On the Existence of Ecological Coloring*

Pilu Crescenzi[1], Miriam Di Ianni[2], Federico Greco[3], Gianluca Rossi[2], and Paola Vocca[4]

[1] Dipartimento di Sistemi e Informatica, Università di Firenze, Firenze, Italy
piluc@dsi.unifi.it
[2] Dipartimento di Matematica, Università di Roma "Tor Vergata", Roma, Italy
{miriam.di.ianni,gianluca.rossi}@uniroma2.it
[3] Dipartimento di Matematica, Università di Perugia, Perugia, Italy
greco@dipmat.unipg.it
[4] Dipartimento di Matematica "Ennio De Giorgi", Università del Salento, Lecce,Italy
paola.vocca@unile.it

Abstract. In this paper, we study the problem of ecologically coloring a graph. Intuitively, an ecological coloring of a graph is a role assignment to the nodes of the graph, such that two nodes surrounded by the same set of roles must be assigned the same role (Borgatti and Everett, 1992). We prove that, for any graph G with n_G distinct neighborhoods and for any integer k with $1 \leq k < n_G$, G admits an ecological coloring which uses exactly k roles, and that this coloring can be computed in polynomial time. Our result strongly contrasts with the NP-completeness result of the regular coloring problem, where it is required that two nodes with the same role must be surrounded by the same set of roles (Fiala and Paulusma, 2005). Hence, we conclude that not only the ecological coloring is easier to understand as a model of social relationships (Borgatti and Everett, 1994), but it is also feasible from a computational complexity point of view.

1 Introduction

One of the main goals of the analysis of a social network consists of determining patterns of relationships and interactions among social actors (such as persons and groups) in order to identify the social structure of the network [BE05,B04]. To this aim, a social network is usually represented as a graph, whose nodes denote the network members and whose edges denote their relationships, which is analyzed from a structural point of view by means of methods that broadly fall into one of the following two categories: *relational analysis* methods that are often used in order to identify central members or to partition the graph into clusters, and *positional analysis* methods that examine the similarity between the connection of two network members with the other members. *Role assignment* is one of the main positional analysis methods, whose goal consists of classifying the members of a social network, so that members which are equally classified can be considered to behave in a similar way or to play a similar role. If the number of roles is limited, this kind of classification can turn out to be extremely useful while trying to understand the overall structure of very complex social networks.

Different kinds of role assignment have been introduced in the literature. A *strong structural* role assignment, for example, imposes that if two actors play the same role, then they must have the same neighborhood [LW71], while a *regular* role assignment imposes that if two actors play the same role, then they must be surrounded by the same set of roles [WR83,BE89]. In this paper we are interested in another kind of role assignment, that is, *ecological* role assignments, according to which if two netowrk members are surrounded by the same set of roles, then they must play the same role: in other words, the role played by a social actor is completely determined by the roles played by its neighbors [BE92]. The relationship imposed by an ecological role assignment is the opposite of the one imposed by a regular role assignment: a

* This work was partially supported by the EU under the EU/IST Project 15964 *AEOLUS*.

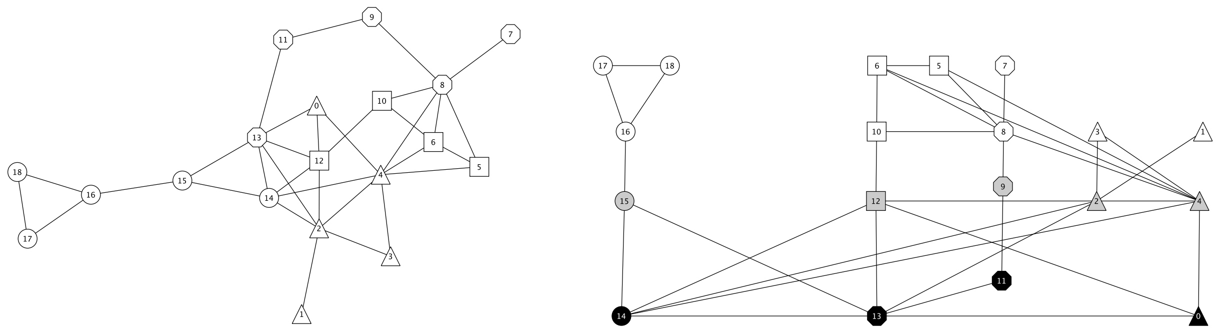


Fig. 1. The network of the 19 hijackers on September 11, 2001 (left) and an ecological coloring of the network itself (right)

role assignment that satisfies both constraints is called *perfect* [BE94]. As stated by Borgatti and Everett, an ecological role assignment is *easier to understand* as a model of social relationships, in which a member's neighborhood tends to shape this member into that or this kind.

As an example of application of the ecological role assignment method, let us consider the terrorist network containing the 19 hijackers that participated to the events of September 11, 2001 [K02,XC05a,XC05b]: this network is depicted in the left part of Figure 1, where different node shapes correspond to different flights (see [K02], where a more complicated network is also analyzed, which includes other terrorists who did not get on the planes). The right part of the figure shows an ecological role assignment of the terrorist network that uses three roles (represented by three different colors): it is interesting to observe that this role assignment clearly partitions the network into three groups and that one of these three groups (that is, the gray one) acts as an interface between the other two groups. It also worth noting that this assignment is clearly not strong structural, it is not regular (for instance, node 16 and node 17 play the same role but are surrounded by two different set of roles) and that it does not even correspond to a normal coloring of the network (since several members of the network playing the same role are connected to each other). Indeed, it is easy to prove that this network is not three-colorable, since it contains K_4 (see nodes 4, 5, 6 and 8); moreover, by exhaustive search it is also possible to show that this network does not admit any regular role assignment which uses exactly three roles. In other words, the ecological role assignment is the *only positional method* that can be applied to this network, if we insist on requiring that the number of roles is exactly three.

