MASTER'S THESIS

Interuniversity Master in Statistics and Operations Research UPC-UB

Title: Markov chains and applications in the generation of combinatorial designs

Author: Julia Calatayud Gregori

Advisor: Oriol Serra Albo

Department: Department of Mathematics

University: Universitat Politècnica de Catalunya

Academic year: 2017/2018

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

UNIVERSITAT DE BARCELONA

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Master's final thesis
Master's degree in Statistics and Operations Research

Markov chains and applications in the generation of combinatorial designs

Julia Calatayud Gregori

Advisor: Oriol Serra Albo

Department of Mathematics

TO MY DAD

Resum

Paraules clau: Cadena de Markov, Metropolis, K_n, disseny combinatori, coloració pròpia, potencial, entropia.

MSC2000: 60J10.

Aquesta Tesi tracta cadenes de Markov discretes i les seves aplicacions en la generació de dissenys combinatoris. Una conjectura sobre la generació de coloracions pròpies d'arestes del graf complet K_n, per a n parell, és abordada. Una coloració pròpia de les arestes, o simplement una coloració pròpia, és una coloració de les arestes tal que dues arestes adjacents no tenen el mateix color. Les coloracions pròpies d'arestes estan caracteritzades per minimitzar el potencial i maximitzar l'entropia. Implementem un algorisme al programari `R` per a generar coloracions pròpies començant per qualsevol coloració de K_n, identificant les coloracions com a estats en una cadena de Markov, on les probabilitats de transició es defineixen de manera que el potencial decreix o, alternativament, l'entropia creix. La conjectura afirma que l'algorisme convergeix en temps polinomial. Donem proves originals de la conjectura en K_4 i K_6, i proporcionem nous resultats i idees que podrien utilitzar-se per a una prova de la conjectura en el cas general K_n.

Abstract

Keywords: Markov chain, Metropolis, K_n, combinatorial design, proper coloring, potential, entropy.

MSC2000: 60J10.

This Thesis deals with discrete Markov chains and their applications in the generation of combinatorial designs. A conjecture on the generation of proper edge colorings of the complete graph K_n, for n even, is tackled. A proper edge coloring, or simply a proper coloring, is an edge coloring such that no two adjacent edges have the same color. Proper edge colorings are characterized by minimizing the potential and maximizing the entropy. We implement an algorithm in the software `R` to generate proper colorings from any arbitrary coloring of K_n, by identifying the colorings as nodes in a Markov chain, where transition probabilities are defined so that the potential decreases or, alternatively, the entropy increases. The conjecture states that the algorithm converges in polynomial time. We give original proofs of the conjecture in K_4 and K_6, and we provide new results and ideas that could be used to prove the conjecture in the general case K_n.

Notation

$\mathbb{Z}^+$: Positive integers: $1, 2, \ldots$
$\mathbb{R}$: Real numbers
$\forall$: For all
$a|b$: a divides b
$n \equiv k \bmod m$: The integer division between n and m has remainder k
$a \wedge b$: $\min\{a, b\}$
$\lfloor a \rfloor$: Integer part of the real number a
$\lceil a \rceil$: Ceiling of the real number a
gcd: Greatest common divisor
$A \cup B$: Union of the sets A and B
$A \uplus B$: Disjoint union of the sets A and B
$A \cap B$: Intersection of the sets A and B
$A \backslash B$: Set A minus the elements of the set B
A^c: Complementary of the set A
$A \subseteq B$: Set A is contained in set B
$A \times B$: Cartesian product of A and B
$|A|$: Cardinality of the set A, its number of elements
$\in$: Belongs to
$\notin$: Does not belong to
$[n]$: Set $\{1, \ldots, n\}$, $n \in \mathbb{Z}^+$
$\mathbb{1}_A$: Characteristic function on the set A
$|x|$: Absolute value of the real number x
Ω: Sample space
$X \sim \mu$: The random variable X follows the distribution μ
$\mathbb{P}$: Probability measure
$\mathbb{E}$: Expectation operator
ω: Generic element of Ω, an arbitrary outcome in our probability space
$\|\mu - \nu\|_{\mathrm{TV}}$: Total variation distance between the probability measures μ and ν
$\mathcal{X}$: Generic state space of our Markov chains
P: Generic transition matrix
π: Generic stationary or invariant distribution of P
$M(i, j)$: Entry (i, j) of the matrix M
$M^n(i, j)$: Entry (i, j) of the matrix M^n
M^T: Transpose of the matrix M

M^{-1}: Inverse of the matrix M
G: Generic graph
V, $V(G)$: Set of vertices of the graph G
E, $E(G)$: Set of edges of the graph G
$\mathcal{X}'(G)$: Chromatic index of the graph G
$\mathcal{X}(G)$: Chromatic number of the graph G
K_n: Complete graph with n vertices
A_G: Adjacency matrix of the graph G
$d(x)$: Degree of a vertex x (number of its neighbors)
C: Generic proper coloring of the edges of a graph
$a_{\nu,\mu}(C)$: In the coloring C, number of edges incident to vertex ν of color μ
ϕ: Potential function
H: Entropy function
$c_C(\mu)$: Number of μ-colored edges in the coloring C

Contents

Chapter 1
Introduction

One of the most important topics in Probability Theory is Markov chains. An introduction to this area is taught in the compulsory subject *Probability and Stochastic Processes* of the Master of Statistics and Operations Research. This Thesis was thought to increase my knowledge on this topic, which belongs to the foundations of Statistics. As a mathematician, I am primarily interested in the theory behind the applied Statistics that I learned in the Master. Our idea was to apply the theory of Markov chains to a recent open problem.

A Markov chain in discrete time is a sequence of random variables such that at each time, there is dependence only at the previous time. This is a generalization of the usual i.i.d. (independent and identically distributed) hypothesis for samples. There are two fundamental theorems in discrete Markov chains: the convergence theorem, which states, under some hypotheses, that the limiting behavior of the Markov chain corresponds to its stationary distribution; and the ergodic theorem, which states that the sample mean of the Markov chain tends to the expectation of its stationary distribution (generalization of the Law of Large Numbers) [**5**, **16**]. In this Thesis, Chapter 2 deals with the convergence theorem. Several proofs of this theorem are provided: using the total variation distance of two distributions, using the coupling of a pair of distributions and finally in the setting of graphs by using spectral techniques. Each proof provides a specific insight into the convergence of Markov chains. In the application we discuss in Chapter 4, we still use another approach which is suitable to analyze convergence in the so-called absorbing Markov chains, and the necessary material is presented in that chapter. For the material in Chapter 2 on convergence theorems and rapid mixing, we relied on the recent monograph by Levin and Peres [**12**], the standard textbook by Norris [**16**] and the survey on random walks in graphs by Lovász [**13**].

Markov chains have a great number of applications in different branches, some of them taught in subjects of the Master, such as bioinformatics, survival time analysis, epidemiology, modeling, etc. Markov chains are applied in computer science as well, with one of the most celebrated algorithms being the Metropolis algorithm. Invented by physicists in the middle of the 20th century, it turned out to be the key tool in the computational development of Bayesian statistics at the end of the 20th century [**11**, Chapter 6]. Algorithms related to Markov chains are studied in Chapter 3.

One of the most notable applications of Markov chains appears in the random generation of combinatorial objects. The seminal paper of Jerrum and Sinclair [**19**] and the monography by A. Sinclair which followed it [**20**] provided the theoretical ground for systematic use of Markov chains in random generation of combinatorial objects. The general idea is to

perform a random walk through a class of objects and deliver the object reached after a reasonable number of steps. A wealth of applications have followed with applications in statistics, biology, physics and various areas of mathematics.

Among the applications in statistics we find the problem of generating combinatorial designs. These are combinatorial structures which were introduced by Fisher in the early 40's in the context of experimental designs. The goal is to design statistical tests covering a range of varieties and a range of conditions in order to achieve a good balance in its distribution. The handbook by Colbourn and Dinitz [4] gives a panoramic view of the state of the art on the topic and its many applications in statistics and computer science.

One important class of combinatorial designs can be expressed in terms of the so-called Latin squares, square $n \times n$ matrices with pairwise distinct entries in each row and each column using n symbols. By themselves, Latin squares have been the object of extensive study and their applications in Statistics are wide, see for instance the monograph by Dénes and Keedwell [10]. The tight structure of Latin squares has posed a problem related to random generation. One of the first attempts was proposed by McKay and Wormald [14] and was followed by a Markov chain approach by Jacobson and Mattews [9] in the context of random generation of contingency tables. One of the problems related to the random generation of Latin squares is that its total number is only known asymptotically and only few exact values are known for small values of n.

In 2017 Dotan and Linial [7] suggested a new simpler approach to the random generation of symmetric Latin squares by using a Metropolis algorithm. The authors suggest several variations of the algorithms and describe some statistical behavior for small values of n. They manage to prove the convergence of one of these variations which involves a more intricate procedure, but they conjecture [7, Conjecture 1] that the simplest approach does converge in polynomial time (a detailed statement of the conjecture is given in Chapter 4). The study of this conjecture is one of the main purposes of this work. The problem is stated in the language of one-factorizations of complete graphs, a topic with extensive literature in the area of combinatorial designs (see e.g. the monograph by Wallis [21] on the topic): the goal is to generate a proper edge coloring of the complete graph K_n of even order n with $n-1$ colors (see Chapter 4 for precise definitions).

Taking as main reference the paper by Dotan and Linial [7], we develop an algorithm in the statistical software `R` to obtain a proper coloring from any arbitrary coloring of K_n with $n-1$ colors, being based on the potential and the entropy of colorings. Numerical experiments on the generation of proper edge colorings have been performed for over 30 years, but no mathematical proof has been achieved to justify the numerical results. One of the main goals of Chapter 4 is to tackle one of the conjectures arisen from the numerical experiments: Conjecture 1 from [7]. According to our numerical simulations, we restate the conjecture: the algorithm to generate proper colorings always converges for any order n in polynomial time. New proofs for K_4 and K_6 are obtained in our work, both computational and analytical, and we give ideas on how to tackle the conjecture in K_n. The theory of Markov Chains is essential to face the problem.

In the last chapter, conclusions on the original results obtained in this Thesis are drawn and further conjectures on Graph coloring are proposed.

Chapter 2
Markov chains. A convergence theorem

This chapter is concerned with Markov chains and a very important convergence theorem. In the first section, we will present basic theory about Markov chains, the necessary to understand the statement of the convergence theorem. Roughly speaking, under some conditions, the distribution of the Markov chain tends to behave as a stationary distribution as time increases. After this, in the second, third and fourth sections of this chapter, we will show three different proofs of this convergence theorem: in the second section, using the total variation distance between distributions; in the third section, making use of the coupling of a pair of distributions; and in the fourth section, using spectral theory to prove the convergence theorem in the particular case of a connected graph. In the fifth section of this chapter, we will see examples of applications of the convergence theorem. In the sixth and last section, we will study the so-called mixing times, which are related to how fast the distribution of the Markov chain tends to the stationary distribution.

1. Markov Chains

Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables in a common probability space with probability measure $\mathbb{P}$, that take values in a finite or countable set $\mathcal{X}$. This set $\mathcal{X}$ is called the state space, whereas the subscript n is called the time.

Definition 2.1. *We say that the sequence $\{X_n\}_{n\geq 0}$ is a Markov chain if*

$$\mathbb{P}(X_n = j | X_0 = i_0, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i) = \mathbb{P}(X_n = j | X_{n-1} = i).$$

We say that the Markov chain is homogeneous if $\mathbb{P}(X_n = j | X_{n-1} = i) = P(i,j)$ (i.e. the probability does not depend on the time n).

We say that $P = (P(i,j))_{i,j\in\mathcal{X}}$ is the transition matrix.

Our Markov chains will always be homogeneous and we will refer to the Markov chain just by specifying the transition matrix P.

Notice that the definition of Markov chain is equivalent to the following fact: Let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ be the σ-algebra generated by $X_0, \ldots, X_n$, $n \geq 0$. Then, $\mathbb{P}(X_n = j | \mathcal{F}_{n-1}) = \mathbb{P}(X_n = j | X_{n-1})$.

The intuitive idea of a Markov chain is that the future conditioned to the past and the present only depends on the present.

Let $\lambda_i = \mathbb{P}(X_0 = i)$, $i \in \mathcal{X}$, and $\lambda = (\lambda_i)_{i\in\mathcal{X}}$ be a row vector. By the Total Probability Theorem and the Markov condition,

$$\begin{aligned}\mathbb{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i_n) &= P(i_{n-1}, i_n)\mathbb{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ &= \dots = P(i_{n-1}, i_n)P(i_{n-2}, i_{n-1})\cdots P(i_0, i_1)\lambda_{i_0}.\end{aligned}$$

Then,

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_n = i_n) &= \sum_{i_1,\dots,i_{n-1}\in\mathcal{X}} \mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n) \\ &= \sum_{i_1,\dots,i_{n-1}\in\mathcal{X}} \lambda_{i_0} P(i_0, i_1)\cdots P(i_{n-1}, i_n).\end{aligned}$$

From this, we can compute the law and the conditional probability of a Markov chain in terms of the transition matrix P:

$$\begin{aligned}\mathbb{P}(X_n = i_n | X_0 = i_0) &= \frac{\mathbb{P}(X_n = i_n, X_0 = i_0)}{\mathbb{P}(X_0 = i_0)} \\ &= \sum_{i_1,\dots,i_{n-1}\in\mathcal{X}} P(i_0, i_1)\cdots P(i_{n-1}, i_n) = P^n(i_0, i_n),\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X_n = i) &= \sum_{i_0,i_1,\dots,i_{n-1}\in\mathcal{X}} \mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) \\ &= \sum_{i_0,i_1,\dots,i_{n-1}\in\mathcal{X}} \lambda_{i_0} P(i_0, i_1)\cdots P(i_{n-1}, i) = (\lambda P^n)(i).\end{aligned}$$

Notice also that $\mathbb{P}(X_{n+m} = j | X_n = i) = \mathbb{P}(X_m = j | X_0 = i)$. Indeed,

$$\begin{aligned}\frac{\mathbb{P}(X_{n+m} = j, X_n = i)}{\mathbb{P}(X_n = i)} &= \frac{\sum_{i_{n+1},\dots,i_{n+m-1}\in\mathcal{X}} P(i, i_{n+1})P(i_{n+1}, i_{n+2})\cdots P(i_{n+m-1}, j)\mathbb{P}(X_n = i)}{\mathbb{P}(X_n = i)} \\ &= \frac{\mathbb{P}(X_m = j, X_0 = i)}{\mathbb{P}(X_0 = i)}.\end{aligned}$$

An important and easy result is Kolmogorov's equation:

$$P^{n+m}(i, j) = \sum_{k\in\mathcal{X}} P^n(i, k)P^m(k, j).$$

Definition 2.2. *We say that the state i connects with the state j if the Markov chain can reach j starting at i. Mathematically, $\mathbb{P}(X_n = j | X_0 = i) > 0$ for some n.*

We say that i communicates with j if i connects with j and j connects with i. This defines an equivalence relation in the state space $\mathcal{X}$.

We say that the state i is an absorbing state if $P(i, i) = 1$ (equivalently, if i is connected with j, then $j = i$). A finite Markov chain is said to be absorbing if there exists at least one absorbing state.

We say that P is irreducible if there is only one equivalence class in $\mathcal{X}$ (every state can be reached from any other state in one or more steps).

Definition 2.3. *A state i is aperiodic if* $\gcd\{n \geq 0 : P^n(i, i) > 0\} = 1$.

We say that P is aperiodic if every state is aperiodic.

Proposition 2.1. *Aperiodicity is a class property.*

PROOF. Consider $i \in \mathcal{X}$ and denote

$$\tau(i) = \{n \geq 0 : P^n(i,i) > 0\}.$$

Suppose that i communicates with j, $j \in \mathcal{X}$. We show that $\gcd\{\tau(i)\} = \gcd\{\tau(j)\}$. Assume that $i \neq j$, otherwise this equality is obvious. Since i connects with j and j connects with i, $P^r(i,j) > 0$ and $P^l(j,i) > 0$ for certain $r, l \geq 1$. Let $m = r + l$, then $m \in \tau(i) \cap \tau(j)$ since

$$P^m(i,i) = \sum_{k \in \mathcal{X}} P^r(i,k)P^l(k,i) \geq P^r(i,j)P^l(j,i) > 0.$$

We can see that if $n \in \tau(i)$, then $n + m \in \tau(j)$. Indeed,

$$P^{n+m}(j,j) = \sum_{k,s \in \mathcal{X}} P^l(j,k)P^n(k,s)P^r(s,j) \geq P^l(j,i)P^n(i,i)P^r(i,j) > 0.$$

Denote $d_i = \gcd\{\tau(i)\}$. We use the notation $|$ for divisibility. Since $d_j | l \ \forall l \in \tau(j)$ and $n + m \in \tau(j) \ \forall n \in \tau(i)$, we have that $d_j | n + m \ \forall n \in \tau(i)$. As $d_j | m$ too (because $m \in \tau(j)$), we have that $d_j | n \ \forall n \in \tau(i)$, thus $d_j | d_i$. By symmetry $d_i | d_j$, and with this we conclude that $d_i = d_j$, as we wanted to show. □

PROPOSITION 2.2. *A state i is aperiodic if and only if there exists $n_i \geq 1$ such that $P^n(i,i) > 0$ for all $n \geq n_i$.*

PROOF. We prove both implications:

[←] It is clear, because the only common factor of n and $n+1$ is 1 (for example if $n_i = 10$, then $P^{10}(i,i) > 0$ and $P^{11}(i,i) > 0$, therefore $10, 11 \in \{n \geq 0 : P^n(i,i) > 0\}$. Since $\gcd\{10, 11\} = 1$, we have $\gcd\{n \geq 0 : P^n(i,i) > 0\} = 1$, which is the definition of i aperiodic).

[→] Notice that if $S \subseteq \mathbb{Z}^+$ with $\gcd\{S\} = g_S$, then there exists $F \subseteq S$ finite satisfying $\gcd\{F\} = g_S$. Indeed, the sequence $\{\gcd\{S \cap [1,n]\}\}_{n \geq 1}$ is bounded below by g_S (each term is $\geq g_S$, in fact g_S divides each term) and is monotonically decreasing (when you add more terms to a set, the gcd is less or equal). If there is an n_0 from which the sequence is constant, let $d = \gcd\{S \cap [1,n]\}$ for all $n \geq n_0$. Given $s \in S$, then $s \leq n$ for some $n \geq n_0$, therefore $d | s$. Hence, $d | g_S$ and therefore $d = g_S$, so $F = S \cap [1, n_0]$. If the sequence is never constant, there is a subsequence $\{\gcd\{S \cap [1,n_k]\}\}_{k \geq 1}$ strictly decreasing, but this is a contradiction because the (integer) sequence is bounded below by g_S.

Denote the finite set F by $\{f_1, \ldots, f_m\}$. By Bézout's Lemma, we know that there exist $\alpha_1, \ldots, \alpha_m \in \mathbb{Z}$ such that

$$g_S = \alpha_1 f_1 + \ldots + \alpha_m f_m.$$

Schur Lemma says that there exists m_0 such that for all $k \geq m_0$, kg_S is a linear combination of $f_1, \ldots f_m$ with non-negative integers. This is an elementary number theory statement that is proved in Lemma 1.30 in [**12**].

In our case, take $S = \{n \geq 0 : P^n(i,i) > 0\}$ and $g_S = 1$. Notice that S is closed for the sum: if $n, m \in S$, then $n + m \in S$, because by Kolmogorov's equation $P^{n+m}(i,i) = \sum_k P^n(i,k)P^m(k,i) \geq P^n(i,i)P^m(i,i) > 0$. By what we proved, there exists a finite set $F \subseteq S$ such that $\gcd\{F\} = g_S = 1$. There exists n_i such that, for all $n \geq n_i$, the number $n = ng_S$ is a linear combination of F with non-negative integers, which belongs to S because S is closed under the sum. Hence, $n \in S$ for all $n \geq n_i$, as wanted. □

PROPOSITION 2.3. *If P is irreducible and aperiodic, then there exists $N \geq 1$ such that P^n has all entries positive for all $n \geq N$.*

PROOF. Since P is aperiodic, by Proposition 2.2, for all $i \in \mathcal{X}$ there exists a number n_i such that $P^n(i,i) > 0$ for every $n \geq n_i$. Let $N = \max\{n_i : i \in \mathcal{X}\}$. Then $P^n(i,i) > 0$ for all $n \geq N$ and $i \in \mathcal{X}$.

Take two distinct states i and j. As P is irreducible, i is connected with j, therefore there exists a number n_{ij} such that $P^{n_{ij}}(i,j) > 0$. Let $N_0 = \max\{n_{ij} : i, j \in \mathcal{X} \text{ distinct}\} + N$.

Given two arbitrary states i and j and $n \geq N_0$,

$$P^n(i,j) = (P^{n-n_{ij}}P^{n_{ij}})(i,j) = \sum_{k\in\mathcal{X}} P^{n-n_{ij}}(i,k)P^{n_{ij}}(k,j) \geq P^{n-n_{ij}}(i,i)P^{n_{ij}}(i,j) > 0$$

(notice that, since $n - n_{ij} \geq N_0 - n_{ij} \geq N$, we have $P^{n-n_{ij}}(i,i) > 0$). □

DEFINITION 2.4. *We say that the row vector $\pi = \{\pi(i)\}_{i\in\mathcal{X}}$ is an invariant or a stationary distribution for the transition matrix P if it satisfies that $\pi = \pi P$, that is, $\pi(j) = \sum_{i\in\mathcal{X}} \pi(i)P(i,j)$.*

The name invariant or stationary comes from the fact that if $P(X_0 = i) = \pi(i)$ for all $i \in \mathcal{X}$, then $\mathbb{P}(X_0 = j) = \sum_{i\in\mathcal{X}} \mathbb{P}(X_0 = i)P(i,j) = \sum_{i\in\mathcal{X}} \mathbb{P}(X_1 = j, X_0 = i) = \mathbb{P}(X_1 = j)$, so all the X_n share the same distribution given by π.

PROPOSITION 2.4. *If $\pi(i)P(i,j) = \pi(j)P(j,i)$ for all $i, j \in \mathcal{X}$, then π is a stationary distribution.*

PROOF. We have

$$\pi(j) = \sum_{i\in\mathcal{X}} \pi(j)P(j,i) = \sum_{i\in\mathcal{X}} \pi(i)P(i,j),$$

which is the definition of stationary distribution. □

PROPOSITION 2.5. *If $\mathcal{X}$ is finite, then P has a stationary distribution.*

PROOF. Notice that 1 is an eigenvalue of P, since each row of P sums 1. Therefore, 1 is an eigenvalue of P^T. Since the spectral radius r of a matrix is less or equal than any multiplicative norm, $r \leq \|P^T\|_1 = 1$ (recall that $\|\cdot\|_1$ is the maximum of the sums of each column)[2], therefore 1 is the greatest eigenvalue in absolute value of P^T. Since P^T has all entries non-negative, by Perron-Frobenius Theorem (see page 670 in [15]) there exists an eigenvector $u = (u_i)_i$ of 1 with non-negative entries. We have $P^T u = u$, so $u^T P = u$. Let $\pi = (1/\sum_i u_i)u$. Then $\pi P = \pi$, π has all entries non-negative and with sum 1, so π is a stationary distribution for P. □

PROPOSITION 2.6. *If $\mathcal{X}$ is finite and P is irreducible, then P has a stationary distribution with all entries positive.*

PROOF. The proof is as Proposition 2.5, but since in this case P is irreducible, for all i, j there exists $m \geq 0$ such that $(P^T)^m(i,j) > 0$. A version of Perron-Frobenius Theorem (see page 673 in [15]) guarantees that the eigenvector u has positive entries, therefore π has positive entries. □

THEOREM 2.1. *Let P be an irreducible and aperiodic Markov chain with stationary distribution π. Then*

$$P^n(i,j) \overset{n\to\infty}{\longrightarrow} \pi(j),$$

for all $i, j \in \mathcal{X}$.

In the following three sections, we will prove this theorem in three different ways. However, we will restrict to the case of a finite state space. The reason for this is that, for all our proofs to work, we have to deal necessarily with a finite state space, and moreover, in terms of applications, we will just work with finite Markov chains.

Concerning notation, we will denote our finite state space by $\mathcal{X}$, and the states will be denoted by the last letters of the alphabet: x, y, etc.

2. Convergence of Markov Chains. Total Variation distance

Denote by $\mathcal{X}$ our state space. Notice that, for each $x \in \mathcal{X}$, $P^n(x,\cdot)$ is a probability measure on $\mathcal{X}$, as well as the stationary distribution π. Thus, to prove Theorem 2.1, we have to study the distance between two probability distributions on $\mathcal{X}$. This motivates the following definition.

Definition 2.5. *The total variation distance between two probability distributions μ and ν on $\mathcal{X}$ is defined as:*

$$\|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)|.$$

Notice that the total variation distance is indeed a distance: it is greater or equal than zero, being zero if and only if $\mu = \nu$, it is symmetric and it satisfies the triangular inequality because the modulus satisfies it in $\mathbb{R}$.

The total variation distance measures the maximum difference between two probabilities assigned by μ and ν in $\mathcal{X}$. Although this is an intuitive definition, it is difficult to check in practice, since one has to consider every possible event contained in $\mathcal{X}$. The following proposition allows us to have a more useful formula for the total variation distance in terms of computations.

Proposition 2.7. *The total variation distance can be expressed as follows:*

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|.$$

Proof. Let $B = \{x \in \mathcal{X} : \mu(x) \geq \nu(x)\}$. Fix any event $A \subseteq \mathcal{X}$.

If $x \in A \cap B^c$, then $\mu(x) - \mu(x) < 0$. Summing for all $x \in A \cap B^c$, we arrive at $\mu(A \cap B^c) - \nu(A \cap B^c) \leq 0$ (it is not strictly less than zero because it could happen that $A \cap B^c = \varnothing$). Therefore,

$$\mu(A) - \nu(A) = \mu(A \cap B) - \nu(A \cap B) + \mu(A \cap B^c) - \nu(A \cap B^c) \leq \mu(A \cap B) - \nu(A \cap B). \quad (2.1)$$

Similarly, if $x \in B \cap A^c$, then $\mu(x) - \nu(x) \geq 0$. Summing for all $x \in B \cap A^c$, we obtain $\mu(B \cap A^c) - \nu(B \cap A^c) \geq 0$. This implies that

$$\mu(B) - \nu(B) = \mu(B \cap A) - \nu(B \cap A) + \mu(B \cap A^c) - \nu(B \cap A^c) \leq \mu(B \cap A) - \nu(B \cap A). \quad (2.2)$$

Combining inequalities (2.1) and (2.2),

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B).$$

As A is an arbitrary event, taking maximums on A gives $\|\mu - \nu\|_{\mathrm{TV}} \leq \mu(B) - \nu(B)$. But that maximum is achieved at the event B, thus we have equality:

$$\|\mu - \nu\|_{\mathrm{TV}} = \mu(B) - \nu(B). \quad (2.3)$$

Since $\mu(B) - \nu(B) = 1 - \mu(B^c) - 1 + \nu(B^c) = \nu(B^c) - \mu(B^c)$,

$$\begin{aligned}\|\mu - \nu\|_{\mathrm{TV}} &= \frac{1}{2}[\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)] \\ &= \frac{1}{2}\left[\sum_{x\in B}(\mu(x) - \nu(x)) + \sum_{x\in B^c}(\nu(x) - \mu(x))\right] \\ &= \frac{1}{2}\sum_{x\in\mathcal{X}}|\mu(x) - \nu(x)|.\end{aligned}$$

□

COROLLARY 2.1. *The total variation distance can be expressed as*

$$\|\mu - \nu\|_{\mathrm{TV}} = \sum_{\substack{x\in\mathcal{X}\\ \mu(x)\geq\nu(x)}}(\mu(x) - \nu(x)).$$

PROOF. In the proof of Proposition 2.7 above, we defined $B = \{x \in \mathcal{X} : \mu(x) \geq \nu(x)\}$ and we showed in (2.3) that

$$\|\mu - \nu\|_{\mathrm{TV}} = \mu(B) - \nu(B).$$

With this the proof is completed. □

EXAMPLE 2.1. Let μ and ν be two probability measures on $\{1, 2, 3\}$ with $\mu(1) = 3/4$, $\mu(2) = 1/8$ and $\mu(3) = 1/8$, $\nu(1) = 1/3$, $\nu(2) = 1/3$ and $\nu(3) = 1/3$. Then

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\left(\left|\frac{3}{4} - \frac{1}{3}\right| + \left|\frac{1}{8} - \frac{1}{3}\right| + \left|\frac{1}{8} - \frac{1}{3}\right|\right) = \frac{5}{12}.$$

Without this Proposition 2.7 and trying to apply directly the definition of total variation distance, we would have to take every possible subset $A \subseteq \{1, 2, 3\}$ and compute $|\mu(A) - \nu(A)|$. The number of computations required to achieve this is exponential in the size n of $\mathcal{X}$ (2^n computations), whereas Proposition 2.7 and Corollary 2.1 require only n computations.

We are ready to prove Theorem 2.1, even a stronger result.

THEOREM 2.2. *Let P be an irreducible and aperiodic transition matrix, with stationary distribution π. Then there exist $\alpha \in (0, 1)$ and $C > 0$ such that*

$$\max_{x\in\mathcal{X}}\|P^n(x, \cdot) - \pi\|_{\mathrm{TV}} \leq C\alpha^n.$$

PROOF. Since P is irreducible and aperiodic, by Proposition 2.3 there exists $r \geq 1$ such that P^r has strictly positive entries. Let Π be the matrix such that each row is the vector π (stationary distribution). Since $P^r(x, y) > 0$, $\forall x, y \in \mathcal{X}$, there exists m such that $P^r(x, y) \geq m > 0$ $\forall x, y \in \mathcal{X}$. We choose $0 < \delta < \frac{m}{\max_{y\in\mathcal{X}}\{\pi(y)\}}$, so that $P^r(x, y) \geq \delta\pi(y)$, $\forall x, y \in \mathcal{X}$. Let $\theta = 1 - \delta$ and consider the matrix Q defined by the following equation:

$$P^r = (1 - \theta)\Pi + \theta Q. \tag{2.4}$$

Notice that $P^r(i, j) = \mathbb{P}(X_r = j|X_0 = i)$, then $\sum_{j\in\mathcal{X}} P^r(i, j) = 1$. Since by (2.4)

$$P^r(i, j) = (1 - \theta)\underbrace{\Pi(i, j)}_{\pi(j)} + \theta Q(i, j),$$

applying $\sum_{j\in\mathcal{X}}$ at both sides of the last equality and taking into account that $\sum_{j\in\mathcal{X}}\pi(j)=1$, we get

$$1=(1-\theta)+\theta\sum_{j\in\mathcal{X}}Q(i,j).$$

Then $\sum_{j\in\mathcal{X}}Q(i,j)=1$, that is, Q is stochastic.

