

# Un Método Exacto para el Problema de Equiparticionamiento de Grafos en Componentes Conexas

Viteri, Estéfano <sup>1,\*</sup>  ; Torres, Ramiro <sup>2</sup> 

<sup>1</sup>Escuela Politécnica Nacional, Facultad de Ciencias, Quito, Ecuador

<sup>2</sup>Escuela Politécnica Nacional, Departamento de Matemática, Quito, Ecuador

**Resumen:** En el presente trabajo, el problema de equiparticionamiento de grafos en componentes conexas es estudiado. El problema consiste en particionar un grafo no dirigido con costos sobre las aristas en un número fijo de componentes conexas, tal que el número de nodos en cada componente difiera en a lo más una unidad y el costo total de las aristas con nodos finales en la misma componente sea minimizado. Se presentan varios modelos de programación lineal entera usando diferentes enfoques (maximización de los costos de las aristas del corte y minimización de los costos de las aristas en cada componente conexa) y sus resultados son comparados. Además, se exponen varias familias de desigualdades válidas asociadas a los poliedros de estas formulaciones, junto con un algoritmo exacto tipo Branch & Cut. Finalmente, se reportan resultados computacionales basados en instancias simuladas de diferentes tamaños.

**Palabras clave:** Equiparticionamiento de grafos, Programación entera, Branch & Cut

## An Exact Approach for the Graph Equipartitioning Problem in Connected Components

**Abstract:** In the present work, the graph equipartitioning problem in connected components is studied. The problem consists of partitioning an undirected graph with cost on the edges into a fixed number of connected components, such that the number of nodes in each component differs in at most one unit and the total cost of the edges with end-nodes in the same connected component is minimized. Several integer linear programming models using different approaches (maximizing the edges in the cut or minimizing the edges in the connected components) are presented and the results are compared. Moreover, several families of valid inequalities associated to the polytope of these formulations, together with a Branch & Cut algorithm are exposed. Finally, computational results based on simulated instances of different sizes are reported.

**Keywords:** Graph equipartitioning, Integer programming, Branch & Cut

### 1. INTRODUCCIÓN

Una tarea frecuente en múltiples áreas consiste en particionar un conjunto de objetos en varios subconjuntos tal que cada subconjunto cumpla con ciertos requerimientos como número mínimo o máximo de elementos u homogeneidad sobre los atributos cuantificados. Al asociar los objetos a nodos, la relación entre pares de objetos con la existencia de una arista y la similitud o disimilitud con una función de costo sobre dichas aristas, se podría pensar que el problema puede ser modelado usando teoría de grafos.

Un problema específico consiste en particionar el conjunto de nodos de un grafo de modo que el número de nodos en cada partición no exceda al resto de partes en más de una unidad y que en cada parte exista al menos un camino entre cada par de nodos. A esta variante se la conoce como el problema de

k-equiparticionamiento en componentes conexas. Enfrentar este problema desde el punto de vista matemático es una tarea compleja, ya que se conoce que es un problema NP-duro, incluso en el caso en el de que se elimine la restricción de la cantidad de nodos en cada subconjunto de la partición (Hojny et al., 2021).

El problema de particionamiento de grafos ha sido ampliamente aplicado en diferentes campos, tales como procesamiento de imágenes (Camilus & V K, 2012), diseño de distritos territoriales y políticos (Mehrotra et al., 1998), procesamiento en paralelo (Hendrickson & Kolda, 2000), diseño de circuitos VLSI (Kahng et al., 2011), deportes (Mitchell, 2003), entre otros. Múltiples aplicaciones y una completa revisión bibliográfica del problema de particionamiento de grafos pueden ser encontradas en Kalyanaraman et al. (2011) y Buluç et al. (2016).

