood (and thus also the same colourhood) as u, to colour it x (as any correct algorithm must), and to state that no further vertices will be added that are adjacent to v. Hence all future graphs obtained by adding additional vertices will also induce the rule $A \to x$.

We use this method immediately: we protect each of the rules induced by (G, c). In this way, we ensure that the set of induced rules for any future graph is valid for (G, c).

While α is being applied, let G^* denote the current graph. If C is given, then let C^* be the set of colours used on G^*; otherwise let $C^* = c(V)$. Let R^* be the set of rules induced by G^*. As long as R^* is not full on C^*, we add a new vertex as follows:

> Let $B \subseteq C^*$ be a set for which R^* does not contain a rule. Add to G^* a new vertex u with colourhood B and use the algorithm α to obtain an ecological colouring. Add vertices to protect any rule induced by the new graph not in R^*.

Note that it is possible to add such a vertex u without making it adjacent to vertices that have been used for protection (so the rules induced by the new graph are a superset of the rules induced by G^*). There is at least one rule induced by the new graph not induced by the previous graph, namely $B \to y$, where y is the colour α assigns to u. So, if we continue in this way, the number of rules will increase. In the case that C is given, a full set of rules R_F will eventually be obtained for some set $C' \subseteq C$. In the case that C is not given and $C^* = c(V)$), we will eventually obtain a set of rules that contains a rule for every subset of $c(V)$; in this case, let R_F be the subset of those rules that contains one rule $A \to x$ for each $A \subseteq c(V)$. In each case, let G_F be the final graph obtained.

In each case, R_F is valid for (G, c) and if C is specified it is full for C'. Thus it only remains to prove that R_F is good. If the rules are not good, then they include rules $A \to x$, $B \to y$, $A + y \to z$ such that $x \in B$ and $x \neq z$. Let u be a vertex in G_F coloured x with colourhood A. Choose a set of vertices $S \ni u$ with $c(S) = B$ such that each vertex, except possibly u, is not one that was created to protect a rule. Add a new vertex adjacent to the vertices of S. This must be coloured y by α. But now the neighbourhood of u is $A + y$ and the colouring is not ecological (since no vertex has been added adjacent to the vertex protecting the rule $A + y \to z$); this contradicts the definition of α.

($\Longleftarrow$) Suppose that R is a set of rules that is valid for (G, c), good, and, if C is given, is full on C' where $C \supseteq C' \supseteq c(V)$. We must show that this implies that (G, c) is online extendible (with C). That is, we must describe how to colour vertices as they are added to G in such a way that all colourings obtained are ecological.

Let $G_0 = G$, and suppose that graphs $G_1, G_2, \ldots, G_r$ are obtained by the successive addition of vertices $v_1, v_2, \ldots, v_r$ (that is, G_i is obtained from G_{i-1} by adding the vertex v_i and some incident edges). We will describe how to colour each v_i to obtain a colouring c_i of G_i, and also describe a set of rules R_i that is good and valid for (G_i, c_i). By Observation 1, this will prove that c_i is an ecological colouring of G_i.

If C is given, then we simply let each $R_i = R$ and colour each new vertex using R; this can be done since R is full on $C' \subseteq C$. We note that each R_i is good and will prove validity below. If C is not given, we colour v_i and construct R_i as follows. Set $R_0 = R$, which we know is a set of rules that is valid for (G_0, c) and good. Given (G_{i-1}, c_{i-1}) and the set of rules R_{i-1}, let S be the colourhood of v_i in G_{i-1}. If R_{i-1} contains a rule for S, use that rule to colour v_i and set $R_i = R_{i-1}$. If no rule exists for S, colour v_i with a new (that is, as yet unused) colour k, and set

$$R_i = R_{i-1} \cup \{S \to k\} \cup \{(T + k \to z) | \, (T \to z) \in R_{i-1}, z \in S\}. \qquad (1)$$

Note that R_i is a set of rules on $c_i(V_i)$ and that R_i contains at most one rule for each subset of $c_i(V_i)$ (since each of the added rules involves a set that does not have a rule in R_{i-1}).

We show that R_i is good, assuming by induction that R_{i-1} is good. If $R_i = R_{i-1}$, then R_i is good, so we may assume that R_i is constructed according to (1). Suppose that $(A \to x) \in R_i$ and $(B \to y) \in R_i$ with $x \in B$. We show by case analysis that $(A + y \to x) \in R_i$.