By induction on $k\geq 1$ we show that

$$P^{rk}=(1-\theta^k)\Pi+\theta^kQ^k. \tag{2.5}$$

The first case, $k=1$, is true by definition of Q (2.4). Assume that the equation (2.5) is true for $k=n$ and we prove that it holds for $k=n+1$.

$$P^{r(n+1)}=P^{rn}P^r\underbrace{=}_{\text{induction}}[(1-\theta^n)\Pi+\theta^nQ^n]P^r,$$

then

$$P^{r(n+1)}=(1-\theta^n)\Pi P^r+\theta^nQ^nP^r.$$

Applying (2.4) to the right-hand side of the sum of the last expression,

$$\begin{aligned}P^{r(n+1)}&=(1-\theta^n)\Pi P^r+\theta^nQ^n[(1-\theta)\Pi+\theta Q]\\&=(1-\theta^n)\Pi P^r+\theta^nQ^n(1-\theta)\Pi+\theta^{n+1}Q^{n+1}.\end{aligned}$$

Since π is the stationary distribution of P, one has $\Pi P^r=\Pi$. Moreover, since Q is stochastic we know that Q^n is stochastic, so $Q^n\Pi=\Pi$. Hence,

$$\begin{aligned}P^{r(n+1)}&=(1-\theta^n)\Pi+\theta^n(1-\theta)\Pi+\theta^{n+1}Q^{n+1}\\&=\pi(1-\theta^n+\theta^n-\theta^{n+1})+\theta^{n+1}Q^{n+1}\\&=\pi(1-\theta^{n+1})+\theta^{n+1}Q^{n+1},\end{aligned}$$

and (2.5) is true. Multiplying at both sides of (2.5) by P^j one has,

$$P^{rk+j}=(1-\theta^k)\Pi P^j+\theta^kQ^kP^j=(1-\theta^k)\Pi+\theta^kQ^kP^j.$$

Finally, we have that

$$P^{rk+j}-\Pi=-\theta^k\Pi+\theta^kQ^kP^j=\theta^k(-\Pi+Q^kP^j).$$

By Proposition 2.7,

$$\begin{aligned}\|P^{rk+j}(x_0,\cdot)-\pi\|_{\mathrm{TV}}&=\frac{1}{2}\sum_{x\in\mathcal{X}}|P^{rk+j}(x_0,x)-\underbrace{\pi(x)}_{\Pi(x_0,x)}|\\&=\frac{1}{2}\theta^k\sum_{x\in\mathcal{X}}|-\Pi(x_0,x)+(Q^kP^j)(x_0,x)|\\&\leq\frac{1}{2}\theta^k\Bigg(\underbrace{\sum_{x\in\mathcal{X}}\Pi(x_0,x)}_{=1}+\underbrace{\sum_{x\in\mathcal{X}}(Q^kP^j)(x_0,x)}_{=1}\Bigg)=\theta^k,\end{aligned}$$

where the last inequality comes from applying the triangular inequality, and

$$\sum_{x\in\mathcal{X}}(Q^kP^j)(x_0,x)=1$$

holds because the product of stochastic matrices is stochastic[1]. Let $n \geq 1$. Define k_n and j_n such that $n = rk_n + j_n$, where $k_n = \lfloor \frac{n}{r} \rfloor$ (i.e. $k_n r \leq n < (k_n + 1)r$) and $j_n = n - rk_n$. Then

$$\|P^n(x_0, \cdot) - \pi\|_{\mathrm{TV}} = \|P^{rk_n + j_n}(x_0, \cdot) - \pi\|_{\mathrm{TV}} \leq \theta^{k_n} \leq \theta^{\frac{n}{r}} = (\theta^{\frac{1}{r}})^n.$$

To end up, consider $C = 1$ and $\alpha = \theta^{\frac{1}{r}}$ and we have the theorem proved.

□

From this theorem, Theorem 2.1 follows. Indeed,

$$0 = \lim_{n\to\infty} \max_{x\in\mathcal{X}} \|P^n(x, \cdot) - \pi\|_{\mathrm{TV}} = \lim_{n\to\infty} \max_{x\in\mathcal{X}} \frac{1}{2} \sum_{y\in\mathcal{X}} |P^n(x, y) - \pi(y)|,$$

whence $\lim_{n\to\infty} P^n(x, y) = \pi(y)$, for all $x, y \in \mathcal{X}$.

3. Convergence of Markov Chains. Coupling.

There is a very close relation between probability distributions and random variables: every random variable has a probability distribution, and for every probability distribution there exist a probability space and a random variable with the law specified by the probability distribution. This makes us think on expressing the total variation distance between two probability distributions as some sort of distance between two random variables. This fact motivates the definition of coupling.

Definition 2.6. *A coupling of μ and ν is a pair of random variables (X, Y) defined on the same probability space such that the distribution of X is μ and the distribution of Y is ν. That is, in a finite probability space with probability measure $\mathbb{P}$, a coupling (X, Y) satisfies $\mathbb{P}(X = x) = \mu(x)$ and $\mathbb{P}(Y = y) = \nu(y)$.*

Notice that, for any pair of distributions μ and ν, a coupling (X, Y) always exists, since we can always find a probability space and two independent random variables X and Y such that the law of X is μ and the law of Y is ν (by Kolmogorov's Extension Theorem, see [17, Theorem 2.4]). In fact, this is a stronger result than the definition of coupling, since a coupling does not require the pair of random variables to be independent.

Example 2.2. Let μ and ν be two probability measures on $\{0, 1\}$ such that $\mu(0) = \nu(0) = 1/2$ and $\mu(1) = \nu(1) = 1/2$. Essentially, $\mu, \nu \sim \mathrm{Ber}(1/2)$. We can find a probability space and two *independent* random variables X and Y such that $X, Y \sim \mathrm{Ber}(1/2)$. Then the pair (X, Y) is a coupling of μ and ν. Notice that (X, X) is also a coupling of μ and ν.

Concerning notation when we use couplings, we will use the notation $\mathbb{P}$ for the probability measure of the probability space that arises when defining the coupling. We will not specify the connection between the couplings and the corresponding probability space or measure, it should be understood by the context.

The following proposition gives the total variation distance between two probability distributions in terms of their couplings.

[1] Let A and B be two stochastic matrices and consider $\mathbb{1}$ a vector of ones. By definition of stochastic matrix, we have that $A\mathbb{1} = \mathbb{1}$ and $B\mathbb{1} = \mathbb{1}$. Hence $(AB)\mathbb{1} = A\mathbb{1} = \mathbb{1}$.

PROPOSITION 2.8. *Let μ and ν be two probability distributions in a finite probability space. Then*

$$\|\mu-\nu\|_{\mathrm{TV}} = \min\{\mathbb{P}(X\neq Y) : (X,Y) \text{ coupling}\}.$$

The coupling for which that minimum is attained is called ***optimal coupling****.*

PROOF. Consider any coupling (X,Y) of μ and ν and any event $A\subset\mathcal{X}$. Notice that

$$\begin{aligned}\mathbb{P}(X\in A, Y\notin A)+\mathbb{P}(Y\in A) &\geq \mathbb{P}(X\in A, Y\notin A)+\mathbb{P}(X\in A, Y\in A)\\ &\underbrace{=}_{\text{disjoint}} \mathbb{P}(\{X\in A, Y\notin A\}\cup\{X\in A, Y\in A\}) = \mathbb{P}(X\in A),\end{aligned}$$

where the first inequality is because $\{X\in A, Y\in A\}\subset\{Y\in A\}$. Hence, we have that

$$\mathbb{P}(X\in A)\leq \mathbb{P}(X\in A, Y\notin A)+\mathbb{P}(Y\in A).$$

Taking this into account,

$$\begin{aligned}\mu(A)-\nu(A) = \mathbb{P}(X\in A)-\mathbb{P}(Y\in A) &\leq \mathbb{P}(X\in A, Y\notin A)\\ &\leq \mathbb{P}(X\neq Y).\end{aligned}$$

Taking the infimum in last expression, one has:

$$\mu(A)-\nu(A)\leq \inf\{\mathbb{P}(X\neq Y) : (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\},$$

and taking the supremum on A we get:

$$\|\mu-\nu\|_{\mathrm{TV}}\leq \inf\{\mathbb{P}(X\neq Y) : (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\}.$$

Now we construct a coupling (X,Y) such that $\|\mu-\nu\|_{\mathrm{TV}} = \mathbb{P}(X\neq Y)$. Let

$$p = \sum_{x\in\mathcal{X}} \mu(x)\wedge\nu(x),$$

where $\wedge$ denotes the minimum. We decompose $\sum_{x\in\mathcal{X}}\mu(x)\wedge\nu(x)$ into two parts:

$$\sum_{x\in\mathcal{X}}\mu(x)\wedge\nu(x) = \sum_{\substack{x\in\mathcal{X}\\ \mu(x)\leq\nu(x)}}\mu(x)+\sum_{\substack{x\in\mathcal{X}\\ \mu(x)>\nu(x)}}\nu(x).$$

Add and subtract

$$\sum_{\substack{x\in\mathcal{X}\\ \mu(x)>\nu(x)}}\mu(x)$$

to the right-hand side of this last expression, to derive

$$\sum_{x\in\mathcal{X}}\mu(x)\wedge\nu(x) = 1-\sum_{\substack{x\in\mathcal{X}\\ \mu(x)>\nu(x)}}(\mu(x)-\nu(x)).$$

By Corollary 2.1, we have that

$$p = \sum_{x\in\mathcal{X}}\mu(x)\wedge\nu(x) = 1-\|\mu-\nu\|_{\mathrm{TV}}.$$

Toss a coin with probability of heads equal p.

- If it comes up a head, we choose a value of Z according to
$$\gamma_1(x) = \frac{\mu(x) \wedge \nu(x)}{p}$$
(this is obviously a probability taking into account the definition of p) and set $X = Y = Z$.
- If it comes up a tail, we choose a value of X according to the distribution
$$\gamma_2(x) = \begin{cases} \frac{\mu(x)-\nu(x)}{\|\mu-\nu\|_{\mathrm{TV}}} & \text{if } \mu(x) > \nu(x), \\ 0 & \text{otherwise,} \end{cases}$$
and independently, we choose a value of Y according to the distribution
$$\gamma_3(x) = \begin{cases} \frac{\nu(x)-\mu(x)}{\|\mu-\nu\|_{\mathrm{TV}}} & \text{if } \nu(x) > \mu(x), \\ 0 & \text{otherwise.} \end{cases}$$

Let us see that X has distribution μ and Y has distribution ν. Notice that, since $p = 1 - \|\mu - \nu\|_{\mathrm{TV}}$,
$$p\gamma_1 + (1-p)\gamma_2 = \mu$$
and
$$p\gamma_1 + (1-p)\gamma_3 = \nu.$$
Indeed, let us see that $p\gamma_1 + (1-p)\gamma_2 = \mu$ (the case $p\gamma_1 + (1-p)\gamma_3 = \nu$ is analogous):

- suppose that $\mu(x) > \nu(x)$, then
$$p\gamma_1 + (1-p)\gamma_2 = p\frac{\mu(x) \wedge \nu(x)}{p} + \|\mu - \nu\|_{\mathrm{TV}} \frac{\mu(x) - \nu(x)}{\|\mu - \nu\|_{\mathrm{TV}}} = \mu(x);$$
- suppose that $\mu(x) \leq \nu(x)$, then
$$p\gamma_1 + (1-p)\gamma_2 = p\frac{\mu(x)}{p} + \|\mu - \nu\|_{\mathrm{TV}} \cdot 0 = \mu(x).$$

Call H the event of getting a head and T the event of getting a tail. According to this notation and the last two equations,
$$\mathbb{P}(X = x) = \mathbb{P}(X = x|H)\mathbb{P}(H) + \mathbb{P}(X = x|T)\mathbb{P}(T).$$

- If we obtain a head, we choose X according to γ_1, then $\mathbb{P}(X = x|H) = \gamma_1(x)$ and $\mathbb{P}(H) = p$,
- If we obtain a tail, we choose X according to γ_2, then $\mathbb{P}(X = x|T) = \gamma_2(x)$ and $\mathbb{P}(T) = 1 - p$.

Thus, $\mathbb{P}(X = x) = p\gamma_1(x) + (1-p)\gamma_2(x) = \mu(x)$, and with this we conclude that $X \sim \mu$ (analogously, we deduce that $Y \sim \nu$). Therefore, (X, Y) is a coupling. Notice that if the coin comes up tails, $X \neq Y$, since then
$$\mathrm{support}(X) \subseteq \{x : \mu(x) > \nu(x)\},$$
$$\mathrm{support}(Y) \subseteq \{x : \nu(x) > \mu(x)\},$$
so $\mathrm{support}(X)$ and $\mathrm{support}(Y)$ are disjoint. Moreover, from this last result we have that $X = Y$ if and only if the coin comes up heads. Then,
$$\mathbb{P}(X \neq Y) = \mathbb{P}(T) = 1 - p = 1 - (1 - \|\mu - \nu\|_{\mathrm{TV}}) = \|\mu - \nu\|_{\mathrm{TV}}.$$

□

Coupling will allow us to prove Theorem 2.1 in an easy way. Define

$$d(n) = \max_{x\in\mathcal{X}} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}}, \quad \bar{d}(n) = \max_{x,y\in\mathcal{X}} \|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}}.$$

In the following proposition, we establish the relation between $d(n)$ and $\bar{d}(n)$.

PROPOSITION 2.9. *It holds*

$$d(n) \le \bar{d}(n) \le 2d(n).$$

PROOF. The inequality $\bar{d}(n) \le 2d(n)$ is clear by the triangular inequality.

Let us prove the other inequality, $d(n) \le \bar{d}(n)$. By definition of π, $\pi = \pi P$. Componentwise, this means that $\pi(x) = \sum_{y\in\mathcal{X}} \pi(y)P^n(y,x)$. This gives the probability of any event $A \subseteq \mathcal{X}$ under the probability π:

$$\pi(A) = \sum_{x\in A} \pi(x) = \sum_{y\in\mathcal{X}} \pi(y) \left(\sum_{x\in A} P^n(y,x) \right) = \sum_{y\in\mathcal{X}} \pi(y) P^n(y,A).$$

Then,

$$\begin{aligned} |P^n(x,A) - \pi(A)| &= \left| \sum_{y\in\mathcal{X}} \pi(y)(P^n(x,A) - P^n(y,A)) \right| \le \sum_{y\in\mathcal{X}} \pi(y)|P^n(x,A) - P^n(y,A)| \\ &\le \sum_{y\in\mathcal{X}} \pi(y)\|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}} \\ &\le \max_{y\in\mathcal{X}} \|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}} \sum_{y\in\mathcal{X}} \pi(y) \\ &= \max_{y\in\mathcal{X}} \|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}}. \end{aligned}$$

Taking supremum on A,

$$\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \le \max_{y\in\mathcal{X}} \|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}}.$$

Finally, taking maximum on $x \in \mathcal{X}$, we arrive at the desired inequality $d(t) \le \bar{d}(t)$. □

From this proposition, we realize that to prove Theorem 2.2 it suffices to bound $\bar{d}(n) \le C\alpha^n$, for certain $C > 0$ and $\alpha \in (0,1)$. Thus, we need to study more deeply the function $\bar{d}(n)$. The key result is that $\bar{d}(n)$ is a multiplicative function. This fact is proved in the following proposition, using as a tool the concept of coupling.

PROPOSITION 2.10. *It holds*

$$\bar{d}(n+m) \le \bar{d}(n)\bar{d}(m).$$

PROOF. For fixed $x, y \in \mathcal{X}$, consider the probability measures $P^n(x,\cdot)$ and $P^n(y,\cdot)$. Let (X_n, Y_n) be an optimal coupling for $P^n(x,\cdot)$ and $P^n(y,\cdot)$:

$$\|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}} = \mathbb{P}(X_n \ne Y_n).$$

By Kolmogorov's equation,

$$P^{n+m}(x,w) = \sum_{z\in\mathcal{X}} P^n(x,z)P^m(z,w) = \sum_{z\in\mathcal{X}} \mathbb{P}(X_n = z)P^m(z,w) = \mathbb{E}[P^m(X_n,w)].$$

Analogously,

$$P^{n+m}(y,w) = \sum_{z\in\mathcal{X}} P^n(y,z)P^m(z,w) = \sum_{z\in\mathcal{X}} \mathbb{P}(Y_n = z)P^m(z,w) = \mathbb{E}[P^m(Y_n,w)].$$

Then

$$P^{n+m}(x,A) - P^{n+m}(y,A) = \mathbb{E}[P^m(X_n,A) - P^m(Y_n,A)],$$

for every event $A \subseteq \mathcal{X}$.

For a given outcome ω in our probability space, if $X_n(\omega) \neq Y_n(\omega)$, then

$$|P^m(X_n(\omega),A) - P^m(Y_n(\omega),A)| \leq \|P^m(X_n(\omega),\cdot) - P^m(Y_n(\omega),\cdot)\|_{\mathrm{TV}} \leq \bar{d}(m).$$

This implies that,

$$|P^m(X_n,A) - P^m(Y_n,A)| \leq \bar{d}(m)\mathbb{1}_{\{X_n \neq Y_n\}}.$$

By the monotonicity of the expectation,

$$\begin{aligned} |P^{n+m}(x,A) - P^{n+m}(y,A)| &= \mathbb{E}[|P^m(X_n,A) - P^m(Y_n,A)|] \leq \mathbb{E}[\bar{d}(m)\mathbb{1}_{\{X_n \neq Y_n\}}] \\ &= \bar{d}(m)\mathbb{P}(X_n \neq Y_n) = \bar{d}(m)\|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}} \\ &\leq \bar{d}(m)\bar{d}(n). \end{aligned}$$

Taking the maximum on $A \subseteq \mathcal{X}$,

$$\|P^{n+m}(x,\cdot) - P^{n+m}(y,\cdot)\|_{\mathrm{TV}} \leq \bar{d}(m)\bar{d}(n).$$

Finally, taking the maximum on $x, y \in \mathcal{X}$ the inequality $\bar{d}(n+m) \leq \bar{d}(m)\bar{d}(n)$ follows. □

We are in position of proving again Theorem 2.2.

THEOREM 2.3. *Let P be an irreducible and aperiodic transition matrix, with stationary distribution π. Then there exist $\alpha \in (0,1)$ and $C > 0$ such that*

$$\max_{x \in \mathcal{X}} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \leq C\alpha^n.$$

PROOF. Since P is irreducible and aperiodic, there exists $n \geq 1$ such that P^n has all entries greater than zero. Then, for all $x, y \in \mathcal{X}$ and $A \subsetneq \mathcal{X}$, we have that $P^n(x,A) \in (0,1)$ and $P^n(y,A) \in (0,1)$, so $|P^n(x,A) - P^n(y,A)| < 1$ (this last inequality also holds when $A = \mathcal{X}$). Taking maximum on $A \subseteq \mathcal{X}$, we get $\|P^n(x,\cdot) - P^n(y,\cdot)\|_{\mathrm{TV}} < 1$. Finally, taking maximum on $x, y \in \mathcal{X}$, $\beta := \bar{d}(n) < 1$.

Using the multiplicative property of $\bar{d}$, we obtain that for any $m \geq 1$

$$\bar{d}(n+m) \leq \bar{d}(n)\bar{d}(m) \leq \beta\bar{d}(m),$$

$$\bar{d}(n+2m) \leq \bar{d}(n+m)\bar{d}(m) \leq \beta^2\bar{d}(m),$$

$$\dots$$

$$\bar{d}(n+km) \leq \beta^k\bar{d}(m) \leq 2\beta^k$$

(the 2 comes from the fact that $|P^m(x,A) - P^m(y,A)| \leq 2$ by the triangular inequality).

Fix $m \geq 1$. Write $m = m_0 + k_0 n$, where $k_0 = \lfloor m/n \rfloor$ is the unique integer satisfying $k_0 n \leq m < (k_0+1)n$ and $m_0 = m - k_0 n$. Then,

$$d(m) \leq \bar{d}(m) = \bar{d}(m_0 + k_0 n) \leq 2\beta^{k_0} \leq 2\beta^{m/n} = C\alpha^m,$$

where $C = 2$ and $\alpha = \beta^{1/n} < 1$. □

4. Convergence of Markov Chains in graphs. Spectral tools.

Let $G = (V, E)$ be a connected graph, with $|V| = n$ vertices and $|E| = m$ edges. We are going to define a Markov chain on G in the following manner: if we are at a vertex $x \in V$ with degree $d(x)$, we move to a neighbor of x with probability $1/d(x)$. In this way, the state space is $\mathcal{X} = V$ and the transition probability matrix is defined as

$$P(x,y) = \begin{cases} 1/d(x), & xy \in E, \\ 0, & xy \notin E. \end{cases}$$

Notice the relation between the transition matrix and the adjacency matrix A_G of G: $P = DA_G$, where D is a diagonal matrix such that $D(x,x) = 1/d(x)$. Recall that the adjacency matrix A_G is an $n \times n$ matrix such that $A_G(x,y) = 1$ if $xy \in E$ and $A_G(x,y) = 0$ otherwise.

The stationary distribution π of P is the following:

$$\pi(x) = \frac{d(x)}{2m}. \tag{4.1}$$

This can be checked in two different ways. First, directly with the definition $\pi = \pi P$. Second, realizing that $\pi(x)P(x,y) = \pi(y)P(y,x)$, which implies that π is stationary.

Our goal will be to prove Theorem 2.1 for this particular Markov chain. We will use spectral tools in the proof, which means to take advantage of the eigenvalues of the transition matrix P.

First, we need to check if the hypotheses of Theorem 2.1 hold, that is, if P is irreducible and aperiodic. The irreducibility of P follows from the connectivity of G, since for any pair of states there is a path in G between them. The aperiodicity requires more work to be done, since not every connected graph gives the aperiodicity of the transition matrix. In fact, we have a characterization of aperiodicity by means of the concept of bipartite graph. A graph G is said to be bipartite if there is a partition of the set of vertices $V = V_1 \cup V_2$ such that every edge connects a vertex in V_1 to one in V_2, that is, $E \subseteq V_1 \times V_2$.

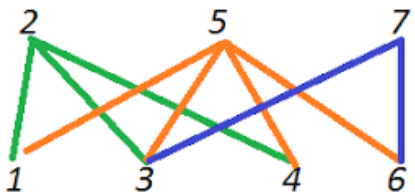


FIG. 1. Example of a bipartite graph, with $V_1 = \{1,3,4,6\}$ and $V_2 = \{2,5,7\}$.

THEOREM 2.4. *The graph G is non-bipartite if and only if the Markov chain is aperiodic.*

PROOF. We prove both implications:

[$\leftarrow$] If G is bipartite, you can come back to the same vertex only in an even number of steps. Therefore $\gcd\{n \geq 0 : P^n(x,x) > 0\}$ is even, for every vertex x. Hence, the Markov chain is not aperiodic.

[→] As any edge "leaves and comes back", we have $P^{2k}(x,x) > 0$ for all $k \geq 0$ and every vertex x. Thus, to prove the aperiodicity, it is enough to ckeck that you can come back to a vertex x in an odd number of steps. Suppose by contradiction that you can come back to x only in an even number of steps. Given a vertex y, let $D(y)$ be the length of the shortest path from x to y (this path does exist because our graph G is connected). Paint y red if $D(y)$ is even, and blue if $D(y)$ is odd. Let $V_1 = \{y : y \text{ is red}\}$ and $V_2 = \{y : y \text{ is blue}\}$. Trivially, $x \in V_1$ (because $D(x) = 0$ even). Suppose that there is an edge between two blue vertices y_1 and y_2. Then you can come back to x in an odd number of steps (odd path from x to y_1, union edge from y_1 to y_2, union odd path from y_2 to x), which is not possible by assumption. Suppose similarly that there is an edge between two red vertices y_1 and y_2. None of the two vertices is x, since a red vertex cannot be adjacent to x. Hence, there is an odd cycle starting and ending at x (even path from x to y_1, union edge from y_1 to y_2, union even path from y_2 to x), which again is not possible.

Thus, V_1 and V_2 bipartite G, but G is not bipartite by assumption. Thus, it is possible to come back to x in an odd number of steps. □

Now we are able to prove Theorem 2.1.

THEOREM 2.5. *If G is connected and non-bipartite (that is, P is irreducible and aperiodic), then*

$$\lim_{n\to\infty} P^n(x,y) = \pi(y),$$

for all $x, y \in \mathcal{X} = V$.

PROOF. Denote the vertices of G by $V = \{1, \ldots, n\}$. We want to use spectral tools to prove this result. The problem is that P is not symmetric (unless G is regular). However, we will use the fact that $P = DA$, where D is a diagonal matrix with entries $D(i,i) = 1/d(i)$ and $A = A_G$ is the adjacency matrix of G. Consider the matrix $N = D^{1/2}AD^{1/2} = D^{-1/2}PD^{1/2}$. Since A is symmetric, N is symmetric, so we can take advantage of the spectrum of N. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of (column) eigenvectors of N associated to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. We can write $N = \sum_{k=1}^n \lambda_k v_k v_k^T$. Notice that the vector w defined as $w(i) = \sqrt{d(i)}$ is an eigenvector of N with eigenvalue 1. Indeed,

$$\begin{aligned}(Nw)(i) &= \sum_{j=1}^n N(i,j)w(j) = \sum_{j=1}^n \frac{1}{\sqrt{d(i)}} A(i,j) \frac{1}{\sqrt{d(j)}} \sqrt{d(j)} = \frac{1}{\sqrt{d(i)}} \sum_{j=1}^n A(i,j) \\ &= \frac{1}{\sqrt{d(i)}} d(i) = \sqrt{d(i)} = w(i).\end{aligned}$$

Then the normalized vector $v = w/\sqrt{2|E|}$ is a unit eigenvector of N with eigenvalue 1, so we can assume that v belongs to the set $\{v_1, \ldots, v_n\}$.

Since the entries of N are non-negative, Perron-Frobenius Theorem (see e.g. [15, page 670]) asserts that the greatest eigenvalue of N in modulus, λ_1, is non-negative, and moreover an eigenvector (we can assume v_1) with non-negative entries can be chosen. As v has positive entries, $\lambda_1 = 1$ and $v_1 = v$ (otherwise $v \cdot v_1 > 0$, which is not possible by orthogonality).

Now, this version of Perron-Frobenius Theorem does not ensure $\lambda_2 < \lambda_1$, because we only know that the entries of N are non-negative, and not strictly positive. To solve this problem, we will use another version of Perron-Frobenius Theorem (see [15, page 673]) for the so-called irreducible matrix. Motivated by the definition for Markov chains, a square

non-negative matrix M is said to be irreducible if for any entry (i,j) there exists a power r such that $M^r(i,j) > 0$. Perron-Frobenius Theorem says that, for an irreducible matrix M, the greatest eigenvalue is simple. In our case, given an entry (i,j), we know that there is a power r such that $P^r(i,j) > 0$ (because the graph G is connected), therefore

$$N^r(i,j) = (D^{-1/2}PD^{1/2})^r(i,j) = \sqrt{d(i)}P^r(i,j)\frac{1}{\sqrt{d(j)}} > 0.$$

Therefore N is irreducible in the sense of this last general definition, and by Perron-Frobenius Theorem λ_1 is simple, that is, $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$.

From the non-bipartite condition of G, we prove directly that $\lambda_n > -1$. Suppose by contradiction that $\lambda_n = -1$. From $Nv_n = \lambda_n v_n$, multiplying to the left by v_n^T we obtain $\lambda_n = v_n^T N v_n$. Applying modulus,

$$1 = |\lambda_n| = \left|\sum_{i,j} v_n(i)N(i,j)v_n(j)\right| \leq \sum_{i,j} N(i,j)|v_n(i)||v_n(j)| = y^T N y \leq \lambda_1 y^T y = \lambda_1 = 1,$$

where $y(i) \coloneqq |v_n(i)|$. Then $y^T N y = \lambda_1 y^T y$. This implies $y = v_1$. Also,

$$\left|\sum_{i,j} v_n(i)N(i,j)v_n(j)\right| = \sum_{i,j} N(i,j)|v_n(i)||v_n(j)|.$$

This is equivalent to the fact that $N(i,j)v_n(i)v_n(j)$ has constant sign for all i,j. In fact, since $\sum_{i,j} N(i,j)v_n(i)v_n(j) = v_n^T N v_n = \lambda_n = -1$, we obtain $N(i,j)v_n(i)v_n(j) \leq 0$ for all i,j. Let $V_1 = \{i : v_n(i) > 0\}$ and $V_2 = \{i : v_n(i) < 0\}$. We have $V_1 \cup V_2 = V$, because we saw that $|v_n(i)| = v_1(i) > 0$ for all i. On the other hand, if there were two vertices $k, l \in V_1$ or $k, l \in V_2$ joined with an edge, $N(k,l) = (D^{1/2}AD^{1/2})(k,l) = (1/\sqrt{d(k)})A(k,l)(1/\sqrt{d(l)}) > 0$, therefore $N(i,j)v_n(i)v_n(j) > 0$, which is not possible. Then V_1 and V_2 bipartite G, and this is a contradiction. Hence, $\lambda_n > -1$, as wanted.

In conclusion, $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n > -1$. From this, we compute $\lim_{n\to\infty} P^n(i,j)$. We have

$$P^n = D^{1/2}N^n D^{-1/2} = \sum_{k=1}^{n} \lambda_k^n D^{1/2} v_k v_k^T D^{-1/2} = Q + \sum_{k=2}^{n} \lambda_k^n D^{1/2} v_k v_k^T D^{-1/2}.$$

We compute $P^n(i,j)$ as a sum of

$$\begin{aligned} Q(i,j) &= \lambda_1^n (D^{1/2} v_1 v_1^T D^{-1/2})(i,j) = \lambda_1^n (1/\sqrt{d(i)}) v_1(i) v_1(j) \sqrt{d(j)} \\ &= (1/\sqrt{d(i)})\sqrt{d(i)}\sqrt{d(j)}\sqrt{d(j)}/(2|E|) = \pi(j) \end{aligned}$$

and

$$\left(\sum_{k=2}^{n} \lambda_k^n D^{1/2} v_k v_k^T D^{-1/2}\right)(i,j) = \sum_{k=2}^{n} \lambda_k^n (D^{1/2} v_k v_k^T D^{-1/2})(i,j) = \sum_{k=2}^{n} \lambda_k^n \sqrt{\frac{d(j)}{d(i)}} v_k(i) v_k(j).$$

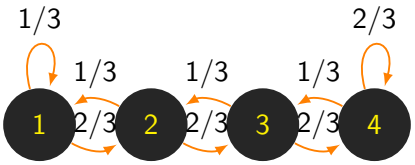
Since $|\lambda_k| < 1$ for $k = 2, \ldots, n$, we conclude that $\lim_{n\to\infty} P(i,j) = \pi(j)$. □

5. Examples

In this section we will show applications of the convergence theorem. The main result of interest is that we are able to compute the asymptotic distribution of the Markov chain.