El problema de particionamiento de grafos ha sido ampliamente

\*estefano.viteri@epn.edu.ec

Recibido: 18/01/2022

Aceptado: 01/08/2022

Publicado en línea: 02/02/2023

10.33333/tp.vol51n1.09

CC BY 4.0

estudiado y diferentes técnicas han sido usadas para su solución. Por ejemplo, Kernighan & Lin (1970) presentaron un método heurístico que permite particionar grafos arbitrarios con costos sobre las aristas en subconjuntos de nodos de tamaño conocido. Extensiones del trabajo anterior pueden ser encontradas en métodos heurísticos recientes como METIS (Karypis & Kumar, 1998) o Kahip (Sanders & Schulz, 2012) que permiten particionar grafos de gran tamaño considerando costos sobre las aristas y múltiples pesos sobre los nodos. Mediante el uso de la programación cuadrática, Fan & Pardalos (2010) reportaron un modelo que utiliza variables binarias e incluyeron varias reformulaciones y métodos de solución para aplicaciones prácticas. Con la aplicación de la programación semidefinida, Sotirov (2014) propone una nueva relajación para el problema de particionamiento general de grafos y compara los resultados teóricos y numéricos con trabajos previos. Por otro lado, Alpert et al. (1999) reportan una heurística de ordenamiento glotón basado en particionamiento espectral, donde los vectores propios del grafo son usados para construir una representación geométrica del mismo.

Nuestro interés se enfoca en métodos exactos basados en programación lineal entera. Así, Grötschel & Wakabayashi (1989) estudian el problema sobre grafos completos introduciendo el problema de particionamiento en cliques, donde los autores proponen una formulación lineal y un algoritmo tipo Branch & Cut. Enfocados en el problema anterior, Labbé & Özsoy (2010) imponen restricciones de tamaño sobre el número de nodos en cada conjunto de la partición y aportan con un estudio poliedral del mismo. Chopra & Rao (1993) presentaron un modelo de programación lineal entera para el problema de  $k$ -particionamiento junto con varias familias de desigualdades válidas y facetes para el poliedro asociado. En el caso de que se fije el número de particiones a dos, el problema de bisección puede ser identificado en Dellling et al. (2015).

Requerimientos de conectividad son frecuentes en múltiples trabajos reportados en la literatura. Por ejemplo, Wang et al. (2017) estudian desigualdades válidas que inducen facetes para el problema de encontrar el subgrafo conexo con peso máximo en un grafo. Jünger et al. (1985) proveen de condiciones necesarias para obtener conexidad en un subgrafo y de una herramienta para identificar si un grafo admite un particionamiento donde el peso en cada partición no exceda un valor fijo y se induzca un subgrafo conexo. Para el caso que se requiere particionar un grafo en componentes conexas, Hojny et al. (2021) buscan maximizar los costos de las aristas que pertenecen al corte. Los autores utilizan variables binarias para la elección de los arcos usados en la solución y variables continuas de flujo para garantizar la conexidad de las componentes. Adicionalmente, Miyazawa et al. (2021) plantean dos formulaciones para una versión del  $k$ -particionamiento conexo balanceado, es decir, se busca particionar un grafo con pesos sobre los nodos en un número fijo de componentes conexas de pesos similares. Los autores proponen una formulación basada en conjuntos separadores que incluye un gran número de restricciones que potencialmente pueden ser implementadas en tiempo polinomial. El trabajo anterior es complementado con una segunda formulación basada en flujos.

El problema de equiparticionamiento en componentes conexas usando Programación Lineal Entera es estudiado en este artículo, donde tres formulaciones para resolver el problema junto con varias familias de desigualdades válidas para el poliedro asociado son propuestas. Las desigualdades encontradas fueron incluidas como cortes en un algoritmo exacto de solución tipo Branch & Cut y son probadas en varias instancias simuladas.

El presente trabajo se encuentra organizado de la siguiente forma: En la Sección 2 se introduce la notación y tres modelos de programación lineal entera mixta que describen el problema del equiparticionamiento de grafos en  $k$ -componentes conexas. En la Sección 3 se detallan varias familias de desigualdades válidas para las formulaciones diseñadas. Finalmente, en la Sección 4 se exponen ampliamente los resultados computacionales obtenidos.

