Balanced Paths in Colored Graphs⋆,⋆⋆

A. Bianco, M. Faella, F. Mogavero, and A. Murano

Università degli Studi di Napoli "Federico II", Italy
{alessandrobianco, mfaella, mogavero, murano}@na.infn.it

Abstract. We consider finite graphs whose edges are labeled with elements, called *colors*, taken from a fixed finite alphabet. We study the problem of determining whether there is an infinite path where either *(i)* all colors occur with the same asymptotic frequency, or *(ii)* there is a constant which bounds the difference between the occurrences of any two colors for all prefixes of the path. These two notions can be viewed as refinements of the classical notion of fair path, whose simplest form checks whether all colors occur infinitely often. Our notions provide stronger criteria, particularly suitable for scheduling applications based on a coarse-grained model of the jobs involved. We show that both problems are solvable in polynomial time, by reducing them to the feasibility of a linear program.

1 Introduction

In this paper, a colored graph is a finite directed graph whose edges are labeled with tags belonging to a fixed finite set of colors. For an infinite path in a colored graph, we say that the asymptotic frequency of a color is the long-run average number of occurrences of that color. Clearly, a color might have no asymptotic frequency on a certain path, because its long-run average oscillates. We introduce and study the problem of determining whether there is an infinite path in a colored graph where each color occurs with the same asymptotic frequency. We call such a path *balanced*. The existence of such a path in a given colored graph is called the *balance problem*.

Then, we consider the following stronger property: a path has the *bounded difference* property if there is a constant c such that, at all intermediate points, the number of occurrences of any two colors up to that point differ by at most c. The existence of such a path is called the *bounded difference problem* for a given graph. It is easy to prove that each bounded difference path is balanced. Moreover, each path that is both balanced and ultimately periodic (i.e., of the form $\sigma_1 \cdot \sigma_2^{\omega}$, for two finite paths σ_1 and σ_2) is also a bounded difference path. However, there are paths that are balanced but do not have the bounded difference property, as shown in Example 1.

We provide a loop-based characterization for each one of the mentioned decision problems. Both characterizations are based on the notion of *balanced set of loops*. A set of simple loops in the graph is balanced if, using those loops as building blocks, it is possible to build a finite path where all colors occur the same number of times.

⋆ Work partially supported by MIUR PRIN Project 2007-9E5KM8.

⋆⋆ This is the improved and extended version of the paper by the same title appeared in MFCS09: 34th International Symposium on Mathematical Foundations of Computer Science.

We prove that a graph satisfies the balance problem if and only if it contains a balanced set of loops that are mutually reachable (Theorem 1). Similarly, a graph satisfies the bounded difference problem if and only if it contains a balanced set of loops that are *overlapping*, i.e., each loop has a node in common with another loop in the set (Theorem 2).

Using the above characterizations, for each problem we devise a linear system of equations whose feasibility is equivalent to the solution of the problem. Since the size of these linear systems is polynomial, we obtain that both our problems are decidable in polynomial time. Further, we can compute in polynomial time a finite representation of a path with the required property. We also provide evidence that the problems addressed are non-trivial, by showing that a closely related problem is NP-hard: the problem of checking whether there is a perfectly balanced finite path connecting two given nodes in a graph.

We believe that the two problems that we study and solve in this paper are natural and canonical enough to be of independent theoretical interest. Additionally, they may be regarded as instances of the well established notion of *fairness*.

Balanced paths as fair paths. Colored graphs as studied in this paper routinely occur in the field of computer science that deals with the analysis of concurrent systems [MP91]. In that case, the graph represents the transition relation of a concurrent program and the color of an edge indicates which one of the processes is making progress along that edge. One basic property of interest for those applications is called *fairness* and essentially states that, during an infinite computation, each process is allowed to make progress infinitely often [Fra86]. Starting from this core idea, a rich theory of fairness has been developed, as witnessed by the amount of literature devoted to the subject (see, for instance, [LPS81, Kwi89, dA00]).

Cast in our abstract framework of colored graphs, the above basic version of fairness asks that, along an infinite path in the graph, each color occurs infinitely often. Such requirement does not put any bound on the amount of steps that a process needs to wait before it is allowed to make progress. As a consequence, the asymptotic frequency of some color could be zero even if the path is fair. Accordingly, several authors have proposed stronger versions of fairness. For instance, Alur and Henzinger define *finitary fairness* roughly as the property requiring that there be a fixed bound on the number of steps between two occurrences of any given color [AH98]. A similar proposal, supported by a corresponding temporal logic, was made by Dershowitz et al. [DJP03]. On a finitarily fair path, all colors have positive asymptotic frequency [1].

Our proposed notions of balanced paths and bounded difference paths may be viewed as two further refinements of the notion of fair path. Previous definitions treat the frequencies of the relevant events in isolation and in a strictly qualitative manner. Such definitions only distinguish between zero frequency (not fair), limit-zero frequency (fair, but not finitarily so), and positive frequency (finitarily fair). The current proposal, instead, introduces a quantitative comparison between competing events.

Technically, it is easy to see that bounded difference paths are special cases of finitarily fair paths. On the other hand, finitarily fair paths and balanced paths are incomparable notions.

We believe that the two proposed notions are valuable to some applications, perhaps quite different from the ones in which fairness is usually applied. Both the balance property and

[1] For the sake of clarity, we are momentarily ignoring those paths that have *no* asymptotic frequency.

the bounded difference property are probably too strong for the applications where one step in the graph represents a fine-grained transition of unknown length in a concurrent program. In that case, it may be of little interest to require that all processes make progress with the same (abstract) frequency.

On the other hand, consider a context where each transition corresponds to some complex or otherwise lengthy operation. As an example, consider the model of a concurrent program where all operations have been disregarded, except the access to a peripheral that can only be used in one-hour slots, such as a telescope, which requires some time for re-positioning. Assuming that all jobs have the same priority, it is certainly valuable to find a scheduling policy that assigns the telescope to each job with the same frequency.

As a non-computational example, the graph may represent the rotation of cultures on a crop, with a granularity of 6 months for each transition [Wik09]. In that case, we may very well be interested not just in having each culture eventually planted (fairness) or even planted within a bounded amount of time (finitarily fair), but also occurring with the same frequency as any other culture (balanced or bounded difference).

The rest of the paper is organized as follows. Section 2 introduces the basic definitions. Section 3 establishes connections between the existence of balanced or bounded difference paths in a graph and certain loop-based properties of the graph itself. Section 4 (respectively, Section 5) exploits the properties proved in Section 3 to define a system of linear equations whose feasibility is equivalent to the balance problem (resp., the bounded difference problem).

2 Preliminaries

Let X be a set and i be a positive integer. By X^i we denote the Cartesian product of X with itself i times. By $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ we respectively denote the set of non-negative integer, relative integer, rational, and real numbers. Given a positive integer k, let $[k] = \{1,\dots,k\}$ and $[k]_0 = [k] \cup \{0\}$.

A *k-colored graph* (or simply *graph*) is a pair $G = (V,E)$, where V is a set of *nodes* and $E \subseteq V \times [k] \times V$ is a set of *colored edges*. We employ integers as colors for technical convenience. All the results we obtain also hold for an arbitrary set of labels. An edge (u,a,v) is said to be *colored* with a. In the following, we also simply call a k-colored graph a *graph*, when k is clear from the context. For a node $v \in V$ we call ${}_vE = \{(v,a,w) \in E\}$ the set of edges exiting from v, and $E_v = \{(w,a,v) \in E\}$ the set of edges entering in v. For a color $a \in [k]$, we call $E(a) = \{(v,a,w) \in E\}$ the set of edges colored with a. For a node $v \in V$, a *finite v-path* ρ is a finite sequence of edges $(v_i,a_i,v_{i+1})_{i\in\{1,\dots,n\}}$ such that $v_1 = v$. The length of ρ is n and we denote by $\rho(i)$ the i-th edge of ρ. Sometimes, we write the path ρ as $v_1v_2\dots v_n$, when the colors are unimportant. A finite path $\rho = v_1v_2\dots v_n$ is a *loop* if $v_1 = v_n$. A loop $v_1v_2\dots v_n$ is *simple* if $v_i \neq v_j$, for all $i \neq j$, except for $i = 1$ and $j = n$. An *infinite v-path* is defined analogously, i.e., it is an infinite sequence of edges. Let ρ be a finite path and π be a possibly infinite path, we denote by $\rho \cdot \pi$ the *concatenation* of ρ and π. By ρ^ω we denote the infinite path obtained by concatenating ρ with itself infinitely many times. A graph G is *strongly connected* if for each pair (u,v) of nodes there is a finite u-path with last node v and a finite v-path with last node u.