Case 1: $x = k$
By the definition of R_i, if $x = k$ then $A = S$. Also, using the definition of R_i, if $x = k \in B$ and $(B \to y) \in R_i$, then $y \in S$. Thus $(A + y \to x) \in R_i$.

Case 2: $x \neq k, y = k$
By the definition of R_i, if $y = k$ then $B = S$. Also, using the definition of R_i, since $x \in S = B$ and $(A \to x) \in R_i$ then $(A + k \to x) \in R_i$, that is, $(A + y \to x) \in R_i$.

Case 3: $x \neq k, y \neq k$
Since $x, y \neq k$, then $A \neq S$ and $B \neq S$. By the definition of R_i, since $(A \to x) \in R_i$

then $(A-k \to x) \in R_{i-1}$. Similarly, since $(B \to y) \in R_i$, then $(B-k \to y) \in R_{i-1}$. Since $x \neq k$, then $x \in B-k$. Since R_{i-1} is good, we have that $(A+y-k \to x) \in R_{i-1}$. If $k \notin A$, then the previous statement gives precisely that $(A+y \to x) \in R_{i-1} \subset R_i$ as required. If $k \in A$, then since $(A \to x) \in R_i$, we have $x \in S$ (using the definition of R_i). Thus from our deduction that $(A+y-k \to x) \in R_{i-1}$ and $x \in S$, we see that $(A+y \to x) \in R_i$ (from the definition of R_i).

Finally, we show that R_i is valid for (G_i, c_i) (in both the case where C is unspecified and each R_i is as constructed just above, and also the case where C is given and each $R_i = R$, the given set of rules). We know that R_i contains at most one rule $A \to x$ for each set A, so we need only to show that R_i contains the rules induced by (G_i, c_i). Assume, for induction, that R_{i-1} is valid for (G_{i-1}, c_{i-1}). Let B be the colourhood of v_i in G_i. Now R_i contains a rule for B, say $B \to y$, by which v_i is coloured y, giving the colouring c_i of G_i. We must check that there are rules in R_i to represent each vertex of G_i. Clearly the rule $B \to y$ represents v_i. Let $v \neq v_i$ be a vertex of G_i. Suppose, as a vertex of G_{i-1}, it is represented by the rule $(A \to x) \in R_{i-1} \subset R_i$. If v is not adjacent to v_i, then in G_i, v is still represented by the rule $(A \to x) \in R_i$. If v is adjacent to v_i, then v is represented by the rule $A + y \to x$; this rule is present in R_i since R_i is good and contains the rules $A \to x$ and $B \to y$, where $x \in B$. □

3 Computational Complexity of Online Extendibility

We now consider the computational complexity of deciding whether or not (G, c) is online extendible.

Theorem 2. *Given an input graph $G = (V, E)$ with an ecological colouring c, the problem of deciding if (G, c) is ecologically online extendible is solvable in polynomial time.*

The theorem considers the case where the set of colours to be used is not specified. Whether or not (G, c) is online extendible with a fixed set C of cardinality ℓ is also solvable in polynomial time: enumerate all full sets of rules on all $C' \subseteq C$ and check whether they are good and valid for (G, c); the number of sets of rules to be checked depends only on ℓ and is independent of $|G|$, and checking a set of rules requires polynomial time. So far, we have not been able to prove the computational complexity if C is part of the input rather than fixed.

Proof. By Theorem 1, to decide whether (G, c) is online extendible, we must check whether there exists a set of rules that is valid for (G, c) and good. We successively reduce this condition to simpler but equivalent conditions, eventually reaching a condition that can be checked in time polyonomial in $|G|$.

We begin by observing that there exists a set of rules that is valid for (G, c) and good if and only if there exists a set of rules on $c(V)$ that is valid for (G, c) and good. In one direction this is obvious. For the other direction, assume we have a set of rules R that is valid for (G, c) and good. Note that R may contain rules $(A \to x)$ with $A \not\subseteq c(V)$. However, let $\hat{R}$ be the set of rules on $c(V)$ given by

$$\hat{R} = \{(A \to x) \in R : A \subseteq c(V), x \in c(V)\}.$$

It is easy to verify that $\hat{R}$ is good given that R is good.