## 2. NOTACIÓN Y FORMULACIONES ENTERAS

### 2.1 Notación

Sea  $G = (V, E)$  un grafo no dirigido, donde  $V = \{1, \dots, n\}$  es el conjunto de nodos y  $E \subset \{\{i, j\} : i, j \in V, i \neq j\}$  es el conjunto de aristas. Además, sobre el conjunto de aristas se define una función de costos  $d : E \rightarrow \mathbb{R}^+$  y sea  $k$  un entero mayor o igual a dos que denota el número de componentes en las que será particionado el grafo  $G$ . Notaremos por  $[k] = \{1, \dots, k\}$ .

Sobre un grafo  $G$ , se define un camino  $P$  de  $v_1$  a  $v_k$  ( $v_1$ - $v_k$ -path) como una secuencia  $v_1, v_2, \dots, v_k \in V$  tal que  $\{v_i, v_{i+1}\} \in E$  para  $i = 1, \dots, k-1$  y  $k \geq 0$ . El número de aristas en  $P$  se conoce como la longitud del camino. Así, para dos nodos  $u, v \in V$  se define  $d(u, v)$  como la longitud más corta desde el nodo  $u$  hasta el nodo  $v$ . El grafo  $G$  es llamado conexo si existe un  $u$ - $v$ -path, para todo  $u, v \in V$ . Por otro lado, para un subconjunto de nodos  $X \subset V$ , se define el conjunto de aristas  $d(X) = \{\{u, v\} \in E : u \in X, v \in V \setminus X\}$  también llamado el corte de  $X$ . Si tomamos un subconjunto  $X \subset V$ , entonces podemos ver que si eliminamos el conjunto de aristas  $d(X)$  se generan al menos dos componentes conexas.

Lo anterior nos permite definir la idea de conjuntos separadores. Sea  $u, v \in V$  dos nodos no adyacentes de  $G$ , entonces diremos que el conjunto  $X \subset V \setminus \{u, v\}$  es un  $(u, v)$ -separador si  $u$  y  $v$  pertenecen a diferentes componentes conexas en el subgrafo inducido por  $V \setminus X$ . Si el nodo  $w$  es el único elemento del conjunto separador, entonces diremos que  $w$  es un nodo de articulación. Se define  $G_1(u, v)$  como el conjunto de todos los  $(u, v)$ -separadores de cardinalidad mínima en  $G$ .

Una  $k$ -partición de  $V$  es definida por  $k$  subconjuntos  $\{V_1, V_2, \dots, V_k\}$  donde  $V_i \cap V_j = \emptyset$  para todo  $i \neq j$ ,  $\bigcup_{c=1}^k V_c = V$ , y  $V_c \neq \emptyset$  para todo  $c \in [k]$ .

El problema de equiparticionamiento en componentes conexas ( $PECC$ ) consiste en encontrar una  $k$ -partición  $\{V_1, V_2, \dots, V_k\}$  tal que cada subconjunto  $V_i$  induce una componente conexa,  $\lfloor n/k \rfloor \leq |V_i| \leq \lceil n/k \rceil$ , y el costo total de las aristas en la misma componente es minimizado.

Es fácil notar que, si el número de particiones divide exactamente al número de nodos del grafo, todas las particiones tendrán el mismo número de nodos. Por otro lado, si dicha división no es exacta, entonces existirá una o varias particiones que posean un nodo adicional. El siguiente resultado permite conocer el número de nodos contenidos en cada subconjunto de la partición.

**Teorema 1.** Dado un grafo  $G = (V, E)$ , un entero  $k \geq 2$  y  $r = n \bmod k$ . Entonces en una  $k$ -partición se tienen  $r$  componentes con  $\lfloor n/k \rfloor + 1$  nodos y  $k - r$  componentes con  $\lfloor n/k \rfloor$  nodos.