It should now be clear that computing a role assignment for a given network is equivalent to computing a coloring of the network's nodes, such that the constraint imposed by the role assignment is satisfied by the colors assigned to the nodes. For instance, an *ecological coloring* of a graph is an assignment of colors to the nodes of the graph such that if two nodes "see" the same set of colors, then they are assigned the same color. Our main contribution is proving that, *for any graph G with n_G distinct neighborhoods and for any integer k with $1 \leq k \leq n_G$, G admits an ecological coloring which uses exactly k colors*, and that *this coloring can be computed in polynomial time* by means of a bottom-up approach and by making use of some combinatorial properties of graphs whose nodes have all distinct neighborhoods.

Our results strongly contrast with the results obtained in [RS01,FP05] according to which deciding whether a graph can be regularly colored with k colors is NP-complete, for any $k \geq 2$. This somehow implies that an ecological role assignment not only is easier to understand but it is also more useful from a computational complexity point of view than a regular role assignment, since it can be always efficiently computed. This allows the network analyzer to reduce the size of complex social networks to a desired size: this feature is extremely important in network analysis for which different aspects can be studied depending on network dimension (for example, the degree of relevance in information retrieval and the degree of relationship in e-communities discovery).

Making role assignment feasible is useful not only in the field of social network analysis, but also in the field of graph drawing. Indeed, once a role assignment r of a graph G has been computed, a graph drawer can focus on its corresponding *color graph* $C^{G,r}$, that is, the graph whose nodes denote roles and where there is an edge between two nodes u and v if and only if two adjacent nodes of G have been assigned role u and v, respectively. For example, the drawing of the terrorist network shown in the right part of Figure 1 is very natural once we realize that the color graph of the ecological role assignment is a path of three nodes with self-loops. Observe that if the number of roles is very small compared to the number of nodes in G, then the graph drawer is allowed to make use of any graph drawing algorithm (even a non-efficient one).

Another field of application of our feasibility result is in the field of *mobility models* for mobile ad hoc wireless networks (in short, *MANET*). In fact, considering that different nodes may move according to different mobility models and that the mobility behavior of a node may vary during time because of changes of its environment, nodes of a network can move according to mobility models that are determined by the roles played by the nodes themselves: these roles, in turn, can be determined by computing ecological role assignments of the graph induced by the communication network [BCDRV07]. Observe that prior applications of social network analysis to the development of MANET mobility models assume that the structure of the social network is known *a priori* and that this structure does not change over time [MHM07,MM07]: in the role assignment based approach, instead, the social network structure is determined by the topology of the MANET, which in turn changes over time due to the movement of the nodes.

The paper is structured as follows. In the rest of this section, we give some preliminary definitions and results concerning the ecological coloring of a graph. In Section 2 we introduce the notion of neighborhood distinct graph, we prove how we can restrict our attention to this kind of graphs and we show some interesting structural properties of these graphs. In Section 3 we prove our main result, that is, that any graph can be ecologically colored by using any reasonable number of colors. Finally, in Section 4 we conclude by stating our main open question.

1.1 Preliminaries

Given a graph $G = (V, E)$, for any node $u \in V$, $N(u)$ denotes the neighborhood of u. A *coloring* of G which uses k colors is a surjective function $r : V \to [k]$[5]. The *color graph* $C^{G,r} = ([k], E^{G,r})$ includes the edge (i, j) if and only there exist $u, v \in V$ such that $r(u) = i$, $r(v) = j$ and $(u, v) \in E$.

Given a coloring r of G which uses the k colors $\{c_1, \ldots, c_k\}$, the *colorhood* of a node $u \in V$ with respect to r is defined as the set

$$C_r(u) = \{c_i : i \in [k] \wedge \exists v \in N(u)[r(v) = c_i]\}$$

A coloring r of a graph $G = (V, E)$ is *regular* [BE89] if, for any $u, v \in V$,

$$r(u) = r(v) \Rightarrow C_r(u) = C_r(v)$$

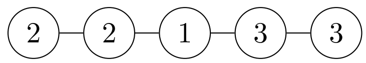
Observe that any graph with no isolated nodes can be regularly colored with one color and with n colors, where n is the number of nodes of the graph. The k-ReRA decision problem consists in deciding whether a graph G admits a regular coloring which uses k colors: in [RS01,FP05] it is proved that k-ReRA is NP-complete for any $k \geq 2$.

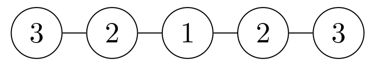
[5] In the following, for any positive integer n, $[n]$ will denote the set $\{1, 2, \ldots, n\}$.

A coloring r of a graph $G = (V, E)$ is *ecological* [BE92] if, for any $u, v \in V$,

$$C_r(u) = C_r(v) \Rightarrow r(u) = r(v)$$

Observe that any graph can be ecologically colored with one color. However, it is not true that any graph can be ecologically colored with n colors, where n is the number of nodes of the graph (see the results of Section 2). In general, an ecological coloring is not necessarily regular and a regular coloring is not necessarily ecological. The left part of the following picture shows an example of an ecological coloring of the path formed by five nodes which is not regular, while the right part shows an example of a regular coloring of the same graph which is not ecological.