For a finite or infinite path ρ and an integer i, we denote by $\rho^{\leq i}$ the *prefix* of ρ containing i edges. For a color $a \in [k]$, we denote by $|\rho|_a$ the number of edges labeled with a occurring in ρ. For two colors $a, b \in [k]$, we denote the difference between the occurrences of edges labeled with a and b in ρ by $\mathit{diff}_{a,b}(\rho) = |\rho|_a - |\rho|_b$. An infinite path π is *periodic* iff there exists a finite path ρ such that $\pi = \rho^\omega$. A loop σ is *perfectly balanced* iff $\mathit{diff}_{a,b}(\sigma) = 0$ for all $a, b \in [k]$. Finally, we denote by **0** and **1** the vectors containing only 0's and 1's, respectively. We can now define the following two decision problems.

The balance problem. Let G be a k-colored graph. An infinite path ρ in G is *balanced* if for all $a \in [k]$,

$$\lim_{i \to \infty} \frac{|\rho^{\leq i}|_a}{i} = \frac{1}{k}.$$

The *balance problem* is to determine whether there is a balanced path in G.

The bounded difference problem. Let G be a k-colored graph. An infinite path ρ in G has the *bounded difference property* (or, is a *bounded difference path*) if there exists a number $c \geq 0$, such that, for all $a, b \in [k]$ and $i > 0$,

$$|\mathit{diff}_{a,b}(\rho^{\leq i})| \leq c.$$

The *bounded difference problem* is to determine whether there is a bounded difference path in G.

3 Basic Properties

In this section, we assume that $G = (V, E)$ is a finite k-colored graph, i.e., both V and E are finite. In the following lemma, the proof of item 1 is trivial, while the proof of item 2 can be found in the Appendix.

Lemma 1. *The following properties hold:*

1. *if a path has the bounded difference property, then it is balanced;*
2. *a path ρ is balanced if and only if for all $a \in [k-1]$,*

$$\lim_{i \to \infty} \frac{\mathit{diff}_{a,k}(\rho^{\leq i})}{i} = 0.$$

The following example shows that the converse of item 1 of Lemma 1 does not hold.

Example 1. For all $i > 0$, let $\sigma_i = (1 \cdot 2)^i \cdot 1 \cdot 3 \cdot (1 \cdot 3 \cdot 2 \cdot 3)^i \cdot 1 \cdot 3 \cdot 3$. Consider the infinite sequence $\sigma = \prod_{i=1}^{\omega} \sigma_i$ obtained by a hypothetic 3-colored graph. On one hand, it is easy to see that for all $i > 0$ it holds $\mathit{diff}_{3,1}(\sigma_i) = 1$. Therefore, $\mathit{diff}_{3,1}(\sigma_1 \sigma_2 \dots \sigma_n) = n$, and σ is not a bounded difference path.

On the other hand, since the length of the first n blocks is $\Theta(n^2)$ and the difference between any two colors is $\Theta(n)$, in any prefix $\sigma^{\leq i}$ the difference between any two colors is in $O(\sqrt{i})$. According to item 2 of Lemma 1, σ is balanced. □

Two loops σ, σ' in G are *connected* if there exists a path from a node of σ to a node of σ', and vice-versa. A set $\mathcal{L}$ of loops in G is connected if all pairs of loops in $\mathcal{L}$ are connected. Two loops in G are *overlapping* if they have a node in common. A set $\mathcal{L}$ of loops in G is overlapping if for all pairs of loops $\sigma, \sigma' \in \mathcal{L}$ there exists a sequence $\sigma_1, \ldots, \sigma_n$ of loops in $\mathcal{L}$ such that *(i)* $\sigma_1 = \sigma$, *(ii)* $\sigma_n = \sigma'$, and *(iii)* for all $i = 1, \ldots, n-1$, (σ_i, σ_{i+1}) are overlapping. Given a set of loops $\mathcal{L}$ in G, the subgraph *induced by* $\mathcal{L}$ is $G' = (V', E')$, where V' and E' are all and only the nodes and the edges, respectively, belonging to a loop in $\mathcal{L}$.

Lemma 2. *Let G be a graph, $\mathcal{L}$ be a set of loops in G, and $G' = (V', E')$ be the subgraph of G induced by $\mathcal{L}$, then the following statements are equivalent:*

1. *$\mathcal{L}$ is* overlapping.
2. *The subgraph G' is strongly connected.*
3. *There exists $u \in V'$ such that for all $v \in V'$ there exists a path in G' from u to v.*

Proof. $[1 \Rightarrow 2]$ If $\mathcal{L}$ is overlapping, then, for all pairs of loops σ_1, σ_2, there exists a sequence of loops that links σ_1 with σ_2. Thus, from any node of σ_1, it is possible to reach any node of σ_2. Hence, G' is strongly connected.

$[2 \Rightarrow 3]$ Trivial.

$[3 \Rightarrow 2]$ Let $u \in V'$ be a witness for (3). Let $v, w \in V'$, we prove that there is a path from v to w. We have that u is connected to both v and w. Since all edges in G' belong to a loop, for all edges $(u', \cdot, v')$ along the path from u to v there is a path from v' to u'. Thus, there is a path from v to u, and, as a consequence, a path from v to w, through u.

$[2 \Rightarrow 1]$ If G' is strongly connected, for all $\sigma_1, \sigma_2 \in \mathcal{L}$ there is a path ρ in G' from any node of σ_1 to any node of σ_2. This fact holds since G' is induced by $\mathcal{L}$, so ρ uses only edges of the loops in $\mathcal{L}$. While traversing ρ, every time we move from one loop to the next, these two loops must share a node. Therefore, all pairs of adjacent loops used in ρ are overlapping. Thus $\mathcal{L}$ is overlapping. □

The above lemma implies that if $\mathcal{L}$ is overlapping then it is also connected, since G' is strongly connected.

For all finite paths ρ of G, with a slight abuse of notation let $\mathit{diff}(\rho) = (\mathit{diff}_{1,k}(\rho), \ldots, \mathit{diff}_{k-1,k}(\rho))$ be the vector containing the differences between each color and color k, which is taken as a reference. We call $\mathit{diff}(\rho)$ the *difference vector* of ρ [2]. For all finite and infinite paths ρ we call *difference sequence* of ρ the sequence of difference vectors of all prefixes of ρ, i.e., $\{\mathit{diff}(\rho^{\leq n})\}_{n \in \mathbb{N}}$. Given a finite set of loops $\mathcal{L} = \{\sigma_1, \ldots, \sigma_l\}$ and a tuple of positive natural numbers $c_1, \ldots, c_l$, we call *natural linear combination* (in short, *n.l.c.*) of $\mathcal{L}$ with coefficients $c_1, \ldots, c_l$ the vector $x = \sum_{i=1}^{l} c_i \mathit{diff}(\sigma_i)$.

A loop is a *composition* of a finite tuple of simple loops $\mathcal{T}$ if it is obtained by using all and only the edges of $\mathcal{T}$ as many times as they appear in $\mathcal{T}$. Formally, for a loop σ and an edge e, let $n(e, \sigma)$ be the number of occurrences of e in σ. The loop σ is a composition of $(\sigma_1, \ldots, \sigma_l)$ if, for all edges e, it holds $n(e, \sigma) = \sum_{i=1}^{l} n(e, \sigma_i)$.

[2] The difference vector is related to the Parikh vector [Par66] of the sequence of colors of the path. Precisely, the difference vector is equal to the first $k-1$ components of the Parikh vector, minus the k-th component.

Lemma 3. *Let σ be a loop of length n containing m distinct nodes. Then, σ is a composition of at least $\lceil \frac{n}{m} \rceil$ simple loops.*

Proof. We use a decomposition algorithm on σ: the algorithm scans the edges of σ from the beginning to the end. As soon as a simple loop is found, i.e., as soon as a node is repeated, such a simple loop is removed from σ and added to the tuple. The tuple given by the removed loops is the decomposition we are looking for. Since a simple loop contains at most m nodes, the tuple contains at least $\lceil \frac{n}{m} \rceil$ loops. □

3.1 The Balance Problem

The following lemma, whose proof can be found in the Appendix, shows that a sequence of integral vectors has sum in $o(n)$ only if there is a finite set of vectors that occur in the sequence which have an n.l.c. with value zero.

Lemma 4. *Let $A \subset \mathbb{Z}^d$ be a finite set of vectors such that there is no subset $A' \subseteq A$ with an n.l.c. of value zero. Let $\{(a_{n,1}, \dots, a_{n,d})\}_{n \in \mathbb{N}}$ be an infinite sequence of elements of A, and $S_{n,i} = \sum_{j=0}^{n} a_{j,i}$ be the partial sum of the i-th component, for all $n \in \mathbb{N}$ and $i \in [d]$. Then, there exists at least an index h such that $\lim_{n \to \infty} \frac{S_{n,h}}{n} \neq 0$.*

The following result provides a loop-based characterization for the balance problem.