EXAMPLE 2.3. Consider the points 1, 2, 3 and 4 in the real line. Let X_n be a Markov chain that goes to the right with probability 2/3 and moves to the left with probability 1/3, but with the condition that if it tries to go the left of 1 or to the right of 4 it remains at the same place.

According to this information the Markov chain is



We have that $\mathcal{X} = \{1, 2, 3, 4\}$ and the transition matrix that models this situation is

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}.$$

The chain is irreducible since we can reach any state from any other state. To check the aperiodicity, since it is a class property, it is enough to check it for a single state, say 1. When we are at state 1, in the next step we can return to 1 with probability 1/3, so $\gcd\{n \geq 0 : P^n(1,1) > 0\} = 1$ since $P(1,1) = 1/3 > 0$. By the convergence theorem, $\lim_{n\to\infty} P^n(i,j) = \pi(j)$, where π is the stationary distribution of P. To compute $\pi = (\pi(1), \pi(2), \pi(3), \pi(4))$, we have to solve the linear system $\pi = \pi P$, which is

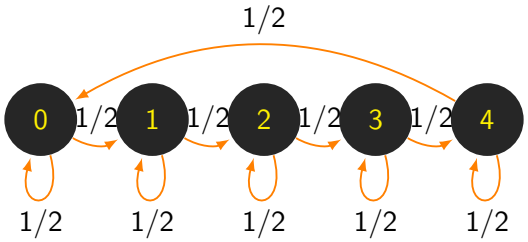
$$\begin{cases} \pi(1) = \frac{1}{3}\pi(1) + \frac{1}{3}\pi(2) \\ \pi(2) = \frac{2}{3}\pi(1) + \frac{1}{3}\pi(3) \\ \pi(3) = \frac{2}{3}\pi(2) + \frac{1}{3}\pi(4) \\ \pi(4) = \frac{2}{3}\pi(3) + \frac{2}{3}\pi(4) \end{cases}.$$

Solving this system and taking into account that $\pi(1) + \pi(2) + \pi(3) + \pi(4) = 1$, we get $\pi = (1/15, 2/15, 4/15, 8/15)$. Then

$$\begin{aligned} &\lim_{n\to\infty} (\mathbb{P}(X_n = 1), \mathbb{P}(X_n = 2), \mathbb{P}(X_n = 3), \mathbb{P}(X_n = 4)) \\ &= \lim_{n\to\infty} (\mathbb{P}(X_0 = 1), \mathbb{P}(X_0 = 2), \mathbb{P}(X_0 = 3), \mathbb{P}(X_0 = 4)) P^n \\ &= (\mathbb{P}(X_0 = 1), \mathbb{P}(X_0 = 2), \mathbb{P}(X_0 = 3), \mathbb{P}(X_0 = 4)) \begin{pmatrix} \pi \\ \pi \\ \pi \\ \pi \end{pmatrix} \\ &= \pi = (1/15, 2/15, 4/15, 8/15) \end{aligned}$$

EXAMPLE 2.4. Let N_n be the number of heads observed in the first n tosses of a fair coin and let $X_n \equiv N_n \mod 5$.

Then the Markov chain is the following



We have that X_n is a Markov chain with state space $\mathcal{X} = \{0,1,2,3,4\}$. Since we have a fair coin, the probability of having heads is $1/2$, then we have a transition probability matrix of the form

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}.$$

Notice that P is irreducible, since we can go from one state to any other state, and it is aperiodic, since $\gcd\{n \geq 0 : P^n(0,0) > 0\} = 1$ because $P(0,0) = 1/2 > 0$. By the convergence theorem, $\lim_{n\to\infty} P^n(i,j) = \pi(j)$. Since each column of P sums 1, we can take $\pi = (1/5, 1/5, 1/5, 1/5, 1/5)$ (a transition matrix with this property is called doubly stochastic). Thus we have that $\lim_{n\to\infty} \mathbb{P}(X_n = k) = 1/5$, for $k = 0,1,2,3,4$.

6. Mixing time

In the previous sections we saw that the distribution of the Markov chain tends to be close (in the total variation distance) to the stationary distribution as time increases. The idea now is to study how rapid this convergence is, in the sense of the total variation distance. Recall that

$$d(n) = \max_{x\in\mathcal{X}} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}}$$

is bounded by $C\alpha^n$ for certain $C > 0$ and $\alpha \in (0,1)$ whenever P is irreducible and aperiodic. In general, we do not know C nor α.

To study the speed of the convergence, define

$$t_{\mathrm{mix}}(\epsilon) = \min\{n : d(n) \leq \epsilon\}.$$

This term is called the *mixing time*, and it refers to the time needed by the Markov chain to behave as the stationary distribution. The goal in this section is to obtain a lower-bound for $t_{\mathrm{mix}}(\epsilon)$ under some conditions.

Let X_n be an irreducible and aperiodic Markov chain, and let P be its transition matrix with uniform stationary distribution π. Define

$$\mathcal{X}_n^x = \{y : P^n(x,y) > 0\}$$

and

$$\Delta = \max_{x\in\mathcal{X}} |\mathcal{X}_1^x|.$$

By induction, $|\mathcal{X}_n^x| \leq \Delta^n$. Indeed, for $n = 1$ this is true by definition of Δ. Suppose that this holds for $n-1$, that is, $|\mathcal{X}_{n-1}^x| \leq \Delta^{n-1}$, and we prove it for n. By Kolmogorov's equation, $P^n(x,y) = \sum_{z\in\mathcal{X}} P^{n-1}(x,z)P(z,y)$. Then, $P^n(x,y) > 0$ if and only if there exists $z \in \mathcal{X}$

such that $P^{n-1}(x,z) > 0$ and $P(z,y) > 0$. Thus, $y \in \mathcal{X}_n^x$ if and only if there exists $z \in \mathcal{X}_{n-1}^x$ such that $y \in \mathcal{X}_1^z$, so

$$\mathcal{X}_n^x = \bigcup_{z \in \mathcal{X}_{n-1}^x} \mathcal{X}_1^z.$$

Taking cardinalities,

$$|\mathcal{X}_n^x| \leq \sum_{z \in \mathcal{X}_{n-1}^x} |\mathcal{X}_1^z| \leq \Delta |\mathcal{X}_{n-1}^x| \leq \Delta \Delta^{n-1} = \Delta^n.$$

Then we can lower-bound $\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}}$:

$$\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \geq P^n(x, \mathcal{X}_n^x) - \pi(\mathcal{X}_n^x) = 1 - \frac{|\mathcal{X}_n^x|}{|\mathcal{X}|} \geq 1 - \frac{\Delta^n}{|\mathcal{X}|}.$$

If

$$n < \frac{\log((1-\epsilon)|\mathcal{X}|)}{\log(\Delta)},$$

we have $\Delta^n < (1-\epsilon)|\mathcal{X}|$, so $\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} > \epsilon$. Thus,

$$t_{\mathrm{mix}}(\epsilon) \geq \frac{\log((1-\epsilon)|\mathcal{X}|)}{\log(\Delta)}. \tag{6.1}$$

THEOREM 2.6. *Let P be the transition matrix of an irreducible and aperiodic Markov chain with finite state space $\mathcal{X}$. Let $\Delta = \max_{x \in \mathcal{X}} |\{y \in \mathcal{X} : P(x,y) > 0\}|$. Then*

$$t_{mix}(\epsilon) \geq \frac{\log((1-\epsilon)|\mathcal{X}|)}{\log(\Delta)}.$$

This result can be applied in Section 4, in the case where the graph $G = (V, E)$ is d-regular, that is, every vertex of G has degree d. In such a case, the stationary distribution is given by

$$\pi(x) = \frac{d}{2|E|}$$

for all $x \in \mathcal{X}$ (see (4.1)), so π is uniform on $\mathcal{X}$. By (6.1),

$$t_{\mathrm{mix}}(\epsilon) \geq \frac{\log((1-\epsilon)|V|)}{\log d}.$$

THEOREM 2.7. *Let $G = (V, E)$ be a d-regular graph. Consider the Markov chain that moves from a vertex x to an adjacent vertex y with probability $1/d$. Then*

$$t_{mix}(\epsilon) \geq \frac{\log((1-\epsilon)|V|)}{\log d}.$$

Chapter 3
Generation of Markov chains

In this chapter we will study examples of Markov chain Monte Carlo (MCMC) algorithms. MCMC methods are a type of algorithms to sample from a given probability distribution, based on constructing a Markov chain with stationary distribution the given probability distribution. By the convergence Theorem 2.1, as time passes this Markov chain will move approximately as the given probability distribution. It is clear that, as more steps, better quality of the approximation. The difficult problem is to find out how many steps are needed to converge to the stationary distribution so that the error is small. A good chain will have to be rapid mixing, meaning that t_{mix} will have to be small.

In what follows, we will present two examples of MCMC algorithms: Metropolis and Glauber algorithms.

1. Metropolis algorithm

In the Metropolis algorithm, instead of starting with a transition matrix P and obtaining a stationary (or invariant) distribution π, we start with a certain distribution π and the goal is to arrive at a transition matrix P whose stationary distribution is π.

Let π be a distribution of $\mathcal{X}$ with all its components positive and let ψ be a symmetric transition matrix. We analyze how to modify ψ in such a way that it becomes a transition matrix P with stationary distribution π.

Given $x \in \mathcal{X}$, the row $\psi(x,\cdot)$ is a probability measure indicating the chances of moving from $x \in \mathcal{X}$ to any other state in $\mathcal{X}$. Thus, given $y \in \mathcal{X}$, $\psi(x,y)$ gives the probability of moving from x to y. Imagine that with probability $a(x,y)$ this movement is accepted, otherwise, with probability $1-a(x,y)$, the movement is not accepted and we remain at x. Obviously, this implies a new behavior of the Markov chain described by a new transition matrix P whose entries are:

$$P(x,y) = \begin{cases} \psi(x,y)a(x,y), & y \neq x, \\ 1 - \sum_{z \neq x} \psi(x,z)a(x,z), & y = x. \end{cases}$$

We want to choose the acceptance probability $a(x,y)$ in such a way that P has as stationary distribution π. By Proposition 2.4, if $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x,y \in \mathcal{X}$, then π is an invariant distribution for P. Assume that we have $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x,y \in \mathcal{X}$. This means that $\pi(x)\psi(x,y)a(x,y) = \pi(y)\psi(y,x)a(y,x)$. If $\psi(x,y) = \psi(y,x) \neq$

0, then we can cancel out ψ so that $\pi(x)a(x,y) = \pi(y)a(y,x)$. Define

$$a(x,y) = \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}.$$

This acceptance probability satisfies $\pi(x)a(x,y) = \pi(y)a(y,x)$ for all $x, y \in \mathcal{X}$. Indeed, if $\pi(y)/\pi(x) \leq 1$, that is, $\pi(y) \neq \pi(x)$, $a(x,y) = \pi(y)/\pi(x)$, then we have that $\pi(x)a(x,y) = \pi(x)\pi(y)/\pi(x) = \pi(y)$. On the other hand, since $a(x,y) = \pi(y)/\pi(x)$, $a(y,x) = 1$ because $1 \leq \pi(x)/\pi(y)$, then $\pi(y)a(y,x) = \pi(y)$.

Thus, the transition matrix P with entries

$$P(x,y) = \begin{cases} \psi(x,y)\min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}, & y \neq x, \\ 1 - \sum_{z\neq x}\psi(x,z)\min\left\{\frac{\pi(z)}{\pi(x)}, 1\right\}, & y = x, \end{cases} \tag{1.1}$$

has as stationary distribution π. A Markov chain having transition matrix P is called *Metropolis chain*.

The Metropolis chain can also be defined in the general case, when the initial transition matrix ψ need not be symmetric. The above computations can be adapted to obtain a transition matrix P having π as a stationary distribution. Indeed, the only step that changes is the simplification of $\psi(x,y)$ and $\psi(y,x)$, since now ψ is not necessarily symmetric. Then, defining

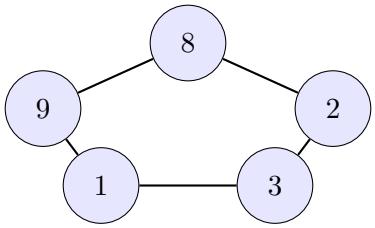
$$a(x,y) = \min\left\{\frac{\pi(y)\psi(y,x)}{\pi(x)\psi(x,y)}, 1\right\},$$

we deduce that the transition matrix P with entries

$$P(x,y) = \begin{cases} \psi(x,y)\min\left\{\frac{\pi(y)\psi(y,x)}{\pi(x)\psi(x,y)}, 1\right\}, & y \neq x, \\ 1 - \sum_{z\neq x}\psi(x,z)\min\left\{\frac{\pi(z)\psi(z,x)}{\pi(x)\psi(x,z)}, 1\right\}, & y = x, \end{cases}$$

has as stationary distribution π.

EXAMPLE 3.1. *Imagine that we have a finite vector and we move among its components. Suppose that, when some time is passed we want to stop in the maximum (in number) of this components and remain there as the time increases, that is, we want that the Markov chain whose states are the positions of the components of the vector has a specific stationary distribution. We will use a Metropolis algorithm to see how the path is and to realize that we pass more times at the greatest values. Let us see this by means of a numeric example. Consider the following vector* $f = (8,2,3,1,9)$ *and let us represent it as a* 2*-regular graph, that is,*



Notice that the state space is $\mathcal{X} = \{1,2,3,4,5\}$ *(vector of the positions). Suppose that once we are in a vertex we have the same probability of visiting its neighbors, that is* $1/2$.

According to this the transition matrix of this Markov chain is

$$\psi = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{pmatrix}.$$

We want that, when time passes, the Markov chain remains at the position of the maximum of the vector f. We modify the transition matrix ψ via the Metropolis algorithm to achieve such a Markov chain. Consider the probability distribution

$$\pi(x) = \frac{\lambda^{f(x)}}{\sum_{z\in\mathcal{X}} \lambda^{f(z)}},$$

for a fixed $\lambda > 1$. This probability distribution favors the positions $x \in \mathcal{X}$ such that $f(x)$ is bigger. We construct a Markov chain with transition matrix given by (1.1) and has π as a stationary distribution:

$$P(x,y) = \begin{cases} \psi(x,y)\min\left\{\lambda^{f(y)-f(x)}, 1\right\}, & y \neq x, \\ 1 - \sum_{z\neq x} \psi(x,z)\min\left\{\lambda^{f(z)-f(x)}, 1\right\}, & y = x. \end{cases}$$

Now we use the following `R` *function to apply the Metropolis algorithm. The inputs of the function are: a vector (in our case f), λ (fixed number), the number of steps we want to walk around the 2-regular graph and the state in which we want to start with. The output will be the position of the vector at which we stay in each step.*

```
find_maximum_vector <- function(f,lambda,steps,first_state) {

  n <- length(f)
  psi <- matrix(rep(0, len = n^2), nrow = n) # initialize psi with zeros
  psi[1,2] <- 0.5; psi[1,n] <- 0.5; # first row of psi
  psi[n,1] <- 0.5; psi[n,n-1] <- 0.5; # last row of psi
  for(i in 2:(n-1)) {
    psi[i,i-1] <- 0.5
    psi[i,i+1] <- 0.5
  } # we have constructed psi

  P <- matrix(rep(0, len = n^2), nrow = n) # initialize P
  for(i in 1:n) {
    for(j in 1:n) {
      if(i != j)
        P[i,j] <- psi[i,j] * min(1, lambda^(f[j]-f[i]))
    }
    P[i,i] <- 1 - sum(P[i,])
  }

  visited_states <- numeric(steps)
  visited_states[1] <- first_state
  for(k in 2:steps) {
    visited_states[k] <- sample(1:n,size=1,replace=FALSE,prob=P[visited_states[k-1],])
  }

  return(visited_states)
}
```

Finally, let us see an example of an implementation:

```
find_maximum_vector(f,5,1000,3)
    [1] 3 3 2 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 1 1 5 5 5 5 5 5 5 5
      5 5 5 5 5 5 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
   [77] 1 5 5 5 1 5 5 5 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
      5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 1 5 5 5 5
  [153] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 1 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
        5 5 5 5 5 5 5 5 5 1 1 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 1 5 5 5 5 5
  [229] 1 5 5 5 5 5 1 1 5 5 5 5 5 5 1 1 5 5 5 5 5 1 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5
        5 5 5 5 5 5 5 5 5 1 1 5 5 5 5 5 5 5 5 5 5 1 5 5 1 5 5 1 1 5 1 5 5 5 5 5 5 5
  [305] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
        5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 1 5 5 1 5 1 5 1 5 5 5 5 5 5 5 5 5 1 5 5 5
  [381] 5 5 5 5 5 5 5 1 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 1 1 1 5 5 1
        5 5 5 5 5 1 5 1 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 1 1 1 5 5 5 5 5 5 5 5 5 5 5
  [457] 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 1 1 1 5 5 5
        5 5 5 5 5 1 1 5 5 5 5 1 1 5 5 5 5 5 5 5 5 5 1 1 5 5 5 5 5 5 5 5 5 5 5 5 1 5
  [533] 1 5 5 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 1 1 1
        1 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 1 1 1 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5
  [609] 5 5 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5
        1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 5
  [685] 5 5 5 5 5 1 1 5 5 5 5 5 5 5 5 1 5 1 1 5 5 5 5 5 5 5 5 5 1 1 5 5 5 5 5 5 5 5
        5 5 5 5 5 5 5 5 5 5 5 5 1 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
  [761] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 5 1 5 5 5 5 1 5 5 5 5 5 5 5
        5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5
  [837] 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 1 5 5 5 5 5 5 1 1 1 5 1 5 5 5 5 5 5 5 5 5 1 5
        5 5 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 1 1 5 1 5
  [913] 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 5 5 1 1 5 5 5 1 1 1 1 1 1
        5 5 5 5 1 1 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
  [989] 5 5 5 5 5 5 1 5 5 5 5 5
```

Notice that the position 5*, which is a* 9*, is repeated several times, which shows us that greater numbers have more probability to be reached.*

2. Glauber algorithm

Let us motivate this algorithm with an example, whose scenario will be the same as in Chapter 4. Let $G = (V, E)$ be a graph. A proper q-coloring is a coloring of the edges E with q colors such that two incident edges to the same vertex have different color. Let $\mathcal{X}$ be the finite space of q-colorings of G. The idea is to choose proper colorings from $\mathcal{X}$ in a uniform way. Thus, given a distribution π uniform in $\mathcal{X}$, we want to construct a Markov chain on the state space $\mathcal{X}$ such that its stationary distribution is π. For a large number of steps, the transitions of this Markov chain will follow the uniform distribution π, by Theorem 2.1.

A proper q-coloring can be denoted by $x \in \{1, \dots, q\}^E$ or by $x : E \to \{1, \dots, q\}$. For a given coloring x and a given edge e, call a color $j \in \{1, \dots, q\}$ allowable at e if j is different from the colors of all edges incident to e. In notation, $j \notin \{x(f) : f \text{ not incident to } e\}$.

Construct a Markov chain on the set of q-proper colorings $\mathcal{X}$ via the following algorithm. Given a proper q-coloring $x \in \mathcal{X}$, generate a new coloring by:

- selecting an edge e in E at random;

- selecting a color j allowable at e in a uniform way (i.e. at random from the allowable colors at e);
- changing the color of e to j.

Let us analyze the transition matrix of this Markov chain. If x and y are two proper colorings that differ in more than one edge, then $P(x,y) = P(y,x) = 0$. Otherwise, if x and y agree everywhere except at an edge e, the probability of moving from x to y is

$$P(x,y) = \frac{1}{|E||A_e(x)|},$$

where $A_e(x)$ is the set of allowable colors at e in x. Notice that $P(x,y) = P(y,x)$, since $A_e(x) = A_e(y)$ whenever the colorings x and y agree everywhere except at the edge e. Thus, the transition matrix P is symmetric, therefore the uniform distribution π on $\mathcal{X}$ is a stationary distribution of P. Note that, if the number of edges compared to the number of colors is small, then the chain given by P is irreducible and aperiodic. Indeed, from any proper q-coloring we can reach any other proper q-coloring by changing the colors of the edges one by one, which implies irreducibility; regarding aperiodicity, we can repeat the same proper q-coloring in a next step, because the color of an edge e is allowable.

This algorithm described can be implemented in R. From a proper q-coloring, the algorithm generates another proper q-coloring by changing the color of an arbitrary edge:

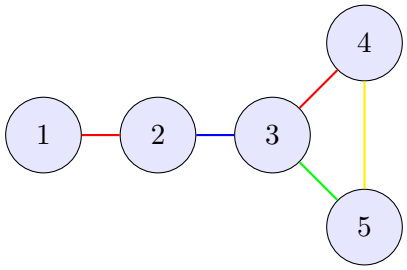
```
generation_proper_coloring <- function(A,q) {
  # Input: q = number of colors;
  #        A = q-proper coloring in the graph, where entry (x,y) is 0 if there is no edge
  #            joining x and y; entry is c in {1,...,q} if there is an edge xy of color c.
  # Output: a proper q-coloring.
  # Goal: to generate q-proper colorings from the one given.

  n <- dim(A)[1] # dimension of the square matrix
  while(TRUE) {
    edge <- sample(1:n,size=2,replace=FALSE,prob=rep(1/n,n)) # we choose a non diagonal
                                                             # entry
    if(A[edge[1],edge[2]] != 0) # we check that there is an edge between the elements of
                                # the entry
      break
  }
  admissible_colors <- setdiff(1:q,c(A[edge[1],],A[,edge[2]])) # admissible colors
  card <- length(admissible_colors) # cardinal (number of admissible colors)
  A[edge[1],edge[2]] <- sample(admissible_colors,size=1,replace=FALSE,prob=rep(1/card,card))
                        # we change the color of the edge among the admissible colors
  A[edge[2],edge[1]] <- A[edge[1],edge[2]] # recall A is symmetric

  return(A)
}
```

Let us see an example of execution of this algorithm considering $q = 7$.

EXAMPLE 3.2. *Consider the following graph:*



We denote by 1 *color red, by* 2 *color blue; by* 3 *color orange; by* 4 *color green; by* 5 *color yellow; by* 6 *color pink and by* 7 *color purple. Notice that this graph picture is an example of a proper coloring using colors* 1, 2, 4 *and* 5.

Now we find another proper coloring of the graph by using the last R *function:*

```
> # matrix of the colorings of the graph
> (A <- cbind(c(0,1,0,0,0),c(1,0,2,0,0),c(0,2,0,1,4),c(0,0,1,0,5),c(0,0,4,5,0)))
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    1    0    0    0
[2,]    1    0    2    0    0
[3,]    0    2    0    1    4
[4,]    0    0    1    0    5
[5,]    0    0    4    5    0
> q <- 7 # number of colors
> # with these two inputs we find another proper coloring.
> generation_proper_coloring(A,q) # call the function
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    1    0    0    0
[2,]    1    0    3    0    0
[3,]    0    3    0    1    4
[4,]    0    0    1    0    5
[5,]    0    0    4    5    0
```

Notice that in this example we obtain another proper coloring by changing the color of edge $(3, 2)$.

Let us extend these ideas in a general setting. Let $\mathcal{S}$ be a finite set (for example, a set $\{1, \ldots, q\}$ of colors). Consider a state space $\mathcal{X} \subseteq \mathcal{S}^E$. A state $x \in \mathcal{X}$ can be viewed as a map $x : E \to \mathcal{S}$ (for instance, a proper coloring).

Let π be a positive probability distribution on $\mathcal{X}$. The goal is to find a Markov chain on $\mathcal{X}$ with stationary distribution π. For $x \in \mathcal{X}$ and $e \in E$, let $\mathcal{X}(x, e) = \{y \in \mathcal{X} : y(f) = x(f), \forall\, f \neq e\}$ (for example, if x is a proper coloring, $\mathcal{X}(x, e)$ is the set of q-proper colorings that agree with x except possibly at e). Define the following probability:

$$\pi^{x,e}(y) = \pi(y|\mathcal{X}(x,e)) = \frac{\pi(\{y\} \cap \mathcal{X}(x,e))}{\pi(\mathcal{X}(x,e))} = \begin{cases} \frac{\pi(y)}{\pi(\mathcal{X}(x,e))}, & y \in \mathcal{X}(x,e), \\ 0, & y \notin \mathcal{X}(x,e). \end{cases}$$

The algorithm proceeds as follows. Given $x \in \mathcal{X}$, generate a new state in x by:

- selecting a random edge e;
- selecting y in $\mathcal{X}(x, e)$ with probability $\pi^{x,e}(y)$.

This procedure generates a Markov chain. Let us compute its transition matrix P. Given two states $x, y \in \mathcal{X}$, if $y \notin \cup_{e \in E} \mathcal{X}(x, e)$ (for instance, if x and y are two proper colorings differing in more than one edge), then $P(x, y) = P(y, x) = 0$. Otherwise, if $y \in \mathcal{X}(x, e)$ for certain $e \in E$ (for example, if x and y are two proper colorings that agree except at most at e),

$$P(x, y) = \frac{\pi^{x,e}(y)}{|E|} = \frac{\pi(y)}{|E|\pi(\mathcal{X}(x, e))}.$$

Notice that $\pi(x)P(x, y) = \pi(y)P(y, x)$, because $\mathcal{X}(x, e) = \mathcal{X}(y, e)$ whenever $y \in \mathcal{X}(x, e)$. This implies that π is a stationary distribution of P. Note also that P is irreducible and aperiodic: aperiodicity follows from the fact that, starting at x, one can select $x \in \mathcal{X}(x, e)$ in a next step.

Chapter 4
An application of Markov chains in combinatorial design theory: proper colorings of the complete graph K_n

This chapter is devoted to a conjecture on the random generation of proper edge colorings of the complete graph K_n. The complete graph K_n is defined as the graph with n vertices such that all vertices are joined with an edge. Taking as main reference the paper by Dotan and Linial [7], we introduce the setting and state the conjecture.

In [7] the authors describe a random procedure to generate a proper edge coloring of the complete graph K_n with even n by using $n-1$ colors. We describe this procedure together with new results which extend the ones in [7]. We also have written an implementation of the resulting algorithm in `R` code. The implementation is new and differs slightly from the one proposed in [7] by modeling more efficiently the same Markov chain. This is done in Section 1.

Numerical experiments on the generation of proper edge colorings have been performed for over 30 years, but no mathematical proof has been achieved to justify the numerical results. One of the main goals of this chapter is to tackle one of the conjectures arisen from the numerical experiments: Conjecture 1 from [7]. Roughly speaking, it says that the algorithm converges asymptotically in polynomial time. Our numerical experiments with the algorithm written in R makes us to believe that the algorithm must always converge, and not just asymptotically, in polynomial time. Thus, our aim will be to tackle the proof of the convergence of the algorithm, so that part of Conjecture 1 will be solved. We obtain new results and an original solution to the conjecture for K_4 and K_6, via both a computational (in `R`) and an analytic proof (Section 2 and Section 3). We give new ideas on how to tackle the conjecture for the general case K_n (Section 4).

The theory of Markov chains will be used for the proofs. The convergence Theorem (2.1) is not enough to deal with this problem, as the Markov chain generated by the colorings is not irreducible. We will show that the proper edge colorings, which are the target of the algorithm, are absorbing states, and we will describe the necessary theory to analyze the convergence of the chain, which will extend the one described in Chapter 3.

The algorithm in `R` is thought in the spirit of Metropolis algorithm (see Chapter 3), in the sense that one moves from a coloring (state of the Markov chain) to another coloring (another state of the Markov chain) if the new coloring is "closer" to a proper coloring, which is, to some extent, an acceptance probability.

1. Description of the problem and new results

We start this section with the definition of proper coloring and the known result that, for n even, the complete graph K_n can be properly colored with just $n-1$ colors. These definitions and Proposition 4.1 are taken from [3].

DEFINITION 4.1. *A proper edge coloring, or simply a proper coloring in this work, of a graph G is an edge coloring such that no two incident edges have the same color. That is, if $e, e' \in E(G)$ with $e \cap e' \neq \emptyset$, then e and e' have different colors.*

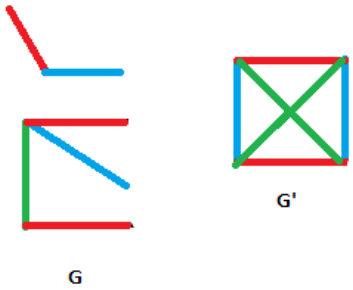


FIG. 1. Examples of proper colorings. To the left, a general graph G with a proper coloring done with colors blue, red and green; to the right, the complete graph K_4, G', with a proper coloring of 3 colors: blue, red and green.

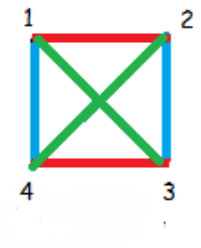
Proper colorings of complete graphs can be viewed as symmetric Latin squares. A Latin square is a square $n \times n$ matrix with entries from a set of n symbols such that the entries in each row and in each column are pairwise distinct (see Figure 2 for an example). From a proper edge coloring of the complete graph K_n, n even, with colors $\{1, 2, \ldots, n-1\}$, we can obtain a symmetric Latin square of order n by placing the color of edge $\{i, j\}$ at the (i, j) entry of a square $n \times n$ matrix, and by placing the symbol n on its main diagonal (see Figure 3 for an illustration).

0	4	8	2	3	9	6	7	1	5
3	6	2	8	7	1	9	5	0	4
8	9	3	1	0	6	4	2	5	7
1	7	6	5	4	8	0	3	2	9
2	1	9	0	6	7	5	8	4	3
5	2	7	4	9	3	1	0	8	6
4	3	0	6	1	5	2	9	7	8
9	8	5	7	2	0	3	4	6	1
7	0	1	9	5	4	8	6	3	2
6	5	4	3	8	2	7	1	9	0

FIG. 2. Example of a Latin square.