**Demostración.** Si se tiene que  $k$  divide exactamente a  $n$ , entonces  $r = 0$  y cada componentes contiene  $n/k$  nodos y el resultado es trivial. Por otro lado, si  $n/k$  no es entero, entonces existen dos números enteros  $q$  y  $r$  tal que  $n = kq + r$  y  $r < k$ . Ahora, si tomamos  $q = \lfloor n/k \rfloor$  tenemos que:

$$\begin{aligned} n &= k \left\lfloor \frac{n}{k} \right\rfloor + r \\ &= k \left\lfloor \frac{n}{k} \right\rfloor + r + r \\ &= r \left\lfloor \frac{n}{k} \right\rfloor + 1 + (k - r) \left\lfloor \frac{n}{k} \right\rfloor \\ &= r \left\lfloor \frac{n}{k} \right\rfloor + (k - r) \left\lfloor \frac{n}{k} \right\rfloor \end{aligned}$$

Esto implica que el conjunto de  $n$  nodos puede ser particionado en  $r$  subconjuntos de tamaño  $\lfloor n/k \rfloor + 1$  y  $k - r$  subconjuntos de tamaño  $\lfloor n/k \rfloor$ , demostrando el teorema.  $\square$

### 2.2 Primer modelo

Para el primer modelo es necesario construir la versión dirigida del grafo de entrada. Así, sea  $G = (V, E)$  un grafo no dirigido y se define el digrafo  $D = (V, A)$  que dispone del mismo conjunto de nodos y sus arcos son creados de la siguiente manera: si la arista  $(u, v) \in E$ , entonces los arcos anti-paralelos  $(u, v), (v, u) \in A$ , es decir, una arista del grafo original esta asociada con dos arcos en el grafo dirigido. Sobre la versión dirigida se define como el conjunto de arcos salientes del nodo  $u$  mientras que  $A_u^+$  representa el conjunto de arcos entrantes al nodo  $u$ .

El primer modelo dispone de las siguientes variables: Se definen las variables  $x_{vi}$  que toman el valor de uno si el nodo  $v$  pertenece a la partición  $i$ , o cero si esto no sucede, y variables  $y_{uv}$  que se les asigna un valor igual a uno si la arista  $(u, v)$  pertenece al corte, o cero si no lo es. Por otro lado, si la arista  $(u, v)$  pertenece al corte, entonces los nodos  $u$  y  $v$  son localizados en la misma partición. Para satisfacer el requerimiento de conexidad, se propone enviar unidades de flujo entre los nodos de la misma componente conexa. Para este efecto se definen variables  $z_{ui}$  que toma el valor de uno si el nodo  $u$  es un nodo sumidero, o cero si no lo es. Además, sobre la versión dirigida se definen variables  $f_{uv}$  que representan la cantidad de flujo sobre el arco  $(u, v)$  que debe ser un número real positivo incluido el cero. El modelo puede ser escrito de la siguiente forma (1):

$$\begin{aligned} \max \quad & \sum_{(u,v) \in E} c_{uv} y_{uv} \\ \text{sujeto a:} \quad & \end{aligned} \tag{1a}$$

$$\sum_{i \in [k]} x_{vi} = 1, \quad \forall v \in V, \tag{1b}$$

$$\sum_{i \in [k]} x_{vi} = \lfloor \frac{n}{k} \rfloor, \quad \forall v \in V, \tag{1c}$$