Theorem 1. *A graph G satisfies the balance problem iff there exists a connected set $\mathcal{L}$ of simple loops of G, with zero as an n.l.c.*

Proof. [if] Let $\mathcal{L} = \{\sigma_0, \dots, \sigma_{l-1}\}$ be a connected set of simple loops having zero as an n.l.c., with coefficients $c_0, \dots, c_{l-1}$. For all $i = 0, \dots, l-1$, let v_i be the initial node of σ_i. Since $\mathcal{L}$ is connected, there exists a path ρ_i from v_i to $v_{(i+1) \bmod l}$. For all $j > 0$, define the loop $\pi_j = \sigma_0^{j \cdot c_0} \rho_0 \sigma_1^{j \cdot c_1} \rho_1 \dots \sigma_{l-1}^{j \cdot c_{l-1}} \rho_{l-1}$. We claim that the infinite path $\pi = \prod_{j>0} \pi_j$ is balanced. Each time a π_j block ends along π, the part of the difference vector produced by the loops of $\mathcal{L}$ is zero. So, when a π_j ends, the difference vector is due only to the paths ρ_i. Since the index of the step $k(j)$ at which π_j ends grows quadratically in j and the difference vector $\mathit{diff}(\pi_1 \dots \pi_j)$ grows linearly in j, we have that $\lim_{j \to \infty} \mathit{diff}(\pi_1 \dots \pi_j)/k(j) = \mathbf{0}$. It can be shown that in the steps between $k(j)$ and $k(j+1)$, the i-th component of the difference vector differs from the one of $\mathit{diff}(\pi_1 \dots \pi_j)$ no more than a function $C_{i,j}$ that grows linearly in j. Specifically, $C_{i,j} = MP_i + jMA_i$, where MP_i is the sum, for all ρ_j, of the maximum modulus of the i-th component of the difference vector along ρ_j, and MA_i is the sum, for all σ_j, of the maximum modulus of the i-th component of the difference vector along σ_j. As a consequence, $\lim_{k \to \infty} \mathit{diff}(\pi^{\leq k})/k = \mathbf{0}$ and π is balanced.

[only if] If there exists an infinite balanced path ρ, since the set of nodes is finite, there is a set V' of nodes occurring infinitely often in ρ. Let ρ' be a suffix of ρ containing only nodes in V'. The path ρ' is balanced and it is composed by an infinite sequence of simple loops on V', plus a remaining simple path (see the proof of Lemma 3 for further details). Let $\mathcal{L}$ be the (finite) set of such simple loops, and let $A \subset \mathbb{Z}^{k-1}$ be the set of difference vectors of the loops in $\mathcal{L}$.

Every time a loop trough V' closes along ρ', the difference vector up to that point is the sum of the difference vectors of the simple loops occurred so far, plus the difference vector of the remaining simple path. Since the remaining simple path cannot have length greater than $|V'|$, the difference vector up to that point differs from a sum of a sequence of elements of A by a constant-bounded term. Let $n(i)$ be the index of the i-th point where a loop is closed along ρ'. Since ρ' is balanced, by statement 2 of Lemma 1, each component of the difference sequence $\{diff(\rho'^{\leq i})\}_{i\in\mathbb{N}}$ is in $o(i)$. Hence, each component of the partial sum of the difference vectors associated to the sequence of loops closed is in $o(n(i))$. By Lemma 4, this is possible only if A has a subset A' with an n.l.c. of value zero. Thus, the set of loops $\mathcal{L}'$ with difference vectors in A' has an n.l.c. with value zero. Moreover, since the loops in $\mathcal{L}'$ are constructed with edges of ρ', they are connected. This concludes the proof. □

3.2 The Bounded Difference Problem

Given a graph, if there exists a perfectly balanced loop σ, it is easy to see that σ^ω is a periodic bounded difference path. Moreover, if ρ is an infinite bounded difference path, then there exists a constant c such that the absolute value of all color differences is smaller than c. Since both the set of nodes and the possible difference vectors along ρ are finite, we can find two indexes $i < j$ such that $\rho(i) = \rho(j)$ and $diff(\rho^{\leq i}) = diff(\rho^{\leq j})$. So, $\sigma' = \rho(i)\rho(i+1)\dots\rho(j)$ is a perfectly balanced loop. Therefore, the following holds.

Lemma 5. *Given a graph G, the following statements are equivalent:*

1. *There exists a bounded difference path.*
2. *There exists a periodic bounded difference path.*
3. *There exists a perfectly balanced loop.*

We now prove the following result.

Lemma 6. *Let G be a graph. There exists a perfectly balanced loop in G iff there exists an overlapping set $\mathcal{L}$ of simple loops of G, with zero as n.l.c.*

Proof. [only if] If there exists a perfectly balanced loop σ, by Lemma 3 the loop is the composition of a tuple $\mathcal{T}$ of simple loops. Let $\mathcal{L}$ be the set of distinct loops occurring in $\mathcal{T}$, and for all $\rho \in \mathcal{L}$, let c_ρ be the number of times ρ occurs in $\mathcal{T}$. Since in the computation of the difference vector of a path it does not matter the order in which the edges are considered, we have $\sum_{\rho\in\mathcal{L}} c_\rho \cdot diff(\rho) = diff(\sigma) = \mathbf{0}$. Finally, since the loops in $\mathcal{L}$ come from the decomposition of a single loop σ, we have that $\mathcal{L}$ is overlapping.

[if] Let $\mathcal{L} = \{\sigma_1,\dots,\sigma_l\}$ be such that $\sum_{i=1}^{l} c_i \cdot diff(\sigma_i) = \mathbf{0}$. We construct a single loop σ such that $diff(\sigma) = \sum_{i=1}^{l} c_i \cdot diff(\sigma_i)$. The construction proceeds in iterative steps, building a sequence of intermediate paths $\rho_1,\dots,\rho_l$, such that ρ_l is the wanted perfectly balanced loop. In the first step, we take any loop $\sigma_{i_1} \in \mathcal{L}$ and we traverse it c_{i_1} times, obtaining the first intermediate path $\rho_1 = \sigma_{i_1}^{c_{i_1}}$. After the j-th step, since $\mathcal{L}$ is overlapping, there must be a loop $\sigma_{i_{j+1}} \in \mathcal{L}$ that is overlapping with one of the loops in the current intermediate path ρ_j, say in node v. Then, we *reorder* ρ_j in such a way that it starts and ends in v. Let ρ'_j be such reordering, we set $\rho_{j+1} = \rho'_j \sigma_{i_{j+1}}^{c_{i_{j+1}}}$. One can verify that ρ_l is perfectly balanced. □

The following theorem is a direct consequence of the previous two lemmas.

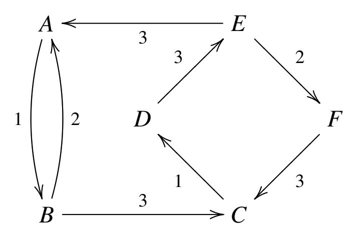

Fig. 1. A 3-colored graph satisfying the balance problem, but not the bounded difference problem.

Theorem 2. *A graph G satisfies the bounded difference problem iff there exists an overlapping set $\mathcal{L}$ of simple loops of G, with zero as n.l.c.*

Example 2. Consider the graph G in Fig. 1. First note that, up to rotation, there are just three simple loops in it: $\sigma_1 = A \cdot B \cdot A$, $\sigma_2 = C \cdot D \cdot E \cdot F \cdot C$, and $\sigma_3 = A \cdot B \cdot C \cdot D \cdot E \cdot A$. It is easy to see that $\mathit{diff}(\sigma_1) = (1,1)$, $\mathit{diff}(\sigma_2) = (-1,-1)$, and $\mathit{diff}(\sigma_3) = (-1,-3)$. On one hand, since the connected set of simple loops $\{\sigma_1, \sigma_2\}$ has zero as n.l.c., we obtain that there is a balanced path in G. Example 1 shows a particular balanced sequence of colors obtained by a non-periodic path of the subgraph G' of G induced by these two loops. On the other hand, for all the three overlapping sets of loops ($\{\sigma_1, \sigma_3\}$, $\{\sigma_2, \sigma_3\}$, and $\{\sigma_1, \sigma_2, \sigma_3\}$) there is no way to obtain a zero n.l.c. with all coefficients different from zero. So, there is no bounded difference path in G. □

3.3 2-Colored Graphs

When the graph G is 2-colored, the difference vector is simply a number. So, if $\mathcal{L}$ is a connected set of simple loops having zero as n.l.c., then there must be either a perfectly balanced simple loop or two loops with difference vectors of opposite sign. Notice that two loops σ, σ' with color differences of opposite sign have the following n.l.c. of value zero: $|\mathit{diff}(\sigma')| \cdot \mathit{diff}(\sigma) + |\mathit{diff}(\sigma)| \cdot \mathit{diff}(\sigma') = 0$. If the two loops are connected but not overlapping, we can construct a sequence of adjacent overlapping simple loops (the details can be found in the Appendix) connecting them. In this sequence, we are always able to find a perfectly balanced simple loop or two overlapping simple loops with difference vectors of opposite sign. Therefore, the following holds.