The k-ECRA decision problem consists in deciding whether a graph G admits an ecological coloring which uses k colors. The next result shows that the complexity of ecologically coloring a graph is significantly different from the complexity of regularly coloring a graph.

Theorem 1. *Any graph with at least two nodes which is not an independent set can be ecologically colored with two colors.*

Proof. Let $G = (V, E)$ be a graph with $|V| > 1$ and let I be a maximal independent set of G. By coloring all nodes in I with color 1 and all nodes in $V - I$ with color 2 we obtain an ecological coloring of G. Indeed, for any node $u \in I$, $C_r(u) = \{2\}$ or $C_r = \emptyset$ while, for any node $u \notin I$, $C_r(u) = \{1\} \vee C_r(u) = \{1, 2\}$. $\square$

As a consequence of the above theorem, we have the following result which contrasts with the NP-completeness of the 2-RERA decision problem proved in [RS01].

Corollary 1. *The* 2-ECRA *decision problem belongs to* P.

2 Neighborhood Distinct graphs

According to the definition of an ecological coloring, two nodes with the same neighborhood must be colored with the same color. The number of distinct neighborhoods contained in a graph is thus an upper bound on the number of colors that can be used by any ecological coloring. The following definition and results formalize this statement.

Definition 1. *A graph $G = (V, E)$ is* neighborhood distinct *(in short,* ND*) if, for any $u, v \in V$, $N(u) \neq N(v)$.*

Even though the notion of neighborhood distinct graphs might seem quite natural, as far as we know no definition of these graphs has been given in the literature: the only definition similar to ours is the one given in [MMPW07] which concerns closed-neighborhood anti-Sperner graphs. Observe that any ND graph with n nodes can clearly be ecologically colored with n colors.

Given a graph $G = (V, E)$, we define an equivalence relation ρ_N on the vertices of G as follows: two vertices $u, v \in V$ are equivalent if and only if $N(u) = N(v)$. The *neighborhood graph* corresponding to G is a ND graph $G_N = (V_N, E_N)$ where V_N is the set of equivalence classes with respect to the relation ρ_N, and $(x, y) \in E_N$ if all nodes in the equivalence class x are adjacent to all nodes in the equivalence class y. The *neighborhood degree* n_G of G is defined as the number of nodes in G_N.

Theorem 2. *A graph can be ecologically colored with k colors if and only if its neighborhood graph can be ecologically colored with k colors.*

Proof. Let $G = (V, E)$ be a graph and let $G_N = (V_N, E_N)$ be its neighborhood graph. If r is an ecological coloring of G, then, for each pair of nodes $u, v \in V$ such that $N_G(u) = N_G(v)$, it must hold $r(u) = r(v)$: this implies that all nodes in the same equivalence class have been assigned the same color. Hence, by assigning to each equivalence class $x \in V_N$ the color $r(u)$ with $u \in x$, we obtain an ecological coloring of G_N which uses the same number of colors as r. Conversely, if G_N can be ecologically colored with k colors, then assigning to each node $u \in V$ the color of the node corresponding to the equivalence class u belongs to yields an ecological coloring of G with k colors. The lemma is thus proved. $\square$

Corollary 2. *Each graph G can be ecologically colored with n_G colors and it cannot be ecologically colored with $k > n_G$ colors.*

Hence, any graph admits an ecological 1-coloring, **an ecological 2-coloring (non e' vero!)** and an ecological n_G-coloring. In Section 3 we will prove that, actually, any graph can be ecologically k-colored, for any $k \in [n_G]$. To this aim, we need to prove a structural property of ND graphs which is stated in Theorem 3. Proof of Theorem 3 makes use of the following lemma.

Lemma 1. *Let $\mathcal{F} = \{N_1, \ldots, N_n\}$ be a family of n distinct subsets of $[n]$. Then, there exists $i \in [n]$ such that, for any pair $N_j, N_k \in \mathcal{F}$, $N_j \neq N_k \cup \{i\}$.*

Proof. The proof is by contradiction. Assume that, for any $i \in [n]$, there exists a pair of two distinct sets $L_i, S_i \in \mathcal{F}$ such that $L_i = S_i \cup \{i\}$: if there exist more than one such pairs of sets, then we arbitrarily choose one of them as the only one associated with i. Observe that, for any distinct $i, j \in [n]$, $L_i \neq L_j$ or $S_i \neq S_j$ since otherwise the distinctness between L_i and S_i would imply that $i = j$.

Let us define a directed graph $G_{\mathcal{F}} = (\mathcal{F}, A_{\mathcal{F}})$ as follows. For any $h, k \in [n]$, $(N_h, N_k) \in A_{\mathcal{F}}$ if and only if there exists $i \in [n]$ such that $S_i = N_h$ and $L_i = N_k$. Clearly, $G_{\mathcal{F}}$ is acyclic, since otherwise there would exist a sequence $X_0, \ldots, X_{h-1}$ of h distinct subsets of $[n]$ such that $X_i \subset X_{i+1}$ for $0 \leq i < h-1$ and $X_{h-1} \subset X_0$: this would imply that $X_0 \subset X_0$.