DEFINITION 4.2. *The chromatic index of a graph G, denoted by $\mathcal{X}'(G)$, is the minimum number of colors required by a proper coloring of G.*

REMARK 4.1. *The notation with the prime symbol in $\mathcal{X}'(G)$ is used to distinguish the chromatic index and the chromatic number, $\mathcal{X}(G)$. The chromatic number $\mathcal{X}(G)$ is defined as the minimum number of colors required to construct a proper vertex coloring of G.*



$$\begin{pmatrix} 4 & 1 & 3 & 2 \\ 1 & 4 & 2 & 3 \\ 3 & 2 & 4 & 1 \\ 2 & 3 & 1 & 4 \end{pmatrix}$$

FIG. 3. Example of proper coloring with its associated Latin square. The complete graph is K_4, with colors red = 1, blue = 2 and green = 3.

PROPOSITION 4.1. *The chromatic index of the complete graph K_n, $\mathcal{X}'(K_n)$, is $n-1$, when n is even.* [1]

PROOF. Let C be a proper coloring of K_n. We prove that the number of colors used in C is greater than or equal to $n-1$. For all vertex $v \in V(K_n)$, its degree is $d(v) = n-1$. Thus, since all edges incident to v must have distinct colors in C, necessarily the number of colors in C is $\geq n-1$. This gives $\mathcal{X}'(K_n) \geq n-1$. In fact, this is a particular case of the inequality $\mathcal{X}'(G) \geq \Delta(G)$, where $\Delta(G)$ is defined as the maximum degree in $V(G)$.

We prove $\mathcal{X}'(K_n) \leq n-1$. By Vizing's Theorem (see e.g. Beşeri [3, Theorem 44]), $\mathcal{X}'(G) \leq \Delta(G)+1$, so we already know that $\mathcal{X}'(K_n) \leq (n-1)+1 = n$. To prove the stronger inequality $\mathcal{X}'(K_n) \leq n-1$, we must search for a specific example. Here is where we use the fact that n is even. Label the vertices of K_n as $0, 1, \ldots, n-1$. Draw a circle with center at 0 and put the remaining $n-1$ vertices on the circle, where vertex 1 is above 0. Let $\alpha_1, \ldots, \alpha_{n-1}$ be $n-1$ colors. We proceed as follows: paint the vertical edge $\{0,1\}$ and all the horizontal edges $\{2, n-1\}, \{3, n-2\}, \ldots, \{n/2, n/2+1\}$ with color α_1. Rotate the circle counterclockwise so that 2 is located above 0. Paint the vertical edge $\{0,2\}$ and all the horizontal edges $\{3,1\}, \{4, n-1\}, \ldots, \{n/2+1, n/2+2\}$ with color α_2. Rotate the circle counterclockwise so that 3 is placed above 0. Paint the vertical edge $\{0,3\}$ and all the horizontal edges $\{4,2\}, \{5,1\}, \ldots, \{n/2+2, n/2+3\}$ with color α_3. Continue this procedure until $n-1$ is located above 0. An outline of the procedure is shown in Figure 4. Notice that we have performed $n-1$ steps. In each step i, the edges painted with color α_i are disjoint, therefore we have exhibited a proper coloring of K_n. Since we have used $n-1$ colors, we deduce $\mathcal{X}'(K_n) \leq n-1$. Thus, $\mathcal{X}'(K_n) = n-1$, as wanted. □

Now we are in position of describing the problem stated in [7]. We are going to work with (edge) colorings of the complete graph K_n, being n even, of $n-1$ colors. The colorings are viewed as vertices or nodes in a Markov chain, where transition probabilities will be described according to an acceptance probability that will depend on the nearness of the coloring to a proper coloring.

If we fix a coloring C of K_n with $n-1$ colors, we will call $a_{\nu,\mu}(C)$ the total number of incident edges to the vertex ν with color μ. Taking the convention that $\nu \in [n] = \{1, \ldots, n\}$

[1] When n is odd, $\mathcal{X}'(K_n) = n$, but we do not prove this fact because we are interested in the even case.

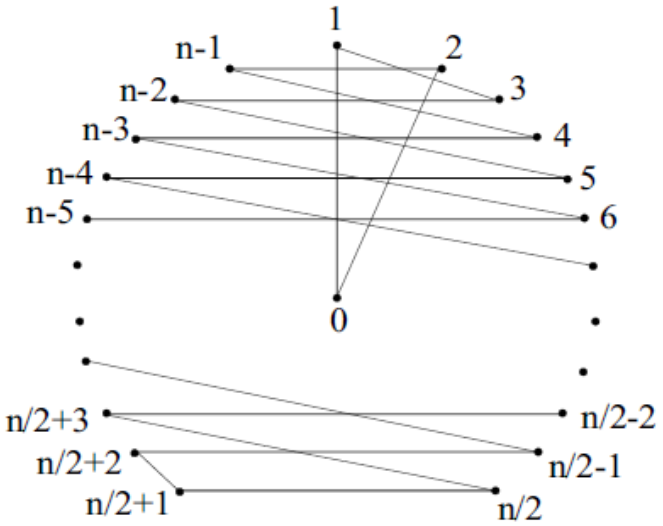


FIG. 4. Schema of the procedure used in the proof of Proposition 4.1. Picture taken from [3].

and $\mu \in [n-1]$, we define the potential of a vertex ν of K_n as

$$\phi(\nu) = \sum_{\mu=1}^{n-1} a_{\nu,\mu}(C)^2.$$

The potential of the coloring C is defined as

$$\phi(C) = \sum_{\nu=1}^{n} \phi(\nu) = \sum_{\nu=1}^{n} \sum_{\mu=1}^{n-1} a_{\nu,\mu}(C)^2.$$

The potential $\phi(\nu)$ measures the number of incident edges to ν colored different, and penalizes raising to the square the fact of having edges incident to ν with the same color. The potential $\phi(C)$ is defined as the sum of the potential of the vertices, so that we are measuring to some extent the distance to a proper coloring. Intuitively, a proper coloring has the least possible potential (see Lemma 4.2).

In our Markov chain, where colorings are placed as vertices, two colorings are connected if C and C' differ by the color of a single edge. This will be referred as C' being a recoloring of C.

DEFINITION 4.3. *Given a coloring C, we say that C' is a recoloring of C if C and C' differ by the color of a single edge.*

Our Markov chain can move from C to a recoloring C' of it if and only if $\phi(C') \leq \phi(C)$. That is, the acceptance probability $a(C, C')$ is $1/|N_\phi(C)|$ if $\phi(C') \leq \phi(C)$, and 0 otherwise (see the analogy with Chapter 3). Here,

$$N_\phi(C) := \{C' : C' \text{ is a recoloring of } C \text{ and } \phi(C') \leq \phi(C)\}$$

and $|N_\phi(C)|$ is the number of neighbors of C with less or equal potential. The idea is that we move on if and only if we do not move away from a proper coloring. The fact that we move even when the potential is equal gives more "flexibility" to our chain, so that we will

not stop moving when arriving at a "local minimum". These concepts will be explained later on.

The above described algorithm is the first one described in Dotan and Linial [7] and, after some experimental analysis, leads them to formulate the following conjecture. In [7], $\mathcal{G}_n$ stands for the graph with nodes being the colorings and edges joining recolorings. The term "mild random walk" means that we may move from C to a recoloring C' if $\phi(C') \leq \phi(C)$. They speak about "strict random walk" when we may move from C to a recoloring C' if $\phi(C') < \phi(C)$.

CONJECTURE 4.1. [7, Conjecture 1] *The mild random walk on $\mathcal{G}_n$ started from a uniformly random starting point asymptotically almost surely reaches a one-factorization in $\mathcal{O}(n^4)$ steps.*

The conjecture has two parts, the first one concerning the convergence with high probability and the second one about the speed of convergence. Being unable to tackle the conjecture, the authors suggest an alternative procedure which changes colors of edges incident to two vertices for which they manage to show convergence and rapid mixing. However each step of this modified Markov chain requires a time polynomial in n. They also leave open the problem to find a true Metropolis algorithm, by replacing the value 0 in the above acceptance probability by some small $\epsilon > 0$, in order to obtain rapid mixing to a uniform distribution on proper edge colorings.

As we have mentioned in the introduction, our main goal is to concentrate on Conjecture 4.2, as we believe that the resulting Markov chain does converge to proper colorings, not just with high probability, and that the running time is $\mathcal{O}(n^p)$ with $2 < p < 3$. The precise statement is formulated in Conjecture 4.2 below after giving some preliminary notation and results.

Now we study some results concerning the potential function.

LEMMA 4.1. *Let C be a coloring of K_n such that the edge uv (which is the edge that joins vertices u and v) is painted in color μ_0. Let C' be a recoloring of C, in such a way that the edge uv of K_n is painted in another color $\mu_1 \neq \mu_0$, that is, C and C' are equal except for the color of the edge uv. Then we have*

$$\phi(C') = \phi(C) + 2(a_{u,\mu_1}(C) + a_{v,\mu_1}(C) + 2 - a_{u,\mu_0}(C) - a_{v,\mu_0}(C)).$$

PROOF. Notice that $a_{\nu,\mu}(C) = a_{\nu,\mu}(C')$ for all $\nu \in [n]$ different from u and v and for all $\mu \in [n-1]$ different from μ_0 and μ_1.

Since $\mu_0 \neq \mu_1$, we have

$$\begin{cases} a_{u,\mu_0}(C) = 1 + a_{u,\mu_0}(C'), \\ a_{v,\mu_0}(C) = 1 + a_{v,\mu_0}(C'), \end{cases}$$

because in C' there is one less edge painted in color μ_0 than in C. Indeed, the only difference between C and C' is the color of one edge. Moreover,

$$\begin{cases} a_{u,\mu_1}(C') = 1 + a_{u,\mu_1}(C), \\ a_{v,\mu_1}(C') = 1 + a_{v,\mu_1}(C), \end{cases}$$

since in C' there is one more edge painted in color μ_1.

Taking this into account and applying the definition of the potential of C' announced before,

$$\phi(C') = \sum_{\nu=1}^{n} \phi(\nu) = \sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}(a_{\nu,\mu}(C'))^2 = \sum_{\substack{\nu=1\\ \nu\neq u,v}}^{n}\sum_{\mu=1}^{n-1}(a_{\nu,\mu}(C'))^2 + \sum_{\nu=1}^{n}\sum_{\substack{\mu=1\\ \mu\neq\mu_0,\mu_1}}^{n-1}(a_{\nu,\mu}(C'))^2$$

$$+(a_{u,\mu_0}(C'))^2 + (a_{v,\mu_0}(C'))^2 + (a_{u,\mu_1}(C'))^2 + (a_{v,\mu_1}(C'))^2 = \sum_{\substack{\nu=1\\ \nu\neq u,v}}^{n}\sum_{\mu=1}^{n-1}(a_{\nu,\mu}(C))^2$$

$$+\sum_{\nu=1}^{n}\sum_{\substack{\mu=1\\ \mu\neq\mu_0,\mu_1}}^{n-1}(a_{\nu,\mu}(C))^2 + (a_{u,\mu_0}(C)-1)^2 + (a_{v,\mu_0}(C)-1)^2 + (a_{u,\mu_1}(C)+1)^2 + (a_{v,\mu_1}(C)+1)^2$$

$$= \phi(C) + 2(a_{u,\mu_1}(C) + a_{v,\mu_1}(C) + 2 - a_{u,\mu_0}(C) - a_{v,\mu_0}(C)).$$

□

COROLLARY 4.1. *Let C be a coloring of K_n, then $\phi(C)$ is even.*

PROOF. Let C an arbitrary coloring of K_n. We want to see that $\phi(C)$ is even. Let us consider a coloring C_0 with a single color. We have that

$$\phi(C_0) = \sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}(a_{\nu,\mu}(C_0))^2 = n(n-1)^2.$$

Notice that we can find a sequence of colorings $C_0, C_1, \ldots, C_m = C$ in such a way that C_{i+1} comes from C_i by changing the color of a single edge. By Lemma 4.1, $\phi(C_{i+1})$ has the same parity as $\phi(C_i)$. Then, since $\phi(C_0)$ is even (because n is even), $\phi(C)$ is also even. □

LEMMA 4.2. *Let C be a coloring of K_n. Then $n(n-1) \leq \phi(C) \leq n(n-1)^2$. Moreover, it holds:*

- *$\phi(C) = n(n-1)$ if and only if C is a proper coloring.*
- *$\phi(C) = n(n-1)^2$ if and only if all edges of K_n are painted with the same color.*

PROOF. Notice that it suffices to prove that $n-1 \leq \phi(\nu) \leq (n-1)^2$ for all vertex ν and

- $\phi(\nu) = (n-1)$ if and only if all the incident edges of ν have different color.
- $\phi(C) = n(n-1)^2$ if and only if all the incident edges of ν have the same color.

- Firstly, we prove by **backwards** induction over the number of colors incident to ν that $n-1 \leq \phi(\nu)$.
  Fix an arbitrary coloring C. Define the set

  $$\mathcal{A} = \{\text{colors incident to } \nu \text{ according to } C\}.$$

  According to this definition, the potential of ν can be written as

  $$\phi(\nu) = \sum_{\mu\in\mathcal{A}}(a_{\nu,\mu}(C))^2.$$

  The first step of the induction is when $|\mathcal{A}| = n-1$, and it is obviously fulfilled, since we have $n-1$ colors to share with $n-1$ edges, so the only possibility is that $\phi(\nu) = 1+\overset{(n-1)}{\ldots}+1 = n-1$. By hypothesis of induction, assume that for $|\mathcal{A}| = k+1$ it holds that $\phi(\nu) \geq n-1$, and we have to prove this condition for $|\mathcal{A}| = k$, $1 \leq k \leq n-2$.

Since k is at the most $n-2$ and ν has $n-1$ incident edges, ν has at least two edges with the same color. Suppose that $\mu_0 \in \mathcal{A}$ is the color repeated in the incident edges of ν. Now we choose one edge colored with μ_0 and change the color by μ_1. Notice that we have a recoloring C' of C with one color more than C. Let us define

$$\mathcal{A}' = \{\text{colors incident to } \nu \text{ according to } C'\}.$$

We have that $\mathcal{A}' = \mathcal{A} \cup \{\mu_1\}$, so $|\mathcal{A}'| = k+1$. By induction,

$$\begin{aligned} n-1 \leq \sum_{\mu\in\mathcal{A}'} (a_{\nu,\mu}(C'))^2 &= \sum_{\substack{\mu\in\mathcal{A}\cap\mathcal{A}'\\ \mu\neq\mu_0}} (a_{\nu,\mu}(C'))^2 + (a_{\nu,\mu_0}(C'))^2 + \underbrace{(a_{\nu,\mu_1}(C'))^2}_{=1} \\ &= \sum_{\substack{\mu\in\mathcal{A}\cap\mathcal{A}'\\ \mu\neq\mu_0}} (a_{\nu,\mu}(C))^2 + (a_{\nu,\mu_0}(C)-1)^2 + 1 \\ &= \sum_{\mu\in\mathcal{A}} (a_{\nu,\mu}(C))^2 + 2(1 - \underbrace{a_{\nu,\mu_0}(C)}_{\geq 2}) \leq \sum_{\mu\in\mathcal{A}} (a_{\nu,\mu}(C))^2 - 2 \\ &< \sum_{\mu\in\mathcal{A}} (a_{\nu,\mu}(C))^2 = \phi(\nu). \end{aligned}$$

- Finally, we prove by induction over the number of colors incident to ν that $\phi(\nu) \leq (n-1)^2$.
  Fix an arbitrary coloring C and consider $\mathcal{A}$ defined as before. The first step of the induction is when $|\mathcal{A}| = 1$, so all the incident edges of ν are of the same color. Then $\phi(\nu) = (n-1)^2$, since one color is repeated $n-1$ times and the other colors are not in $\mathcal{A}$. Suppose by hypothesis of induction that for $|\mathcal{A}| = k$ it holds that $\phi(\nu) < (n-1)^2$, and we prove this condition for $|\mathcal{A}| = k+1$, $1 \leq k \leq n-2$. Notice that $|\mathcal{A}| = k+1 \geq 2$, thereby the incident edges of ν have at least two different colors, call them μ_0 and μ_1. Now we paint all the edges painted in μ_1 with color μ_0 (joining in this way two colors in a single one). Then we have a recoloring of C, C', with one color less than C, that is, using the same notation as before, $\mathcal{A}' = \mathcal{A}\backslash\{\mu_1\}$. Then $|\mathcal{A}'| = |A| - 1 = k$, and by induction hypothesis,

$$\begin{aligned} (n-1)^2 \geq \sum_{\mu\in\mathcal{A}'} (a_{\nu,\mu}(C'))^2 &= \sum_{\substack{\mu\in\mathcal{A}\cap\mathcal{A}'\\ \mu\neq\mu_0}} (a_{\nu,\mu}(C'))^2 + (a_{\nu,\mu_0}(C'))^2 + \underbrace{(a_{\nu,\mu_1}(C'))^2}_{=0} \\ &= \sum_{\substack{\mu\in\mathcal{A}\cap\mathcal{A}'\\ \mu\neq\mu_0}} (a_{\nu,\mu}(C))^2 + (a_{\nu,\mu_0}(C) + a_{\nu,\mu_1}(C))^2 \\ &= \sum_{\mu\in\mathcal{A}} (a_{\nu,\mu}(C))^2 + 2\underbrace{a_{\nu,\mu_0}(C)}_{\geq 1}\underbrace{a_{\nu,\mu_1}(C)}_{\geq 1} \\ &\geq \sum_{\mu\in\mathcal{A}} (a_{\nu,\mu}(C))^2 + 2 \geq \sum_{\mu\in\mathcal{A}} (a_{\nu,\mu}(C))^2 = \phi(\nu). \end{aligned}$$

Hence, we arrive at $n-1 \leq \phi(\nu) \leq (n-1)^2$, and since K_n has n vertices, according to the definition of the potential of a coloring, it holds that the potential of the coloring C satisfies that $n(n-1) \leq \phi(C) \leq n(n-1)^2$. □

The idea of the algorithm is, from a given coloring of K_n with $n-1$ colors, to move in the Markov chain until arriving at a proper coloring. One of the problems of this procedure are the so-called local minimums, which are colorings that have less or equal potential to their neighbors.

If we were using a strict walk on the Markov chain, in the sense that we moved from C to a recoloring C' of it if and only if $\phi(C') < \phi(C)$, then our algorithm would stop when arriving at a local minimum coloring C, so there would not be convergence. Intuitively, the algorithm would converge asymptotically, because the proportion of local minimums seems to decrease as n grows.

In our case, as we allow the chain to move even when $\phi(C')$ is equal to $\phi(C)$, it seems that this problem should not appear, so convergence for any order n should be expected.

We study new and original properties of local minimum colorings in what follows.

DEFINITION 4.4. *We say that a coloring C is a local minimum if $\phi(C) \leq \phi(C')$ for all coloring C' that can be reached from C by changing the color of a single edge (i.e., for all recoloring C').*

LEMMA 4.3. *Let C be a coloring of K_n such that the edge uv is painted in color μ_0. If $\phi(C') \geq \phi(C)$ for all recoloring C' of the edge uv of C, then*

$$a_{u,\mu_0}(C) + a_{v,\mu_0}(C) \leq 3.$$

PROOF. Since we are assuming that $\phi(C') \geq \phi(C)$, by Lemma 4.1

$$a_{u,\mu_1}(C) + a_{v,\mu_1}(C) + 2 - a_{u,\mu_0}(C) - a_{v,\mu_0}(C) \geq 0,$$

that is,

$$a_{u,\mu_1}(C) + a_{v,\mu_1}(C) + 2 \geq a_{u,\mu_0}(C) + a_{v,\mu_0}(C)$$

for all $\mu_1 \neq \mu_0$. Now, applying $\sum_{\mu_1 \neq \mu_0}$ at both sides we obtain

$$\begin{aligned}(n-2)(a_{u,\mu_0}(C) + a_{v,\mu_0}(C)) &\leq \sum_{\mu_1 \neq \mu_0} a_{u,\mu_1}(C) + \sum_{\mu_1 \neq \mu_0} a_{v,\mu_1}(C) + 2(n-2) \\ = \sum_{\mu_1=1}^{n-1} a_{u,\mu_1}(C) - a_{u,\mu_0}(C) &+ \sum_{\mu_1=1}^{n-1} a_{v,\mu_1}(C) - a_{v,\mu_0}(C) + 2(n-2) \\ = (n-1) - a_{u,\mu_0}(C) + (n-1) - a_{v,\mu_0}(C) &+ 2(n-2) = 4n - 6 - (a_{u,\mu_0}(C) + a_{v,\mu_0}(C)).\end{aligned}$$

Then,

$$(n-2)(a_{u,\mu_0}(C) + a_{v,\mu_0}(C)) \leq 4n - 6 - (a_{u,\mu_0}(C) + a_{v,\mu_0}(C)),$$

which is the same as saying

$$(n-1)(a_{u,\mu_0}(C) + a_{v,\mu_0}(C)) \leq 4n - 6.$$

Therefore

$$a_{u,\mu_0}(C) + a_{v,\mu_0}(C) \leq \frac{4n-6}{n-1} < 4,$$

as n increases. Since $a_{u,\mu_0}(C) + a_{v,\mu_0}(C) \in \mathbb{Z}$, we arrive at

$$a_{u,\mu_0}(C) + a_{v,\mu_0}(C) \leq 3.$$

□

Hereinafter, we will denote by $\mu(uv)$ the color of the edge uv of the initial coloring C.

LEMMA 4.4 (Characterization of local minimum colorings of K_n). *Let C be a coloring of K_n. The following conditions are equivalent:*

(1) $\phi(C') \geq \phi(C)$, for all recoloring C' of C.

(2) $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) \leq 3$ *for all edge* uv*, and if for some edge* uv *the equality holds, then* $a_{u,\mu}(C) + a_{v,\mu}(C) \geq 1$ *for each color* μ*.*

Proof.

- (1) implies (2):
  By Lemma 4.3, we know that if $\phi(C') \geq \phi(C)$ for all recoloring C', then $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) \leq 3$, for every edge uv. Suppose that there exists an edge uv such that $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) = 3$. Then, by Lemma 4.1, for all $\mu_1 \neq \mu(uv)$ we have
  $$a_{u,\mu_1}(C) + a_{v,\mu_1}(C) + 2 \geq a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) = 3.$$
  Thus, for all $\mu_1 \neq \mu(uv)$, $a_{u,\mu_1}(C) + a_{v,\mu_1}(C) \geq 3 - 2 = 1$.
- (2) implies (1):
  Suppose by contradiction that there exists C' recoloring of C such that $\phi(C') < \phi(C)$. Assume that C' consists in repainting an edge u_0v_0 with color $\mu_0(u_0v_0) \neq \mu(u_0v_0)$. Since $\phi(C') < \phi(C)$, by Lemma 4.1:
  $$a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) > a_{u_0,\mu_0(u_0v_0)}(C) + a_{v_0,\mu_0(u_0v_0)}(C) + 2.$$
  Now we distinguish two cases:
  - case $a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) = 3$: we have that
    $$3 > a_{u_0,\mu_0(u_0v_0)}(C) + a_{v_0,\mu_0(u_0v_0)}(C) + 2 \underbrace{\geq}_{\text{by hypothesis}} 1 + 2 = 3,$$
    and this is a contradiction.
  - case $a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) \leq 2$: we have that
    $$2 > a_{u_0,\mu_0(u_0v_0)}(C) + a_{v_0,\mu_0(u_0v_0)}(C) + 2,$$
    and this is a contradiction, since $a_{u_0,\mu_0(u_0v_0)}(C) + a_{v_0,\mu_0(u_0v_0)}(C) \geq 0$.

□

Lemma 4.5 (Characterization of proper colorings of K_n). *Let* C *be a coloring of* K_n*. The following statements are equivalent:*

(1) $\phi(C') > \phi(C)$ *for all recoloring* C' *of* C*.*
(2) The following two conditions fulfill:
  (a) $a_{u,\mu}(C) + a_{v,\mu}(C) \geq 1$ *for all edge* uv *and for each color* μ*.*
  (b) $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) \leq 3$ *for all edge* uv*, and if for some edge* uv *the equality holds, then* $a_{u,\mu}(C) + a_{v,\mu}(C) \geq 2$ *for each color* μ*.*
(3) C *is a proper coloring.*

Proof.

- (1) implies (2):
  By Lemma 4.3 we know that $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) \leq 3$ for all edge uv. Suppose that there exists an edge uv such that $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) = 3$. Since $\phi(C') > \phi(C)$ for all recoloring C' of C, by Lemma 4.1, for each color $\mu_1 \neq \mu(uv)$ we have
  $$a_{u,\mu_1}(C) + a_{v,\mu_1}(C) + 2 > a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) = 3.$$

Then $a_{u,\mu_1}(C) + a_{v,\mu_1}(C) > 1$, that is, $a_{u,\mu_1}(C) + a_{v,\mu_1}(C) \geq 2$, which proves (b).
Now we prove (a). Since $\phi(C') > \phi(C)$ for all recoloring C' of C, we have that for all edge uv and for each color $\mu_1 \neq \mu(uv)$,

$$a_{u,\mu_1}(C) + a_{v,\mu_1}(C) + 2 > a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) \geq 2.$$

Then, for all $\mu_1 \neq \mu(uv)$, $a_{u,\mu_1}(C) + a_{v,\mu_1}(C) \geq 1$, which shows (a).

- (2) implies (1):
Suppose by contradiction that there exists C' (recoloring of C) such that $\phi(C') \leq \phi(C)$. Assume that C' consists in repainting the edge u_0v_0 with color $\mu_0(u_0v_0) \neq \mu(u_0v_0)$. We have, by Lemma 4.1,

$$a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) \geq a_{u_0,\mu_0(u_0v_0)}(C) + a_{v_0,\mu_0(u_0v_0)}(C) + 2.$$

We distinguish two cases:
  - case $a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) = 3$: we have that

$$3 \geq a_{u_0,\mu_0(u_0v_0)}(C) + a_{v_0,\mu_0(u_0v_0)}(C) + 2 \underbrace{\geq}_{(b)} 2 + 2 = 4,$$

    and this is a contradiction.
  - case $a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) \leq 2$: we have that

$$2 \geq a_{u_0,\mu_0(u_0v_0)}(C) + a_{v_0,\mu_0(u_0v_0)}(C) + 2 \underbrace{\geq}_{(a)} 1 + 2 = 3,$$

    which is again a contradiction.
- (2) implies (3):
Suppose by contradiction that there exists an edge uv with $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) = 3$. Then, by (b), we have that

$$a_{u,\mu}(C) + a_{v,\mu}(C),$$

for each color μ. Consider Figure 5, where the edges painted in color blue represent any color (or colors) different from $\mu(uv)$, painted in green.

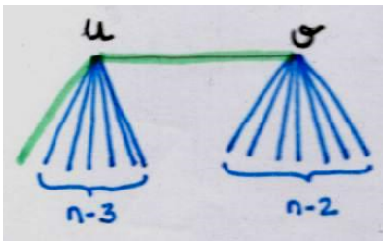


FIG. 5. General form of the edge uv of K_n.

As we can see in 5, the number of edges with color different from $\mu(uv)$ is $2n-5$, and this number, by what we proved, is greater or equal than $2(n-2) = 2n-4$, since $n-2$ is the number of possible colors different from $\mu(uv)$. But this is a contradiction.
- (3) implies (2):
If C is proper, for all edge uv we have that

$$\underbrace{a_{u,\mu(uv)}(C)}_{=1} + \underbrace{a_{v,\mu(uv)}(C)}_{=1} = 2,$$

and this proves (b). Moreover, we also have that, for all edge uv and for each color μ, $a_{u,\mu}(C) + a_{v,\mu}(C) = 2 \geq 1$, and this proves (a).

□

COROLLARY 4.2. *Let C be a coloring of K_n. Then C is proper if and only if $\phi(C) < \phi(C')$ for all recoloring C' of C.*

Next, we are going to see that the distance between a local minimum coloring and a proper coloring depends only on the number of monochromatic 2-paths of the local minimum coloring. Denote by

$$\phi_C(\alpha) = \sum_{\nu=1}^{n} a_{\nu,\alpha}(C)^2, \tag{1.1}$$

$$c_C(\alpha) = \text{number of times color } \alpha \text{ appears in coloring } C,$$

$$V_\alpha(C) = \text{number of 2-paths of color } \alpha$$

and

$$V(C) = \text{number of monochromatic 2-paths} = \sum_{\alpha=1}^{n-1} V_\alpha(C).$$

Notice that

$$\phi(C) = \sum_{\alpha=1}^{n-1} \phi_C(\alpha),$$

$$\sum_{\alpha=1}^{n-1} c_C(\alpha) = \frac{n(n-1)}{2}$$

and

$$\sum_{\nu=1}^{n} a_{\nu,\alpha}(C) = 2c_C(\alpha). \tag{1.2}$$

LEMMA 4.6. *Let C be a local minimum. Then, for each color α,*

$$V_\alpha(C) = \frac{\phi_C(\alpha)}{2} - c_C(\alpha).$$

As a consequence,

$$V(C) = \frac{\phi(C) - n(n-1)}{2}.$$

PROOF. By Lemma 4.4, C is composed by α-colored 2-paths and independent edges. A 2-path of color α contributes 6 to $\phi_C(\alpha)$ in (1.1) and 4 to (1.2). An α-colored independent edge contributes 2 to both (1.1) and (1.2). Then $\phi_C(\alpha) - 2c_C(\alpha) = 2V_\alpha$, as wanted. Summing over all colors α and taking into account the equalities from above, the second equality holds. □

We present our algorithm in `R`. From any coloring of K_n with $n-1$ colors, it arrives at a proper coloring of K_n moving on the Markov chain according to the potential of the neighbors (recolorings).