Lemma 7. *Let G be a 2-colored graph. If there exists a* connected *set of simple loops of G with zero as n.l.c., then there exists an* overlapping *set of simple loops of G with zero as n.l.c.*

Due to the above characterization, both decision problems can be solved efficiently, by using a minimum spanning tree algorithm to find two loops of opposite color difference sign, if such exist.

Theorem 3. *A 2-colored graph $G = (V,E)$ satisfies the bounded difference problem iff it satisfies the balance problem. Both problems can be solved in time $O(|V| \cdot |E| \cdot \log |V|)$.*

3.4 A Related NP-Hard Problem

In this section, we introduce an NP-hard problem similar to the bounded difference problem. Given a k-colored graph G and two nodes u and v, the new problem asks whether there exists a perfectly balanced path from u to v. We call this question the *perfectly balanced finite path problem*. To see that this problem is closely related to the bounded difference problem, one can note that it corresponds to the statement of item 3 in Lemma 5, by changing the word *loop* to *finite path*. The following result can be proved using a reduction from 3SAT. Due to space constraints, the following result is proved in the Appendix.

Theorem 4. *The perfectly balanced finite path problem is NP-hard.*

4 Solving the Balance Problem

In this section, we define a system of linear equations whose feasibility is equivalent to the balance problem for a given strongly connected graph.

Definition 1. *Let $G = (V,E)$ be a k-colored graph. We call* balance system *for G the following system of equations on the set of variables $\{x_e \,|\, e \in E\}$.*

1. for all $v \in V$	$\sum_{e \in E_v} x_e = \sum_{e \in {}_vE} x_e$
2. for all $a \in [k-1]$	$\sum_{e \in E(a)} x_e = \sum_{e \in E(k)} x_e$
3. for all $e \in E$	$x_e \geq 0$
4.	$\sum_{e \in E} x_e > 0.$

Let $m = |E|$ and $n = |V|$, the balance system has m variables and $m+n+k$ constraints. It helps to think of each variable x_e as a load associated to the edge $e \in E$, and of each constraint as having the following meaning.

1. For each node, the entering load is equal to the exiting load.
2. For each color $a \in [k-1]$, the load on the edges colored by a is equal to the load on the edges colored by k.
3. Every load is non-negative.
4. The total load is positive.

The following lemma justifies the introduction of the balance system.

Lemma 8. *There exists a set $\mathcal{L}$ of simple loops in G with zero as n.l.c. iff the balance system for G is feasible.*

Proof. (Sketch) [only if] If there exists an n.l.c. of $\mathcal{L}$ with value zero, let c_σ be the coefficient associated with a loop $\sigma \in \mathcal{L}$. We can construct a vector $x \in R^m$ that satisfies the balance system. First, define $h(e,\sigma)$ as 1 if the edge e is in σ, and 0 otherwise. Then, we set $x_e = \sum_{\sigma \in \mathcal{L}} c_\sigma h(e,\sigma)$. Considering that, for all $\sigma \in \mathcal{L}$ and $v \in V$, it holds that $\sum_{e \in {}_vE} h(e,\sigma) = \sum_{e \in E_v} h(e,\sigma)$, it is a matter of algebra to show that x satisfies the balance system.

[if] If the system is feasible, since it has integer coefficients, it has to have a rational solution. Moreover, all constraints are either equalities or inequalities of the type $a^T x \sim 0$, for $\sim \in \{>, \geq\}$. Therefore, if x is a solution then cx is also a solution, for all $c > 0$.

Accordingly, if the system has a rational solution, it also has an integer solution $x \in \mathbb{Z}^m$. Due to the constraints (3), such solution must be non-negative. So, in fact $x \in \mathbb{N}^m$.

Then, we consider each component x_e of x as the number of times the edge e is used in a set of loops, and we use x to construct such set with an iterative algorithm. At the first step, we set $x^1 = x$, we take a non-zero component x^1_e of x^1, we start constructing a loop with the edge e, and then we subtract a unit from x^1_e to remember that we used it. Next, we look for another non-zero component $x^1_{e'}$ such that e' exits from the node e enters in. It is possible to show that the edge e' can always be found. Then, we add e' to the loop and we subtract a unit from $x^1_{e'}$. We continue looking for edges e' with $x^1_{e'} > 0$ and exiting from the last node added to the loop, until we close a loop, i.e., until the last edge added enters in the node the first edge e exits from. After constructing a loop, we have a residual vector x^2 for the next step. If such vector is not zero, we construct another loop, and so on until the residual vector is zero. In the end we have a set of (not necessarily simple) loops, and we show that it has zero as n.l.c. Finally, we decompose those loops in simple loops with the algorithm of Lemma 3, and we obtain a set $\mathcal{L}$ of simple loops having zero as a natural linear combination. □

Since in a strongly connected graph all loops are connected, from the previous lemma, we have:

Corollary 1. *If G is strongly connected, there exists a balanced path in G iff the balance system for G is feasible.*

In order to solve the balance problem in G, first we compute the maximal connected components of G using the classical algorithm [CLRS01]. This algorithm is polynomial in n and m. Then, in each component we compute whether the balance system is feasible, by using the polynomial algorithm for feasibility of sets defined by linear constraints [NW88]. This second algorithm is used at most n times and it is polynomial in the number of constraints $(n+m+k)$ and in the logarithm of the maximum modulus of a coefficient in a constraint (in our case, the maximum modulus is 1).

Theorem 5. *The balance problem is in P.*

We remark that the feasibility algorithm can also provide the value of a solution to the system in input. By the proof of Lemma 8, such a solution allows us to compute in polynomial time a set of connected simple loops and the coefficients of an n.l.c. of value zero. As shown in the *if* part of the proof of Theorem 1, this in turn allows us to constructively characterize a balanced path in the graph.

5 Solving the Bounded Difference Problem

In this section, we solve the bounded difference problem using the same approach as in Section 4.

Definition 2. *Let $G = (V, E)$ be a k-colored graph with $m = |E|$, $n = |V|$, and $s_G = \min\{n + k - 1, m\}$, and let $u \in V$ be a node. We call* bounded difference system *for (G, u) the following system of equations on the set of variables $\{x_e, y_e \mid e \in E\}$.*

1-4. The same constraints as in the balance system for G

5. for all $v \in V \setminus \{u\}$	$\sum_{e \in E_v} y_e - \sum_{e \in {}_vE} y_e = \sum_{e \in {}_vE} x_e$
6.	$\sum_{e \in {}_uE} y_e - \sum_{e \in E_u} y_e = \sum_{v \in V \setminus \{u\}} \sum_{e \in {}_vE} x_e$
7. for all $e \in E$	$y_e \geq 0$
8. for all $e \in E$	$y_e \leq (m \cdot s_G!) x_e$.

The bounded difference system has $2m$ variables and $3m + 2n + k$ constraints. It helps to think of the vectors x and y as two loads associated to the edges of G. The constraints 1-4 are the same constraints of the balance problem for G, and they ask that x should represent a set of simple loops of G having zero as a natural linear combination.

The constraints 5-8 are *connection constraints*, asking that y should represent a connection load, from u to every other node of the simple loops defined by x, and carried only on the edges of those loops. Thus, constraints 5-8 ask that the loops represented by x should be overlapping, because of Lemma 2.

5. Each node $v \in V \setminus \{u\}$ absorbs an amount of y-load equal to the amount of x-load traversing it. These constraints ensure that the nodes belonging to the x-solution receive a positive y-load.
6. Node u generates as much y-load as the total x-load on all edges, except the edges exiting u.
7. Every y-load is non-negative.
8. If the x-load on an edge is zero, then the y-load on that edge is also zero. Otherwise, the y-load can be at most $m \cdot s_G!$ times the x-load. More details on the choice of this multiplicative constant follow.

In Lemma 9, whose proof can be found in the Appendix, we show that if there is a solution x of the balance system, then there is another solution x' whose non-zero components are greater or equal to 1 and less than or equal to $s_G!$, so that $\sum_{e \in E} x'_e \leq m \cdot s_G!$. In this way, the constraints (8) allow each edge that has a positive x-load to carry as its y-load all the y-load exiting from u.

Lemma 9. *Let $G = (V, E)$ be a k-colored graph, with $|V| = n$, $|E| = m$, and $s_G = min\{n + k - 1, m\}$. For all solutions x to the balance system for G there exists a solution x' such that, for all $e \in E$, it holds $(x_e = 0 \Rightarrow x'_e = 0)$ and $(x_e > 0 \Rightarrow 1 \leq x'_e \leq s_G!)$. As a consequence, $1 \leq \sum_{e \in E} x'_e \leq m \cdot s_G!$.*

The following lemma states that the bounded difference system can be used to solve the bounded difference problem.

