

# Map Coloring Algorithm

Martín Alejandro Castro Álvarez

Universidad Internacional de Valencia (VIU)

Calle Pintor sorolla, 21 46002, Valencia (España)

[martincastro.10.5@gmail.com](mailto:martincastro.10.5@gmail.com)

**Abstract.** Este estudio presenta un algoritmo genético para el problema clásico de colorear mapas, un desafío en la teoría de grafos que consiste en asignar colores a regiones de un mapa de manera que regiones adyacentes no compartan el mismo color. Basado en el Teorema de los Cuatro Colores, el algoritmo propuesto explora soluciones utilizando mecanismos evolutivos como selección, reproducción y mutación para optimizar el coloreo del mapa. Además, se examina el impacto de varios parámetros, incluyendo el tamaño de la muestra, la tasa de mutación, la población y la tasa de reproducción, en la eficiencia y efectividad del algoritmo. Se observó que incrementar el número de conexiones entre países o el tamaño de la población mejora la calidad de las soluciones, aunque aumenta el tiempo de procesamiento. Los resultados demuestran la eficacia del algoritmo en diferentes configuraciones, destacando su potencial en la resolución de problemas de optimización, en general

**Keywords:** Genetic Algorithm, Graph Coloring, Optimization, Population Dynamics, Mutation, Reproduction, Selection, Fitness Evaluation, Symmetric Adjacency Matrix, Cost Calculation, Algorithm Performance, Computational Complexity, Evolutionary Computing, Node Coloring, Convergence Analysis.

## 1. Introducción

### 1.1. Problema

Este estudio se centra en una cuestión fundamental en la teoría de grafos: el desafío de colorear un mapa de manera eficiente y óptima. En este contexto, el "mapa" se conceptualiza como un grafo, donde cada región del mapa se representa como un nodo. El problema que nos proponemos resolver es el de asignar uno de cuatro colores distintos a cada una de estas regiones de forma que dos regiones adyacentes no compartan el mismo color. Las regiones adyacentes se refieren a aquellas que comparten una frontera común.

### 1.2. Solución Propuesta

El algoritmo genético propuesto como solución se enfoca en explorar el espacio de soluciones posibles de una manera eficiente, utilizando mecanismos inspirados en la evolución natural, como selección, reproducción y mutación. A través de sucesivas generaciones, el algoritmo refina las soluciones, seleccionando y modificando aquellas asignaciones de colores que resulten en el menor número de regiones adyacentes con el mismo color. De esta forma, aunque el algoritmo no garantiza encontrar una solución perfecta en todos los casos, busca aproximarse lo más posible a la solución ideal, minimizando los errores en un tiempo razonable.

El algoritmo genético será evaluado en términos de su eficacia y eficiencia para abordar el problema planteado. Para ello, se evalúa la capacidad del algoritmo de encontrar la solución deseada en el espacio de búsqueda. Luego, en la última sección, se evalúan los efectos de la variación de los distintos parámetros sobre la eficacia y la eficiencia computacional del algoritmo.

## 2. Antecedentes Teóricos

### 2.1. Teorema de los Cuatro Colores

Desde su concepción en 1852 por Francis Guthrie, el Teorema de los Cuatro Colores [1] ha desafiado a la comunidad matemática por más de cien años. Este teorema propone que es posible colorear las regiones de cualquier mapa utilizando un máximo de cuatro colores distintos, asegurando que no existan dos áreas adyacentes con el mismo color.

## 3. Metodología

### 3.1. Modelo Matemático

**Genes:** Los colores del mapa vienen representados por el conjunto de genes  $G = \{g_1, g_2, g_3, g_4\}$  donde cada gen representa, arbitrariamente, un color tal como "rojo", "verde", etc.

**Individuo:** Representa una solución potencial al problema de colorear el mapa y se define como un conjunto ordenado de genes  $I = \{c_1, c_2, c_3, \dots, c_N\}$ .  $N$  corresponde al número total de países o regiones a colorear, y cada gen  $c_i$  perteneciente a  $G$  especifica el color asignado a un país determinado. Por ejemplo,  $c_1$  representa el color del primer país.

**Adyacencia:** La relación de adyacencia entre los países se modela mediante una matriz binaria y simétrica  $A$ , donde la presencia de un 1 en la posición  $A_{ij}$  indica que el país  $i$  comparte una frontera con el país  $j$ . Para evitar redundancias, los elementos de la diagonal principal de esta matriz se establecen en 0.