We now prove that also the undirected graph corresponding to $G_{\mathcal{F}}$ does not contain any cycle. This implies that $G_{\mathcal{F}}$ contains at most $n-1$ arcs: since, for any distinct $i, j \in [n]$, $L_i \neq L_j$ or $S_i \neq S_j$, this contradicts the fact that there must be exactly n arcs in $G_{\mathcal{F}}$, thus proving that there must exist $i \in [n]$ such that, for any pair $N_j, N_k \in \mathcal{F}$, $N_j \neq N_k \cup \{i\}$.

Assume, by contradiction, that the undirected graph corresponding to $G_{\mathcal{F}}$ contains a cycle $X_0, \ldots, X_{h-1}$ with $h \geq 3$. Then, there must exist r with $0 \leq r \leq h-1$ such that $(X_r, X_{r-1}) \in A_{\mathcal{F}} \wedge (X_r, X_{r+1}) \in A_{\mathcal{F}}$ (in the following, we assume that all operations are performed modulo h). Indeed, either $r = 0$ or there exists an incoming arc incident to X_0: in this latter case, we can follow the chain of incoming arcs starting from X_0 and, since G_F is acyclic, we certainly encounter a node X_r with no incoming arcs (see, for example, the cycle in Figure 2 for which $r = 4$).

Let $i \in [n]$ be the element such that $i \in X_{r-1} \wedge i \notin X_r$ and let us prove that $i \in X_s$, for any $s \neq r$ with $0 \leq s \leq h-1$. This is due to the fact that if $i \in X_t$, for some t with $0 \leq t \leq h-1$, and $(X_t, X_s) \in A_{\mathcal{F}}$ or $(X_s, X_t) \in A_{\mathcal{F}}$, then $i \in X_s$. Indeed, if $(X_t, X_s) \in A_{\mathcal{F}}$, then $X_t \subset X_s$; otherwise, if $(X_s, X_t) \in A_{\mathcal{F}}$, then $X_t - X_s \neq \{i\}$ since $s \neq r$ and, for any $i \in [n]$, there exist only two adjacent nodes of $G_{\mathcal{F}}$ whose difference is equal to $\{i\}$. Hence, in both cases we have that $i \in X_s$. We then have that $i \in X_{r+1}$ (see, for example, the cycle in Figure 2 where $i \notin X_4 \wedge i \in X_5$). On the other hand, there must exist $j \neq i$ such that $j \in X_{r+1} \wedge j \notin X_r$: hence, $X_{r+1} - X_r \supseteq \{i, j\}$ which contradicts the fact that $(X_r, X_{r+1}) \in A_{\mathcal{F}}$ (in the

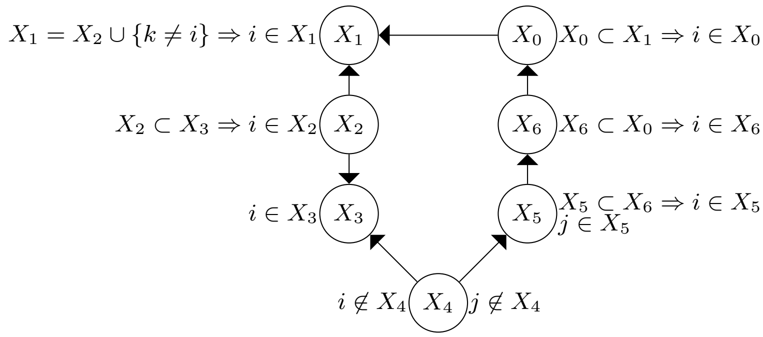


Fig. 2. The proof of Lemma 1

example, we have that $X_5 - X_4 \supseteq \{i, j\}$). This completes the proof of the fact that the undirected graph corresponding to $G_{\mathcal{F}}$ is acyclic and, hence, the proof of the lemma. □

Theorem 3. *Let $G = (V, E)$ be a ND graph with n nodes. Then, there exists a node $u \in V$, such that the graph induced by $V - \{u\}$ is a ND graph with $n - 1$ nodes.*

Proof. Without loss of generality, assume that $V = [n]$ and let $\mathcal{F} = \{N_1, \ldots, N_n\}$ be the family of neighborhoods of the n nodes. Since G is a ND graph, $\mathcal{F}$ satisfies the hypothesis of Lemma 1: hence, there exists a node i such that, for any pair of two other nodes j and k, the neighborhoods of these two nodes do not differ for i only. This implies that graph induced by $V - \{i\}$ is a ND graph. □

Theorem 4. *Let $G = (V, E)$ be a ND graph with n nodes. Then, G can be ecologically colored with $n - 1$ colors.*

Proof. Let $G = (V, E)$ be a ND graph with n nodes. From Theorem 3 it follows that there exist two nodes u and v such that the graph $G^{u,v}$ induced by $V - \{u, v\}$ is a ND graph with $n - 2$ nodes. Let r be any coloring of G that assigns $n - 2$ different colors to the nodes of $G^{u,v}$ and that assigns the same new color to u and v. In order to prove that r is ecological, we proceed by contradiction and assume that there exist two nodes p and q such that $r(p) \neq r(q)$ and $C_r(p) = C_r(q)$. We then distinguish the following cases.