Lemma 10. *There exists an overlapping set of simple loops in G, passing through a node u and having zero as n.l.c. iff the bounded difference system for (G, u) is feasible.*

Proof. [only if] Let $\mathcal{L}$ be an overlapping set of simple loops having an n.l.c. of value zero. Let c_σ be the coefficient associated with the loop $\sigma \in \mathcal{L}$ in such linear combination. We start by constructing a solution $x \in \mathbb{R}^m$ to the balance system as follows. Define $h(e, \sigma) \in \{0, 1\}$ as 1 if the edge e belongs to the loop σ, and 0 otherwise. We set $x_e = \sum_{\sigma \in \mathcal{L}} c_\sigma h(e, \sigma)$. We have that x is a solution to the balance system for G, or equivalently that it satisfies constraints (1)-(4) of the bounded difference system for (G,u).

By Lemma 9, there exists another solution $x' \in \mathbb{R}^m$ to the balance system, such that $x_e = 0 \Rightarrow x'_e = 0$ and $x_e > 0 \Rightarrow 1 \leq x'_e \leq s_G!$. If any loop of the overlapping set $\mathcal{L}$ passes through u, by Lemma 2, there exists a path ρ_v from u to any node v occurring in $\mathcal{L}$. We set $y_e = \sum_{v \in V' - \{u\}} (h(e, \rho_v) \sum_{e \in_v E} x'_e)$. Simple calculations show that (x', y) is a solution to the bounded difference system for (G,u).

[if] If there exists a vector $(x,y) \in \mathbb{R}^{2m}$ satisfying the bounded difference system, then like we did in the second part of Lemma 8, using x, we can construct a set of simple loops $\mathcal{L}$ having zero as n.l.c. Since $\sum_{e \in_u E} y_e - \sum_{e \in E_u} y_e = \sum_{v \in V - \{u\}} \sum_{e \in_v E} x_e$, we have that u belongs to at least one edge used in the construction of $\mathcal{L}$. If we set $G' = (V', E')$ as the subgraph of G induced by $\mathcal{L}$, we are able to show by contradiction that there is a path in G' from u to every other node of V'. Indeed if for some $v \in V' - \{u\}$ there is no path in G' from u to v then there is some load exiting from u that cannot reach its destination using only edges of G'. Since the constraints (8) make it impossible to carry load on edges of G that are not used in $\mathcal{L}$, the connection constraints cannot be satisfied. So, for all $v \in V'$ there is a path in G' from u to v. By Lemma 2, $\mathcal{L}$ is overlapping. □

In order to solve the bounded difference problem in G, for all $u \in V$ we check whether the bounded difference system for (G,u) is feasible, by using a polynomial time algorithm for feasibility of linear systems [NW88]. This algorithm is used at most n times and it is polynomial in the number of constraints $(2n+3m+k)$ and in the logarithm of the maximum modulus M of a coefficient in a constraint. In our case, $M = m \cdot s_G!$. Using Stirling's approximation, we have $\log(m \cdot s_G!) = \log(m) + \Theta(s_G \log(s_G))$. Therefore, we obtain the following.

Theorem 6. *The bounded difference problem is in P.*

Acknowledgements. The second author would like to thank Luca de Alfaro and Krishnendu Chatterjee for the fruitful discussions on a preliminary version of this work.

References

[AH98] R. Alur and T.A. Henzinger. Finitary fairness. *ACM Trans. on Programming Languages and Systems*, 20(6), 1998.

[CLRS01] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 2001. Second Edition.

[dA00] L. de Alfaro. From fairness to chance. *ENTCS*, 22, 2000.

[DJP03] N. Dershowitz, D. N. Jayasimha, and S. Park. Bounded fairness. In *Verification: Theory and Practice*, LNCS 2772, pages 304–317. Springer, 2003.

[Fra86] N. Francez. *Fairness*. Springer, 1986.

[Kwi89] M. Kwiatkowska. Survey of fairness notions. *Information and Software Technology*, 31(7):371–386, 1989.

[LPS81] D. Lehmann, A. Pnueli, and J. Stavi. Impartiality, justice and fairness: The ethics of concurrent termination. In *Proc. 8th Int. Colloq. Aut. Lang. Prog.*, LNCS 115, pages 264–277. Springer, 1981.

[MP91] Z. Manna and A. Pnueli. *The Temporal Logic of Reactive and Concurrent Systems: Specification*. Springer, 1991.

[NW88] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization*. Wiley-Interscience, 1988.

[Par66] R.J. Parikh. On Context-Free Languages. *Journal of the ACM*, 13(4):570–581, 1966.

[Sch86] A. Schrijver. *Theory of linear and integer programming*. John Wiley and Sons, 1986.

[Wik09] Wikipedia. Crop rotation, 2009. http://www.wikipedia.com.

A Additional Proofs

A.1 Proof of Item 2 of Lemma 1

In the following lemma, we use the trivial property that, given a sequence of real numbers $r_1, \ldots, r_n$ such that $\sum_{i=1}^{n} r_i = C$, then at least one r_i is greater than or equal to $\frac{C}{m}$.

Proof. [only if] The property is easy to prove since when the inner limits exist, the limit of the difference is equal to the difference of the inner limits.

[if] We first show that *(i)* $\lim_{j\to\infty} |\rho^{\leq j}|_k / j = l = 1/k$. Then, we show that *(ii)* for all $a \in [k-1]$, the sequence $\{|\rho^{\leq j}|_a / j\}_j$ also converges to l, or else the difference sequence $\{\mathit{diff}_{a,k}(\rho^{\leq j})/j\}_j$ would not converge to zero.

First we show (i). Indeed, if by contradiction the sequence is not convergent to l we have that

$$\exists \varepsilon > 0 \,.\, \forall m \in \mathbb{N} \,.\, \exists n_m \geq m \,.\, \left(\frac{|\rho^{\leq n_m}|_k}{n_m} > l + \varepsilon \;\text{ or }\; \frac{|\rho^{\leq n_m}|_k}{n_m} < l - \varepsilon \right).$$

The points $\{n_m\}_m$ form a sequence, from which we can extract two subsequences $\{n_{m_i}\}_i$, given by all the points such that $|\rho^{\leq n_{m_i}}|_k / n_{m_i} > l + \varepsilon$, and $\{n_{m'_i}\}_i$, given by all the points such that $|\rho^{\leq n_{m'_i}}|_k / n_{m'_i} < l - \varepsilon$. We know that at least one subsequence is infinite. Moreover, for all $i \in \mathbb{N}$, we have that $m_i \geq i$ and $m'_i \geq i$. We remind that for every path ρ, we have $\sum_{a=1}^{k} |\rho|_a = |\rho|$, since all edges are colored. Then, for all $j \in \mathbb{N}$, we have $\sum_{a=1}^{k-1} |\rho^{\leq j}|_a = j - |\rho^{\leq j}|_k$. Now there are two possible situations: $\{n_{m_i}\}_i$ is infinite or $\{n_{m'_i}\}_i$ is infinite. The two situations are dual and we only discuss the first one.

Assume that $\{n_{m_i}\}_i$ is infinite. For all $i \in \mathbb{N}$, we have that $|\rho^{\leq n_{m_i}}|_k > (l+\varepsilon) n_{m_i}$. Then, $\sum_{a=1}^{k-1} |\rho^{\leq n_{m_i}}|_a \leq n_{m_i} - n_{m_i}(l+\varepsilon) = n_{m_i}(1 - l - \varepsilon)$. So, there exists at least one color $a \in [k-1]$ such that $|\rho^{\leq n_{m_i}}|_a \leq n_{m_i}(1-l-\varepsilon)/(k-1) = n_{m_i}(l - (\varepsilon/(k-1)))$. Then, there is a color $a \in [k-1]$ and a subsequence $\{n_{m_i^a}\}_i$ of $\{n_{m_i}\}_i$ such that for all $i \in \mathbb{N}$ we have that $|\rho^{\leq n_{m_i^a}}|_a \leq n_{m_i^a}(l - (\varepsilon/(k-1)))$. Moreover, for all $i \in \mathbb{N}$, we have that $m_i^a \geq i$. For all $i \in \mathbb{N}$, we then have that

$$\frac{\mathit{diff}_{k,a}(\rho^{\leq n_{m_i^a}})}{n_{m_i^a}} = \frac{|\rho^{\leq n_{m_i^a}}|_k - |\rho^{\leq n_{m_i^a}}|_a}{n_{m_i^a}} \geq (l+\varepsilon) - \left(l - \frac{1}{k-1}\varepsilon\right) = \frac{k}{k-1}\varepsilon.$$

Then, due to $\varepsilon' = k\varepsilon/(k-1)$ the following holds.

$$\exists \varepsilon' > 0 \,.\, \forall i \in \mathbb{N} \,.\, \exists m_i^a \geq i \,.\, \frac{\mathit{diff}_{k,a}(\rho^{\leq n_{m_i^a}})}{n_{m_i^a}} \geq \varepsilon'.$$

So, the sequence $\{\mathit{diff}_{k,a}(\rho^{\leq n_{m_i^a}})/n_{m_i^a}\}_i$ does not converge to zero, so does not the sequence $\{\mathit{diff}_{k,a}(\rho^{\leq j})/j\}_j$, since $\{\mathit{diff}_{k,a}(\rho^{\leq n_{m_i^a}})/n_{m_i^a}\}_i$ is one of its subsequences.