Next we prove a useful lemma that gives equivalent conditions for goodness.

Lemma 1. *Let $G = (V, E)$ be a graph with ecological colouring c, and let R be a set of rules on $c(V)$ that is valid for (G, c). Then R is good if and only if the following conditions hold.*

(1) If $(A \to x) \in R$ with $y \in A$, then there exists $B \subseteq c(V)$ such that $(B \to y) \in R$ with $x \in B$.
(2) If $(A \to x) \in R$ and $(B \to x) \in R$, then $(A \cup B \to x) \in R$.
(3) If $(A \to x) \in R$ and $(B \to x) \in R$ and $A \subseteq X \subseteq B$, then $(X \to x) \in R$.

Proof. ($\Longrightarrow$) Suppose R is a set of rules on $c(V)$ that is valid for (G, c) and good.

To prove (1), suppose $y \in A$ and $(A \to x) \in R$. Since R is valid for (G, c), there exists B' such that $(B' \to y) \in R$ (for every colour z, there exists $Z \subseteq c(V)$ such that $(Z \to z) \in R$). Since R is good, $(B' + x \to y) \in R$, and setting $B = B' + x$ proves (1).

We prove (2) and (3) together by showing that if $(A \to x) \in R$ and $(B \to x) \in R$ and $y \in B$, then $(A + y \to x) \in R$. Thus adding the elements of B one by one to A yields both (2) and (3). If $(B \to x) \in R$ and $y \in B$, then by (1), there exists B' such that $(B' \to y) \in R$ and $x \in B'$. Then since R is good and $(A \to x) \in R$, $(B' \to y) \in R$, and $x \in B'$, we have $(A + y \to x) \in R$ as required.

($\Longleftarrow$) To prove the converse, we assume that R is valid for (G, c) and satisfies (1), (2), and (3), and we must show that R is good.

Suppose $(A \to x) \in R$ and $(B \to y) \in R$ with $x \in B$. Then by (1), there exists B' such that $(B' \to x) \in R$ and $y \in B'$. By (2), $(A \cup B' \to x) \in R$, and by (3), since $y \in B'$, $(A + y \to x) \in R$, as required. This completes the proof of Lemma 1 □

So far (using the above lemma), we have deduced that (G, c) is online extendible if and only if there exists a set of rules on $c(V)$ that is valid for (G, c) and satisfies conditions (1), (2), and (3) of Lemma 1. We reduce this condition further.

Let R be the set of rules induced by (G, c) and let R' be a set of rules on $c(V)$ that is valid for (G, c) and good (if such a set of rules exists). Clearly $R' \supseteq R$, and R' satisfies conditions (1), (2), and (3) of Lemma 1. For each $x \in c(V)$, let

$$A(x) = \{A | (A \to x) \in R\},$$

and for each $x \in c(V)$, let $M_x \subseteq c(V)$ be defined by

$$M_x = \bigcup A(x).$$

By condition (2) of Lemma 1, R' must contain the rule $M_x \to x$ for each $x \in c(V)$. For each $x \in c(V)$, let $R(x)$ be the set of rules defined by

$$R(x) = \{B \to x | A \subseteq B \subseteq M_x, A \in A(x)\}.$$

By condition (3) of Lemma 1, R' must contain all the rules in $R(x)$ for every $x \in c(V)$. Setting

$$R^* = \bigcup_{x \in c(V)} R(x),$$

we have that $R' \supseteq R^*$.

Notice that R^* satisfies conditions (2) and (3) of Lemma 1, by construction. One can check that R^* also satisfies condition (1) of Lemma 1 as follows. Suppose $(A \to x) \in R^*$ and $y \in A$. By construction of R^*, there exists $A' \in A(x)$ such that $y \in A'$ and $(A' \to x) \in R$. Since R is the induced set of rules for (G, c), there exists a vertex v of G coloured y with colourhood A'. Thus the neighbourhood of v contains a vertex w coloured y. If the rule induced by w is $B \to y$, then $x \in B$ (since v, coloured x, is in the neighbourhood of w) and $(B \to y) \in R \subseteq R^*$. Thus (3) holds.

Thus, if R^* is valid for (G, c), then R^* is good by Lemma 1; if R^* is not valid for (G, c) (the only way this can happen is if there is more than one rule for some set), then there is no set of rules that is good and valid for (G, c) because any such set of rules must contain R^*. Thus (G, c) is online extendible if and only if R^* is valid for (G, c).