1. $\{p, q\} = \{u, v\}$. In this case, $r(p) = r(q)$, and, hence, we get a contradiction.
2. $\{p, q\} \subseteq V - \{u, v\}$. In this case, since $G^{u,v}$ is ND and since all its nodes are colored with different colors, $C_r(p)$ must be different from $C_r(q)$, and, hence, we get a contradiction.
3. $p \in \{u, v\} \wedge q \in V - \{u, v\}$. In this case, since all the nodes of $G^{u,v}$ are colored with different colors and since $C_r(p) = C_r(q)$, we have that $N(p) \cap (V - \{u, v\}) = N(q) \cap (V - \{u, v\})$. Moreover, p and q cannot be adjacent, since otherwise $r(q) \in C_r(p) - C_r(q)$, contradicting the fact that $C_r(p) = C_r(q)$. Since G is ND, exactly one node among p and q must be adjacent to the node in $\{u, v\} - \{p\}$: this implies that either $r(p) \in C_r(p) - C_r(q)$ or $r(p) \in C_r(q) - C_r(p)$, and, hence, we get a contradiction.
4. $p \in V - \{u, v\} \wedge q \in \{u, v\}$. This case is symmetric to the previous one.

It follows that there cannot exist two nodes p and q such that $r(p) \neq r(q)$ and $C_r(p) = C_r(q)$: that is, r is an ecological coloring of G with $n - 1$ colors, and the theorem is proved. □

We conclude this section by stating another interesting property of ND graphs. To this aim, we need the following definition.

Definition 2. *Let $G = (V, E)$ be a graph and $\mathcal{I} = \{I_1, \ldots, I_{h-1}, I_h\}$ be a partition of V into non $h \geq 1$ empty sets such that $I_1, \ldots, I_{h-1}$ are independent sets. We say $\mathcal{I}$ is an ecological family for G if either $h = 1$ or for any $i, j \in [h]$ with $i \neq j$ and for any $u \in I_i$ and $v \in V_j$ there exists $t \in [h]$ such that u is adjacent to some node in I_t and v is not adjacent to any node in I_t. We call h the size of $\mathcal{I}$.*

Lemma 2. *Let $G = (V, E)$ be an ND graph of n nodes and let $\mathcal{I} = \{I_1, \ldots, I_h\}$ be an ecological family for G of size $h \geq 1$. Then, there exists an ecological family $\mathcal{I}'$ for G of size $h + 1$ or $h + 2$. Furthermore, $\mathcal{I}'$ can be computed in polynomial time.*

Proof. If I_h is not an independent set, then let I_{h1} be a maximal independent set for the subgraph of G induced by I_h, $I_{h2} = I_h - I_{h1}$ and $\mathcal{I}' = \mathcal{I} - \{I_h\} \cup \{I_{h1}, I_{h2}\}$. Let $u \in I_i \in \mathcal{I}$ and $v \in I_j \in \mathcal{I}$ with $i \neq j$ and let $I_t \in \mathcal{I}$ be such that u is adjacent to some node in I_t and v is not adjacent to any node in I_t. This is still true in $\mathcal{I}'$ with I_t eventually replaced I_{h1} or I_{h2}. Hence, assume $i = j = h$, that is, $u \in I_{h1}$ and $v \in I_{h2}$. in this case, since I_{h1} is a maximal independent set for the subgraph of G induced by I_h, v has to be adjacent to some node in I_{h1} while u is not adjacent to any other node in I_{h1}.

Assume now that I_h is an independent set. In this case, we first show that there exist $p, q \in [h]$ such that the subgraph of G induced by $I_p \cup I_q$ is not an independent set and is not a bipartite complete graph. The proof proceeds by contradiction. Assume that all pairs of independent sets in $\mathcal{I}$ induce either a complete bipartite graph or an independent set. This, in turn, implies that, since at least one independent set is not a singleton, there must exist at least two nodes u and v in the same independent set which, for any $i \in [h]$, are adjacent either to all nodes or to no node in I_i. That is, u and v have the same neighborhood, thus contradicting the ND property of G.

Hence, let $p, q \in [h]$ such that the subgraph of G induced by $I_p \cup I_q$ is not an independent set and is not a bipartite complete graph. We now find a partition of I_p into $I_{p1} \cup I_{p2}$ and a partition of I_q into $I_{q1} \cup I_{q2}$ such that $I_{p1} \neq \emptyset$, $I_{q1} \neq \emptyset$, $I_{p2} \cup I_{q2} \neq \emptyset$ and $I_{p1} \cup I_{q1}$ is an independent set.

Since , the subgraph of G induced by $I_p \cup I_q$ is not an independent set and is not a bipartite complete graph, it has to contain at least 3 nodes; without loss of generality, assume $|I_p| \geq 2$. Furthermore, there exist three nodes $u_1, u_2 \in I_p$ and $v_1 \in I_q$ such that $(u_1, v_1) \in E$ and $(u_2, v_1) \notin E$. Let $I_{q1} = \{x \in I_q : (u_2, x) \notin E\}$, $I_{p1} = \{x \in I_p : (x, y) \notin E \forall y \in I_{q1}\}$, $I_{p2} = I_p - I_{p1}$ and $I_{q2} = I_q - I_{q1}$: notice that $v_1 \in I_{p1}$, $u_2 \in I_{q1}$ and $u_1 \in I_{p2}$. Finally, $I_{p1} \cup I_{q1}$ is an independent set by construction. Notice that I_{q2} could be empty: this happens if no node in I_q is adjacent to v_1, that is, only if G is not connected and V_1 is an isolated node in the subgraph induced by $I_p \cup I_q$.

We now define $\mathcal{I}'$ as follows:

1. $\mathcal{I}' = \mathcal{I} - \{I_p, I_q\} \cup \{I_{p1}, I_{p2}, I_{q1}, I_{q2}\}$ if $I_{q2} \neq \emptyset$;
2. $\mathcal{I}' = \mathcal{I} - \{I_p, I_q\} \cup \{I_{p1}, I_{p2}, I_{q1}\}$ if $I_{q2} = \emptyset$.