$$x_{ui} + x_{vi} + y_{uv} = 2, \quad \forall (u,v) \in E, i \in [k], \tag{1d}$$

$$x_{ui} + x_{vi} + y_{uv} = 2, \quad \forall (u,v) \in E, i \in [k], \tag{1e}$$

$$x_{ui} + x_{vi} + y_{uv} = 2, \quad \forall (u,v) \in E, i \in [k], \tag{1f}$$

$$\sum_{u \in V} z_{ui} = 1, \quad \forall i \in [k], \tag{1g}$$

$$z_{ui} + x_{ui} = 1, \quad \forall u \in V, i \in [k], \tag{1h}$$

$$f_{uv} + f_{vu} \leq g(1 - y_{uv}), \quad \forall (u,v) \in E, \tag{1i}$$

$$\sum_{(u,v) \in A_u^+} f_{uv} - \sum_{(v,u) \in A_u^+} f_{vu} = 1 - g z_{ui}, \quad \forall u \in V, i \in [k], \tag{1j}$$

$$x_{vi} \in \{0, 1\}, \quad \forall v \in V, i \in [k], \tag{1k}$$

$$y_{uv} \in \{0, 1\}, \quad \forall (u,v) \in E, \tag{1l}$$

$$z_{ui} \in \{0, 1\}, \quad \forall u \in V, i \in [k], \tag{1m}$$

$$f_{uv}, f_{vu} \geq 0, \quad \forall (u,v) \in E, \tag{1n}$$

donde  $g = \lfloor n/k \rfloor + 1$ .

La función objetivo (1a) maximiza el costo total de las aristas en el corte, produciendo que las aristas de menor costo no sean seleccionadas y se ubiquen en los subgrafos inducidos por la partición. Se pueden identificar un bloque de restricciones que se concentran en equiparticionar el grafo. Así, (1b) asegura que cada nodo sea asignado exactamente a un solo subconjunto, (1c) limita la cantidad de nodos en cada subconjunto a  $\lfloor n/k \rfloor$  y (1d) y (1e) obligan a que si la arista  $(u, v)$  no pertenece al corte, entonces los nodos  $u$  y  $v$  se encuentren en el mismo subconjunto. Además, la restricción (1f) asegura que si los nodos  $u$  y  $v$  se encuentran en la misma componente, entonces la arista  $(u, v)$  no debe encontrarse en el corte. Por otro lado, un segundo grupo de restricciones asegura la conexidad de la partición. Así, la restricción (1g) exige que exactamente un solo nodo sea asignado como sumidero para cada subconjunto, mientras que (1h) es una restricción de acoplamiento de variables e indica que únicamente los nodos que pertenecen a un subconjunto son candidatos para ser sumideros. Finalmente, las restricciones (1i) y (1j) permiten calcular el flujo en los arcos. En (1i) se tiene que si la arista  $(u, v)$  pertenece al corte, entonces no debe existir un flujo por sus arcos asociados en el grafo dirigido. Por otro lado, si la arista no es asignada al corte, entonces puede existir flujo sobre dichos arcos y la suma no debe superar el  $\lfloor n/k \rfloor + 1$  unidades. Finalmente, (1j) es una restricción de conservación de flujo. Las restricciones antes mencionadas producen que los nodos en un mismo subconjunto de la partición se encuentren conectados y se conviertan en componentes conexas a través del flujo. El primer modelo es derivado de Hojny et al. (2021).

### 2.3 Segundo modelo

La siguiente formulación plantea modelar la conexidad usando conjuntos separadores. Las variables usadas en esta formulación están asociadas a los nodos y las aristas del grafo. Las variables  $x_{vi}$  toman el valor de uno si el nodo  $v$  pertenece a la partición  $i$ , o cero en caso contrario, mientras que la variable  $y_{uv}$  toma en valor

de uno si la arista  $uv$  pertenece al corte, o cero en otro caso.