We can determine in time polynomial in $|G|$ whether R^* is valid (without a priori constructing R^*). To see this, consider that the only way R^* is not valid is if there exist rules $A \to x$ and $A \to y$ in R^* with $x \neq y$. In that case, we must have

$$\begin{aligned} & M_x \supseteq A \supseteq A_x \quad \text{for some } A_x \in A(x), \\ \text{and} \quad & M_y \supseteq A \supseteq A_y \quad \text{for some } A_y \in A(y). \end{aligned}$$

The above holds if and only if $M_x \cap M_y \supseteq A_x \cup A_y$. Thus, for each pair of distinct colours x and y, and for each $A_x \in A(x)$ and $A_y \in A(y)$, we simply check whether $M_x \cap M_y \supseteq A_x \cup A_y$ (this can be done in time polynomial in n). If no such containment exists then R^* is good and valid for (G, c) and so (G, c) is online extendible; otherwise, there is no set of rules that is good and valid for (G, c) and so (G, c) is not online extendible. This completes the proof of Theorem 2. $\square$

4 Online Extensions of Twin-Free Graphs

We noted in the previous section that if we are given a graph $G = (V, E)$ ecologically coloured with c and a set of colours $C \supseteq c(V)$, we do not know the complexity of deciding whether (G, c) is online extendible with C (that is, whether (G, c) is online extendible with a specified number of extra colours). A natural question to start with is to consider the case in which all vertices of G have distinct colours. Thus we assume that G is twin-free else the colouring would not be ecological. Theorem 3 solves this case by showing that any such pair (G, c) is online extendible using one extra colour in addition to $c(V)$. We show in the second part of this theorem that the above is tight by characterizing those pairs (G, c) for which we always need the extra colour. The simple necessary and sufficient conditions in our characterization can easily be checked in polynomial time.

Theorem 3. *Let $G = (V, E)$ be a twin-free graph on k vertices and let c be a colouring of G with $|c(V)| = k$ (thus c is an ecological colouring of G).*

1. *(G, c) is online extendible with $c(V)$ and one extra colour.*
2. *(G, c) is online extendible with $c(V)$ if and only if G contains a vertex u^* such that*
   (i) the neighbourhood of u^* is maximal in G, that is $N(u^*)$ is not a proper subset of $N_G(v)$ for all $v \in V$, and

(ii) the graph $G - u^*$ is twin-free.

The smallest twin-free graph that does not satisfy the two conditions (i) and (ii) in Theorem 3 is a graph on two components, one of which is an isolated vertex and the other is an edge. The smallest connected twin-free graph that does not satisfy these two conditions is obtained from a complete graph on four vertices u_1, u_2, u_3, u_4 after adding two new vertices v_1, v_2 with edges (u_1, v_1), (u_2, v_1), (u_3, v_2), (u_4, v_2) and (v_1, v_2). We can construct an infinite family of such examples as follows. Take two disjoint copies, H and H', of the complete graph K_{2n} on $2n$ vertices with a perfect matching removed. Let $(v_1, w_1), (v_2, w_2), \ldots, (v_n, w_n)$ be the perfect matching removed from H, and let $(v_1', w_1'), (v_2', w_2'), \ldots, (v_n', w_n')$ be the perfect matching removed from H'. Let G be the graph obtained by adding the edges $(v_1, v_1'), (v_2, v_2'), \ldots, (v_n, v_n')$ to $H \cup H'$. Clearly, the vertices with maximal neighbourhoods are $v_1, \ldots, v_n, v_1', \ldots, v_n'$, but removing v_i (resp. v_i') from G leaves twins v_i', w_i' (resp. v_i, w_i).

Proof. We restate that G is a twin-free graph on k vertices and that c is an ecological colouring of G with $|c(V)| = k$. Define $C := c(V) = \{1, 2, \ldots, k\}$. To prove each part of the theorem, we must find a valid, good, full set of rules R for (G, c). We know that R must contain rules that represent each vertex of G; we must describe how to define the remaining rules. Here is a useful technique.