It remains to show that $\mathcal{I}'$ is an ecological family for G. To this aim, consider a pair of nodes $u \in I_i \in \mathcal{I}$ and $v \in I_j \in \mathcal{I}$ with $i \neq j$ and let $I_t \in \mathcal{I}$ be such that u is adjacent to some node in I_t and v is not adjacent to any node in I_t. This is still true in $\mathcal{I}'$ with I_t eventually replaced by one set out of $I_{p1}, I_{p2}, I_{q1}, I_{q2}$. Hence, assume $i = j$. One of the following cases may occur:

- $u \in I_{p1}$ and $v \in I_{p2}$: in this case, by construction, u is not adjacent to any node in I_{q1} and v has to be adjacent to some node in I_{q1} (otherwise, v should be contained in I_{p1});
- $u \in I_{q1}$ and $v \in I_{q2}$: this case may occur only if $I_{q2} \neq \emptyset$ and, if so, it is similar to the previous one.

Hence, $\mathcal{I}'$ is an ecological family for G and its size is either $h + 1$ or $h + 2$. $\square$

3 The ecological coloring algorithm

In this section we prove the existence of ecological colorings with any feasible number of colors. According to Theorem 2 we can state our main result in terms of ND graphs.

Let $G = (V, E)$ be an ND graph of n nodes and $k < n$ the number of colors we are interested in. Our algorithm works in two phases. During the first phase, an ecological family for G of size either $k-1$ or k is computed by using Lemma 2. If the size of such a family is k an ecological coloring is directly derived from the family. Conversely, if the size of the ecological family is smaller than k then the second phase is started and the informations conveyed by the ecological family are used to compute the k-ecological coloring of G. The algorithm is described in Figure 3.

Input: ND graph $G = (V, E)$ with $|V| = n$ and integer k with $1 \leq k \leq n$.
Output: An ecological coloring r for G which uses k colors.

```
1: Phase 1: apply Algorithm Partitioning with input G and k to compute an ecological family I for G of size h ≤ k;
2: if h = k then
3:     for (i = 1; i ≤ k; i++) do
4:         ∀u ∈ I_i: r(u) ← i;
5: else
6:     Phase 2: apply Algorithm Refining with input I and k to compute an ecological coloring r for G which uses k colors;
```

Fig. 3. The ecological coloring algorithm

In the next two subsections the two phases are detailed and the proof that they are in fact correct is drawn. This allows us to prove our main result, stated in the following theorem.

Theorem 5. *For any ND graph $G = (V, E)$ with n nodes and for any $k \in [n]$, G admits an ecological coloring which uses k colors. Such a coloring can be computed in polynomial time.*

Proof. Lemma 3 insures that Algorithm `Partitioning` computes indeed an ecological family $\mathcal{I}$ for G.

If $|\mathcal{I}| = k$, lines 2–4 are executed. We now show that they compute an ecological coloring for G. Let u and v be two nodes such that $r(u) = p \neq r(v) = q$. Hence, $u \in I_p \in \mathcal{I}$ and $v \in I_q \in \mathcal{I}$. Since $\mathcal{I}$ is an ecological family for G, then there exists t such that u is adjacent to some node in I_t and v is not adjacent to any node in I_t. Since $r(x) = t$ if and only if $x \in I_t$, $C_r(u) \neq C_r(v)$. Hence, r is an ecological coloring of G.

Conversely, if the `else` statement is executed then the computed coloring is ecological by Lemma 4.

The theorem is completely proved. □

3.1 Phase 1: computing an ecological family

This phase iteratively applies Lemma 2 in order to construct an ecological family for G. More formally, algorithm `Partitioning` performing this task is described in Figure 4.

Lemma 3. *Let G be a ND graph with n nodes and k a positive integer such that $k \leq n$. Algorithm* `Partitioning` *computes an ecological family for G of size either $k-1$ or k.*

Proof. The condition of the `while` loop (line 2) requires $|\mathcal{I}| < k - 1$: since $k \leq n$, then Lemma 2 can be actually applied to set $\mathcal{I}$ at each iteration.

Given an ecological family of size h for a graph G, Lemma 2 computes an ecological family for G of size either $h+1$ or $h+2$. The lemma is thus proved. □

Input: ND graph $G = (V, E)$ with $|V| = n$ and integer k with $1 \le k \le n$.
Output: An ecological family $\mathcal{I}$ for G of size at most k.

$\mathcal{I} = \{V\}$;
while $|\mathcal{I}| < k - 1$ **do**
    Let $\mathcal{I}'$ the ecological family for G obtained by applying Lemma 2 to $\mathcal{I}$;
    $\mathcal{I} \leftarrow \mathcal{I}'$;

Fig. 4. `Partitioning`: Phase 1 of the ecological coloring algorithm

3.2 Phase 2: the last refinement

If Phase 1 ends with an ecological family for G of size $k-1$, the second phase is started and the size $(k-1)$ ecological family for G is used to compute the k-ecological coloring of G. Algorithm `Refining` shown in Figure 5 first tries to increase by 1 the size of the ecological family (according to Lemma 2) in order to color it in the same way as in lines 2–4 of Algorithm in Figure 3. If this is not possible, then it both increases the size of $\mathcal{I}$ by 2 (always according to Lemma 2) and it computes a $(k-2)$-coloring r_2 of $\mathcal{I}$. In order to perform the last step, it builds the color graph C^{G,r_1} (see Subsection 1.1) of G with respect to the $(k-1)$-coloring r_1 deriving from $\mathcal{I}$ and applies to it Theorem 3. Finally, by exploiting both r_2 and the ecological family of size $k+1$, it computes the ecological coloring r for G that uses k colors.