**Población:** En cada generación  $t$ , se define una población  $P_t$ , compuesta por un conjunto de individuos, representados como  $P_t = \{I_1, I_2, I_3, \dots, I_M\}$ , donde  $M$  es el número total de individuos. En otras palabras, es un conjunto de soluciones posibles, que el algoritmo evaluará y sobre la cual se aplicarán las operaciones genéticas de selección, reproducción y mutación.

**Reproducción:** Durante la etapa de reproducción, que se produce en cada generación  $t$ , los individuos de la población actual  $P_t$  generan descendencia en proporción a una tasa de reproducción positiva  $R$ . Este factor determina el crecimiento de la población, asegurando que el número de individuos en la generación subsecuente  $P_{t+1}$  sea un múltiplo de la generación actual, según la ecuación (1).

$$|P_{t+1}| = R|P_t| \quad (1)$$

**Mutación:** En el proceso de reproducción, los individuos heredan inicialmente los genes de sus progenitores. Sin embargo, para introducir variabilidad genética: existe una probabilidad  $p$  de que cualquier gen sea sustituido por otro gen seleccionado al azar del conjunto disponible  $G$ .

**Selección:** Al concluir cada generación  $t$ , se evalúa a cada miembro de la población recién formada  $P_{t+1}$  otorgándole un valor de aptitud  $L(I)$ , que refleja su desempeño frente al problema de colorear el mapa. Basándose en estos valores, se elige un subconjunto de  $M$  individuos más prometedores, es decir, aquellos cuyo coste es el más bajo, asegurando su avance a la siguiente generación.

**Novedad:** Se introduce un componente estocástico en la evaluación de la aptitud de las soluciones: a cada individuo  $I$  en una generación  $t$ , se le suma un valor aleatorio  $N(I)$  a su costo original  $L(I)$ . Esta técnica, expresada por la ecuación (2), tiene como objetivo proporcionar a todas las soluciones una oportunidad equitativa de ser seleccionadas para la reproducción, incluso si no son las más aptas. De esta manera, se está asegurando una exploración más completa del espacio de soluciones y se aumenta las posibilidades de escapar de mínimos locales durante el proceso evolutivo.

$L'(I) = L(I) + N(I)$	(2)
-----------------------	-----

### 3.2. Optimización

**Inicialización de la Población:** En la primera generación  $t = 0$ , se crea una población inicial compuesta por  $M$  individuos, cuyos colores se asignan de manera aleatoria y con igual probabilidad a partir del conjunto de genes  $G$ . Esta inicialización aleatoria es muy importante, ya que un mayor número  $M$  de individuos en la población inicial  $P_0$  aumenta las posibilidades de explorar más el espacio de soluciones.

**Función de Coste:** La función de costo viene dada por  $L(I)$  para cada individuo  $I$  y mide la aptitud de cada individuo cuantificando el número de países adyacentes que comparten el mismo color en su configuración genética. Por tanto, individuos con un costo más bajo, indicando menor número de adyacencias de colores repetidos, son preferidos durante el proceso de selección.

**Criterio de Finalización:** El proceso se detiene bajo dos posibles condiciones: la primera es el alcance de un número predefinido de generaciones,  $T$ , evitando que se ejecute indefinidamente. En ese caso, se elige aquella solución que tiene menor costo, dentro de las regiones exploradas hasta el momento (condición necesaria). La segunda condición es la consecución de un individuo cuyo costo  $L(I)$  es igual a cero, ya que implica que se ha encontrado una solución ideal donde ningún país adyacente comparte el mismo color (condición suficiente).

**Análisis de Convergencia:** Partiendo de un costo inicial  $L_0(I)$  para cada individuo, el algoritmo selecciona para reproducción aquellos individuos cuyo costo mejora en, al menos, una cantidad mínima  $\epsilon$  mayor o igual a cero en cada generación. A medida que el número de generaciones  $t$  se acerca a un valor alto y arbitrario  $E$ , la diferencia entre el costo inicial y la mejora acumulada,  $L_0(I) - \epsilon$ , se aproxima a un valor constante  $K$ . Es decir, converge.