$$8u \in V, \quad (3h)$$

El segundo modelo para el problema de equiparticionamiento en componentes conexas (FG -2) puede ser escrito de la siguiente forma:

$$\begin{aligned} x_{ui} &\in \{0, 1\}, & 8u \in V, i \in [k], & (3i) \\ y_{uv} &\in \{0, 1\}, & 8f \in E, & (3j) \\ z_{ui} &\in \{0, 1\}, & 8u \in V, i \in [k], & (3k) \\ f_{uv}, f_{vu} &\in \mathbb{R}_+, & 8f \in E, & (3l) \end{aligned}$$

$$\max_{f \in E} \sum_{uv \in E} c_{uv} y_{uv} \quad (2a)$$

sujeto a:  
(1b) (1f)

$$x_{ui} + x_{vi} - \sum_{w \in S} x_{wi} = 1, \quad 8f \in E, S \subseteq G_1(u, v), i \in [k], \quad (2b)$$

$$x_{vi} \in \{0, 1\}, \quad 8v \in V, i \in [k], \quad (2c)$$

$$y_{uv} \in \{0, 1\}, \quad 8f \in E. \quad (2d)$$

donde  $g = n - k + 1$  y  $r = n \bmod k$ .

La función objetivo (3a) apunta a minimizar el costo total de las aristas con nodos finales en la misma componente conexa. La restricción (3b) garantiza que cada nodo sea asignado exactamente a una sola componente conexa. La desigualdad (3c) muestra que si dos nodos incidentes  $u$  y  $v$  son asignados a un mismo subconjunto, entonces la arista que los conecta debe ser incluida en la solución. Adicionalmente, en (3d) se asegura que, si dos nodos incidentes se encuentran en diferentes subconjuntos de la partición, entonces la variable asociada a dicha arista debe tomar el valor de cero. Con estas dos últimas restricciones describimos completamente la relación existente entre nodos y aristas dentro y fuera de una componente conexa. La restricción (3e) asigna un único nodo fuente en cada subconjunto de la partición, mientras que la restricción (3f) permite a un nodo ser fuente si y solo si primero pertenece a dicho subconjunto. Finalmente, (3g) es una restricción de conservación de flujo y muestra que se seleccionan nodos fuente que envían  $k$  unidades de flujo y  $r$  nodos fuente que envían  $k - 1$  unidades de flujo. Además, si un nodo no es fuente, entonces será un sumidero con demanda igual a una unidad. El lado derecho de la restricción claramente es justificado por el Teorema 2. Es fácil ver que, si divide exactamente, entonces es igual a uno y la restricción (3h) puede ser escrita de la siguiente forma:

De forma similar que en el modelo FG -1, la función objetivo (2a) maximiza las aristas del corte y las restricciones (1b) - (1f) buscan equiparticionar el grafo. Por otro lado, la restricción (2b) impone que si dos nodos no adyacentes  $u, v$  se encuentran en el mismo subconjunto, entonces necesariamente debe existir al menos un nodo de cada conjunto separado  $G_1(u, v)$  en el mismo subconjunto para asegurar que exista comunicación entre ellos. Esta formulación descende de Miyazawa et al. (2021).

### 2.4 Tercer modelo

A diferencia de los modelos anteriores, en la presente sección se introduce una nueva formulación que considera desigualdades diferentes a las reportadas en la literatura que permiten modelar la conectividad de cada partición. Así, en este modelo se cambia el sentido de la optimización y se propone minimizar el costo total de las aristas en las componentes conexas. Este hecho provoca una modificación en la naturaleza de las variables asociadas a las aristas y se define  $y_{uv} = 1$  si la arista  $uv$  pertenece a alguna de las componentes conexas y  $y_{uv} = 0$  en otro caso. Por el contrario, las variables binarias  $x_{vi}$  mantienen su interpretación y toman el valor de uno si el nodo  $v$  se encuentra en la partición  $i$ , o cero en caso contrario. Además, las variables  $z_{ui}$  representan la cantidad de flujo enviado por el arco  $(u, v)$  en la versión dirigida. El modelo puede ser expresado de la siguiente forma (3):

$$\min_{f \in E} \sum_{uv \in E} c_{uv} y_{uv} \quad (3a)$$

$$\sum_{(u,v) \in 2d_u} f_{uv} - \sum_{(v,u) \in 2d_u^+} f_{vu} = \sum_{i=1}^n \sum_{j=1}^k z_{ui} - 1, \quad 8u \in V. \quad (3b)$$