Let $\mathcal{A}$ contain the subsets $A \subseteq C$ for which $R_{(G,c)}$ contains a rule involving A. To *propagate* $R_{(G,c)}$ apply the following to obtain a set of rules $R^*_{(G,c)}$:

- for each $1 \leq i \leq |C|$, fix an ordering for the collection of sets in $\mathcal{A}$ of cardinality i;
- for each subset $A \subseteq C$, let, if possible, A^* be the smallest member of $\mathcal{A}$, first ordered, that is a superset of A (possibly $A^* = A$). If A^* exists and $A^* \to x$ is a rule in $R_{(G,c)}$, then add $A \to x$ to $R^*_{(G,c)}$.

We make two claims. The first is a simple observation.

Claim 1. We have that $R^*_{(G,c)}$ is valid for (G, c). Furthermore $R^*_{(G,c)}$ is a full set of rules on C if and only if $R_{(G,c)}$ contains a rule $C \to x$ for some x.

Claim 2. We have that $R^*_{(G,c)}$ is good.

We prove Claim 2 as follows. If $R^*_{(G,c)}$ is not good, then there are rules $A \to x$ and $B \to y$ in $R^*_{(G,c)}$, where $x \in B$, but $A + y \to x$ is not in $R^*_{(G,c)}$. By definition, $R_{(G,c)}$ contains a rule $A^* \to x$. Notice that A^* is the set of colours used on the neighbours of the vertex in G coloured x. Similarly $R_{(G,c)}$ must contain a rule $B^* \to y$, where $x \in B \subseteq B^*$ and B^* is the set of colours used on the neighbours of the vertex in G coloured y. So the vertices in G coloured x and y are adjacent and so $y \in A^*$. But then A^* contains $A + y$ so we must have $A^* = (A + y)^*$. Thus $A + y \to x$ is in $R^*_{(G,c)}$. This proves Claim 2.

We now prove the first part of the theorem. Let G' be obtained from G by adding a new vertex v^* adjacent to all existing vertices and to itself (we could avoid having a loop by adding two new vertices adjacent to every vertex in G and each other; but allowing the loop makes the analysis a little tidier). Colour v^* with colour $k + 1$ to obtain a colouring c' of G', and write $C' = \{1, \ldots, k + 1\}$. Note that G' is twin-free.

As $R^*_{(G',c')}$ contains a rule involving C', Claim 1 tells us that it is a full and valid set of rules on C' for (G',c'). By Claim 2, $R^*_{(G',c')}$ is also good. It remains only to show that $R^*_{(G',c')}$ is valid for (G,c).

Note that each vertex v of G has the colour $k+1$ in its G'-neighbourhood. Therefore, as a vertex of G', v is represented in $R_{(G',c')}$ by a rule $A+(k+1) \to x$ (where A is the set of colours in the G-neighbourhood of v). Observe that, since A^* is a minimal set containing A that is involved in a rule of $R_{(G',c')}$, and since all rules $B \to y$ in $R_{(G',c')}$ satisfy $k+1 \in B$, we have $A^* = A+(k+1)$. Thus $R^*_{(G',c')}$ contains the rule $A \to x$, which represents the vertex v of G. This is true for all vertices of G, and so $R^*_{(G',c')}$ is also valid for (G,c). Thus $R^*_{(G',c')}$ is a full set of rules on $C' = \{1,\dots,k+1\}$ that is good and valid for (G,c). Thus (G,c) is online extendible with $\{1,\dots,k+1\}$ by Theorem 1. This completes the proof of the first part of Theorem 3.

Now we prove the second part of the theorem.

($\Longrightarrow$) We begin by showing that if G contains a vertex u^* such that $G-u^*$ is twin-free and the neighbourhood of u^* in G is maximal (that is, it is not a proper subset of the neighbourhood of another vertex in G), then (G,c) is online extendible with $c(V) = C = \{1,\dots,k\}$. If we can construct a full set of rules on C that is good and valid for (G,c) then we are done by Theorem 1.

We may assume that u^* is coloured k. Let G' be obtained from G by adding edges to G so that u^* is adjacent to every vertex in G, including itself. Note that G' is twin-free: u^* is the only vertex adjacent to every vertex in the graph and if two other vertices both have neighbourhoods $A+u^*$, then in G one must have neighbourhood A and the other $A+u^*$, contradicting that $G-u^*$ is twin-free.