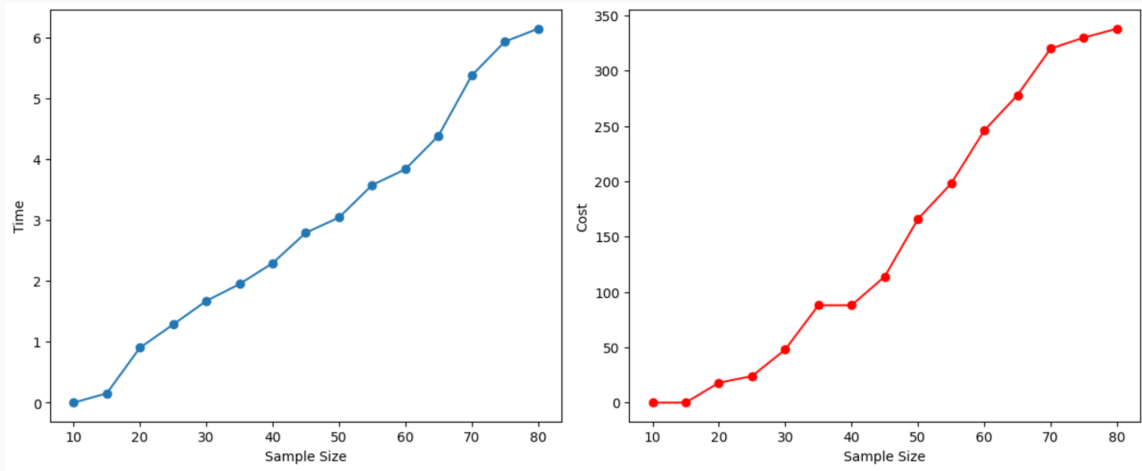
## 4. Resultados

### 4.1. Eficacia

Para resolver el problema de colorear mapas satisfactoriamente, se establecieron parámetros cuidadosamente seleccionados para equilibrar la eficiencia y la eficacia. El tamaño de la muestra  $N$  se fijó en 10, lo que indica que cada individuo representa una solución potencial que involucra 10 países o regiones. Se utilizaron 4 genes distintos, correspondientes a los 4 colores distintos. Es decir,  $|G| = 4$ . La tasa de mutación  $p$  se estableció en un 10%. Por otra parte, el algoritmo se ejecutó por un máximo de  $T = 100$  generaciones. Se configuró una población de  $M = 20$  individuos. La tasa de reproducción  $R$  se fijó en 2, implicando que cada individuo generaría dos descendientes en cada generación. Para fomentar la exploración y evitar mínimos locales, se introdujo una tasa de novedad 1%. Por último, para generar una matriz de adyacencia aleatoria, se utilizó un modelo generativo cuya tasa de conectividad se estableció en 50%, determinando que, en promedio, cada país es vecino del 50% de los otros países.

### 4.2. Variación en la cantidad de países

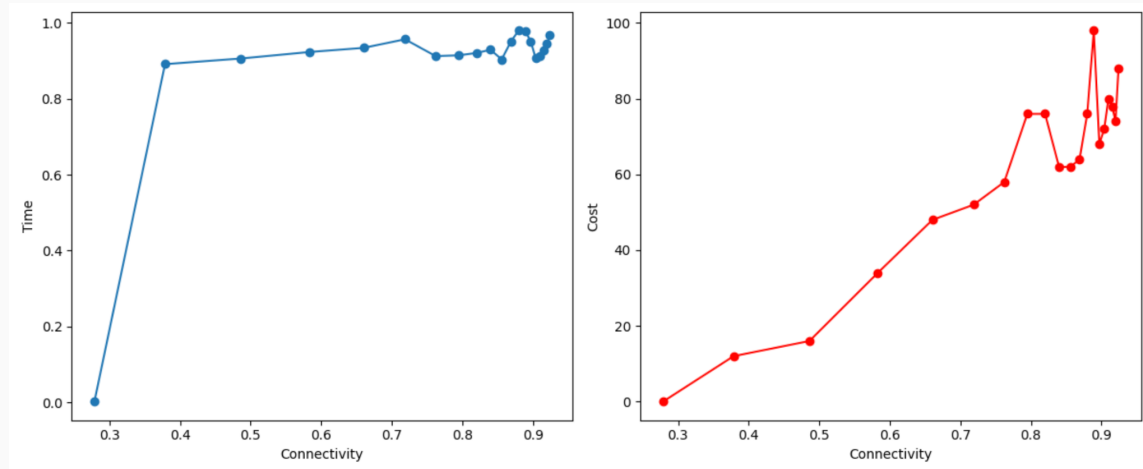
Al mantener constantes los parámetros del algoritmo genético pero aumentando el tamaño de cada individuo, es decir, el número de países en el modelo, observamos que, como era de esperar, el tiempo necesario para que el algoritmo converja hacia una solución se incrementa progresivamente. Interesantemente, cuando la cantidad de países excede los 20, la limitación de 100 generaciones para encontrar una solución óptima ya no es adecuada (Ver Fig. 1), lo que resulta en un aumento significativo del costo final de las soluciones encontradas. Esta observación indica que la cantidad de países es un factor crítico que influye notablemente en la efectividad y la eficiencia del algoritmo, destacando la necesidad de ajustar otros parámetros del algoritmo, como el número de generaciones, para manejar conjuntos de datos más grandes.