The following function creates an arbitrary coloring in matrix form. The entry (i, j) of the matrix indicates the color of the edge joining the vertices i and j of K_n. Recall that we use $n-1$ colors. The matrix is symmetric, and we place n in the diagonal by convention. This matrix is a symmetric Latin square with n symbols and with constant diagonal equal to n.

```
arbitrary_coloring <- function(n) {

  # Goal: with n-1 colors (1,2,...,n-1), to color uniformly the edges of the
  #       complete graph Kn.
  # Input: n=size of the complete graph Kn, we understand n even.
  # Output: matrix C such that C(i,j) is the color of the edge {i,j}, i distinct
  #         j. The coloring is uniform (each color has probability 1/(n-1)). We
  #         have considered C(i,i)=n (just notation).

  C <- matrix(0,n,n) # initialize matrix
  diag(C) <- rep(n,n) # put n at the diagonal (just notation, there are no
                      # loops in Kn)
  for(column in 1:(n-1)) # fill the lower part of C with colors
    C[(column+1):n,column] <- sample(1:(n-1),size=n-column,replace=TRUE,
                           prob=rep(1/(n-1),n-1))
  for(row in 1:(n-1)) # fill the upper part of C so that C is symmetric
                      # (Kn is an undirected graph)
    C[row,(row+1):n] <- C[(row+1):n,row]

  return(C)
}
```

The following function computes the potential of any coloring given in matrix form:

```
potential_function <- function(C) {

  # Goal: to compute the potential function of the coloring given by the matrix C.
  # Input: coloring matrix C.
  # Output: potential function.

  n <- dim(C)[1]
  phi <- 0
  for(vertex in 1:n)
    for(color in 1:(n-1))
      phi <- phi + sum(C[,vertex]==color)^2

  return(phi)
}
```

In the next function, the algorithm is implemented. The authors of [7] propose an implementation of the algorithm which, from a coloring C, one moves to a recoloring C' with probability $1/|N_\phi(C)|$ if $C' \in N_\phi(C)$. That is, one needs to know $N_\phi(C)$ explicitly. In practice, this implementation is computationally expensive (we have to compute the potential of all recolorings). We suggest an alternative, which makes the Markov chain to move in the same way, but more efficiently in terms of computations. The alternative algorithm consists in:

i) Choose randomly a recoloring C' of C;
ii) If $C' \in N_\phi(C)$, move to it. Otherwise, do step i) again

(if after doing step ii) we do step i) again, the probability of choosing a recoloring C' in step i) is independent of the choices done before). By Lemma 4.5, if C is not proper, we

will find a recoloring $C' \in N_\phi(C)$ at some instant. On the other hand, let us see that performing steps i) and ii) (approach 2) is equivalent to moving from a coloring C to a recoloring C' with probability $1/|N_\phi(C)|$ if $C' \in N_\phi(C)$ (approach 1, the one explained in the previous pages). We will call *attempt* the coloring we obtain each time we do step ii) in approach 2. Let $N(C)$ be the set of recolorings of C'. We have

$$\mathbb{P}(\text{move to } C'|\text{approach 1}) = \frac{1}{|N_\phi(C)|}$$

and

$$\mathbb{P}(\text{move to } C'|\text{approach 2}) = \sum_{n=0}^{\infty} \mathbb{P}(\text{attempt}_1 \notin N_\phi(C), \ldots, \text{attempt}_n \notin N_\phi(C), \text{attempt}_{n+1} = C')$$

$$\overset{\text{indep.}}{=} \sum_{n=0}^{\infty} \underbrace{\mathbb{P}(\text{attempt}_1 \notin N_\phi(C))}_{=\frac{|N(C)|-|N_\phi(C)|}{|N(C)|}} \cdots \underbrace{\mathbb{P}(\text{attempt}_n \notin N_\phi(C))}_{=\frac{|N(C)|-|N_\phi(C)|}{|N(C)|}} \underbrace{\mathbb{P}(\text{attempt}_{n+1} = C')}_{=\frac{1}{|N(C)|}}$$

$$= \frac{1}{|N(C)|} \sum_{n=0}^{\infty} \left(1 - \frac{|N_\phi(C)|}{|N(C)|}\right)^n = \frac{1}{|N(C)|} \frac{1}{1 - \left(1 - \frac{|N_\phi(C)|}{|N(C)|}\right)} = \frac{1}{|N_\phi(C)|}.$$

Thus, both approaches are equivalent, but computationally approach 2 is considerably more efficient. In the algorithm below we obtain as an output the proper coloring we arrive at, a logical variable indicating if there has been convergence and the number of steps required (to check numerically that the computational cost is polynomial on n).

```
find_proper_coloring_2 <- function(n) {

  # Goal: to find a proper coloring of the complete graph Kn.
  # Input: n=size of the complete graph Kn. n must be EVEN.
  # Output: a list having a coloring C, a boolean that is TRUE if C is a proper
  #         coloring and FALSE otherwise, and the number of steps to find C.

  C <- arbitrary_coloring(n)  # start with an arbitrary coloring
  phiC <- potential_function(C) # potential of coloring C
  steps <- 0
  proper <- FALSE
  while(proper == FALSE & steps < n^10) {
    edge <- sample(1:n,size=2,replace=FALSE,prob=rep(1/n,n)) # choose an
                                                             # arbitrary edge
    color <- sample((1:(n-1))[-C[edge[1],edge[2]]],size=1,replace=FALSE,
         prob=rep(1/(n-2),n-2)) # choose a color
    C1 <- C # recolor
    C1[edge[1],edge[2]] <- color
    C1[edge[2],edge[1]] <- color # the matrix C1 represents changing the color
                                 # of an edge from C, i.e., C1 is a recoloring (step i))
    phiC1 <- potential_function(C1) # potential of coloring C1
    if(phiC1 <= phiC) { # step ii)
      C <- C1 # if the potential of C1 is smaller or equal, we move to C1,
            # otherwise remain at C
      phiC <- phiC1
      steps <- steps + 1 # count step whenever we move
    }

    # case when we finish the procedure:
    if(phiC == n*(n-1))
```

```
      proper <- TRUE
  }

  return(list(C,proper,steps))
}
```

After executing this algorithm for different n, we conjecture that it should always converge, and not only asymptotically, as [7] states. We analyze the number of steps taken by `find_proper_coloring_2` to arrive at a proper coloring. We plot the number of steps in a logarithmic scale to analyze if the number of steps is polynomial in n (in this case, it is said that the algorithm converges polynomial in time). We show the code to obtain Figure 6.

```
x <- sort(rep(seq(10,22,2),30))
y <- c()
for(size in x) {
  continue <- TRUE
  while(continue == TRUE) {
    result <- find_proper_coloring_2(size)
    if(result[[2]]== TRUE) {
      y <- c(y,result[[3]])
      continue <- FALSE
    }
  }
}
logx <- log(x)
logy <- log(y)

plot(logx,logy,xlab='log(size of graph)',ylab='log(number of steps)',
     main='log convergence rate to a proper coloring') # draw points of
                                                       # the numerical experiments

coeff <- lm(logy ~ logx)$coefficients # linear model
lines(logx,coeff[[1]]+coeff[[2]]*logx,col=2) # plot the regression line
text(1.9,6,paste0('y=',round(coeff[[1]],2),'+',
     round(coeff[[2]],2),' x'),col=2) # write the line$
```

We notice several differences with the [7]. First, as we explained, we conjecture that the algorithm always converges, and not just asymptotically. It would converge asymptotically if we moved only when the potential is strictly smaller, but this is not our setting. On the other hand, we conjecture that the polynomial time is smaller than $\mathcal{O}(n^4)$, in fact $\mathcal{O}(n^p)$, where $2 < p < 3$ and $p \approx 2.51$, see the regression line in Figure 6.

CONJECTURE 4.2. *Consider K_n, with n even, and all possible colorings of it with $n-1$ colors. Consider the algorithm in which we move from a coloring C to a recoloring C' of it if and only if $\phi(C') \leq \phi(C)$. Then the algorithm always converges for all n and, moreover, the number of steps required by the algorithm is $\mathcal{O}(n^p)$, where $2 < p < 3$.*

In the following sections, we tackle this conjecture. We start in Section 2 and Section 3 with K_4 and K_6. We use R and theory of Markov chains to get new results and an original solution to the conjecture for K_4 and K_6. In Section 4, we try to extend our findings to the general case of K_n by presenting new results. The conjecture in its general form is left open.

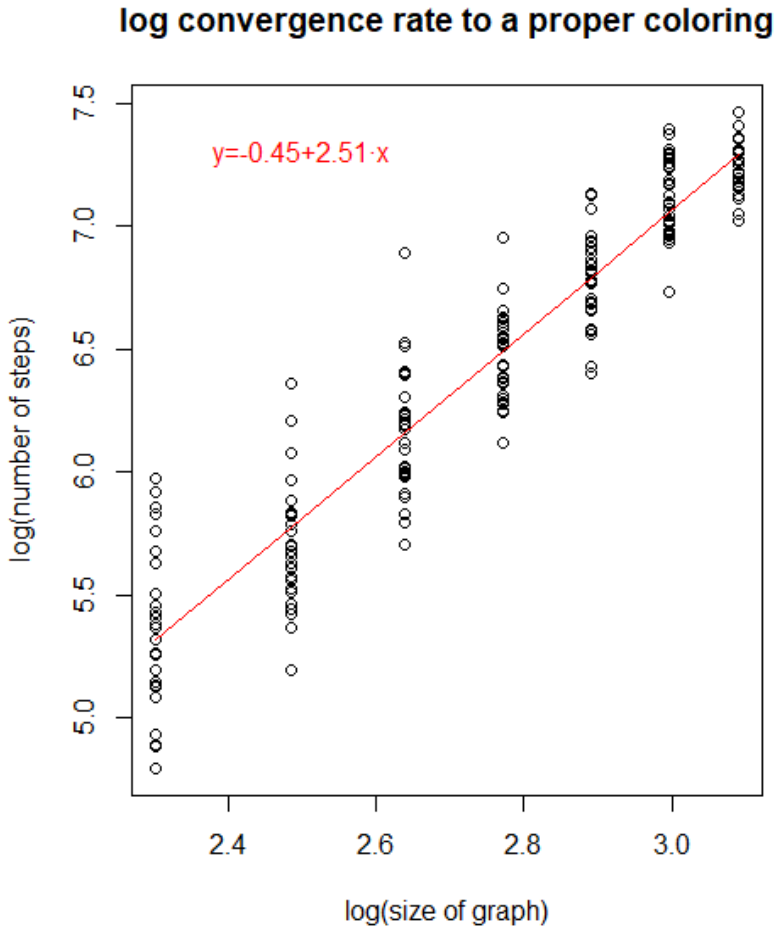


FIG. 6. Steps required in the Markov chain to arrive at a proper coloring of K_n. In red, a regression line to study the cost in polynomial time.

2. Proofs of the conjecture in the case K_4

2.1. Computational proof.

We show that the algorithm converges in the case of K_4.

To prove the conjecture we are going to construct explicitly in `R` the Markov chain associated to the colorings of K_4, and we will see that independently of the coloring we start with, we will get absorbed by the states associated to the proper colorings. To do this we will construct the transition matrix P with the help of `R` and we will prove that the expected number of steps until getting absorbed by a proper coloring is finite.

Since we have 6 edges and 3 colors, we start by computing all $3^6 = 729$ possible permutations (colorings) which will be the states of the Markov chain we are dealing with, using the following `R` command:

```
colorings <- expand.grid(rep(list(1:3),6))
```

Let us see an example of the first 6 permutations:

```
#Example we show the first $6$
> colorings[1:6,]
  Var1 Var2 Var3 Var4 Var5 Var6
1    1    1    1    1    1    1
2    2    1    1    1    1    1
3    3    1    1    1    1    1
4    1    2    1    1    1    1
5    2    2    1    1    1    1
6    3    2    1    1    1    1
```

Let us call by m all possible colorings, that is, $m = 729$:

```
m <- dim(colorings)[1] #3^6
> m
[1] 729
```

We write each coloring as a matrix, and we save them in a list of matrices called `list_matrices_colorings`. From now on, we are going to work with colorings via 4×4 symmetric matrices which would have the number 4 in the diagonal (this 4 can be seen as a symbol, since there are no loops in the graph) and the other entries will be 1, 2 and 3, which are our 3 possible colors. To do this, we use the following `R` command:

```
list_matrices_colorings <- list()
for(i in 1:m) {
  C <- matrix(0,4,4)
  diag(C) <- rep(4,4)
  C[2,1] <- colorings[i,1]; C[1,2] <- C[2,1]
  C[3,1] <- colorings[i,2]; C[1,3] <- C[3,1]
  C[4,1] <- colorings[i,3]; C[1,4] <- C[4,1]
  C[3,2] <- colorings[i,4]; C[2,3] <- C[3,2]
  C[4,2] <- colorings[i,5]; C[2,4] <- C[4,2]
  C[4,3] <- colorings[i,6]; C[3,4] <- C[4,3]
  list_matrices_colorings[[i]] <- C
}
```

Let us see as an example the first coloring:

```
#Example: list_matrices_colorings[[1]]
> list_matrices_colorings[[1]]
     [,1] [,2] [,3] [,4]
[1,]    4    1    1    1
[2,]    1    4    1    1
[3,]    1    1    4    1
[4,]    1    1    1    4
```

Notice that this coloring assigns color 1 to all of the edges of the graph K_4.

Now, once we have the 729 matrices that represent all the possible colorings of K_4, we compute the potential of all these matrices (colorings) taking into account the definition

of potential we stated at the beginning of the chapter, and then we will keep all the 729 potentials in a vector called `potentials` via the following `R` code:

```
potential_function <- function(C) {

  # Goal: to compute the potential function of the coloring
  #        given by the matrix C.
  # Input: coloring matrix C.
  # Output: potential function.

  n <- dim(C)[1]
  phi <- 0
  for(vertex in 1:n)
    for(color in 1:(n-1))
      phi <- phi + sum(C[,vertex]==color)^2

  return(phi)
}


potentials <- numeric(m) # vector of potentials
for(i in 1:m)
  potentials[i] <- potential_function(list_matrices_colorings[[i]])
```

To build the transition matrix that characterizes the Markov chain associated to the colorings of the graph K_4, we change the order of the colorings in `list_matrices_colorings` and we put at the beginning all the proper colorings, that is, all the colorings that have the value $12 = 4 \times (4-1)$ as a potential (see Lemma 4.2).

```
positions_proper <- which(potentials==12) # positions proper colorings
> positions_proper
[1] 157 205 317 413 525 573

for(i in 1:length(positions_proper)) {
  aux <- list_matrices_colorings[[ positions_proper[i] ]]
  list_matrices_colorings[[ positions_proper[i] ]] <- list_matrices_colorings[[i]]
  list_matrices_colorings[[i]] <- aux.
} # put proper colorings at the beginning (to simplify)

# reordered vector of potentials
potentials <- numeric(m)
for(i in 1:m)
  potentials[i] <- potential_function(list_matrices_colorings[[i]])

# check that the proper colorings are at the beginning
which(potentials==12)

> which(potentials==12)
[1] 1 2 3 4 5 6
```

Taking into account that in the first 6 places of `list_matrices_colorings` we have the proper colorings, which are absorbing states since if we start in a proper coloring we do not do any more step, and also considering that from one coloring we move uniformly among

the colorings whose potential is smaller or equal, we construct the transition matrix P as follows:

```
# The matrix P is constructed as follows.
P <- matrix(0,m,m)
#we want the proper colorings to be absorbing states.
for(i in 1:length(positions_proper))
  P[i,i] <- 1
for(i in 1:m) {
  for(j in 1:m) {
# P(i,j)=(probability of going from coloring i to coloring j)
    A <- list_matrices_colorings[[i]]
    B <- list_matrices_colorings[[j]] # compare colorings A and B
    if( sum(A != B) == 2 ) { # one colored edge different
      if(potential_function(B)<=potential_function(A))
        P[i,j] <- 1 # we can move
      if(potential_function(A)<=potential_function(B))
        P[j,i] <- 1 # we can move
    }
  }
}

for(i in 1:m)
  P[i,] <- P[i,] / sum(P[i,]) # normalize so that P is stochastic
```

Since the proper colorings are absorbing states and are located at the beginning, the transition matrix P has the following structure:

$$P = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix},$$

where I is the 6×6 identity matrix, 0 is the 6×723 matrix of zeros, R represents the probabilities of moving from a non-proper coloring to a proper coloring, and Q represents the probabilities of going from a non-proper coloring to a non-proper coloring.

Let us see that, from Q, we can compute the expected number of steps until the chain gets absorbed by a proper coloring. Indeed, the number of visits to a non-proper state j is $\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=j\}}$. Therefore, the expected number of visits to a non-proper state j having started at a non-proper state i is

$$\mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=j\}} | X_0 = i\right] = \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{1}_{\{X_n=j\}} | X_0 = i] = \sum_{n=0}^{\infty} \mathbb{P}(X_n = j | X_0 = i) = \sum_{n=0}^{\infty} Q^n(i,j) = \left(\sum_{n=0}^{\infty} Q^n\right)(i,j).$$

The expected number of steps until getting absorbed by a proper coloring having started at a non-proper state i is

$$\sum_{j=1}^{723} \left(\sum_{n=0}^{\infty} Q^n\right)(i,j).$$

Notice that, in principle, some of these values could be ∞. Nevertheless, if we prove that all these values are finite (this is our goal), the convergence of the algorithm will be ensured (recall that if a random variable has finite expectation, then it is finite).

In `R`, we observe that there exists $N := (I - Q)^{-1}$.

```
n_proper <- length(positions_proper)#number of proper colorings
Q <- P[(n_proper+1):m,(n_proper+1):m]
```

```
I <- diag(dim(Q)[1])
N <- solve(I-Q)
```

In the following propositions, we prove that $\sum_{n=0}^{\infty} Q^n$ is indeed convergent and $N = \sum_{n=0}^{\infty} Q^n$. This will imply that the algorithm converges for any initial coloring and that the expected number of steps until being absorbed is the sum of the corresponding row of N.

PROPOSITION 4.2. *If there exists $N = (I-Q)^{-1}$ and $N(i,j) \geq 0$ for all i,j, then every eigenvalue of Q has modulus less than 1.*

PROOF. Since Q has all entries nonnegative, by Perron-Frobenius Theorem (see Chapter 8 in [15]) the maximum eigenvalue of Q in modulus is a nonnegative real number, call it λ. Assume by contradiction that $\lambda \geq 1$. It cannot be 1, because in such a case $I-Q$ has zero as eigenvalue, therefore $I-Q$ is not invertible, which is a contradiction. Thus, $\lambda > 1$. Let x be an eigenvector of Q corresponding to λ. By Perron-Frobenius Theorem, we can assume that all components of x are nonnegative. From $Qx = \lambda x$ it follows $Nx = 1/(1-\lambda)x$. Now, Nx has all its components nonnegative, but $1/(1-\lambda) < 0$, so this is a contradiction. This proves that $\lambda < 1$. □

In fact, the result of this proposition can be checked in R:

```
eigenvaluesQ <- eigen(Q)$values
max(Mod(eigenvaluesQ))
> eigenvaluesQ <- eigen(Q)$values
> max(Mod(eigenvaluesQ))
[1] 0.8430703
```

PROPOSITION 4.3. *If every eigenvalue of Q has modulus less than 1, then $\lim_n Q^n = 0$ (where the limit is understood componentwise).*

PROOF. The idea of this proof comes from Proposition 6.1.8 of [2]. Write the matrix Q as $Q = HJH^{-1}$, where H is an invertible matrix and J is the Jordan form of Q. Since $Q^n = HJ^nH^{-1}$, it suffices to show that $J^n \to 0$ as $n \to 0$. The structure of J is the following:

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_t \end{pmatrix},$$

where each box J_t is an $n_i \times n_i$ lower-triangular matrix of the form

$$J_i = \begin{pmatrix} \lambda_i & & & \\ 1 & \lambda_i & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_i \end{pmatrix},$$

where $\lambda_1, \ldots, \lambda_t$ are the eigenvalues of Q. It suffices to prove that $J_i^n \to 0$ as $n \to \infty$, for each $i = 1, \ldots, t$. The key fact is that each box J_i can be decomposed in the following manner: $J_i = \lambda_i I_{n_i} + N_i$, where N_i is a nilpotent matrix: $N_i^{n_i-1} = 0$. Using the Newton binomial and the fact that the two matrices $\lambda_i I_{n_i}$ and N_i commute, we have

$$J_i^n = (\lambda_i I_{n_i} + N_i)^n = \lambda_i^n I_{n_i} + \lambda_i^{n-1} N_i + \ldots + \lambda_i^{n-(n_i-2)} N_i^{n_i-2} \overset{n\to\infty}{\longrightarrow} 0,$$

since $|\lambda_i| < 1$ by hypothesis. This completes the proof of the proposition. □

PROPOSITION 4.4. *If* $\lim_n Q^n = 0$*, then* $I - Q$ *is invertible and* $N = \sum_{n=0}^{\infty} Q^n$.

PROOF. Although we already know that $I - Q$ is invertible by our computations in R, let us prove it just by knowing that $\lim_n Q^n = 0$. We need to check that, if $(I - Q)x = 0$ for some vector x, then $x = 0$. If $(I - Q)x = 0$, then $Qx = x$. Multiplying by Q at both sides, $Q^2x = Qx = x$. Continuing in this way, $Q^n x = x$ for all $n \geq 1$. Now, let $n \to 0$: $x = Q^n x \to 0$, therefore $x = 0$. This shows that $I - Q$ is invertible.

It remains to prove that $N = (I - Q)^{-1}$ is equal to $\sum_{n=0}^{\infty} Q^n$. Notice that $I - Q^{n+1} = (I - Q)(I + Q + \ldots + Q^n)$. Multiply by N at both sides: $N - NQ^{n+1} = I + Q + \ldots + Q^n$. Let $n \to \infty$ and use the fact that $Q^n \to 0$: $N = \sum_{n=0}^{\infty} Q^n$. □

Since we computed N in R, the algorithm converges (because $(\sum_{n=0}^{\infty} Q^n)(i,j) = N(i,j) < \infty$) and the expected number of steps until getting absorbed is the sum of each of the rows of N:

```
expected_steps_arrive_proper <- numeric(dim(Q)[1])
for(i in 1:dim(Q)[1])
  expected_steps_arrive_proper[i] <- sum(N[i,])
```

To end up, we show the first 6 components of the last vector as an example

```
> expected_steps_arrive_proper[1:6]
[1] 11.49876 10.05000 11.16202 11.49876 11.16202 10.05000
```

Computational experiments with the algorithm with particular instances of the initial coloring show that the number of steps is close to the the expected value computed above. We next describe the code of the algorithm and some examples which turn out to fit with the theoretical value.

```
find_proper_coloring <- function(mat) {

  # Goal: to find a proper coloring of the complete graph Kn from the matrix mat.
  # Input: mat=matrix to start from. Its size n must be EVEN.
  # Output: a list having a coloring C, a boolean that is TRUE if C is a proper
  #         coloring and FALSE otherwise, and the number of steps to find C.

  n <- dim(mat)[1]
  C <- mat # start with matrix mat
  phiC <- potential_function(C) # potential of coloring C
  steps <- 0
  proper <- FALSE
  while(proper == FALSE & steps < n^10) {
    edge <- sample(1:n,size=2,replace=FALSE,prob=rep(1/n,n)) #choose an arbitrary edge
  # choose a different color
    color <- sample((1:(n-1))[-C[edge[1],edge[2]]],size=1,
                replace=FALSE,prob=rep(1/(n-2),n-2))
    C1 <- C # recolor
    C1[edge[1],edge[2]] <- color
    C1[edge[2],edge[1]] <- color
  # the matrix C1 represents changing the color of an edge from C
    phiC1 <- potential_function(C1) # potential of coloring C1
    if(phiC1 <= phiC) {
```

```
      C <- C1 # if the potential of C1 is smaller or equal, we move to C1,
        #otherwise remain at C
      phiC <- phiC1
      steps <- steps + 1
    }

    # case when we finish the procedure:
    if(phiC == n*(n-1))
      proper <- TRUE
  }

  return(list(C,proper,steps))
}
```

For example,

```
# Example with coloring number 7 (counting the first six proper colorings),
# with expected value of steps 11.498760:
example_steps <- numeric(10000)
for(i in 1:10000)
  example_steps[i] <- find_proper_coloring(list_matrices_colorings[[7]])[[3]]
(example_expected_steps <- mean(example_steps))
expected_steps_arrive_proper[[1]]
> (example_expected_steps <- mean(example_steps))
[1] 11.4414
> expected_steps_arrive_proper[[1]]
[1] 11.49876

# Example with coloring number 74 (counting the first six proper colorings),
# with expected value of steps 6:
example_steps <- numeric(10000)
for(i in 1:10000)
  example_steps[i] <- find_proper_coloring(list_matrices_colorings[[74]])[[3]]
(example_expected_steps <- mean(example_steps))
expected_steps_arrive_proper[[68]]
> (example_expected_steps <- mean(example_steps))
[1] 5.9432
> expected_steps_arrive_proper[[68]]
[1] 6

# Example with coloring number 157 (counting the first six proper colorings),
# with expected value of steps 12.498760:
example_steps <- numeric(10000)
for(i in 1:10000)
  example_steps[i] <- find_proper_coloring(list_matrices_colorings[[157]])[[3]]
(example_expected_steps <- mean(example_steps))
expected_steps_arrive_proper[[151]]

> (example_expected_steps <- mean(example_steps))
[1] 12.5198
> expected_steps_arrive_proper[[151]]
[1] 12.49876
```

2.2. Graphical proof.

To prove the conjecture in the case of K_4, or maybe in general for K_n (n even), we could use the following idea. As we did in the previous subsection, we can see the colorings as states in a Markov chain, where the proper colorings are absorbing states and the transition matrix P has the structure

$$P = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \tag{2.1}$$

(we put the proper colorings at the beginning of P). Let C be an arbitrary non-proper coloring. If C is not a local minimum, then at the next step we are able to reach a new coloring C' with $\phi(C') < \phi(C)$. If C is a local minimum, then it suffices to prove that there exists a finite sequence of steps in our Markov chain until we reach a coloring C' such that $\phi(C') < \phi(C)$. In such a case, we will have that there is a positive probability for the initial non-proper coloring C to be absorbed by a proper coloring at some step. The forthcoming Proposition 4.5 and Proposition 4.4 will give as a consequence that the algorithm converges.

PROPOSITION 4.5. *Consider a finite and absorbing Markov chain $\{X_n\}_{n=0}^{\infty}$ with transition probability matrix of the form (2.1). If every non-absorbing state has a positive probability of being absorbed at some step, then* $\lim_{n\to\infty} Q^n = 0$.

PROOF. Let $\mathcal{T}$ be the set of non-absorbing states. For each $j \in \mathcal{T}$, let m_j be the minimum number of steps until j reaches an absorbing state (by hypothesis, $m_j < \infty$). Let $0 \leq p_j < 1$ be the probability that j does not reach an absorbing state in m_j steps. Define $m = \max_{j\in\mathcal{T}} m_j$ and $p = \max_{j\in\mathcal{T}} p_j < 1$.

Notice that, starting at any non-absorbent state, the probability of not reaching an absorbing state in m steps is less or equal than p. Fix $i, j \in \mathcal{T}$. Claim:

$$Q^{km+l}(i,j) = \mathbb{P}(X_{km+l} = j | X_0 = i) \leq p^k, \quad l = 0, \ldots, m-1. \tag{2.2}$$

From (2.2) and the fact that $\lim_{k\to\infty} p^k = 0$, we deduce that $\lim_{n\to\infty} Q^n(i,j) = 0$ for all $i, j \in \mathcal{T}$, as desired.

Thus, we need to prove the claim (2.2). We do so by induction on $k \geq 1$.

Case $k = 1$. For the case $l = 0$, just notice that

$$\{X_m = j\} \subseteq \{\text{in } m \text{ steps an absorbing state has not been reached}\},$$

whence $\mathbb{P}(X_m = j | X_0 = i) \leq p_i \leq p = p^k$. For the case $l \in \{1, \ldots, m-1\}$, we have:

$$\begin{aligned}
\mathbb{P}(X_{m+l} = j | X_0 = i) &= \sum_{s\in\mathcal{T}} \mathbb{P}(X_{m+l} = j, X_m = s | X_0 = i) + \sum_{s\notin\mathcal{T}} \underbrace{\mathbb{P}(X_{m+l} = j, X_m = s | X_0 = i)}_{=0} \\
&= \sum_{s\in\mathcal{T}} \mathbb{P}(X_{m+l} = j, X_m = s | X_0 = i) \\
&\underbrace{=}_{\text{TPT}} \sum_{s\in\mathcal{T}} \mathbb{P}(X_{m+l} = j | X_m = s) \mathbb{P}(X_m = s | X_0 = i) \\
&\leq \sum_{s\in\mathcal{T}} \mathbb{P}(X_m = s | X_0 = i) = \mathbb{P}(X_m \in \mathcal{T} | X_0 = i) \leq p,
\end{aligned}$$

where the initials TPT stand for the Total Probability Theorem. Suppose (2.2) true for $k-1$ and we prove it for k, $k \geq 2$. For $l \in \{0, \ldots, m-1\}$, we proceed in an similar manner as before:

$$\begin{aligned}
\mathbb{P}(X_{km+l} = j | X_0 = i) &= \sum_{s \in \mathcal{T}} \mathbb{P}(X_{km+l} = j, X_m = s | X_0 = i) + \sum_{s \notin \mathcal{T}} \underbrace{\mathbb{P}(X_{km+l} = j, X_m = s | X_0 = i)}_{=0} \\
&= \sum_{s \in \mathcal{T}} \mathbb{P}(X_{km+l} = j, X_m = s | X_0 = i) \\
&\underbrace{=}_{\text{TPT}} \sum_{s \in \mathcal{T}} \mathbb{P}(X_{km+l} = j | X_m = s) \mathbb{P}(X_m = s | X_0 = i) \\
&= \sum_{s \in \mathcal{T}} \mathbb{P}(X_{(k-1)m+l} = j | X_0 = s) \mathbb{P}(X_m = s | X_0 = i) \\
&\underbrace{\leq}_{\text{induction}} p^{k-1} \sum_{s \in \mathcal{T}} \mathbb{P}(X_m = s | X_0 = i) = p^{k-1} \mathbb{P}(X_m \in \mathcal{T} | X_0 = i) \leq p^{k-1} p = p^k.
\end{aligned}$$

□

Thus, to derive that the algorithm converges, our main goal is to prove that, for any arbitrary non-proper local minimum C, there exists a finite sequence of steps in our Markov chain until we reach a coloring C' such that $\phi(C') < \phi(C)$.

This first part of the reasoning will hold for any K_n (n even). We will inform when we restrict to K_4 later on. Consider K_n and let C be a non-proper local minimum coloring. By Lemma 4.4, $a_{u,\mu(uv)}(C) + a_{v,\mu(uv)}(C) \leq 3$ for every edge uv. As C is non-proper, there exists an edge u_0v_0 such that $a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) = 3$. For every such edge u_0v_0, by Lemma 4.4, $a_{u_0,\mu}(C) + a_{v_0,\mu}(C) \geq 1$ for each color μ.