So, we have $\lim_{j\to\infty}(|\rho^{\leq j}|_k)/j = l$, i.e.,

$$\forall \varepsilon > 0 \,.\, \exists m \in \mathbb{N} \,.\, \forall n \geq m \,.\quad l - \varepsilon < \frac{|\rho^{\leq n}|_k}{n} < l + \varepsilon. \tag{1}$$

Now, we show (ii). Assume by contradiction that $\{|\rho^{\leq j}|_a/j\}$ does not converge to l, for a certain $a \in [k-1]$. Then, we have

$$\exists \varepsilon > 0 \,.\, \forall m \in \mathbb{N} \,.\, \exists n_m \geq m \,.\, \left(\frac{|\rho^{\leq n_m}|_a}{n_m} > l + \varepsilon \text{ or } \frac{|\rho^{\leq n_m}|_a}{n_m} < l - \varepsilon \right). \tag{2}$$

Let $\bar{\varepsilon}$ be a witness for (2). By (1), there is $\bar{m} \in \mathbb{N}$ such that for all $n \geq \bar{m}$, we have $l - \bar{\varepsilon}/2 < (|\rho^{\leq n}|_k)/n < l + \bar{\varepsilon}/2$. So, for all $m \geq \bar{m}$, there exists $n_m \geq m$ such that one of the two following conditions holds (depending which disjunction in (2) holds).

1. $\frac{|\rho^{\leq n_m}|_a}{n_m} > l + \bar{\varepsilon}$, and $l - \frac{\bar{\varepsilon}}{2} < \frac{|\rho^{\leq n_m}|_k}{n_m} < l + \frac{\bar{\varepsilon}}{2}$. So, $\frac{diff_{a,k}(\rho^{\leq n_m})}{n_m} > \frac{\bar{\varepsilon}}{2}$.
2. $\frac{|\rho^{\leq n_m}|_a}{n_m} < l - \bar{\varepsilon}$, and $l - \frac{\bar{\varepsilon}}{2} < \frac{|\rho^{\leq n_m}|_k}{n_m} < l + \frac{\bar{\varepsilon}}{2}$. So, $\frac{diff_{a,k}(\rho^{\leq n_m})}{n_m} < -\frac{\bar{\varepsilon}}{2}$.

Summarizing,

$$\exists \bar{\varepsilon} > 0 \,.\, \exists \bar{m} \in \mathbb{N} \,.\, \forall m \geq \bar{m} \,.\, \exists n_m \geq m \,.\, \left| \frac{diff_{a,k}(\rho^{\leq n_m})}{n_m} \right| > \frac{\bar{\varepsilon}}{2}.$$

Thus, we have that $\{diff_{a,k}(\rho^{\leq n})/n\}_n$ does not converge to 0, which is a contradiction. □

A.2 Proof of Lemma 4

We first introduce a preliminary lemma.

Lemma 11. *Let $Ax = \mathbf{0}$ be a linear homogeneus system with $A \in \mathbb{Q}^{n \times m}$. If the system has a solution x such that $x \geq \mathbf{0}$ and $\sum_{i=1}^m x_i = 1$, then it has a solution with all natural components, and at least one strictly positive component.*

Proof. Let S be the set containing all and only the solutions x of $Ax = \mathbf{0}$, with all non-negative components and such that $\sum_{i=1}^m x_i = 1$. By hypothesis, S is not empty. Let $A' = (\frac{A}{1,1,\ldots,1})$ and $b' = (\frac{\mathbf{0}}{1})$, we have $S = \{x \in \mathbb{R}^m \,|\, A'x = b', x \geq \mathbf{0}\}$. By a well known result in linear programming (see, for instance, Theorem 3.5 of [NW88]), S contains a *basic* solution, i.e., there exists a non-singular submatrix $C \in \mathbb{R}^{r \times r}$ of A', given by the columns $i_1, \ldots, i_r$ and the rows $j_1, \ldots, j_r$ of A', such that there is a point $(z_1, \ldots, z_m) \in S$ such that $z' = (z_{i_1}, \ldots, z_{i_r})$ is the unique solution to the system of linear equations $Cz' = b$ where $b = (b'_{j_1}, \ldots, b'_{j_r})^T$, and for all $i \notin \{i_1, \ldots, i_r\}$, $z_i = 0$. By Cramer's theorem, for all $k \in [r]$, we have $z_{i_k} = \frac{det(C'_{i_k})}{det(C')}$ where C'_{i_k} is the matrix obtained from C' by replacing the i_k-th column with the column vector b. Since the determinant of a rational matrix is rational, z is a point of S with all rational coefficients, i.e., z is a solution of $Ax = \mathbf{0}$ with all non-negative rational coefficients such that $\sum_{i=1}^m z_i = 1$. Clearly, z has at least one positive coefficient. Now, since the system $Ax = \mathbf{0}$ is homogeneus, by multiplying each component of z by the least common denominator of all components, we obtain the thesis. □

Now we are ready to prove Lemma 4.

Proof. Let $A = \{(x_{1,1},\dots,x_{1,d}),\dots,(x_{m,1},\dots,x_{m,d})\}$ and $f : \mathbb{R}^m \mapsto \mathbb{R}_+$ be the function $f(c_1,\dots,c_m) = \max_{1\le i\le d}\{|\sum_{j=1}^m c_j \cdot x_{j,i}|\}$. By construction, f is a continuous function. Let now $K \subset \mathbb{R}^m$ be the set $\{(c_1,\dots,c_m) \in [0,1]^m \mid \sum_{i=1}^m c_i = 1\}$. Note that $\mathbf{0} \notin K$ and that K is compact, since it is a finite dimensional space defined by a linear equation. Hence, by Weierstrass' Theorem, f admits a minimum value M on K. Now, since A is a set of convexly independent vectors, M must be strictly positive. Indeed, if by contradiction $M = 0$, there is a non-null vector $(c_1,\dots,c_m) \in K$ such that $\sum_{j=1}^m c_j \cdot x_{j,1} = \dots = \sum_{j=1}^m c_j \cdot x_{j,d} = M = 0$. By Lemma 11 the system $\sum_{j=1}^m c_j \cdot x_{j,i} = 0$ has a natural solution c with at least one positive component. This solution gives rise to an n.l.c. of some vectors of A (those corresponding to positive components of c) with value zero, which contradicts the hypotheses on A.

Then, consider the sequence $\{(a_{n,1},\dots,a_{n,d})\}_n$ and its partial sums $S_{n,i} = \sum_{j=0}^n a_{j,i}$. Moreover, let $\delta_{i,n}$ be the number of times for which the vector $(x_{i,1},\dots,x_{i,d})$ occurs in the previous sequence up to position n and let $c_{i,n} = \delta_{i,n}/n$. Then, $(S_{n,1},\dots,S_{n,d}) = \sum_{i=1}^m \delta_{i,n} \cdot (x_{i,1},\dots,x_{i,d}) = n \cdot \sum_{i=1}^m c_{i,n} \cdot (x_{i,1},\dots,x_{i,d})$. Since we have $\sum_{i=1}^m \delta_{i,n} = n$ for all $n \in \mathbb{N}$, it is obvious that $(c_{1,n},\dots,c_{m,n}) \in K$. By the convex independence hypothesis on A, it holds that for all $n \in \mathbb{N}$ there exists at least an index i, with $1 \le i \le d$, such that $S_{n,i} \ne 0$. Let $\{j_n\}_n$ be an index sequence such that $|S_{n,j_n}| = \max_{1\le i\le d}\{|S_{n,i}|\} > 0$, for all $n \in \mathbb{N}$. Since $\{j_n\}_n$ can assume at most d different values, there exists a value h which occurs infinitely often in it. Let $\{h_i\}_i$ be the index sequence such that $j_{h_i} = h$ and there is no $l \in]h_i, h_{i+1}[$ with $j_l = h$. Then, from $\{S_{n,h}\}_n$ we can construct the extracted sequence $\{S_{h_i,h}\}_i$. Now, we have that $|S_{h_i,h}| = \max_{1\le j\le d}\{|S_{h_i,j}|\} = h_i \cdot \max_{1\le j\le d}\{|\sum_{k=1}^m c_{k,h_i} \cdot x_{k,j}|\} = h_i \cdot f(c_{1,h_i},\dots,c_{1,h_i}) \ge h_i \cdot M > 0$. Hence, $\lim_{i\to\infty} \frac{|S_{h_i,h}|}{h_i} \ge M > 0$. Since $\left\{\frac{|S_{h_i,h}|}{h_i}\right\}_i$ is an extracted sequence of $\left\{\frac{|S_{n,h}|}{n}\right\}_n$, we finally obtain that $\lim_{n\to\infty} \frac{S_{n,h}}{n} \ne 0$. □

A.3 Proof of Lemma 7

Proof. In a 2-colored graph, the difference vector of any path ρ is simply an integer. Let $\mathcal{L}$ be a connected set of simple loops with zero as a n.l.c. If $\mathcal{L}$ contains a simple loop σ such that $diff(\sigma) = 0$, then $\{\sigma\}$ is trivially an overlapping set.