Input: An ecological family $\mathcal{I} = \{I_1, I_2, \ldots, I_{k-1}\}$ for an ND graph $G = (V, E)$ with $|V| = n$, and a positive integer $k \le n$.
Output: An ecological coloring r of G which uses k colors.

1: Let r_1 be the coloring according to which all nodes in I_i are colored with color i, $i = 1, \ldots, k-1$;
2: **if** $\exists 1 \le p \le k-1$ and $u, v \in I_p$ such that $C_{r_1}(u) \ne C_{r_1}(v)$ **then**
3:     $I_{p1} \leftarrow \{x \in I_p : C_{r_1}(x) = C_{r_1}(u)\}$;
4:     $\forall u \notin I_{p1}, r(u) \leftarrow r_1(u); \forall u \in I_{p1}, r(u) \leftarrow k$;
5: **else**
6:     Let $I_p, I_q \in \mathcal{I}$ such that $I_p \cup I_q$ induces a non complete bipartite graph with at least two nodes in I_p and at least two nodes in I_q;
7:     Partition I_p into I_{p1}, I_{p2} and I_q into I_{q1}, I_{q2} such that all of them are not empty, $I_{p1} \cup I_{q1}$ is an independent set, and both $I_{p1} \cup I_{q2}$ and $I_{p2} \cup I_{q1}$ are not independent sets (see proof of Lemma 2);
8:     $C^{G,r_1} \leftarrow$ color graph of G with respect to r_1;
9:     Color G_{r_1} with $k-2$ colors, let r_2 be such a coloring;
10:     Transform r_2 into a coloring r_3 of G: $\forall I \in V_{r_1}, \forall u \in I : r_3(u) \leftarrow r_2(I)$;
11:     Compute r: $\forall u \notin I_{p1} \cup I_{q2} : r(u) \leftarrow r_3(u); \forall u \in I_{p1} : r(u) \leftarrow k-1; \forall u \in I_{q2} : r(u) \leftarrow k$;

Fig. 5. `Refining`: Phase 2 of the ecological coloring algorithm

Lemma 4. *Let G be a graph and $\mathcal{I}$ an ecological family for G of size $k-1$. Then, Algorithm* `Refining` *computes an ecological coloring of G with k colors.*

Proof. Notice that the coloring computed at line 1 is the same as that computed at line 4 of the Algorithm in Figure 3 and, hence, it is ecological.

When the `if`-statement is executed (line 2), the partition $\mathcal{I} - \{I_p\} \cup \{I_{p1}, I_{p2}\}$ computed at line 3 is ecological by Lemma 2. Hence, the coloring r computed at line 4 can be easily proved to be ecological by the same arguments in the proof of Lemma 3.

Conversely, assume the **else** statement is executed (lines 5–11). In this case, since each pair of nodes contained in the same independet set has the same colorhood, then any pair of independent sets that induces a non complete bipartite graph contains at least four nodes, with at least two nodes in each independent set.

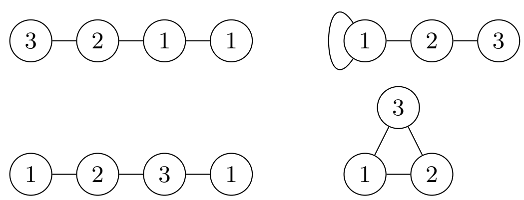


Fig. 6. Two different ecological colorings of the same graph

Hence, the pair of sets I_p and I_q at line 6 is in fact found. By the same arguments in the proof of Lemma 2 it follows that the partition described at line 7 can be computed and should result into an ecological partition for G of size $k+1$. In turn, this sholud result in a $(k+1)$-coloring of G.

Hence, it is now needed to decrease the number of colors. To this aim, the color graph C^{G,r_1} is considered. Observe that, since r_1 is ecological, C^{G,r_1} is a ND graph with $k-1$ nodes. From Theorem 4 it follows that $C^{G,r}$ can be ecologically colored with $k-2$ colors: hence, coloring r_2 of line 9 can be computed. Line 10 then computes a new coloring r_3 of G which uses $k-2$ colors as follows: for any node x of G, $r_3(x) = r_2(r_1(x))$. It is easy to prove that r_3 is ecological. Indeed, since all nodes contained in the same independent set have the same colorhood, then for any node $u \in V$ it holds that $C_{r_3}(u) = C_{r_2}(r_1(u))$. As a consequence, since for any $u, v \in V$ such that $r_3(u) \neq r_3(v)$ it holds that $r_2(r_1(u)) \neq r_2(r_1(v))$ and r_2 is ecological, hence $C_{r_3}(u) = C_{r_2}(r_1(u)) \neq C_{r_2}(r_1(v)) = C_{r_3}(v)$. Finally, coloring r is computed by modifying r_3 at line 11. Nodes $u \in I_{p1}$ are assigned color $r(u) = k-1$ and nodes $u \in I_{q1}$ are assigned color $r(u) = k$; colors of all the remaining nodes are left unchanged. This still results in an ecological coloring for G. In order to prove this last assertion, let us consider two nodes u and v such that $r(u) \neq r(v)$ and prove that $C_r(u) \neq C_r(v)$.