Suppose by contradiction that $a_{u_0,\mu}(C) + a_{v_0,\mu}(C) \geq 2$ for all $\mu \neq \mu(u_0v_0)$. Figure 7 shows that the number of edges incident to uv with color distinct from $\mu(u_0v_0)$ is $2n-5$ (in the figure, blue represents any color distinct from $\mu(u_0v_0)$, which is pink). Then,

$2n-5 =$ (number of edges incident to uv with color distinct from $\mu(u_0v_0)$) $\geq 2(n-2) = 2n-4$,

which is a contradiction. Therefore, there must exist $\bar{\mu}(u_0v_0)$ distinct from $\mu(u_0v_0)$ such that $a_{u_0,\bar{\mu}(u_0v_0)}(C) + a_{v_0,\bar{\mu}(u_0v_0)}(C) = 1$.

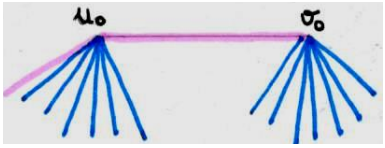


FIG. 7. General form of the edge u_0v_0 of K_n.

In order to imagine C visually, if u_0v_0 is an edge with $a_{u_0,\mu(u_0v_0)}(C) + a_{v_0,\mu(u_0v_0)}(C) = 3$, then the neighborhood of u_0v_0 has the structure given by Figure 8 (the edges painted in pink represent the color $\mu(u_0v_0)$, the edge painted in green represents the color $\bar{\mu}(u_0v_0)$ and the edges painted in blue mean whatever color (or colors) different from $\mu(u_0v_0)$ and $\bar{\mu}(u_0v_0)$).

This structure for non-proper local minimum colorings allows us to find all non-proper local minimum colorings for the particular case of K_4. Consider such an edge u_0v_0 in

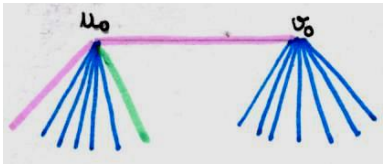


FIG. 8. Form of the edge u_0v_0 of K_n according to its neighbors.

K_4 and consider colors orange, green and pink. Without loss of generality (by making a permutation of the colors, changing the role of u_0 and v_0 and rotating the graph), we may assume that $\mu(u_0v_0)$ is pink and that u_0 has a pink incident edge, as Figure 9 shows.

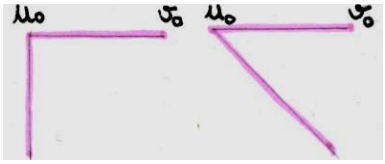


FIG. 9. Possible options (without loss of generality).

By what we proved and without loss of generality (by making a permutation of the colors and automorphisms of K_4), there exists a unique green edge incident to either u_0 or v_0. The other two edges incident to u_0 and v_0 are orange (see Figure 10).

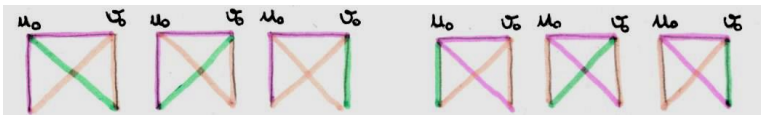


FIG. 10. Six possible cases.

It remains to find suitable colors for the non-intersecting edge with u_0v_0. See Figure 11 for the non-proper local minimum colorings of K_4 (except bijections between the colors, change in the roles of u_0 and v_0 and rotations).

Notice that, taking into account Lemma 4.4, in order to have a non-proper local minimum coloring, the only permissible color for the non-intersecting edge with u_0v_0 is green in all colorings of Figure 11, because:

- if in the first coloring (starting from the left) the non-intersecting edge with u_0v_0 were pink, then the vertices associated to the pink edge of the left-hand side would have in total 4 incident edges with color pink; and if it were orange, the vertices associated to the non-intersecting edge with u_0v_0 would have in total 4 incident edges with color orange.
- if in the second coloring (starting from the left) the non-intersecting edge with u_0v_0 were pink, then the vertices associated to the pink edge of the left-hand side would have in total 4 incident edges with color pink; and if it were orange, the vertices associated to the non-intersecting edge with u_0v_0 would have in total 4 incident edges with color orange.
- if in the third coloring (starting from the left) the non-intersecting edge with u_0v_0 were pink, then the vertices associated to the pink edge of the left-hand side would

have in total 4 incident edges with color pink; and if it were orange, the vertices associated to the non-intersecting edge with u_0v_0 would have in total 4 incident edges with color orange.

- if in the fourth coloring (starting from the left) the non-intersecting edge with u_0v_0 were pink, then the vertices associated to the orange edge of the right-hand side would not have any incident edge with color green; and if it were orange, the vertices associated to the non-intersecting edge with u_0v_0 would have in total 4 incident edges with color orange.
- if in the fifth coloring (starting from the left) the non-intersecting edge with u_0v_0 were pink, then the vertices associated to the the centered pink edge would have in total 4 incident edges with color pink; and if it were orange, the vertices associated to the non-intersecting edge with u_0v_0 would have in total 4 incident edges with color orange.
- Finally, if in the sixth coloring (starting from the left) the non-intersecting edge with u_0v_0 were pink, then the vertices associated to the orange edge at the left-hand side would not have any incident edge with color green; and if it were orange, the vertices associated to the non-intersecting edge with u_0v_0 would have in total 4 incident edges with color orange.

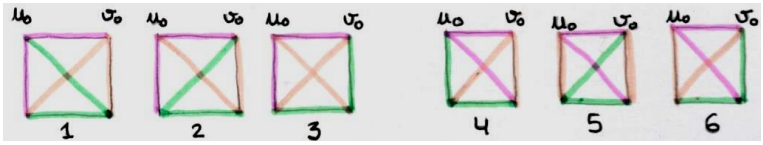


FIG. 11. Six allowed cases.

Now we check by hand that, after a finite number of steps (in fact 2 steps), we can strictly decrease the potential of each coloring from Figure 11. See Figure 12 (below each graphic of K_4, its potential according to the coloring is written).

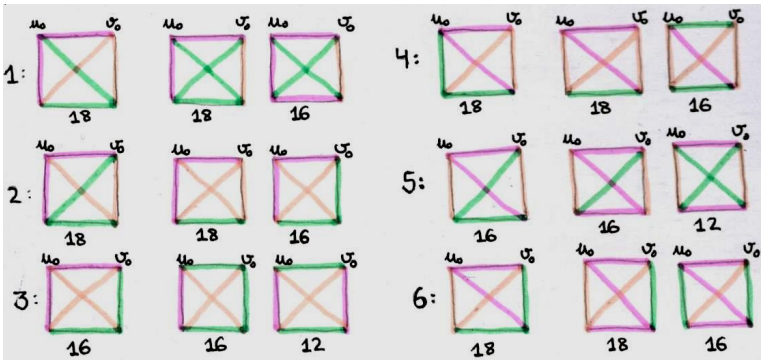


FIG. 12. Finding a coloring with smaller potential from the six allowed cases.

By the comments at the beginning of this subsection, the algorithm converges, and we are done.

Notice that one could try to generalize this proof for the case of K_n, whereas the computational proof from the previous section becomes intractable when n is large.

3. Proofs of the conjecture in the case K_6

Now we tackle the conjecture for the complete graph K_6. The problem now is that, since we have 5 colors, there are

$$5^{\binom{6}{2}} = 5^{15} = 30,517,578,125$$

colorings of the edges of K_6. This is a notable difference with the case K_4, in which we have, with 3 colors,

$$3^{\binom{4}{2}} = 3^6 = 729$$

colorings.

Actually, if we consider colorings of K_6 up to permutations of colors and permutations of labeled vertices, there are $5^{15}/(5!6!) = 211,927,625$ colorings. But, taking into account the symmetries arising from permutations of colors or automorphisms of K_6 has an equivalent computational cost.

This growth in the number of colorings makes it impossible to tackle the problem via a computational proof, as the one presented in Subsection 2.1 for K_4. If we try to create a list containing all possible colorings of K_6, we get an error from R, as vectors with such a big length cannot be allocated in R.

```
> colorings <- expand.grid(rep(list(1:5),15)) # combinations with the 5 colors
Error: cannot allocate vector of size 113.7 Gb
In addition: Warning messages:
1: In rep.int(rep.int(seq_len(nx), rep.int(rep.fac, nx)), orep) :
  Reached total allocation of 3963Mb: see help(memory.size)
2: In rep.int(rep.int(seq_len(nx), rep.int(rep.fac, nx)), orep) :
  Reached total allocation of 3963Mb: see help(memory.size)
3: In rep.int(rep.int(seq_len(nx), rep.int(rep.fac, nx)), orep) :
  Reached total allocation of 3963Mb: see help(memory.size)
4: In rep.int(rep.int(seq_len(nx), rep.int(rep.fac, nx)), orep) :
  Reached total allocation of 3963Mb: see help(memory.size)
```

A direct graphical proof, as the one done in Subsection 2.2 for K_4, is not feasible again. With the information given by Lemma 4.4, the case analysis we carried over in the case of K_4 becomes unpractical due to the large number of local minimum colorings.

Thus, our idea is, first, to obtain new theoretical results in the spirit of Subsection 2.2, so that we will understand better the structure of local minimum colorings in K_6. Then, we will be able to divide the set of non-proper local minimum colorings in three groups, according to the number of times each one of the five colors is used. At this step, it will be possible to check both computationally and graphically that the potential of the non-proper local minimum colorings can be decreased. This will give two subsections, Subsection 3.1 and Subsection 3.2, which will provide two different proofs of the conjecture for K_6. Finally, Proposition 4.5 will give as a consequence the convergence of the algorithm.

We present two lemmas, which will allow us to divide non-proper local minimum colorings of K_6 in three groups, according to the number of times each color appears in the coloring. After these lemmas, we will deal with each one of the three groups both in the computer, to have one proof of the conjecture for K_6, and graphically, to have a second proof of the conjecture for K_6.

LEMMA 4.7. *Let C be a non-proper local minimum coloring of K_n, n being even. Let α be a color such that there is a 2-path in K_n of color α. Then, either C contains a unique 2-path of color α, which implies that there exists another 2-path of color $\gamma \neq \alpha$ such that its center does not have any adjacent edge of color α; or there is another two path of color α. See Figure 13.*

FIG. 13. Two possible options.

PROOF. Notice that, as C is non-proper, there are a color α and a 2-path in K_n of color α, by Lemma 4.4. There are two cases:

- There is a unique 2-path of color α. Suppose by contradiction that every vertex has an incident edge of color α. Then we have the picture of Figure 14.

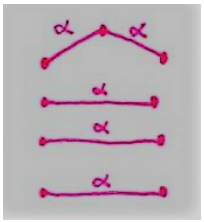


FIG. 14. Form of the edges with color α, which is represented in color pink.

This implies that there is an odd number of vertices, but n is even, which gives a contradiction. Hence, there exists a vertex w_0 in K_n such that it has no incident edge of color α. Then there is a 2-path of color $\gamma \neq \alpha$ centered at w_0 (because w_0 has $n-1$ incident edges and $n-2$ remaining colors, therefore a color must be repeated, say γ). See Figure 15.

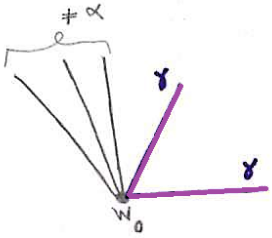


FIG. 15. Incident edges to the vertex w_0. In this picture, color γ is purple and the gray color represents another color different from α.

- There is another two path of color α.

□

LEMMA 4.8. *Let C be a local minimum coloring of K_6. If*

$$c_C(\mu) := \textit{number of times color } \mu \textit{ appears in coloring } C,$$

then $2 \leq c_C(\mu) \leq 4$, for each color μ.

PROOF. If C is proper, then each color is repeated 3 times. Thus, we may assume that C is a non-proper local minimum coloring of K_6.

Notice that every color has to appear at least once, otherwise we could change the color of an edge by the absent color and the potential would decrease (another reasoning consists in using Lemma 4.4, which says that any edge of a monochromatic 2-path must have all colors incident).

Clearly, no color can appear 5 times, since by Lemma 4.4 this would imply that, either there are two 2-paths and a single edge of this color, but this would cover 8 vertices, which is not possible in K_6; or there is one 2-path and three disjoint edges, but this would cover 9 vertices, which is not possible in K_6.

It remains to prove that $c_C(\mu) \geq 2$, for each color μ. Suppose by contradiction that there is a color $\tilde{\mu}$ among the 5 colors such that $c_C(\tilde{\mu}) = 1$. If there were 3 colors among the four remaining ones that are repeated in K_6 at most 3 times, then these 3 colors would encompass at most 9 edges, and together with the $\tilde{\mu}$ colored edge, at most 10 edges, therefore the remaining color would be repeated at least 5 times (because there are 15 edges in K_6), which is not possible. Hence, there are at least 2 colors that are repeated 4 times. There are two possible sequences of color repetitions: 13344 and 12444. As there are two colors repeated 4 times, say red and blue, by Lemma 4.4 there must be two red colored 2-paths and two blue colored 2-paths. By Lemma 4.4 again, the center of these 2-paths must have the $\tilde{\mu}$-colored edge incident. If we paint with green the color $\tilde{\mu}$, we have the structure given in Figure 16. There is a color $\mu' \notin \{\text{red}, \text{blue}\}$ that is repeated at least 3 times (recall the sequences of repetitions, 13344 and 12444). The edges with color μ' cover at least 5 vertices (because either there is a 2-path and a single edge of color μ', which cover 5 vertices; or there are 3 μ'-colored independent edges, which cover 6 vertices). However, in Figure 16, we clearly see that the μ'-colored edges can only be incident to the 4 vertices at the bottom (the ones not touching the green edge), because the eight incident edges to the green edge are already painted with red and blue. This gives a contradiction, because, as we explained, the μ'-colored edges must cover at least 5 vertices, but there are only 4 available vertices (the ones not touching the green edge) to be incident with.

□

As a consequence, changing the name of the colors if necessary, the structure of the non-proper local minimum colorings C of K_6 is given in Table 1.

An example of each one of the three structures given in Table 1 is presented in Figure 17.

From Table 1, we divide the non-proper local minimum colorings of K_6 in three groups. From now on we will denote the vertices of K_6 as in Figure 18.

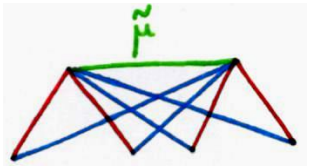


FIG. 16. Schema of a non-proper local minimum coloring of K_6 with a color (green) repeated once. The red and blue colors must be repeated 4 times.

colors	color 1	color 2	color 3	color 4	color 5
c_C(color)	3	3	3	3	3
	2	3	3	3	4
	2	2	3	4	4

TABLE 1. Structure of the non-proper local minimum colorings of K_6 according to the number of times each color appears.

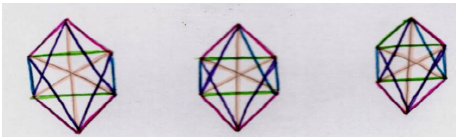


FIG. 17. Examples of the three possible cases. To the right-hand side option 33333, in the center option 23334 and to the left-hand side option 22344 from Table 1. Color 1 is green, color 2 is blue, color 3 is orange, color 4 is purple and color 5 is pink.

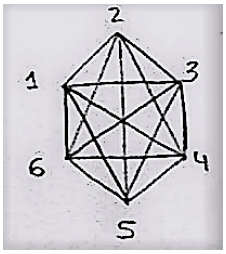


FIG. 18. Notation of K_6.

We handle each of the three groups separately, both in the computer and graphically. This gives rise to two proofs of the conjecture for K_6 in Subsection 3.1 and Subsection 3.2, respectively.

3.1. Computational proof.

In what follows we will check that, for each local minimum of the potential, our algorithm finds a way out to a coloring with smaller potential. As we have mentioned, this shows that the algorithm converges to a proper coloring.

We analyze each of the three groups arisen from Table 1:

- Group 2, 2, 3, 4 and 4. Let α be the fifth color, which is repeated 4 times. Notice that there must be two 2-paths of color α in K_6, otherwise the condition of Lemma 4.4 for non-proper local minimum colorings does not hold. By changing the name of the vertices of K_6 if necessary, we may assume that the two 2-paths of color α are at the top and at the bottom of the hexagon K_6, see Figure 19.

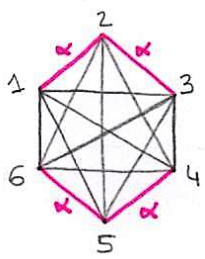


FIG. 19. Structure of K_6 in cases 22344 and 23334 from Table 1. Color α is color 5, which is represented pink.

Thus, given the other four colors, which are repeated 2, 2, 3 and 4 times, respectively, there are

$$\frac{11!}{2!2!3!4!} = 69,300$$

colorings. Among this 69,300 colorings of K_6, we have the non-proper local minimum colorings of K_6 with the first four colors repeated 2, 2, 3 and 4 times, respectively. If we check that the potential of the non-proper colorings among the 69,300 colorings can be decreased, then, in particular, the potential of the non-proper local minimum colorings of K_6 among the 69,300 colorings will be decreased as well. These 69,300 colorings are handled in R. This is possible because 69,300 is a small number in terms of memory allocation in R.
First, recall our function to compute potentials of colorings C in matrix form:

```
potential_function <- function(C) {

  # Goal: to compute the potential function of the coloring given
  #       by the matrix C.
  # Input: coloring matrix C.
  # Output: potential function.

  n <- dim(C)[1]
  phi <- 0
  for(vertex in 1:n)
    for(color in 1:(n-1))
      phi <- phi + sum(C[,vertex]==color)^2

  return(phi)
}
```

We also need a function to check that, given a non-proper coloring C, its potential can be decreased:

```
potential_decreases <- function(mat) {

  # Goal: to check that the potential function of the coloring given
```

```
  #       by the matrix C
  #       can be decreased.
  # Input: coloring matrix C.
  # Output: TRUE if the potential can be decreased, FALSE otherwise.

  n <- dim(mat)[1]
  C <- mat
  phiC <- potential_function(C)
  steps <- 0
  decrease <- FALSE
  while(decrease == FALSE & steps < n^10) {
    edge <- sample(1:n,size=2,replace=FALSE,prob=rep(1/n,n))
    color <- sample((1:(n-1))[-C[edge[1],edge[2]]],size=1,replace=FALSE,
                    prob=rep(1/(n-2),n-2))
    C1 <- C # recolor
    C1[edge[1],edge[2]] <- color
    C1[edge[2],edge[1]] <- color
    phiC1 <- potential_function(C1)
    if(phiC1 < phiC)
      decrease <- TRUE
    else if(phiC1 == phiC) {
      C <- C1
      phiC <- phiC1
      steps <- steps + 1
    }
  }

  return(decrease)
}
```

We need a function to compute permutations of a set with repeated elements. Since, at least to our knowledge, there is no built-in function in R to perform such task, we create the following function:

```
permutation_duplicates <- function(numbers) {

  # Goal: to compute all possible permutations of a vector c() of
  #       elements (colors), with repeated elements.
  # Input: the vector of elements.
  # Output: all possible permutations.

   all_permutations <- c()
   n <- length(numbers)
   if(n > 1) {
    positions_unique_elements <- order(numbers)[!duplicated(sort(numbers))]
    for(position in positions_unique_elements) {
      first_element <- numbers[position]
      A <- permutation_duplicates(numbers[-position])
      if(is.matrix(A))
        rows <- dim(A)[1]
      else
        rows <- 1
      perm_starting_at_position <- cbind(rep(first_element,rows),A)
      all_permutations <- rbind(all_permutations,perm_starting_at_position)
    }
```

```
    return(all_permutations)
   }
   else
     return(numbers)
}
```

Now, we are in a position to create the 69,300 colorings of Figure 19 with the first four colors repeated 2, 2, 3 and 4 times, respectively.

```
# Case 1,1,2,2,3,3,3,4,4,4,4

colorings_22344 <- c(1,1,2,2,3,3,3,4,4,4,4)
colorings_22344 <- permutation_duplicates(colorings_22344)
m <- dim(colorings_22344)[1]
```

Finally, we check that the potential of the non-proper colorings among the 69,300 colorings can be decreased (in the code, the last line checks that the variable `decrease` is `TRUE`). As a consequence, the potential of the non-proper local minimum colorings among the 69,300 colorings can be decreased.

```
decrease <- TRUE

for(i in 1:m) {
  C <- matrix(0,6,6)
  diag(C) <- rep(6,6)
  C[2,1] <- 5; C[1,2] <- C[2,1]
  C[3,1] <- colorings_22344[i,1]; C[1,3] <- C[3,1]
  C[4,1] <- colorings_22344[i,2]; C[1,4] <- C[4,1]
  C[5,1] <- colorings_22344[i,3]; C[1,5] <- C[5,1]
  C[6,1] <- colorings_22344[i,4]; C[1,6] <- C[6,1]
  C[3,2] <- 5; C[2,3] <- C[3,2]
  C[4,2] <- colorings_22344[i,5]; C[2,4] <- C[4,2]
  C[5,2] <- colorings_22344[i,6]; C[2,5] <- C[5,2]
  C[6,2] <- colorings_22344[i,7]; C[2,6] <- C[6,2]
  C[4,3] <- colorings_22344[i,8]; C[3,4] <- C[4,3]
  C[5,3] <- colorings_22344[i,9]; C[3,5] <- C[5,3]
  C[6,3] <- colorings_22344[i,10]; C[3,6] <- C[6,3]
  C[5,4] <- 5; C[4,5] <- C[5,4]
  C[6,4] <- colorings_22344[i,11]; C[4,6] <- C[6,4]
  C[6,5] <- 5; C[5,6] <- C[6,5]

  if(potential_function(C)>30) { # non-proper
    if(potential_decreases(C)==FALSE) {
      decrease <- FALSE
      break
    }
  }
  print(m-i) # to know for how long we have to wait :)
}

decrease # TRUE, then all potentials of non-proper colorings have decreased.
```

```
> decrease
[1] TRUE
```

- Group 2, 3, 3, 3 and 4. Let α be the fifth color, which is repeated 4 times. Exactly as in the previous case, we may assume that K_6 is painted as schematized in Figure 19. Given the other four colors, which are repeated 2, 3, 3 and 3 times, respectively,

there are

$$\frac{11!}{2!3!3!3!} = 92,400$$

colorings. Among this $92,400$ colorings of K_6, we have the non-proper local minimum colorings of K_6 with the first four colors repeated 2, 3, 3 and 3 times, respectively. If we check computationally that the potential of the non-proper colorings among the $92,400$ colorings can be decreased, then, in particular, the potential of the non-proper local minimum colorings of K_6 among the $92,400$ colorings will be decreased as well. We create the $92,400$ colorings of Figure 19 with the first four colors repeated 2, 3, 3 and 3 times, respectively.

```
# Case 1,1,2,2,2,3,3,3,4,4,4

colorings_23334 <- c(1,1,2,2,2,3,3,3,4,4,4)
colorings_23334 <- permutation_duplicates(colorings_23334)
m <- dim(colorings_23334)[1]
```

Eventually, we prove that the potential of the non-proper colorings among the $92,400$ colorings can be decreased (in the code, the last line checks that the variable `decrease` is `TRUE`). This fact implies that the potential of the non-proper local minimum colorings among the $92,400$ colorings can be diminished.

```
decrease <- TRUE

for(i in 1:m) {
  C <- matrix(0,6,6)
  diag(C) <- rep(6,6)
  C[2,1] <- 5; C[1,2] <- C[2,1]
  C[3,1] <- colorings_23334[i,1]; C[1,3] <- C[3,1]
  C[4,1] <- colorings_23334[i,2]; C[1,4] <- C[4,1]
  C[5,1] <- colorings_23334[i,3]; C[1,5] <- C[5,1]
  C[6,1] <- colorings_23334[i,4]; C[1,6] <- C[6,1]
  C[3,2] <- 5; C[2,3] <- C[3,2]
  C[4,2] <- colorings_23334[i,5]; C[2,4] <- C[4,2]
  C[5,2] <- colorings_23334[i,6]; C[2,5] <- C[5,2]
  C[6,2] <- colorings_23334[i,7]; C[2,6] <- C[6,2]
  C[4,3] <- colorings_23334[i,8]; C[3,4] <- C[4,3]
  C[5,3] <- colorings_23334[i,9]; C[3,5] <- C[5,3]
  C[6,3] <- colorings_23334[i,10]; C[3,6] <- C[6,3]
  C[5,4] <- 5; C[4,5] <- C[5,4]
  C[6,4] <- colorings_23334[i,11]; C[4,6] <- C[6,4]
  C[6,5] <- 5; C[5,6] <- C[6,5]

  if(potential_function(C)>30) { # non-proper
    if(potential_decreases(C)==FALSE) {
      decrease <- FALSE
      break
    }
  }
  print(m-i) # to know for how long we have to wait :)
}

decrease # TRUE, then all potentials of non-proper colorings have decreased.
```

```
> decrease
[1] TRUE
```

- Group 3, 3, 3, 3 and 3. Let α be the fifth color, which is repeated 3 times. Reordering the vertices of K_6 if necessary, we may assume that K_6 is painted as outlined in Figure 20.

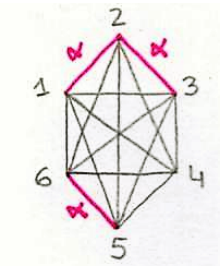


FIG. 20. Structure of K_6 in case 33333 from Table 1. Color α is color 5, which is represented pink.

Given the other four colors, which are repeated 3 times, there are

$$\frac{12!}{3!3!3!3!} = 369,600$$

colorings. Among this 369,600 colorings of K_6, we have the non-proper local minimum colorings of K_6 with the first four colors repeated 3 times. If we verify computationally that the potential of the non-proper colorings among the 369,600 colorings can be diminished, then, in particular, the potential of the non-proper local minimum colorings of K_6 among the 369,600 colorings will be decreased too.
We create the 369,600 colorings of Figure 20 with the first four colors repeated 3 times, respectively.

```
# Case 1,1,1,2,2,2,3,3,3,4,4,4

colorings_33333 <- c(1,1,1,2,2,2,3,3,3,4,4,4)
colorings_33333 <- permutation_duplicates(colorings_33333)
m <- dim(colorings_33333)[1]
```

Finally, we show that the potential of the non-proper colorings among the 369,600 colorings can be decreased (in the implementation, the last line of code checks that the logical variable `decrease` is `TRUE`). This implies that the potential of the non-proper local minimum colorings among the 369,600 colorings can go down.

```
decrease <- TRUE

for(i in 1:m) {
  C <- matrix(0,6,6)
  diag(C) <- rep(6,6)
  C[2,1] <- 5; C[1,2] <- C[2,1]
  C[3,1] <- colorings_33333[i,1]; C[1,3] <- C[3,1]
  C[4,1] <- colorings_33333[i,2]; C[1,4] <- C[4,1]
  C[5,1] <- colorings_33333[i,3]; C[1,5] <- C[5,1]
  C[6,1] <- colorings_33333[i,4]; C[1,6] <- C[6,1]
  C[3,2] <- 5; C[2,3] <- C[3,2]
  C[4,2] <- colorings_33333[i,5]; C[2,4] <- C[4,2]
  C[5,2] <- colorings_33333[i,6]; C[2,5] <- C[5,2]
  C[6,2] <- colorings_33333[i,7]; C[2,6] <- C[6,2]
  C[4,3] <- colorings_33333[i,8]; C[3,4] <- C[4,3]
  C[5,3] <- colorings_33333[i,9]; C[3,5] <- C[5,3]
```

```
  C[6,3] <- colorings_33333[i,10]; C[3,6] <- C[6,3]
  C[5,4] <- colorings_33333[i,11]; C[4,5] <- C[5,4]
  C[6,4] <- colorings_33333[i,12]; C[4,6] <- C[6,4]
  C[6,5] <- 5; C[5,6] <- C[6,5]

  if(potential_function(C)>30) { # non-proper
    if(potential_decreases(C)==FALSE) {
      decrease <- FALSE
      break
    }
  }
  print(m-i) # to know for how long we have to wait :)
}

decrease # TRUE, then all potentials of non-proper colorings have decreased.
> decrease
[1] TRUE
```

In the three cases, we have proved that the potential of the non-proper local minimum colorings of K_6 can go down, which means, in terms of Markov chains, that every non-proper coloring of K_6 has a positive probability of getting absorbed at some step by a proper coloring. Proposition 4.5 combined with Proposition 4.4 give the convergence of the algorithm for K_6, as wanted.

3.2. Graphical proof.

The common feature of non-proper local minimum colorings having the structure 2, 2, 3, 4, 4 and 2, 3, 3, 3, 4 (third and second rows in Table 1) is that there is a color repeated 4 times and another color repeated 2 times. From this fact and Lemma 4.4, we will derive that, in a single step, we can move to a coloring having the structure 3, 3, 3, 3, 3 (first row in Table 1). Thus, it will suffice to check that the non-proper local minimum colorings with structure 3, 3, 3, 3, 3 have a potential that can be decreased. For this goal, Lemma 4.7 will be essential.

We will work with K_6 as Figure 18 shows. Let C be a non-proper local minimum coloring having the structure 2, 2, 3, 4, 4 or 2, 3, 3, 3, 4. We will paint with blue and red a color repeated 4 and 2 times, respectively. By changing the ordering of the vertices 1, 2, 3, 4, 5 and 6 if necessary, we may assume that C has the form given in Figure 21.

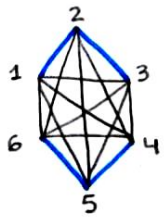


FIG. 21. Positions of color blue, which is repeated 4 times.

According to Lemma 4.4, the red color that is repeated twice has to be at the positions drawn in Figure 22 (except, as usual, rotations, symmetries, etc.). Let us explain why

Figure 22 covers all possibilities. In principle, any vertex may have an incident red edge. Notice however that it suffices to consider the cases in which a red edge is incident to vertex 1 or to vertex 2, by reordering of the vertices (symmetries). In the case in which a red edge is incident to vertex 1, this red edge can be edge 13 (colorings C_1, C_2 and C_3), edge 14 (colorings C_4, C_5 and C_6), edge 15 (colorings C_7, C_8 and C_9) and edge 16 (colorings C_{10} and C_{11}). In the case in which a red edge is incident to vertex 2, this red edge can be edge 26 (colorings C_{12} and C_{13}). Multiple cases in terms of reordering of vertices (symmetries and rotations) have been deleted. All possibilities have been obtained taking into account what Lemma 4.4 tells us in this case: since 123 and 456 form a blue 2-path, the blue edges 12, 23, 45 and 56 must intersect a red edge.