sujeto a:

$$x_{vi} = 1, \quad 8v \in V, \quad (3b)$$

$$x_{ui} + x_{vi} - y_{uv} = 1, \quad 8f \in E, i \in [k], \quad (3c)$$

$$x_{ui} + x_{vj} + y_{uv} = 2, \quad 8f \in E, i, j \in [k], i \neq j, \quad (3d)$$

$$\sum_{u \in V} z_{ui} = 1, \quad 8i \in [k], \quad (3e)$$

$$z_{ui} \leq x_{ui}, \quad 8u \in V, i \in [k], \quad (3f)$$

$$f_{uv} + f_{vu} \leq g y_{uv}, \quad 8f \in E, \quad (3g)$$

$$\sum_{(u,v) \in 2d_u} f_{uv} - \sum_{(v,u) \in 2d_u^+} f_{vu} = \sum_{i=1}^n \sum_{j=1}^k z_{ui} - 1, \quad (3h)$$

### 2.5 Relaciones entre los modelos

Los siguientes resultados establecen una correspondencia entre el conjunto de puntos de los poliedros asociados a cada una de las formulaciones reportadas. Para ello, sean  $P^1 = \{(x, y, z, f) \mid x \in \{0, 1\}^{E_1}, y \in \{0, 1\}^{E_2}, z \in \mathbb{R}_+^{E_3}, f \in \mathbb{R}_+^{E_4}\}$ ,  $P^2 = \{(x, y, z, f) \mid x \in \{0, 1\}^{E_1}, y \in \{0, 1\}^{E_2}, z \in \mathbb{R}_+^{E_3}, f \in \mathbb{R}_+^{E_4}\}$  y  $P^3 = \{(x, y, z, f) \mid x \in \{0, 1\}^{E_1}, y \in \{0, 1\}^{E_2}, z \in \mathbb{R}_+^{E_3}, f \in \mathbb{R}_+^{E_4}\}$  los poliedros asociados a la formulación FG -1, FG -2 y FG -3, respectivamente.

**Teorema 2.** Un punto  $(x, y, z, f) \in P^1$  si y sólo si  $(x, y) \in P^2$ .

**Demostración.** Sea  $(x, y, z, f)$  un punto en  $P^1$  que genera una solución de costo  $C_1$ , entonces podemos elaborar una solución  $(x^0, y^0)$  para  $P^2$  con costo  $C_2$ . Dado que las variables sobre nodos y aristas poseen la misma interpretación se tiene una relación directa, es decir  $x^0 = x$  y  $y^0 = y$ . Notemos que, utilizando dicha relación, las variables  $x^0$  satisfacen la restricción (2b), ya que garantiza conectividad en  $P^1$ . Esto se tiene ya que  $x_{ui}^0 = x_{vi}^0 = 1$ , para nodos  $u, v$  no adyacentes y algún  $i \in [k]$ , entonces cualquier conjunto separado  $S$  debe contener al menos un nodo en  $V_i \cap f \in E$ , asegurando de este modo que  $\sum_{w \in S} x_{wi}^0 = 1$  y por tanto

el cumplimiento de la restricción mencionada. Además, se tiene el resultado anterior puede ser generalizado usando la noción de conjuntos separadores.

La otra implicación sigue inmediatamente tomando  $x^0, y = y^0$  y seleccionando aleatoriamente un nodo sumidero en  $V_2$  para  $i \in [k]$ . Finalmente, los valores de las variables de flujo pueden ser polinomialmente obtenidas asegurando la conservación de flujo sobre cada una de las componentes conexas. □