**Fig. 1:** Relación entre el tiempo de convergencia, y la eficiencia del algoritmo genético en relación al tamaño de la muestra.

### 4.3. Variación en la cantidad de conexiones entre países

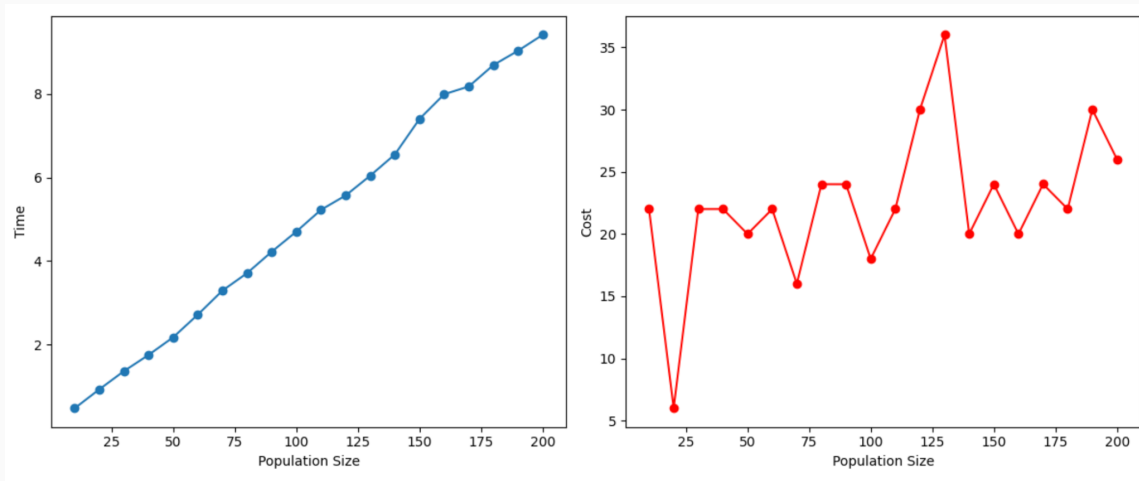
Al variar la cantidad de conexiones entre países mientras se mantienen constantes otros parámetros del algoritmo, se observa que el tiempo de ejecución aumenta de manera cuadrática y converge rápidamente hacia 1 segundo (Ver Fig. 2). Esto se atribuye a que un incremento en las conexiones impide que el algoritmo finalice de manera prematura, llevándolo a completar las 100 generaciones máximas establecidas en cada prueba. Paralelamente, el costo final del algoritmo se incrementa significativamente con el aumento de las conexiones. Esto se debe a que cuantas más conexiones existen entre los países, más desafiante se vuelve mantener colores distintos en países adyacentes, especialmente si no se permite un incremento en la variedad de colores disponibles. La densidad de conexiones se convierte así en un factor crucial que influye en la complejidad y el rendimiento del algoritmo.



**Fig. 2:** Relación entre el tiempo de convergencia, y la eficiencia del algoritmo genético en relación a la cantidad de conexiones entre países.

### 4.4. Variación en el tamaño de la población

Al experimentar con un tamaño de población variable, manteniendo constantes los demás parámetros, emergen resultados interesantes. Como era de esperarse, el tiempo requerido para converger hacia una solución se incrementa de forma lineal con el aumento en el número de individuos de la población; esto se debe probablemente al mayor tiempo computacional necesario para procesar todas las mutaciones posibles. No obstante, la relación entre la función de costo y el tamaño de la población exhibe una tendencia errática (Ver Fig. 3): aunque el costo final se eleva, se mantiene significativamente bajo en comparación con otros escenarios de prueba. Por ejemplo, mientras que el aumento en el número de genes por individuo puede elevar el costo fácilmente por encima de 300, un incremento en el tamaño de la población lo mantiene por debajo de 30. Esto sugiere que ajustar el tamaño de la población podría ser una estrategia efectiva para escalar el modelo y abordar problemas más complejos, como aquellos con alta conectividad o con individuos de gran tamaño.



**Fig. 3:** Relación entre el tiempo de convergencia, y la eficiencia del algoritmo genético en relación a el tamaño de la población.

## 5. Referencias

[1] Xiang, L. (2009). A formal proof of the four color theorem. Department of Information Science, Kyushu Sangyo University, Fukuoka, Japan. Manuscrito del 16 de abril de 2009.