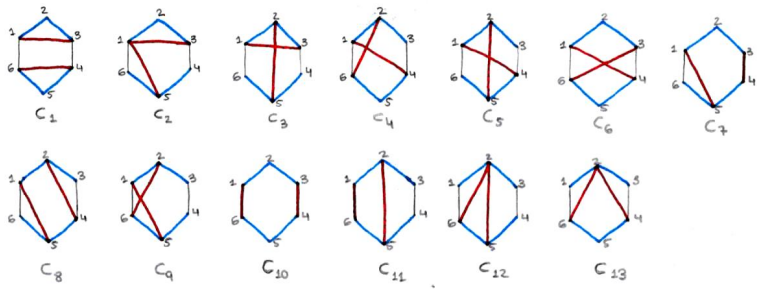


FIG. 22. Positions of color blue and red, which are repeated 4 and 2 times, respectively. Non-proper local minimum colorings of K_6 with structures 23334 and 22344.

The potentials of the 13 colorings of Figure 22 are easily computed. We will denote by Δ_i the remaining part of the potential of C_i that arises from the other three colors. Notice that Δ_i is not affected by the blue and red colors. In the following explicit sums, the order of the addends will be from vertex 1 up to 6, taking into account how many blue and red edges are incident and the definition of potential:

$$\begin{aligned}
\phi(C_1) &= 2+4+2+2+4+2+\Delta_1 = 16+\Delta_1,\\
\phi(C_2) &= 5+4+2+1+5+1+\Delta_2 = 18+\Delta_2,\\
\phi(C_3) &= 2+5+2+1+5+1+\Delta_3 = 16+\Delta_3,\\
\phi(C_4) &= 2+5+1+2+4+2+\Delta_4 = 16+\Delta_4,\\
\phi(C_5) &= 2+5+1+2+5+1+\Delta_5 = 16+\Delta_5,\\
\phi(C_6) &= 2+4+2+2+4+2+\Delta_6 = 16+\Delta_6,\\
\phi(C_7) &= 2+4+2+2+5+1+\Delta_7 = 16+\Delta_7,\\
\phi(C_8) &= 2+5+1+2+5+1+\Delta_8 = 16+\Delta_8,\\
\phi(C_9) &= 2+5+1+1+5+2+\Delta_9 = 16+\Delta_9,\\
\phi(C_{10}) &= 2+4+2+2+4+2+\Delta_{10} = 16+\Delta_{10},\\
\phi(C_{11}) &= 2+5+1+1+5+2+\Delta_{11} = 16+\Delta_{11},\\
\phi(C_{12}) &= 1+8+1+1+5+2+\Delta_{12} = 18+\Delta_{12},\\
\phi(C_{13}) &= 1+8+1+2+4+2+\Delta_{13} = 18+\Delta_{13}.
\end{aligned}$$

We move from each coloring of Figure 22 to an allowable coloring C' (in terms of Markov chains, to a neighbor C' of C) having the same potential as C, but with the structure 3, 3, 3, 3, 3 (C' is not necessarily a local minimum). We will change a blue edge by a red edge

so that the potential remains the same. See Figure 23. We have that C'_i is a neighbor of C_i with the same potential, for $i = 1, \ldots, 13$.

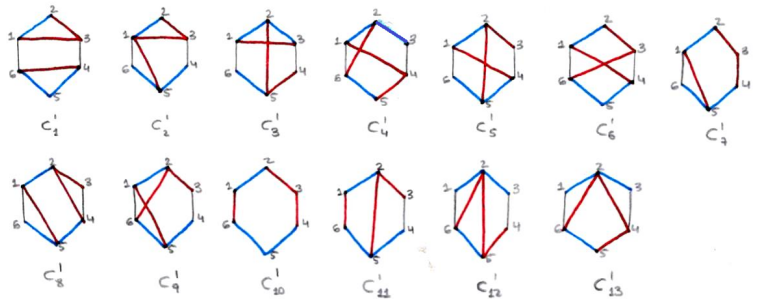


FIG. 23. Neighbors of the colorings from Figure 22, with structure 33333.

Let us compute the potentials of C'_i, and notice that they are equal to $\phi(C_i)$, for $i = 1, \ldots, 13$:

$$\begin{aligned}
\phi(C'_1) &= 2+2+4+2+4+2+\Delta_1 = 16+\Delta_1 = \phi(C_1),\\
\phi(C'_2) &= 5+2+4+1+5+1+\Delta_2 = 18+\Delta_2 = \phi(C_2),\\
\phi(C'_3) &= 2+5+2+1+5+1+\Delta_3 = 16+\Delta_3 = \phi(C_3),\\
\phi(C'_4) &= 2+5+1+4+2+2+\Delta_4 = 16+\Delta_4 = \phi(C_4),\\
\phi(C'_5) &= 2+5+1+2+5+1+\Delta_5 = 16+\Delta_5 = \phi(C_5),\\
\phi(C'_6) &= 2+2+4+2+4+2+\Delta_6 = 16+\Delta_6 = \phi(C_6),\\
\phi(C'_7) &= 2+2+4+2+5+1+\Delta_7 = 16+\Delta_7 = \phi(C_7),\\
\phi(C'_8) &= 2+5+1+2+5+1+\Delta_8 = 16+\Delta_8 = \phi(C_8),\\
\phi(C'_9) &= 2+5+1+1+5+2+\Delta_9 = 16+\Delta_9 = \phi(C_9),\\
\phi(C'_{10}) &= 2+2+4+2+4+2+\Delta_{10} = 16+\Delta_{10} = \phi(C_{10}),\\
\phi(C'_{11}) &= 2+5+1+1+5+2+\Delta_{11} = 16+\Delta_{11} = \phi(C_{11}),\\
\phi(C'_{12}) &= 1+8+1+1+5+2+\Delta_{12} = 18+\Delta_{12} = \phi(C_{12}),\\
\phi(C'_{13}) &= 1+8+1+4+2+2+\Delta_{13} = 18+\Delta_{13} = \phi(C_{13}).
\end{aligned}$$

Thus, we need to check that the potential of the non-proper local minimum colorings of the form 3, 3, 3, 3, 3 can be decreased at some step. This will give as a consequence that, given any non-proper local minimum coloring of K_6, its potential can go down at some step.

To analyze all non-proper local minimum colorings of the form 3, 3, 3, 3, 3 in a clever way, let us paint with blue any of the five colors that is repeated 3 times. By reordering the vertices 1, 2, 3, 4, 5 and 6 of K_6 if necessary, we may assume that the coloring has the form specified in Figure 24.

We have a blue 2-path and a single blue edge. Thus, there is a unique 2-path of color blue. By Lemma 4.7, there must exist another 2-path, say of color red, such that its center has no incident edge of color blue. With this information, given Figure 21, the red 2-path must have its center at vertex 4. Once we have all possible red 2-paths with center at vertex 4, see Figure 25, the remaining red edge has to be painted according to the conditions of Lemma 4.4. All the possibilities are given in Figure 26. In Figure 26, we can see the schemas drawn in Figure 25, together with a vertical arrow that leads to

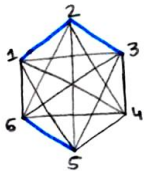


FIG. 24. Position of the color blue, which is repeated 3 times. Non-proper local minimum colorings of K_6 with structure 33333.

the possible positions of the alone red edge. Thereby, Figure 26 gives all possible schemas for a non-proper local minimum coloring of the form 3, 3, 3, 3, 3. The colorings have the name $C_1, \ldots, C_{21}$.

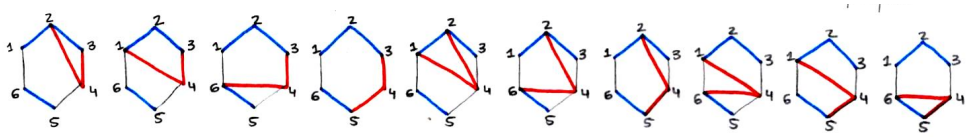


FIG. 25. Position of the color blue, which is repeated 3 times, and position of the red 2-path. Non-proper local minimum colorings of K_6 with structure 33333.

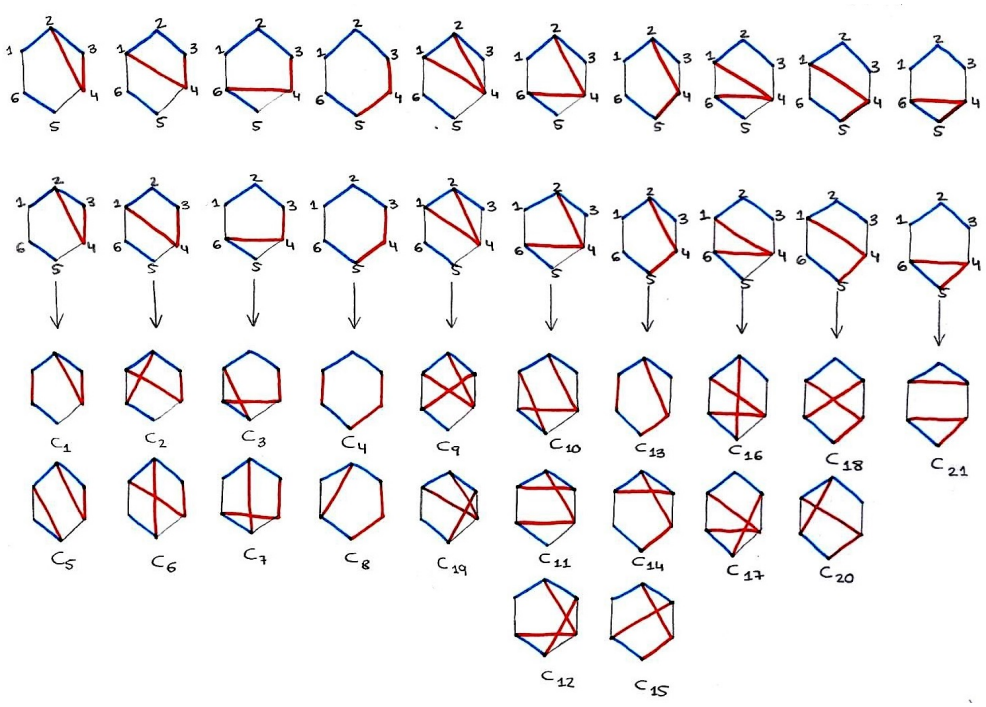


FIG. 26. Position of the color blue, which is repeated 3 times; position of the red 2-path; and position of the alone red edge. Non-proper local minimum colorings of K_6 with structure 33333.

Now we need to prove that, at some step, the potential of $C_1, \ldots, C_{21}$ decreases strictly. In Figure 27, we move from C_i to either a coloring with lower potential (indicated with

(LP)) or to a coloring that is not a local minimum (indicated by **(NLM)**), by Lemma 4.4 (there is a monochromatic 3-path), for $i = 1, \ldots, 20$, by interchanging colors blue and red appropriately. In Figure 27, the potential of each coloring is written below each hexagon. The case C_{21} is special, in the sense that we cannot move to a coloring with less potential by interchanging colors blue and red. This is due to the fact that C_{21} has two unconnected triangles, each one with a color repeated twice. Thus, a different reasoning is required for C_{21}. The idea is that, if the blue and red colors appear in C_{21} in the form of Figure 26 (two unconnected triangles), then there are two other colors from the three remaining ones, say green and orange, that are repeated three times each one and that, reordering the vertices again, they are located in edges following the structure of $C_1, \ldots, C_{20}$ (with green and orange instead of blue and red, so that the green and orange colored edges do not form two unconnected triangles, basically because there are no more uncolored triangles in the coloring C_{21} of Figure 26). Thus, working with the colors green and orange in C_{21}, we are able to decrease its potential at some step, as this case belongs to the previous colorings $C_1, \ldots, C_{20}$ (because the green and orange edges do not form two unconnected triangles in C_{21}).

Hence, we conclude that any non-proper local minimum coloring of the form 3, 3, 3, 3, 3 can be decreased in terms of its potential function. As any non-proper local minimum coloring can reach a coloring of the form 3, 3, 3, 3, 3, we derive that the potential of any non-proper local minimum coloring of K_6 can be decreased at some step. In the language of Markov chains, every non-proper coloring of K_6 has a positive probability of getting absorbed at some step by a proper coloring. Proposition 4.5 together with Proposition 4.4 give the convergence of the algorithm for K_6, and we are done.

4. Tackling the conjecture in K_n

The goal of this section is to propose ideas in order to tackle the conjecture in the general case of K_n, n even. These ideas will be in the spirit of our proofs for K_4 and K_6. In the case of K_n, a computational proof cannot be done, as n is arbitrary. Hence, via theoretical results, we will try to understand the structure of the local minimum colorings of K_n.

The key auxiliary results for the proof of the conjecture in K_4 and K_6 have been Lemma 4.4, Lemma 4.7 and Lemma 4.8. Both Lemma 4.4 and Lemma 4.7 hold for the general case K_n. By contrast, Lemma 4.8 is subject to K_6. As we saw, Lemma 4.8 allowed us to understand better the structure of local minimum colorings. Thus, we attempt to generalize Lemma 4.8 for K_n.

LEMMA 4.9. *Let C be a local minimum coloring of K_n, n even. If*

$$c_C(\mu) := \textit{number of times color } \mu \textit{ appears in coloring } C,$$

then

$$c_C(\mu) \leq \left\lfloor \frac{2n}{3} \right\rfloor,$$

for each color μ.

PROOF. If C is proper, then each color is repeated $n/2$ times. Thus, we may assume that C is a non-proper local minimum coloring.

Suppose by contradiction that $c_C(\mu) > \lfloor 2n/3 \rfloor$, for certain color μ. We distinguish cases according to the value $n \bmod 3$:

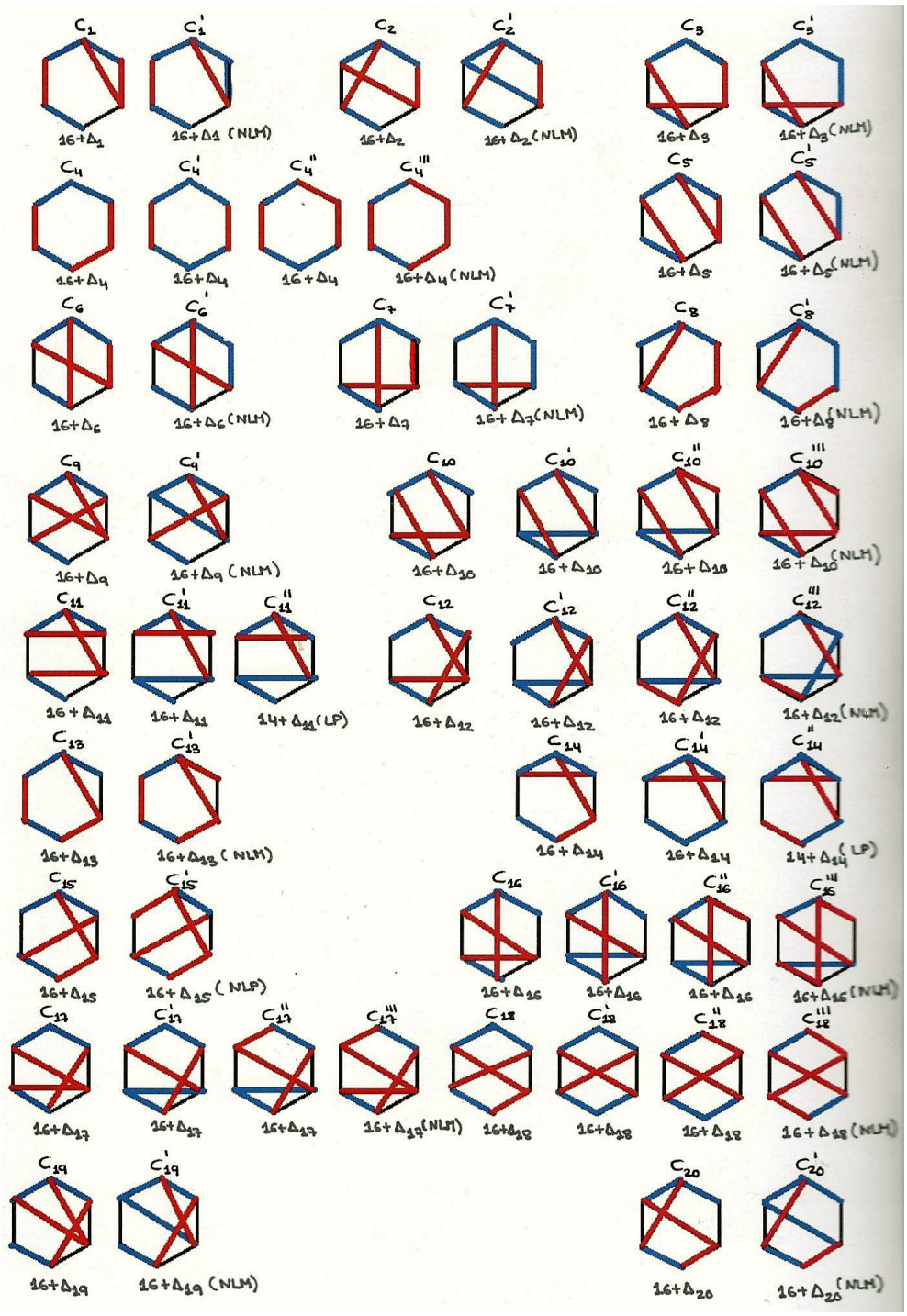


FIG. 27. Decreasing the potential of the non-proper local minimum colorings of K_6 with structure 33333. The initials **(LP)** and **(NLM)** stand for less potential and non local minimum coloring, respectively. Below each hexagon, the potential is written.

- Case $n \equiv 0 \bmod 3$. In this case, $\lfloor 2n/3 \rfloor = 2n/3$, therefore, our assumption becomes $c_C(\mu) > 2n/3$. At least $2n/3+1$ edges are painted μ. These edges can go alone or forming 2-paths, by Lemma 4.4. The way these $2n/3+1$ cover less vertices is when there are $n/3$ 2-paths and an isolated edge, which takes up $3 \cdot n/3 + 2 = n+2$ vertices. This is a contradiction, as K_n has n vertices.
- Case $n \equiv 1 \bmod 3$. In this case, $\lfloor 2n/3 \rfloor = (2n-2)/3$ (since $2n-2 \equiv 0 \bmod 3$). Our assumption on μ becomes $c_C(\mu) > (2n-2)/3$. At least $(2n-2)/3+1$ edges are painted μ. These edges can go alone or forming 2-paths, by Lemma 4.4. The manner these $(2n-2)/3+1$ edges cover less vertices is with $(n-1)/3$ 2-paths and a single edge, which take up $3 \cdot (n-1)/3 + 2 = n+1$ vertices, and this gives a contradiction.
- Case $n \equiv 2 \bmod 3$. In this case, $\lfloor 2n/3 \rfloor = (2n-1)/3$ (notice that $2n-1 \equiv 0 \bmod 3$). Our surmise on μ becomes $c_C(\mu) > (2n-1)/3$. At least $(2n-1)/3+1$ edges are painted μ. These edges can go alone or forming 2-paths, by Lemma 4.4. The way these $(2n-1)/3+1 = (2n+2)/3$ edges occupy less vertices is with $(n+1)/3$ 2-paths, which cover $3 \cdot (n+1)/3 = n+1$ vertices, and this gives once again a contradiction.

This finishes the proof. Notice that this upper-bound is tight, because for $n = 4$ and $n = 6$ it is reached. □

The lower-bound, as in the case of K_6, is harder to prove. By symmetry with $2n/3$ and from the known lower-bound for K_4 and K_6, which is 2 for both cases, we conjecture that the lower-bound for K_n is $\lceil n/3 \rceil$.

CONJECTURE 4.3. *Let C be a local minimum coloring of K_n, n even. Then*

$$c_C(\mu) \geq \left\lceil \frac{n}{3} \right\rceil,$$

for each color μ.

The problem is that, although we knew that this lower-bound is true, the corresponding Table 1 for a general K_n would be extremely large, so considering all cases as we did in K_6 is not feasible.

Thus, in any case, alternatives should be explored. We consider the following lemma, which is a simple and direct consequence of Lemma 4.1, but very useful. Denote by

$$c_{uv,\mu}(C) = a_{u,\mu}(C) + a_{v,\mu}(C)$$

the total number of edges of color μ incident to uv.

LEMMA 4.10. *Let C be a local minimum coloring. Let uv be an edge of K_n of color μ in C. Then, for every color μ',*

$$c_{uv,\mu}(C) - c_{uv,\mu'}(C) \leq 2.$$

Equality holds if and only if $\phi(C') = \phi(C)$, where C' is the recoloring of C constructed by changing the color of the edge uv to μ'.

PROOF. By Lemma 4.1,

$$\phi(C') = \phi(C) + 2(c_{uv,\mu'}(C) + 2 - c_{uv,\mu}(C)).$$

As C is a local minimum,

$$\phi(C) \leq \phi(C') = \phi(C) + 2(c_{uv,\mu'}(C) + 2 - c_{uv,\mu}(C)).$$

Hence, $c_{uv,\mu}(C) - c_{uv,\mu'}(C) \leq 2$, as wanted. Equality holds if and only if $\phi(C) = \phi(C')$. □

Let C be a local minimum coloring. Let uv be any edge of C. By Lemma 4.4, either uv is part of a monochromatic 2-path or it is independent of the other edges with its same color. If uv is part of an α-colored 2-path, then $c_{uv,\alpha}(C) = 3$. By Lemma 4.10, for each color μ', $c_{uv,\mu'} \geq 3 - 2 = 1$. Thus, if C' is a recoloring of C consisting in changing the color α of uv by β, then $\phi(C') = \phi(C)$ if and only if $c_{uv,\beta} = 1$, or in words, there is a single β-colored edge incident to uv. On the other hand, if uv is an independent α-colored edge, then the unique way of having $\phi(C') = \phi(C)$ is by changing α by a color not incident to uv.

Thus, the moves permitted to maintain the potential of a local minimum coloring are:

- if uv is part of a monochromatic 2-path, you can change the color of uv by the color of an edge incident once in uv;
- if uv is an independent edge, you can change the color of uv by a non-incident color to uv.

With these ideas, for a general K_n, one may try to maintain the potential of a local minimum coloring until arriving at a non-local minimum coloring. Then there would be a positive probability of arriving at a proper coloring and, by Proposition 4.5 and Proposition 4.4, the convergence of the algorithm would follow for K_n.

For instance, let us see the following case. Let C be a non-proper local minimum coloring. As C is non-proper, it has a 2-path of color α. As we saw in the reasoning of Figure 7 and Figure 8, there is a color β such that $c_{uv,\beta}(C) = 1$. If β is part of a 2-path of color β, then, by changing the color of uv by β, we obtain a recoloring C' such that $\phi(C') = \phi(C)$ and with a 3-path of color β. By Lemma 4.4, C' is not a local minimum, so its potential can be decreased in a next step, and we are done with this case. See Figure 28.

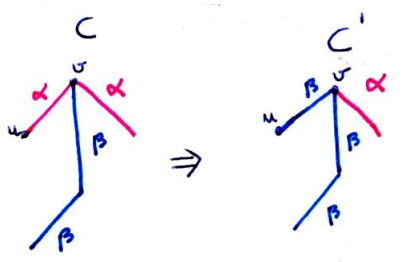


FIG. 28. Example of how to arrive at a non-local minimum coloring C' from a non-proper local minimum coloring C in K_n. Color β is repeated once in uv.

The idea would be to prove this for a general local minimum coloring of K_n. Just to give more ideas, it is easy to see that it suffices to deal with local minimum colorings with a unique 2-path α, for a certain color α. Indeed, consider two 2-paths of color α. Refer to the two 2-paths as P_1 and P_2. Consider one of the two edges of P_1, call it uv. We know that there is a color β incident once to uv. We change the color α of uv to β. Then the 2-path P_1 of color α has disappeared and it remains P_2. See Figure 29.

Repeating this procedure, we arrive at a unique α-colored 2-path. From here, one could think of using Lemma 4.7, which asserts that there exists another 2-path, say of color γ, such that its center does not have any incident α-colored edge.

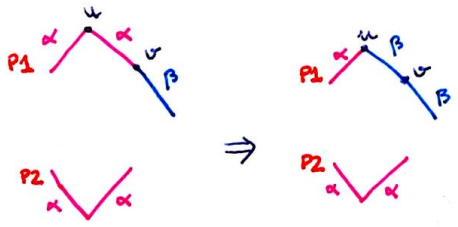


FIG. 29. Proving that one may assume that a non-proper local minimum coloring has a unique α-colored 2-path.

Using these ideas, one could think of a proof for K_n. We leave Conjecture 4.2 open for a general K_n.

5. The potential function from another point of view: the entropy

As we could not finish the proof of the convergence of the algorithm for K_n, let us think in a different way. We will change our definition of potential function by another measure concerning the *information* of the coloring, the entropy, which is a central tool in information theory and in probability theory.

Our definition of the potential of a coloring C of K_n, n even,

$$\phi(C) = \sum_{\nu=1}^{n} \sum_{\mu=1}^{n-1} a_{\nu,\mu}(C)^2,$$

can be seen as an expectation. Let $\Omega = \{(\nu,\mu) : \nu \in [n],\ \mu \in [n-1]\}$ be a sample space and

$$\mathbb{P}(\nu,\mu) = \frac{a_{\nu,\mu}(C)}{n(n-1)}$$

be a probability defined on Ω. This probability is greater for vertices ν and colors μ such that nearly all edges incident to ν are μ-colored. In some way, this probability penalizes vertices and colors that favor C to be proper. Define the random variable $X : \Omega \to \mathbb{R}$ as

$$X(\nu,\mu) = \frac{a_{\nu,\mu}(C)}{n(n-1)}.$$

This random variable contains the *information* given by $\mathbb{P}$. The expected information in C is

$$\mathbb{E}[X] = \sum_{\nu=1}^{n} \sum_{\mu=1}^{n-1} \frac{a_{\nu,\mu}(C)^2}{(n(n-1))^2} = \frac{\phi(C)}{(n(n-1))^2}.$$

However, intuitively, the *information* should be something that, when a very unlikely event occurs, this fact gives a lot of information, whereas when a very probable event happens, this fact does not give much information. For example, imagine that there are five benches in Barcelona painted pink, and they are located in the same small area of Barcelona, say A. In A and outside A, the rest of benches are not pink. I go for a walk in Barcelona. If I find a white bench, then I do not obtain information about where A is; however, if I am lucky and I sit down in one of the five pink benches, I have information about A and about where the other four pink benches are.

Thus, if $\mathbb{P}(\nu,\mu)$ is small, then the occurrence of the event (ν,μ) should give a lot of information, and vice versa. Thus, we redefine the information as

$$X(\nu,\mu) = -\log\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right)$$

since proper colorings are less probable and give as much more information(here, log is the logarithm in base e and $0\log(0)$ is interpreted as 0, by taking a limit). The expectation of X, which is the so-called Shannon entropy [8, Chapter 2] and is denoted by $H(C)$, is

$$H(C) = -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right).$$

The entropy H can be considered as an alternative measure to ϕ.

The entropy is the mean information of the system C. It measures the chaos-uncertainty-surprise of the system. In our case, when K_n is monochromatic, there is no uncertainty. Instead, when K_n is properly colored, the probability distribution $\mathbb{P}$ is uniform and the degree of uncertainty is maximum.

PROPOSITION 4.6. *For every coloring C,*

$$\log(n) \le H(C) \le \log(n(n-1)).$$

Moreover, $H(C) = \log(n)$ if and only if C is monochromatic, and $H(C) = \log(n(n-1))$ if and only if C is proper.

PROOF. To prove the upper bound, we use Jensen's inequality [18, Theorem 3.3]: "if $f : I \subseteq \mathbb{R} \to \mathbb{R}$ is a convex function, $x_1,\dots,x_m \in I$ and $a_1,\dots,a_m > 0$ positive weights such that $\sum_{i=1}^m a_i = 1$, then $f(\sum_{i=1}^m a_i x_i) \le \sum_{i=1}^m a_i f(x_i)$".

Let $f(x) = -x\log(x)$, $x \in (0,1)$. This function is concave, because $f''(x) = -1/x < 0$, so Jensen's inequality works the other way around.

The entropy is expressed as

$$H(C) = \sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1} f\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right).$$

By Jensen's inequality with weights $a_i = 1/(n(n-1))$ for $i = 1,\dots,n(n-1)$,

$$\frac{H(C)}{n(n-1)} = \sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{1}{n(n-1)} f\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right) \le f\left(\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{1}{n(n-1)}\frac{a_{\nu,\mu}(C)}{n(n-1)}\right) = f\left(\frac{1}{n(n-1)}\right),$$

so

$$H(C) \le n(n-1) f\left(\frac{1}{n(n-1)}\right) = -\log\left(\frac{1}{n(n-1)}\right) = \log(n(n-1)).$$

The equality holds when there is equality in Jensen's inequality, that is, if $x_1 = \dots = x_m$, which means $a_{\nu,\mu}(C)$ constant. Hence, $a_{\nu,\mu}(C) = 1$, and the coloring C is proper.

To prove the lower bound, we use Lagrange multipliers Theorem [1, Theorem 13.12]. Given $k = n(n-1)$, we want to minimize

$$H(p_1,\dots,p_k) = -\sum_{i=1}^{k} p_i\log(p_i)$$

on the set $\Omega_0 = \{\vec{p} \in (0,1/n)^k : g(\vec{p}) = 1\}$, being $g(\vec{p}) = \sum_{i=1}^k p_i$ and $\vec{p} = (p_1,\dots,p_k)$. We know that H has an absolute minimum and maximum on $\bar{\Omega}_0 = [0,1/k]^k \cap g^{-1}(\{1\})$ (closure

of Ω_0), by Weierstrass Extreme Value Theorem (because H is a continuous function on a compact set). By Lagrange multipliers Theorem, if a local extremum $\vec{p}_0 = (p_1, \ldots, p_k)$ belongs to Ω_0, then there exists a $\lambda \in \mathbb{R}$ such that

$$\nabla H(\vec{p}_0) + \lambda \nabla g(\vec{p}_0) = 0, \tag{5.1}$$

where ∇ denotes the gradient operator. Doing the computations, $\lambda = \log(p_j) + 1$, for all $j = 1, \ldots, k$, therefore, since λ is the same for all $j = 1, \ldots, k$, we derive $p_1 = \ldots = p_k$, so necessarily $p_i = 1/k = 1/(n(n-1))$, for all $i = 1, \ldots, k$. This point corresponds to the absolute maximum, as we saw using Jensen's inequality. Thus, since every extreme value on Ω_0 must satisfy (5.1), and if it satisfies (5.1) we have seen that it is a maximum, the absolute minimum of H must be reached at the boundary of Ω_0: that is, $p_i \in \{0, 1/n\}$ for all $i = 1, \ldots, k$. As $\sum_{i=1}^{k} p_i = 1$, there must be n terms among $p_1, \ldots, p_k$ that take on the value $1/n$, and the rest of the terms are 0. On the other hand, for p_i of the form $\frac{a_{\nu,\mu}(C)}{n(n-1)}$, this implies $a_{\nu,\mu}(C) = n - 1$, for a fixed color μ and all vertices $\nu = 1, \ldots, n$. Hence, C is monochromatic. □

Although the entropy may seem more adequate than the potential in terms of *information*, the advantage of the potential is that it is always an integer, and the relation between a coloring C and a recoloring C' can be explicitly computed (see Lemma 4.1). The logarithm function is more difficult to deal with, and Mathematical Analysis plays an important role to prove properties of the entropy, as we saw in Proposition 4.6.