If $\mathcal{L}$ contains no perfectly balanced loop, then all the difference vectors of the loops in $\mathcal{L}$ cannot have the same sign, otherwise it is not possible to have a non-trivial natural combination $\sum_{\sigma\in\mathcal{L}} c_\sigma diff(\sigma) = 0$.

Thus, let $diff(\sigma) > 0$ and $diff(\sigma') < 0$, for $\sigma, \sigma' \in \mathcal{L}$. If σ and σ' are overlapping, then $\{\sigma,\sigma'\}$ is the overlapping set we are looking for. If σ and σ' are not overlapping, since they are connected, there exist a path ρ_1 from σ to σ' and a path ρ_2 from σ' to σ. So, there exist four indexes i, i', j, j' such that $\rho_1(i)$ is the last node of ρ_1 in σ, $\rho_1(j)$ is the first node of ρ_1 in σ', $\rho_2(i')$ is the last node of ρ_2 in σ', and $\rho_2(j')$ is the first node of ρ_2 in σ. Then, within the loop σ there exists a simple path ρ from $\rho_2(j')$ to $\rho_1(i)$ and, within the loop σ', there exists a simple path ρ' from $\rho_1(j)$ to $\rho_2(i')$. We then set $\rho'_1 = \rho_1(i)\dots\rho_1(j)$ and $\rho'_2 = \rho_2(i'),\dots,\rho_2(j')$. We observe that the pairs of paths (ρ,ρ'_1), (ρ',ρ'_1), (ρ',ρ'_2), and (ρ,ρ'_1) have only one node in common. Moreover, ρ and ρ' have no node in common since σ and σ' are not overlapping. So, the loop $\sigma'' = \rho\rho'_1\rho'\rho'_2$ is not simple iff ρ'_1 and ρ'_2 have a node in common. Now, observe that two loops π_1 and π_2 with difference vectors of opposite sign have zero as n.l.c. with coefficients $|diff(\pi_2)|$ and $|diff(\pi_1)|$, since $|diff(\pi_2)|diff(\pi_1) + |diff(\pi_1)|diff(\pi_2) = 0$. We conclude with the following case analysis.

1. If σ'' is simple, then (σ, σ'') and (σ', σ'') are pairs of overlapping sets.
   (a) If $\mathit{diff}(\sigma'') = 0$ then $\{\sigma''\}$ is an overlapping set having zero as an n.l.c.
   (b) If $\mathit{diff}(\sigma'') > 0$ then $\{\sigma', \sigma''\}$ is an overlapping set having zero as an n.l.c.
   (c) If $\mathit{diff}(\sigma'') < 0$ then $\{\sigma, \sigma''\}$ is an overlapping set having zero as an n.l.c.
2. If ρ'_1 and ρ'_2 have nodes in common, there exist two indexes k, k' such that $\rho'_1(k) = \rho'_2(k')$. So, we can construct two loops $\sigma'_1 = \rho\rho'_1(1)\dots\rho'_1(k)\dots\rho'_2(|\rho'_2|)$ and $\sigma'_2 = \rho'\rho'_2(1)\dots\rho'_2(k')\dots\rho'_1(|\rho'_1|)$.
   (a) If $\mathit{diff}(\sigma'_i) = 0$, for some $i \in \{0,1\}$, then $\{\sigma'_i\}$ is an overlapping set having zero as an n.l.c.
   (b) If $\mathit{diff}(\sigma'_2) > 0$ then $\{\sigma', \sigma'_2\}$ is an overlapping set having zero as an n.l.c.
   (c) If $\mathit{diff}(\sigma'_1) < 0$ then $\{\sigma, \sigma'_1\}$ is an overlapping set having zero as an n.l.c.
   (d) If $\mathit{diff}(\sigma'_1) > 0$ and $\mathit{diff}(\sigma'_2) < 0$, then $\{\sigma'_1, \sigma'_2\}$ is an overlapping set having zero as an n.l.c.

□

A.4 Proof of Theorem 4

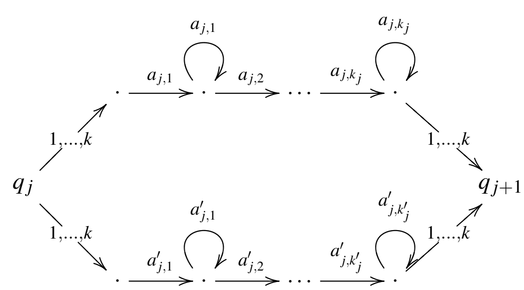

Fig. 2. Proof of Theorem 4: The j-th subgraph G_j of G.

Proof. We prove the statement by a reduction from 3SAT which is known to be NP-hard [CLRS01].

Given a 3SAT formula φ on n variables $x_1, \dots, x_n$ with k clauses $C_1, \dots, C_k$, we construct a k-colored graph G such that each color i is associated with the clause C_i. Precisely, for each variable x_j, we construct a subgraph G_j of G with a starting node q_j and an ending node q_{j+1}, as shown in Figure 2. For $1 \le j \le n$, the labels $a_{j,1}, \dots, a_{j,k_j}$ are the colors corresponding to the clauses in which x_j occurs affirmed and $a'_{j,1}, \dots, a'_{j,k'_j}$ are the colors of the clauses in which x_j occurs negated. Moreover, the edges labeled with $1, \dots, k$ concisely represent a sequence of k edges, each labeled with a different color. Finally, the graph G is obtained by concatenating each graph G_j with G_{j+1}, as they share the node q_{j+1}, for $1 \le j < n$.

We show that the formula φ is satisfiable iff there exists a perfectly balanced path in G from q_1 to q_{n+1}.

First, assume that φ is satisfiable. Then, there exists a truth assignment for the variables that satisfies φ. Using this assignment, we construct a perfectly balanced path in which each color appears exactly $2n+3$ times. In particular, for all subgraphs G_j, the path takes the upper branch if x_j is assigned true and the lower branch otherwise. For each clause C_i, let L_i be the indexes of the variables that render C_i true, under the given truth assignment. We obtain that the constructed path passes through at least $2n+|L_i|$ non-self-loop edges colored with i. This holds because at each subgraph it passes through the edges labeled $1,\dots,k$ once at the beginning and once at the end. Moreover, for all $j \in L_i$, the path passes through another non-self-loop edge labeled with i in G_j. Since $|L_i| \geq 1$, the path may pass through a self-loop labeled with i at least once in the graph. Thus, by taking $3-|L_i|$ times one of those self-loops, we get the desired number $2n+3$ of occurrences of i, for all colors i.

Conversely, assume that there exists a perfectly balanced path from q_0 to q_{n+1}. For all subgraphs G_j the path takes either the upper or the lower branch. Then, there are two possible situations:

1. Each color occurs $2n+l$ times with $l \geq 1$. We define the assignment in the following way: we set x_j to *true* if the path takes the upper branch in the subgraph G_j, and to *false* otherwise. We claim that such assignment satisfies φ. For all colors i the path passes through an i-colored edge α such that it is not a self-loop and it is not a starting or an ending edge of a subgraph G_j (those edges are the first $2n$). Such edge α is on a branch of a subgraph G_j, consequently the assignment for x_j satisfies the clause C_i. Being i arbitrary, all clauses C_i are satisfied by the assignment of the variable.
2. Each color occurs $2n$ times in the path. We define the assignment as follows: we set x_j to *true* if the path takes the lower branch in G_j, and to *false* otherwise. We claim that such assignment satisfies φ. For all colors i there exists at least one variable x_j appearing in the clause C_i. However, the path does not pass through any edge colored with i, except the mandatory edges at the beginning and end of each G_j. Then, in G_j the path takes the branch opposite to the assignment of x_j that makes C_i true Then, the opposite assignment of x_j (the one we choose) makes C_i true. □

A.5 Proof of Lemma 9

We first introduce two preliminary lemmas.

Lemma 12. *Let $t \in \mathbb{N}$ be a natural number and $A \in [t]_0^{m\times m}$ be a square matrix, then $|det(A)| \leq t^m m!$. Moreover, if A is not singular then $|det(A)| \geq 1$.*

Proof. We prove the first statement by induction on m.

1. If $m=1$ then $|det(A)| = |a_{1,1}| \leq t$.
2. If the statement holds for $m-1$, then for any $j \in [m]$ it holds that

$$det(A) = \sum_{i=1}^{m} (-1)^{i+j} a_{i,j} det(M_{i,j}),$$

where $M_{i,j} \in [t]_0^{(m-1)\times(m-1)}$ is a matrix obtained from A by removing the i-th row and the j-th column. So, $|det(A)| \leq |\sum_{i=1}^{m} a_{i,j} det(M_{i,j})| \leq \sum_{i=1}^{m} |a_{i,j}||det(M_{i,j})| \leq \sum_{i=1}^{m} t \cdot t^{m-1}(m-1)! = (tm)t^{m-1}(m-1)! = t^m m!$.