Let $R^*_{(G',c)}$ be obtained from $R_{(G',c)}$ by propagation. As $R_{(G',c)}$ contains the rule $C \to k$, we have that $R^*_{(G',c)}$ is a full set of rules on C that is valid for (G',c) by Claim 1 and that is good for (G',c) by Claim 2.

It remains only to show that $R^*_{(G',c)}$ is valid for (G,c). Note that for each vertex $v \neq u^*$ of G, if $c(N_G(v)) = A$, then $R_{(G',c)}$ contains the rule $A+k \to x$. Also $R^*_{(G',c)}$ contains the rule $A \to x$ as $A^* = A+k$ (since A^* is a minimal superset of A and must contain k). In G, the set of colours in the neighbourhood of v is either A or $A+k$; in either case there is a rule in $R^*_{(G',c)}$ to represent it.

Let B be the colours in the neighbourhood of u^* in G. Then $B^* = C$ as, by the maximality of B, there is no other superset of B involved in a rule of $R_{(G',c)}$. Since $R_{(G',c)}$ contains $C \to k$, $R^*_{(G',c)}$ contains $B \to k$ which represents u^*. So $R^*_{(G',c)}$ is valid for (G,c) as required.

($\Longleftarrow$) Suppose that for every vertex u^* of G, either $G-u^*$ is not twin-free or the neighbourhood of u^* in G is not maximal. We show that (G,c) is not online extendible with $C = \{1,\dots,k\}$.

Suppose, for a contradiction, that there is an online algorithm to extend (G,c). Add vertex v to G adjacent to all vertices in G to form G_1. Without loss of generality, our algorithm assigns colour k to v to give us a colouring c_1 of G_1. Let u be the vertex of $G_0 := G$ that is coloured k. There are two cases to consider: either $G_0 - u$ is not twin-free or $N_{G_0}(u)$ is not maximal.

Suppose $G_0 - u$ has twins, that is two vertices a and b with the same neighbourhood (in $G_0 - u$). The colouring $c_0 = c$, and therefore c_1, colours a and b differently; however we have $c_1(N_{G_1}(a)) = c_1(N_{G_1}(b))$, a contradiction.

Suppose $N_{G_0}(u)$ is not maximal; suppose $S = N_{G_0}(u)$ and $T = N_{G_0}(u')$, where $T = S \cup \{t_1, \ldots, t_r\}$. Let $N_i = N_{G_0}(t_i)$. (Note that $r \neq 0$ since $G_0 = G$ is twin-free.) Add vertices $w_1, \ldots, w_r$ to G_1 one at a time, where w_i is adjacent to each vertex in $N_i \cup \{u\}$. Our online algorithm is forced to assign the colour of t_i to w_i (since they have the same colours in their neighbourhoods). Let G_{r+1} be the graph obtained after addition of $w_1, \ldots, w_r$ and let c_{r+1} be its colouring. In (G_{r+1}, c_{r+1}), we find that u and u' have the same set of colours in their neighbourhoods but are coloured differently (since they were coloured differently by c_0). This is a contradiction. □

Now we show that the two online algorithms implied by statements 1 and 2 of Theorem 3-1, respectively, run in polynomial time. Let G be a twin-free graph with ecological colouring c. We follow the proof of Theorem 3. The procedure is very similar for each of the two algorithms. First we amend G to obtain a graph G'; in the first algorithm we also amend the colouring c to obtain c'. Then we propagate $R_{(G',c')}$ (in fact, $R_{(G',c)}$ for the second algorithm). Next, when a new vertex v_i is presented and needs to be coloured, we first determine the set of colours A in the neighbourhood of v_i. Next we compute the corresponding set A^*, which is the smallest and first ordered member of $\mathcal{A}$ with $A \subseteq A^*$; note that the set A^* exists due to the addition of vertex v^* in the first algorithm and modification of the vertex u^* in the second algorithm, respectively. This gives us the rule by which v_i should be coloured. The total time used is polynomial in the size of G.

5 Ecological H-Colouring

In this section we classify the computational complexity of the problem ECOLOGICAL H-COLOURING for all fixed target graphs H. Before doing this, we must introduce some further terminology. Given a graph H on k vertices, we define the *product graph* H^k. The vertex set of H^k is the Cartesian product

$$V_{H^k} = \underbrace{V_H \times \cdots V_H}_{k \text{ times}}.$$

Thus a vertex u of H^k has k *coordinates* u_i, $1 \leq i \leq k$, where each u_i is a vertex of H (note that these coordinates of u need not be distinct vertices of H). The edge set of H^k, E_{H^k}, contains an edge (u, v) in E_{H^k} if and only if, for $1 \leq i \leq k$, there is an edge (u_i, v_i) in H. For $1 \leq i \leq k$, the *projection* on the ith coordinate of H^k is the function $p_i : V_{H^k} \to V_H$ where $p_i(u) = u_i$. It is clear that each projection is a graph homomorphism.