The algorithm to obtain a proper coloring from any arbitrary coloring works analogously as the potential function. In the Markov chain with its nodes given by the colorings of K_n, we move from C to a recoloring C' if and only if $H(C) \leq H(C')$. We have implemented the algorithm in R.

The following function creates an arbitrary coloring:

```
arbitrary_coloring <- function(n) {

  # Goal: with n-1 colors (1,2,...,n-1), to color uniformly the edges of the
  # complete graph Kn.
  # Input: n=size of the complete graph Kn, we understand n even.
  # Output: matrix C such that C(i,j) is the color of the edge {i,j}, i distinct
  #         j. The coloring is uniform (each color has probability 1/(n-1)).
  #         We have considered C(i,i)=n (just notation).

  C <- matrix(0,n,n) # initialize matrix
  diag(C) <- rep(n,n) # put n at the diagonal (just notation)
  for(column in 1:(n-1)) # fill the lower part of C with colors
    C[(column+1):n,column] <- sample(1:(n-1),size=n-column,replace=TRUE,
                      prob=rep(1/(n-1),n-1))
  for(row in 1:(n-1)) # fill the upper part of C so that C is symmetric
    C[row,(row+1):n] <- C[(row+1):n,row]

  return(C)
}
```

The following function computes the entropy of a coloring given in matrix form:

```
entropy <- function(C) {
```

```
  # Goal: to compute the entropy of the coloring given by the matrix C.
  #       It belongs to the interval [-log(1/n),-log(1/(n*(n-1)))].
  # Input: coloring matrix C.
  # Output: entropy of C.

  n <- dim(C)[1]
  H <- 0
  for(vertex in 1:n) {
    for(color in 1:(n-1)) {
      p <-  sum(C[,vertex]==color)/(n*(n-1)) # a_{vertex,color}(C)=sum(C[,vertex]==color)
      if(p == 0)
        expected_information <- 0 # R does not assign 0 for 0*log(0)
      else
        expected_information <- -p*log(p)
      H <- H+expected_information
    }
  }
  return(H)
}
```

Finally, the following function performs the algorithm. The problem with the entropy is that, as it is not an integer, the infinite decimal part coming from the logarithm evaluation may induce errors in equalities and inequalities, as two numbers that are actually the same may be different from some decimal place due to the numerical errors. Thus, we need to round in order to avoid such a problem:

```
find_proper_coloring_entropy <- function(n) {

  # Goal: to find a proper coloring of the complete graph Kn via the entropy measure.
  # Input: n=size of the complete graph Kn. n must be EVEN.
  # Output: a list having a coloring C, a boolean that is TRUE if C is a proper
  #         coloring and FALSE otherwise, and the number of steps to find C.

  C <- arbitrary_coloring(n) # start with an arbitrary coloring
  entropyC <- entropy(C) # entropy of coloring C
  steps <- 0
  proper <- FALSE
  while(proper == FALSE & steps < n^10) {
    edge <- sample(1:n,size=2,replace=FALSE,prob=rep(1/n,n)) # choose an arbitrary edge
    color <- sample((1:(n-1))[-C[edge[1],edge[2]]],size=1,replace=FALSE,
         prob=rep(1/(n-2),n-2)) # choose color
    C1 <- C # recolor
    C1[edge[1],edge[2]] <- color
    C1[edge[2],edge[1]] <- color # the matrix C1 represents changing the color
                                 # of an edge from C
    entropyC1 <- entropy(C1) # potential of coloring C1
    if(round(entropyC,4) <= round(entropyC1,4)) {
      C <- C1 # if the entropy of C1 is greater or equal, we move to C1,
          # otherwise remain at C
      entropyC <- entropyC1
      steps <- steps + 1 # count step whenever we move
    }
```

```
    # case when we finish the procedure:
    if(round(entropyC,4) == round(log(n*(n-1)),4))
      proper <- TRUE
  }

  return(list(C,proper,steps))
}
```

We can obtain a plot similar to Figure 6 but in this entropy setting. The code to obtain the plot is the following, which gives Figure 30.

```
x <- sort(rep(seq(10,22,2),30))
y <- c()
for(size in x) {
  continue <- TRUE
  while(continue == TRUE) {
    result <- find_proper_coloring_entropy(size)
    if(result[[2]]== TRUE) {
      y <- c(y,result[[3]])
      continue <- FALSE
      print(length(y)) # we know how long is taking
    }
  }
}

logx <- log(x) # logarithmic scale
logy <- log(y)

plot(logx,logy,xlab='log(size of graph)',ylab='log(number of steps)',
     main='log convergence rate to a proper coloring') # draw points of the
                                                       # numerical experiments
coeff <- lm(logy ~ logx)$coefficients # linear model
lines(logx,coeff[[1]]+coeff[[2]]*logx,col=2) # plot the regression line
text(2.4,7,paste0('y=',round(coeff[[1]],2),'+',
    round(coeff[[2]],2),' x'),col=2) #$ write the line
```

Numerical experiments suggest that working with the potential and the entropy is more or less equivalent. Just at an intuitive level, we know that $\log(1+x) \approx x$ for $x \approx 0$, therefore

$$\begin{aligned} H(C) &= -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right) = -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\log\left(1+\frac{a_{\nu,\mu}(C)}{n(n-1)}-1\right) \\ &\approx -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}-1\right) = -\frac{\phi(C)}{(n(n-1))^2}+1. \end{aligned}$$

In fact, this expression with the inequality $\geq$ always holds. One has to use the well-known inequality $\log(1+x) \leq x$, $x \geq -1$:

$$\begin{aligned} H(C) &= -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right) = -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\log\left(1+\frac{a_{\nu,\mu}(C)}{n(n-1)}-1\right) \\ &\geq -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}-1\right) = -\frac{\phi(C)}{(n(n-1))^2}+1. \end{aligned}$$

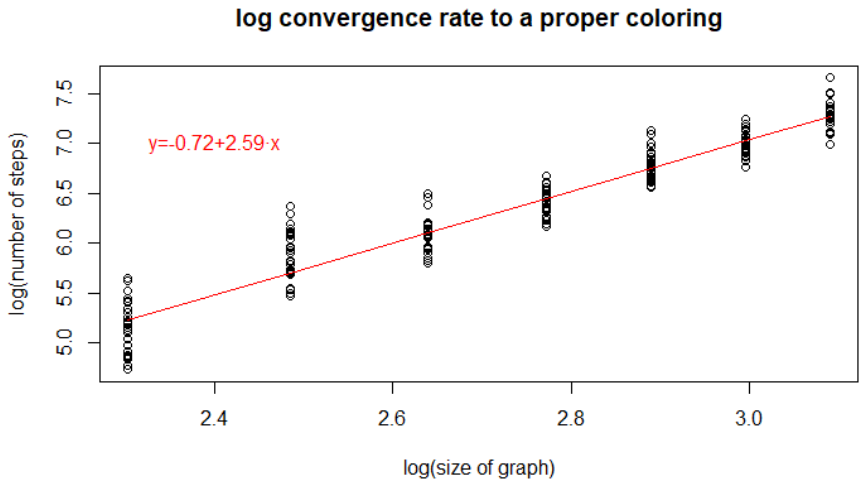


FIG. 30. Steps required in the Markov chain to arrive at a proper coloring of K_n via the entropy algorithm. In red, a regression line to study the cost in polynomial time.

Results concerning entropy in the spirit of the potential function could be attempted, although they are more complicated and require tools from Mathematical Analysis, in the sense that, since the entropy is not an integer, proofs by counting or induction are not so viable (compare Lemma 4.1 with the subsequent development).

Given a coloring C and a recoloring C' of it, we would like to analyze when $H(C') < H(C)$, $H(C') = H(C)$ or $H(C') > H(C)$ holds. Suppose that the edge uv is painted μ_0 in C. Suppose that C' consists in painting uv with color $\mu_1 \neq \mu_0$. As we saw in the proof of Lemma 4.1, $a_{\nu,\mu}(C) = a_{\nu,\mu}(C')$ for all $\nu \in [n]$ different from u and v and for all $\mu \in [n-1]$ different from μ_0 and μ_1. Since $\mu_0 \neq \mu_1$, we have

$$\begin{cases} a_{u,\mu_0}(C) = 1 + a_{u,\mu_0}(C'), \\ a_{v,\mu_0}(C) = 1 + a_{v,\mu_0}(C'), \end{cases}$$

because in C' there is one less edge painted in color μ_0 than in C, and

$$\begin{cases} a_{u,\mu_1}(C') = 1 + a_{u,\mu_1}(C), \\ a_{v,\mu_1}(C') = 1 + a_{v,\mu_1}(C), \end{cases}$$

as in C' there is one more edge painted in color μ_1. Thus,

$$\begin{aligned}
H(C') &= -\sum_{\nu=1}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C')}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C')}{n(n-1)}\right)\\
&= -\sum_{\substack{\nu=1\\ \nu\neq u,v}}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C')}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C')}{n(n-1)}\right)-\sum_{\nu=1}^{n}\sum_{\substack{\mu=1\\ \mu\neq\mu_0,\mu_1}}^{n-1}\frac{a_{\nu,\mu}(C')}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C')}{n(n-1)}\right)\\
&-\frac{a_{u,\mu_0}(C')}{n(n-1)}\log\left(\frac{a_{u,\mu_0}(C')}{n(n-1)}\right)-\frac{a_{v,\mu_0}(C')}{n(n-1)}\log\left(\frac{a_{v,\mu_0}(C')}{n(n-1)}\right)-\frac{a_{u,\mu_1}(C')}{n(n-1)}\log\left(\frac{a_{u,\mu_1}(C')}{n(n-1)}\right)\\
&-\frac{a_{v,\mu_1}(C')}{n(n-1)}\log\left(\frac{a_{v,\mu_1}(C')}{n(n-1)}\right)\\
&= -\sum_{\substack{\nu=1\\ \nu\neq u,v}}^{n}\sum_{\mu=1}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right)\\
&-\sum_{\nu=1}^{n}\sum_{\substack{\mu_1\\ \mu\neq\mu_0,\mu_1}}^{n-1}\frac{a_{\nu,\mu}(C)}{n(n-1)}\log\left(\frac{a_{\nu,\mu}(C)}{n(n-1)}\right)-\frac{a_{u,\mu_0}(C)-1}{n(n-1)}\log\left(\frac{a_{u,\mu_0}(C)-1}{n(n-1)}\right)\\
&-\frac{a_{v,\mu_0}(C)-1}{n(n-1)}\log\left(\frac{a_{v,\mu_0}(C)-1}{n(n-1)}\right)-\frac{a_{u,\mu_1}(C)+1}{n(n-1)}\log\left(\frac{a_{u,\mu_1}(C)+1}{n(n-1)}\right)\\
&-\frac{a_{v,\mu_1}(C)+1}{n(n-1)}\log\left(\frac{a_{v,\mu_1}(C)+1}{n(n-1)}\right).
\end{aligned}$$

Hence,

$$H(C')\overset{<}{=}H(C)$$

if and only if

$$\begin{aligned}
&-\frac{a_{u,\mu_0}(C)-1}{n(n-1)}\log\left(\frac{a_{u,\mu_0}(C)-1}{n(n-1)}\right)-\frac{a_{v,\mu_0}(C)-1}{n(n-1)}\log\left(\frac{a_{v,\mu_0}(C)-1}{n(n-1)}\right)\\
&-\frac{a_{u,\mu_1}(C)+1}{n(n-1)}\log\left(\frac{a_{u,\mu_1}(C)+1}{n(n-1)}\right)-\frac{a_{v,\mu_1}(C)+1}{n(n-1)}\log\left(\frac{a_{v,\mu_1}(C)+1}{n(n-1)}\right)\\
&\overset{<}{=}-\frac{a_{u,\mu_0}(C)}{n(n-1)}\log\left(\frac{a_{u,\mu_0}(C)}{n(n-1)}\right)-\frac{a_{v,\mu_0}(C)}{n(n-1)}\log\left(\frac{a_{v,\mu_0}(C)}{n(n-1)}\right)\\
&-\frac{a_{u,\mu_1}(C)}{n(n-1)}\log\left(\frac{a_{u,\mu_1}(C)}{n(n-1)}\right)-\frac{a_{v,\mu_1}(C)}{n(n-1)}\log\left(\frac{a_{v,\mu_1}(C)}{n(n-1)}\right)
\end{aligned}\tag{5.2}$$

Let us define

$$f(x) = -x\log(x).$$

Then, the last expression (5.2) becomes:

$$\begin{aligned}
&f\left(\frac{a_{u,\mu_0}(C)-1}{n(n-1)}\right)+f\left(\frac{a_{v,\mu_0}(C)-1}{n(n-1)}\right)+f\left(\frac{a_{u,\mu_1}(C)+1}{n(n-1)}\right)+f\left(\frac{a_{v,\mu_1}(C)+1}{n(n-1)}\right)\\
&\overset{<}{=}f\left(\frac{a_{u,\mu_0}(C)}{n(n-1)}\right)+f\left(\frac{a_{v,\mu_0}(C)}{n(n-1)}\right)+f\left(\frac{a_{u,\mu_1}(C)}{n(n-1)}\right)+f\left(\frac{a_{v,\mu_1}(C)}{n(n-1)}\right)
\end{aligned}\tag{5.3}$$

Therefore, rewriting (5.3), we get:

$$\begin{aligned}
&f\left(\frac{a_{u,\mu_1}(C)+1}{n(n-1)}\right)+f\left(\frac{a_{v,\mu_1}(C)+1}{n(n-1)}\right)-f\left(\frac{a_{u,\mu_1}(C)}{n(n-1)}\right)-f\left(\frac{a_{v,\mu_1}(C)}{n(n-1)}\right)\\
&\overset{<}{=}f\left(\frac{a_{u,\mu_0}(C)}{n(n-1)}\right)+f\left(\frac{a_{v,\mu_0}(C)}{n(n-1)}\right)-f\left(\frac{a_{u,\mu_0}(C)-1}{n(n-1)}\right)-f\left(\frac{a_{v,\mu_0}(C)-1}{n(n-1)}\right)
\end{aligned}\tag{5.4}$$

To make computations easily, let us simplify repeated notation (at the end we will redo this change of notation):

$$x = \frac{a_{u,\mu_1}(C)}{n(n-1)}, y = \frac{a_{v,\mu_1}(C)}{n(n-1)}, z = \frac{1}{n(n-1)}, a = \frac{a_{u,\mu_0}(C)}{n(n-1)}, b = \frac{a_{v,\mu_0}(C)}{n(n-1)}$$

Thus, according to this notation introduced, expression (5.4) becomes:

$$f(x+z) + f(y+z) - f(x) - f(y) \lessgtr f(a) + f(b) + f(a-z) + f(b-z) \tag{5.5}$$

Define the following functions, taking into account the definition of f we did before:

$$F(r,s) = f(r) + f(s) = -r\log(r) - s\log(s) \tag{5.6}$$

and

$$G(t) = F(x+t, y+t) - F(x,y) - F(a,b) + F(a-t, b-t). \tag{5.7}$$

Then, since we deduced that

$$H(C') \lessgtr H(C) \text{ if and only if (5.5) holds,}$$

according to the definition of G and H,

$$H(C') \lessgtr H(C) \text{ if and only if } G(z) \lessgtr 0.$$

From this characterization, we can obtain some results concerning the comparison of $H(C')$ and $H(C)$ (considering its corresponding relation $<$, $=$ and $>$). We will present two results for $H(C') < H(C)$. Notice that we are not able to obtain a characterization of $H(C') < H(C)$, so this is a major difference compared with the potential function (see 4.1).

PROPOSITION 4.7. *Let C be a coloring of K_n such that the edge uv is painted with color μ_0. Let C' be a recoloring of C, in such a way that the edge uv of K_n is painted with another color $\mu_1 \neq \mu_0$. If*

$$\log(a_{u,\mu_0}(C)) + \log(a_{v,\mu_0}(C)) < \log(a_{u,\mu_1}(C)) + \log(a_{v,\mu_1}(C)) + \frac{1}{n-1},$$

then $H(C') < H(C)$.

PROOF. In this proof we are going to take advantage of the notation a, b, x, y and z and the definition of the functions F and G (see (5.6 and 5.7), respectively) we introduced before. By Taylor's theorem [1, Theorem 5.19],

$$G(t) = G(0) + G'(0)t + \frac{G''(\epsilon_t)}{2}t^2 \underbrace{=}_{G(0)=0} G'(0)t + \frac{G''(\epsilon_t)}{2}t^2,$$

where $t \in [0,z]$ and $\epsilon_t \in [0,t]$. We know that $H(C') < H(C)$ if and only if $G(z) < 0$, so according to our Taylor's expansion, we have

$$H(C') < H(C) \text{ if and only if } G'(0)z + \frac{G''(\epsilon_z)}{2}z^2 < 0,$$

and simplifying this is equivalent to

$$H(C') < H(C) \text{ if and only if } G'(0) < -\frac{G''(\epsilon_z)}{2}z.$$

Now, let us obtain explicit expressions for G' and G'' differentiating one and two times, respectively, the function G (5.7) [2]:

$$\begin{aligned} G'(t) &= \frac{\partial F}{\partial r}(x+t,y+t) + \frac{\partial F}{\partial s}(x+t,y+t) - \frac{\partial F}{\partial r}(a-t,b-t) - \frac{\partial F}{\partial s}(a-t,b-t) \\ &= (-1-\log(x+t)) + (-1-\log(y+t)) - (-1-\log(a-t)) - (-1-\log(b-t)) \\ &= -\log(x+t) - \log(y+t) + \log(a-t) + \log(b-t) \end{aligned} \tag{5.8}$$

and

$$G''(t) = -\frac{1}{x+t} - \frac{1}{y+t} - \frac{1}{a-t} - \frac{1}{b-t}.$$

Since color μ_0 is different from color μ_1, we have that $a_{u,\mu_1}(C) \le n-2$. Hence, $x \le \frac{n-2}{n(n-1)}$. Moreover, since $t \in [0,z]$,

$$x + t \le x + z \underbrace{\le}_{z=\frac{1}{n(n-1)}} \frac{n-2}{n(n-1)} + \frac{1}{n(n-1)} = \frac{1}{n}.$$

From this last inequality, we obtain $-\frac{1}{x+t} \le -n$. Analogously, we deduce $-\frac{1}{y+t} \le -n$. Then,

$$G''(t) \le -\frac{1}{x+t} - \frac{1}{y+t} \le -2n.$$

Applying this for $t = \epsilon_z$ and the definition of $z = 1/(n(n-1))$,

$$-\frac{G''(\epsilon_z)}{2} z \ge nz = \frac{1}{n-1}.$$

Thus, if $G'(0) < 1/(n-1)$, then $H(C') < H(C)$. So we compute $G'(0)$ and we impose $G'(0) < 1/(n-1)$:

$$\begin{aligned} G'(0) &= -\log(x) - \log(y) + \log(a) + \log(b) \\ &\underbrace{=}_{\substack{\text{definition of} \\ x,y,a,b}} -\log\left(\frac{a_{u,\mu_1}(C)}{n(n-1)}\right) - \log\left(\frac{a_{v,\mu_1}(C)}{n(n-1)}\right) + \log\left(\frac{a_{u,\mu_0}(C)}{n(n-1)}\right) + \log\left(\frac{a_{v,\mu_0}(C)}{n(n-1)}\right) \\ &= -\log(a_{u,\mu_1}(C)) - \log(a_{v,\mu_1}(C)) + \log(a_{u,\mu_0}(C)) + \log(a_{v,\mu_0}(C)) < 1/(n-1). \end{aligned} \tag{5.9}$$

And from (5.9), we obtain that, if

$$\log(a_{u,\mu_0}(C)) + \log(a_{v,\mu_0}(C)) < \log(a_{u,\mu_1}(C)) + \log(a_{v,\mu_1}(C)) + \frac{1}{n-1},$$

then $G'(0) < 1/(n-1)$, and therefore $H(C') < H(C)$. □

PROPOSITION 4.8. *Let C be a coloring of K_n such that the edge uv is painted with color μ_0. Let C' be a recoloring of C, in such a way that the edge uv of K_n is painted with another color $\mu_1 \neq \mu_0$. If $H(C') < H(C)$, then*

$$\log(a_{u,\mu_0}(C)-1) + \log(a_{v,\mu_0}(C)-1) < \log(a_{u,\mu_1}(C)+1) + \log(a_{v,\mu_1}(C)+1).$$

PROOF. As in the proof of Proposition 4.7, in this proof we are going to use the notation a, b, x, y and z and the definition of the functions F and G (see (5.6) and (5.7), respectively).

[2]The notation $\frac{\partial F}{\partial r}$ means differentiating with respect to the first variable and $\frac{\partial F}{\partial s}$ means differentiating with respect to the second variable.

By Taylor's theorem [1, Theorem 5.19],

$$\underbrace{G(0)}_{=0} = G(z) - G'(z)z + \frac{G''(\epsilon_z)}{2}z^2,$$

where $\epsilon_z \in [0, z]$. Since $H(C') < H(C)$ implies $G(z) < 0$, $H(C') < H(C)$ also implies

$$0 < -G'(z)z + \frac{G''(\epsilon_z)}{2}z^2,$$

which is equivalent to $G'(z) < \frac{G''(\epsilon_z)}{2}z < 0$. The important conclusion is that $G'(z) < 0$. Using the computations we did to compute G', see (5.8),

$$G'(z) = -\log(x+z) - \log(y+z) + \log(a-z) + \log(b-z).$$

Substituting a, b, x, y and z by their definition that was introduced in a red box, $x = \frac{a_{u,\mu_1}(C)}{n(n-1)}$, $y = \frac{a_{v,\mu_1}(C)}{n(n-1)}$, $z = \frac{1}{n(n-1)}$, $a = \frac{a_{u,\mu_0}(C)}{n(n-1)}$ and $b = \frac{a_{v,\mu_0}(C)}{n(n-1)}$, we are done, because $H(C') < H(C)$ also implies

$$\begin{aligned} 0 > G'(z) = & -\log\left(\frac{a_{u,\mu_1}(C)+1}{n(n-1)}\right) - \log\left(\frac{a_{v,\mu_1}(C)+1}{n(n-1)}\right) \\ & + \log\left(\frac{a_{u,\mu_0}(C)-1}{n(n-1)}\right) + \log\left(\frac{a_{v,\mu_0}(C)-1}{n(n-1)}\right) = -\log(a_{u,\mu_1}(C)+1) \\ & - \log(a_{v,\mu_1}(C)+1) + \log(a_{u,\mu_0}(C)-1) + \log(a_{v,\mu_0}(C)-1), \end{aligned}$$

as we wanted to show. □

As we do not have any characterization for the comparison of $H(C')$ and $H(C)$, we believe that it is much more difficult to characterize local maximum colorings in terms of entropy (entropy greater than or equal to the entropy of its neighbors).

In any case, numerical experiments make us think that the following statement is true:

CONJECTURE 4.4. *Consider K_n, with n even, and all possible colorings of it with $n-1$ colors. Consider the algorithm in which we move from a coloring C to a recoloring C' of it if and only if $H(C) \leq H(C')$. Then the algorithm always converges for all n and, moreover, the number of steps required by the algorithm is $\mathcal{O}(n^p)$, where $2 < p < 3$.*

Chapter 5
Conclusions

We conclude this work with a summary of the main results obtained and conjectures arisen from them.

The main goal of this work has been to study the application of Markov chains to the random generation of combinatorial objects. The work concentrates on the still open problem of efficient random generation of proper edge colorings of complete graphs. The problem is connected to the more general problem of random generation of Latin squares, the above one corresponding to the case of symmetric Latin squares.

The work was prompted by a conjecture formulated in a recent paper by Dotan and Linial [7] on the convergence of a Markov chain based on a simple appealing algorithm. Given the complete graph K_n, with n even, and all its edge colorings with $n-1$ colors, a mathematical problem in the last decades has consisted in proving that the extant algorithms to generate proper colorings converge. The conjecture states that the algorithm that starts with an arbitrary coloring C, and moves to a recoloring C' if and only if $\phi(C') \leq \phi(C)$, converges asymptotically in polynomial time. Our numerical experiments in `R` suggest that the convergence always holds for all even n, and not just asymptotically, in polynomial time. Thus, we have established Conjecture 4.2. This conjecture has been shown for K_4 and K_6, both computationally and analytically, in Sections 2 and 3 of Chapter 4. The ideas used for K_4 and K_6 have been extended to K_n and new results have been proved in an attempt to tackle the conjecture in the general case. The problem is that our proofs rely on the structure of K_4 and K_6. The new Conjecture 4.2 remains open:

Conjecture 4.2. *Consider K_n, with n even, and all possible colorings of it with $n-1$ colors. Consider the algorithm in which we move from a coloring C to a recoloring C' of it if and only if $\phi(C') \leq \phi(C)$. Then the algorithm always converges for all n and, moreover, the number of steps required by the algorithm is $\mathcal{O}(n^p)$, where $2 < p < 3$.*

A new algorithm using Shannon entropy instead of the potential has been implemented in Section 5 of Chapter 4. Shannon entropy is a tool in probability theory that models the mean information of a system of events. In our case, the system of events is a coloring C, where the events are pairs of vertices and colors (ν, μ). Numerical experiments and analytical approximations suggest that entropy provides the same information as the potential. An open question for us is which is the exact relation between the entropy and the potential of a coloring: Is there an expression that relates both functions? Are the local minimums in terms of potential and the local maximums in terms of entropy the same in both approaches? A conjecture in the spirit of the potential function arises:

CONJECTURE 4.4. *Consider K_n, with n even, and all possible colorings of it with $n-1$ colors. Consider the algorithm in which we move from a coloring C to a recoloring C' of it if and only if $H(C) \leq H(C')$. Then the algorithm always converges for all n and, moreover, the number of steps required by the algorithm is $\mathcal{O}(n^p)$, where $2 < p < 3$.*

We have developed implementations in `R` of the algorithms described in the work, both for the potential and the entropy function. The implementation is accompanied with several additional functions, like the generation of colorings, computation of the number of steps up to convergence and many others. These tools are available for potential interested researchers in the area.

The problems in random generation require not only that the algorithm outputs a desired object (a proper edge coloring of K_n in our case), but also that the stationary distribution is uniform. In this work we concentrated in convergence issues as formulated in Conjecture 4.2. Once convergence is assured, standard tools show that the Metropolis version of the algorithm (by allowing increasing of potential with a small probability) will converge to the uniform distribution.

We have also concentrated on proper edge colorings of complete graphs. Most of the tools and arguments adapt easily to obtain proper edge colorings of other families of graphs. Of particular interest are the proper edge colorings of bipartite complete graphs $K_{n,n}$. Again our interest in analyzing Conjecture 4.2 drove us to focus on the particular case of the complete graph.

References

[1] T. M. Apostol. *Mathematical Analysis*. Addison-Wesley Publishing Company, Second edition, 1981.

[2] F. Aràndiga, R. Donat and P. Mulet. *Mètodes numèrics per a l'àlgebra lineal*. Universitat de València, 2000.

[3] T. Beşeri. *Edge coloring of a graph*. Izmir Institute of Technology, Turkey, June, 2004 (Master's Thesis). Available at http://web.iyte.edu.tr/~tinabeseri/tina-tez.pdf

[4] C. J. Colbourn and J. H. Dinitz. *Handbook of Combinatorial Designs*. Discrete mathematics and its applications, second edition, 2007.

[5] J. M. Corcuera. *A course in Stochastic Processes*. Universitat de Barcelona. Available at http://www.ub.edu/plie/personal_PLiE/corcuera_HTML/Teoria4.pdf

[6] M. Dotan and N. Linial. *Efficient Generation of Random One-Factorizations for Complete Graphs*. ArXiv, 2017. First version, now substituted.

[7] M. Dotan and N. Linial. *Efficient Generation of Random One-Factorizations through Hill Climbing*. ArXiv, 2017. Available at https://arxiv.org/abs/1707.00477

[8] T. S. Han and K. Kobayashi. *Mathematics of Information and Coding*. Translations of Mathematical Monographs, Volume 203, American Mathematical Society, 2002.

[9] M. T. Jacobson and P. Matthews. *Generating uniformly distributed random latin squares*. Journal of Combinatorial Designs, Volume 4, Issue 6 (1996) 405-437.

[10] A. D. Keedwell and J. Dénes. *Latin Squares and their Applications*. New York-London: Academic Press, Second edition, 1974.

[11] E. Lesaffre and A. B. Lawson. *Bayesian Biostatistics*. Statistics in Practice, 2012.

[12] D. A. Levin and Y. Peres. *Markov Chains and Mixing Times*. Second edition. Available at http://pages.uoregon.edu/dlevin/MARKOV/mcmt2e.pdf

[13] L. Lovász. *Random Walks on Graphs: A Survey*. Bolyai Society Mathematical Studies, 2. Combinatorics, Paul Erdős is Eighty (Volume 2), Keszthely (Hungary), 1993, pp. 1-46. Available at http://www.cs.elte.hu/~lovasz/erdos.pdf

[14] B. D. McKay and N. C. Wormald. *Uniform generation of random Latin rectangles*. J. Combin. Math. Combin. Comput. 9 (1991), 179-186.

[15] C. D. Meyer. *Matrix analysis and applied linear algebra*. Society for industrial and applied mathematics (siam), 2000.

[16] J. R. Norris. *Markov Chains*. Cambridge University Press, 1998.

[17] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media, sixth edition, 2003.

[18] W. Rudin. *Real and Complex Analysis*. McGraw-Hill Book Company, Third edition, 1987.

[19] A. Sinclair and M. Jerrum. *Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains*. Information and computation 82, 93-133 (1989).

[20] A. Sinclair. *Algorithms for Random Generation & Counting. A Markov Chain Approach*. Progress in Theoretical Computer Science, 1993.

[21] W. D. Wallis. *One-factorizations*. Springer Science & Business Media, 1997.