Moreover, if A is not singular, since A has an integer determinant it must be $|det(A)| \geq 1$. $\square$

Lemma 13. *Let t be a natural number and $A \in [t]_0^{n\times m}$, $A' \in [t]_0^{n'\times m}$, $B \in [t]_0^{n\times 1}$, and $B' \in [t]_0^{n'\times 1}$ be four matrices. Let $S = \{x \in \mathbb{R}^m \,|\, Ax \geq B, A'x \geq B', x \geq \mathbf{0}\}$ and $M = min\{n+n', n+m\}$. If S is not empty, then there exists a vector $x \in S$ such that $x \in \mathbb{Q}^m$ and every component x_i is less than or equal to $k = M!t^M$.*

Proof. Let $I \in \mathbb{N}^{n\times n}$ be the identity matrix. At first, we convert every inequality of the system $Ax \geq B$ in an equivalent equality by adding a new variable: the inequality $\sum_{j=1}^m a_{i,j}x_j \geq b_i$ becomes $\sum_{j=1}^m a_{i,j}x_j = b_i + y_i$ with $y_i \geq 0$. If we set $C = (\frac{A}{A'} | \frac{-I}{\mathbf{0}}) \in \mathbb{R}^{(n+n')\times(n+m)}$, and $B'' = (\frac{B}{B'})$ we can define the set $S' = \{(x,y) \in \mathbb{R}^{m+n} \,|\, C(x,y) = B'', (x,y) \geq \mathbf{0}\}$. It is easy to see that $S = \{x \in \mathbb{R}^m \,|\, \exists y \in \mathbb{R}^n \,.\, (x,y) \in S'\}$, thus in our hypothesis S' is not empty. We define r the rank of C, so $m \leq r \leq M$ since $-I$ is not singular submatrix. By a well known result in linear programming (see, for instance, Theorem 3.5 of [NW88]), the set S' has a *basic* solution, i.e. there exists a non-singular submatrix $C' \in \mathbb{R}^{r\times r}$ of C, given by the columns $i_1, \ldots, i_r$ and by the rows $j_1, \ldots, j_r$ of C, such that in S there is the point $(z_1, \ldots, z_{m+n}) \in \mathbb{R}^{m+n}$ such that $z' = (z_{i_1}, \ldots, z_{i_r})$ is the unique solution to the system of linear equations $Cz' = (b''_{j_1}, \ldots, b''_{j_r})^T = B'''$, and for all $j \notin \{i_1, \ldots, i_r\}$ $z_j = 0$. By Cramer's theorem, for all $k \in [r]$ we have $z_{i_k} = det(C'_{i_k})/det(C')$ where C'_{i_k} is the matrix obtained from C' by replacing the i_k-th column with the matrix B'''. So z' and z have components in $\mathbb{Q}$. Since $C', C'_{i_1}, \ldots, C'_{i_r} \in [t]_0^{r\times r}$, by Lemma 12 $|det(C_i)| \leq r!t^r$. Moreover, since C' is not singular we have $|det(C)| \geq 1$. In conclusion, $z_{i_k} \leq |det(C_i)|/|det(C)| \leq (r)!t^r \leq M!t^M$, as requested. $\square$

Now, we are ready to prove Lemma 9.

Proof. Let x be a solution to the balance system for G, and let I be the set of all edges e such that $x_e > 0$. By construction, $|I| > 0$. We represent the first two sets of equalities of the balance system in matrix form as $Dx = \mathbf{0}$. Then, the set of points satisfying the balance system is $P = \{y \in \mathbb{R}^m \,|\, Dy = \mathbf{0}, y \geq \mathbf{0}, \sum_{e\in E} y_e > 0\}$. Now the subset of P, $P' = \{y \in P \,|\, \forall e \in I \,.\, y_e \geq 1$ and $\forall e \notin I \,.\, y_e = 0\} = \{y \in P \,|\, \forall e \in E \,.\, (x_e > 0 \Rightarrow y_e > 1)$ and $(x_e = 0 \Rightarrow y_e = 0)\}$ is not empty. Indeed, the vector $z = x(\min_{e\in I} x_e)^{-1}$ is in P', since (i) $Dz = (\min_{e\in I} x_e)^{-1}Dx = \mathbf{0}$, (ii) for all $e \in I$, we have $z_e = x_e(\min_{e\in I} x_e)^{-1} \geq 1$, and (iii) for all $e \notin I$, we have $z_e = 0$.

The set of inequalities "$\forall e \in I \,.\, y_e \geq 1$" can be represented as the system of linear equations $Fy \geq \mathbf{1}$, with $\mathbf{1} \in \{1\}^{l\times 1}$. Similarly, the set of equalities "$\forall e \notin I \,.\, y_e = 0$" can be represented as $F'y = \mathbf{0}$. If we define $D' = (\frac{D}{F'}) \in \{-1,0,1\}^{(2n+k-l-1)\times m}$, we have $P' = \{y \in \mathbb{R}^m \,|\, D'y = \mathbf{0}, Fy \geq \mathbf{1}\}$. Since $D', F, \mathbf{1}, \mathbf{0}$ all have elements in $\{-1,0,1\}$, by Lemma 13 the set P contains an element $x' \in \mathbb{Q}^m$ such that for all $i \in [m]$, $x'_i \leq (\min\{2n+k-1, l+m\})! \leq (\min\{2n+k-1, n+m\})! = s_G!$, which concludes the proof. $\square$

B The Perfectly Balanced Finite Path Problem is NP-Complete

In this appendix, we report the proof sketch of the NP membership for the perfectly balanced finite path problem, as defined in Section 3.4. This is a new result that is not contained into the main paper, since proved after the submission of the final version of this work.

We recall a result of integer programming presented in [Sch86].

Lemma 14. *Let $A \in \mathbb{Z}^{n\times n}$ and $B \in \mathbb{Z}^{n\times 1}$. Let $S = \{x \in \mathbb{Z}^n \,|\, Ax \leq B\}$ be an integer convex set. If S is not empty then there exists a point $x \in S$ such that the sum of the components of x is bounded by $6n^3\varphi$, where φ is the maximum sum of the coefficients of an inequality of the system $Ax \leq B$.*

Definition 3. *Let $G = (V,E)$ be a k-colored graph and $u,w \in V$ be two distinct nodes. We call* perfectly balanced path system *for (G,u,w) the following system of equations on the set of variables $\{x_e, y_e \,|\, e \in E\}$.*

1. for all $v \in V \setminus \{u,w\}$	$\sum_{e\in E_v} x_e = \sum_{e\in_v E} x_e$
2.	$\sum_{e\in E_u} x_e = 1+\sum_{e\in_u E} x_e$
3.	$\sum_{e\in E_w} x_e = -1+\sum_{e\in_w E} x_e$
4. for all $a \in [k-1]$	$\sum_{e\in E(a)} x_e = \sum_{e\in E(k)} x_e$
5. for all $e \in E$	$x_e \geq 0$
6.	$\sum_{e\in E} x_e > 0$
7. for all $v \in V \setminus \{u\}$	$\sum_{e\in E_v} y_e - \sum_{e\in_v E} y_e = \sum_{e\in_v E} x_e$
8.	$\sum_{e\in_u E} y_e - \sum_{e\in E_u} y_e = \sum_{v\in V\setminus\{u\}} \sum_{e\in_v E} x_e$
9. for all $e \in E$	$y_e \geq 0$
10. for all $e \in E$	$y_e \leq (6(d-1)^3\varphi)x_e.$
11. for all $e \in E$	$x_e, y_e \in \mathbb{Z}$

where φ is the maximum sum of the coefficients of an inequality in the first six sets of constraints.

Let $m = |E|$ and $n = |V|$, the perfectly balanced path system has $2m$ variables and $3m+2n+k$ constraints. It helps to think of the vectors x and y as two integer loads associated to the edges of G. The constraints 1-6 are almost the same constraints of the balance problem for G, and they ask that x should represent a path from u to v and a set of simple loops such that the latter have a n.l.c. equal to the inverse of the color difference vector of the path.

The constraints 5-8 are *connection constraints*, asking that y should represent a connection load, from u to every other node of the simple loops defined by x, and carried only on the edges of those loops. Thus, the constraints 5-8 ask that the loops represented by x should be reachable by u, using only edges represented by x, similarly to the bounded difference system of Section 5. The only difference is the bound in the contraints 10, which is directly justified by Lemma 14.

Lemma 15. *There exists a perfectly balanced path in G from u to w iff the perfectly balanced path system (G,u,w) is feasible.*

The proof of the previous lemma is similar to that for the balance problem. Since the feasibility problem for an integer linear system is in NP, and by Theorem 4, we obtain the following.

Theorem 7. *The perfectly balanced finite path problem is NP-complete.*