Theorem 4. *If H is bipartite or contains a loop, then* ECOLOGICAL H-COLOURING *is in P. If H is not bipartite and contains no loops, then* ECOLOGICAL H-COLOURING *is* NP*-complete.*

Proof. The first sentence of the theorem is an easy observation which we briefly justify. If H has no edges, then G has an ecological H-colouring if and only if G has no edges. Suppose H is bipartite and contains at least one edge (x, y). If G is bipartite, then we can find an ecological H-colouring by mapping each vertex of G to either x or y. If G is not bipartite then it is clear that there is no homomorphism from G to H. If H has a loop, then any graph has an ecological H-colouring since we can map every vertex to a vertex with a loop.

We prove that the ECOLOGICAL H-COLOURING problem is NP-complete for loopless non-bipartite H by reduction from H-COLOURING which is known to be NP-complete for loopless non-bipartite H [9].

Let G be an instance of H-colouring and let n be the number of vertices in G. Let k denote the number of vertices in H_N, the neighbourhood graph of H (recall that the neighbourhood graph of H is a graph in which each vertex has a unique neighbourhood and is obtained from H by repeatedly deleting one vertex from any pair with the same neighbourhood). Let π denote a vertex in H_N^k whose k coordinates are the k distinct vertices of H_N (the order is unimportant). Let G' be a graph formed from G and n copies of H_N^k by identifying each vertex u of G with a distinct copy of the vertex π; see Fig. 3. We can distinguish the copies of H_N^k by referring to the vertex of G to which they are *attached*.

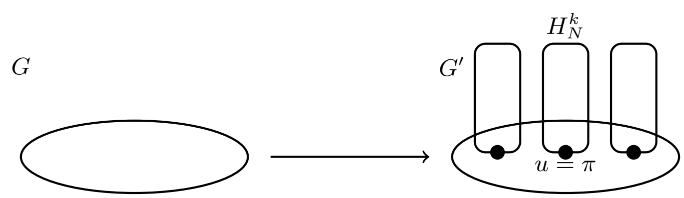

Fig. 3. The graph G' formed by attaching G to copies of H_N^k.

We claim that G has an H-colouring if and only if G' has an ecological H_N-colouring which is clearly equivalent to G' having an ecological H-colouring. As it is clear that if G' has an ecological H_N-colouring, the restriction to V_G provides an H-colouring for G, all we need to prove is that when G has an H-colouring, we can find an ecological H_N-colouring for G'.

If G has an H-colouring, then clearly it also has an H_N-colouring f. We use f to find an ecological H_N-colouring g for G'. For each vertex $u \in V_G$, $f(u) = \pi_i$ for some i (this is possible because of the choice of π as a vertex that has each vertex of H_N as a coordinate). For each vertex v in the copy of H_N^k attached to u, let $g(v) = p_i(v)$. Note that $g(u) = p_i(u) = \pi_i = f(u)$ for each vertex u in V_G.

Certainly g is an H_N-colouring: the edges of E_G are mapped to edges of H_N since g is the same as f on V_G, and the edges of each copy of H_N^k are mapped to edges of H_N as g is the same as one of the projections of H_N^k on these edges.

We must show that it is ecological; that is, for each pair of vertices s and t in G', we must show that

$$g(N_{G'}(s)) = g(N_{G'}(t)) \implies g(s) = g(t). \tag{2}$$

Suppose that $g(N_{G'}(s)) = g(N_{G'}(t))$. We know that $g(s) = p_i(s) = s_i$ for some value of i. Then for each $x \in N_{H_N}(s_i)$, there is a vertex $s' \in N_{G'}(s)$ with $g(s') = x$ (since we can choose as s' a vertex in the same copy of H_N^k as s with $s'_i = x$ and s'_j being any neighbour of s_j, $1 \leq j \leq k$, $j \neq i$).