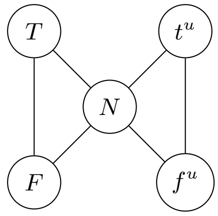
If $u, v \notin I_p \cup I_q$, then the assertion follows since $C_{r_3}(x) \subseteq C_r(x)$ for any $x \notin I_p \cup I_q$. The same reasoning applies if $u, v \in I_p \cup I_q$ and $r_3(u) \neq r_3(v)$. Finally, if $u, v \in I_p \cup I_q$ and $r_3(u) = r_3(v)$, then (without loss of generality) $u \in I_{p1}$ and $v \in I_{q2}$: in this case assume by contradiction that $C_r(u) = C_r(v)$. This might happen only if, in the graph induced by $I_p \cup I_q$, u is adjacent only to nodes in I_{q1} and v is adjacent only to nodes in I_{p2}: this is not possible, since $I_{p1} \cup I_{q1}$ is an independent set.

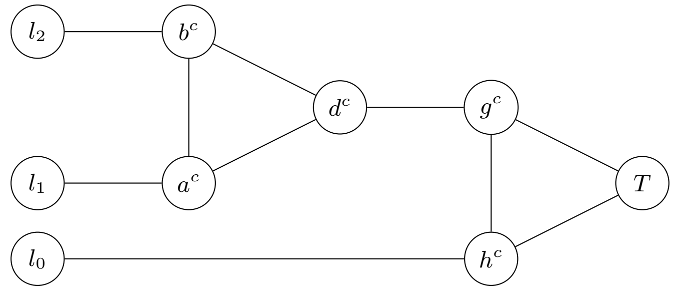
The lemma is thus proved. $\square$

4 Conclusions and open questions

In this paper we have proved that any graph can be ecologically colored by making use of any reasonable number of colors. However, different ecological colorings can be produced for a given graph and for a given number of colors (as shown in Figure 6).

The ECRA(R) decision problem consists in deciding whether a graph G admits an ecological coloring whose color graph is R. By making use of one of the standard NP-completeness proofs of the 3-colorability problem, it is easy to prove that ECRA(K_3) is NP-complete. In particular, the reduction introduces, for any variable u, the gadget shown in the left part of the following picture, while, for any clause $c = \{l_0, l_1, l_2\}$, it introduces the gadget shown in the right part of the picture, where the nodes on the left corresponds to the nodes on the right of the previous gadget and the node T corresponds to the node T of the previous gadget.





It is easy to prove that the Boolean formula is satisfiable if and only if the resulting graph can be colored with three colors. Moreover, since any node of the graph belongs to at least one clique of three nodes, any three coloring of the graphs is an ecological coloring whose color graph is K_3.

It is then a very interesting open question to look for a complete classification of the complexity of the $\mathrm{E}\textsc{c}\mathrm{RA}(R)$ problem similar to the one proposed in [FP05].

References

[BE89] S.P. Borgatti and M.G. Everett. The class of regular equivalences: algebraic structure and computation. *Social Networks*, 11:65–88, 1989.

[BE92] S.P. Borgatti and M.G. Everett. Graph colorings and power in experimental exchange networks. *Social Networks*, 14:287–308, 1992.

[BE94] S.P. Borgatti and M.G. Everett. Ecological and perfect colorings. *Social Networks*, 16:43–55, 1994.

[BCDRV07] S. Boschi, P. Crescenzi, M. Di Ianni, G. Rossi, and P. Vocca. Ecological Role Assignments and Mobility Models for Ad Hoc Wireless Networks. Manuscript in preparation.

[BE05] U. Brandes and T. Erlebach, Eds. Network Analysis: Methodological Foundations. Lecture Notes in Computer Science, Vol. 3418, Springer, Berlin, 2005.

[B04] R.L. Breiger. The Analysis of Social Networks. In *Handbook of Data Analysis*, M.A. Hardy and a. Bryman, Eds. Sage Publications, London, 505–526.

[FP05] J. Fiala and D. Paulusma. A complete complexity classification of the role assignment problem. *Theoretical Computer Science*, 349:67–81, 2005.

[K02] V.E. Krebs. Mapping Networks of Terrorist Cells. *Connections*, 24:43–52, 2002.

[LW71] F. Lorrain and H.C. White. Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1:49–80, 1971.

[MMPW07] J.P. McSorley, A. Marr, T. Porter, and W.D. Wallis. Closed neighborhood anti-Sperner graphs. *Australasian Journal of Combinatorics*, to appear.

[MHM07] M. Musolesi, S. Hailes and C. Mascolo. An Ad Hoc Mobility Model Founded on Social Network Theory. *Proc. 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 20-24, Venice, Italy, 2004.

[MM07] M. Musolesi and C. Mascolo. Designing Mobility Models based on Social Network Theory. *ACM SIGMOBILE Mobile Computing and Communications Review*, to appear.

[RS01] F.S. Roberts and L. Sheng. How hard is to determine if a graph has a 2-role assignment? *Networks*, 37(2):67–73, 2001.

[WR83] D.R. White and K.P. Reitz. Graph and semigroup homomorphisms on networks of relations. *Social Networks*, 5:193–234, 1983.

[XC05a] J.J. Xu and H. Chen. CrimeNet Explorer: A framework for Criminal Network Knowledge Discovery. *ACM Transactions on Information Systems*, 23:201–226, 2005.

[XC05b] J.J. Xu and H. Chen. Criminal Network Analysis and Visualization. *Communications of the ACM*, 48:101–107, 2005.