Thus $g(N_{G'}(s)) \supseteq N_{H_N}(s_i)$ and so, since g is an H_N-colouring, $g(N_{G'}(s)) = N_{H_N}(s_i)$ and then, by (2), $g(N_{G'}(t)) = N_{H_N}(s_i)$. But as the neighbourhoods of vertices in H_N are distinct, we must have $g(t) = s_i = g(s)$. This completes the proof of Theorem 4. $\square$

6 Conclusions and Open Problems

Let $G = (V, E)$ be a graph with ecological colouring c. In the first part of our paper, we showed that checking whether a pair (G, c) is online extendible can be done in polynomial time. We also showed that checking if (G, c) is online extendible with some fixed finite set $C \supseteq c(V)$ is also polynomial-time solvable (although the time taken is a very large function of $|C|$). Determining the computational complexity of this problem when C is part of the input remains an open problem. We obtain a positive result when considering pairs (G, c) in which each vertex of the k-vertex graph G has a distinct colour. For such (G, c), we can check in time polynomial in k if (G, c) is online extendible with any $C \supseteq c(V)$. Indeed, we found that if $|C| = k + 1$, then (G, c) is always online extendible with C, and there are infinitely many examples of (G, c) that are not online extendible with C when $|C| = k$. It would be interesting to know whether there are examples (G, c) that are online extendible with an infinite number of colours but not with a finite number. An answer to this question is also relevant for the problem of finding a good set of rules R that is valid for some pair (G, c) should (G, c) be online extendible. The complexity status of the latter problem is also still open.

In the second part of our paper we gave a complete complexity classification of ECOLOGICAL H-COLOURING. We showed in Theorem 4 that if H is bipartite or contains a loop then ECOLOGICAL H-COLOURING is polynomial-time solvable, and is NP-complete otherwise. Theorem 4 implies that SURJECTIVE ECOLOGICAL H-COLOURING is NP-complete if H is a neighbourhood graph that is neither bipartite nor contains any loops; we can reduce from the corresponding ECOLOGICAL H-COLOURING problem by adding a disjoint copy of H to the given graph G. Moreover, SURJECTIVE ECOLOGICAL H-COLOURING has no yes-instances if H is not a neighbourhood graph. Hence, we are left with the case in which H is a neighbourhood graph that is bipartite or that contains at least one loop. We leave solving this case for future research. We note, however, that classifying the complexity of the SURJECTIVE H-COLOURING problem that asks if a graph allows a surjective homomorphism to a fixed graph H is a notoriously difficult open problem [1].

References

1. M. Bodirsky, J. Kára and B. Martin, The complexity of surjective homomorphism problems – a survey, *Discrete Applied Mathematics* **160** (2012) 1680–1690.
2. S. P. Borgatti and M. G. Everett, Graph colorings and power in experimental exchange networks, Social Networks 14 (1992) 287–308.
3. S.P. Borgatti, M.G. Everett, Ecological and perfect colorings, Social Networks 16 (1994) 43–55.
4. R.S. Burt, STRUCTURE Version 4.2 Reference Manual, Center for the Social Sciences, Columbia University, New York, 1991.
5. P. Crescenzi, M. Di Ianni, F. Greco, G. Rossi and P. Vocca, On the existence of Ecological Colouring, Proc. of the 34th International Workshop on Graph-Theoretic Concepts in Computer Science, Lecture Notes in Computer Science 5344, Springer, Berlin (2008), pp. 90–100.
6. E.H. Erickson, The relational basis of attitudes, Social Structures: A Network Approach, Cambridge, Cambridge University Press, pp. 99–121, 1988.
7. M.G. Everett and S. P. Borgatti, Role colouring a graph, Mathematical Social Sciences 21 (1991) 183–188.
8. A. Gyárfás and J. Lehel, On-line and first-fit colorings of graphs. *Journal of Graph Theory 12* (1988), 217–227.
9. P. Hell and J. Nešetřil, On the complexity of H-colouring, Journal of Combinatorial Theory, Series B 48 (1990) 92–110.
10. N. Hummon and K. Carley, Social networks as normal science, Social Networks 15 (1993), 71–106.
11. H.A. KIERSTEAD, Coloring graphs on-line. In: *Online algorithms: the state of the art* (1998), Lecture Notes in Computer Science 1442, Springer Verlag, pp. 281–305.