**Teorema 3.** Un punto  $(x, y, z, f) \in P^1$  si y sólo si  $(x, y, z, f) \in P^3$ .

**Demostración.** Sea  $(x, y, z, f)$  un punto en  $P^1$  que genera una solución de costo  $C_1$ , entonces podemos construir una solución  $(x^0, y^0, z^0, f^0)$  para  $P^3$  con costo  $C_3$ . Así, se pueden encontrar las aristas en la solución de  $P^3$  tomando  $y_{uv}^0 = 1$  y  $y_{uv}$  para todo  $f(u, v) \in E$ , y para los nodos utilizando  $x_{ui}^0 = x_{ui}$ , para todo  $u \in V$  y  $i \in [k]$ . Además, los nodos fuente pueden ser identificados a partir de los nodos sumideros eligiendo  $z_{ij}^0 = z_{ij}$ , para todo  $u \in V$  y  $i \in [k]$ . Finalmente, las variables  $f_{uv}^0$  son reveladas mediante  $f_{uv}^0 = f_{vu}$  para todo  $u, v \in E$ .

Adicionalmente, el valor de la función objetivo depende de las variables asociadas a las aristas y se tiene:

$$\begin{aligned} C_1 &= \sum_{f(u,v) \in E} c_{uv} y_{uv} \\ &= \sum_{f(u,v) \in E} c_{uv} (1 - y_{uv}^0) \\ &= \sum_{f(u,v) \in E} c_{uv} - \sum_{f(u,v) \in E} c_{uv} y_{uv}^0 \\ &= \sum_{f(u,v) \in E} c_{uv} - C_3 \end{aligned}$$

Por otro lado, si se dispone de un punto factible  $(x^0, y^0, z^0, f^0)$  para  $P^3$  con costo  $C_3$ , se puede notar que mediante las igualdades anteriores es posible construir fácilmente un punto  $(x, y, z, f)$  en  $P^1$  con costo  $C_1$ . □

Por el Teorema 2 y 3 se tiene el siguiente resultado:

**Corolario 1.** Un punto  $(x, y) \in P^2$  si y sólo si  $(x, y, z, f) \in P^3$ .

### 3. DESIGUALDADES VÁLIDAS

En esta sección, varias familias de desigualdades válidas asociadas a los poliedros del problema de equiparticionamiento de grafos en componentes conexas son presentadas. Los primeros resultados se derivan de las desigualdades válidas introducidas en Hojny et al. (2021) y Miyazawa et al. (2021).

**Teorema 4.** Sea  $v \in V$  un nodo de articulación y sean  $u, w \in V$  nodos en diferentes componentes conexas en el subgrafo inducido por  $V \setminus \{v\}$ . Entonces las desigualdades:

$$x_{ui} + x_{vi} - x_i \leq 1, \quad \forall i \in [k],$$

son válidas para  $FG -1$  y  $FG -3$ .

**Teorema 5.** Sean  $u$  y  $v$  dos nodos no adyacentes en  $G(u, v)$  sus conjuntos separadores. Entonces las desigualdades:

$$x_{ui} + x_{vi} - \sum_{w \in S} x_{wi} \leq 1, \quad \forall S \in G_1(u, v),$$

son válidas para  $FG -1$  y  $FG -3$ .

El tercer tipo de desigualdades identifica el número mínimo de aristas necesarias para asegurar la conexidad de una solución factible, lo que permite mejorar la cota inferior de los distintos modelos. La desigualdad tiene la siguiente estructura:

**Teorema 6.** La desigualdad:

$$\sum_{f(u,v) \in E} y_{uv} \geq m - n + k,$$

es válida para  $FG -1$  y  $FG -2$ , y la desigualdad:

$$\sum_{f(u,v) \in E} y_{uv} \geq n - k,$$

es válida para  $FG -3$ .

**Demostración.** Sean  $n_i$  el número de nodos de la componente  $i \in [k]$ , tal que  $\sum_{i \in [k]} n_i = n$  y  $\sum_{i \in [k]} n_i \leq k$ . Es conocido que se necesitan al menos 1 aristas para que la componente sea conexa. Para la formulación  $FG -3$  se tiene:

$$\sum_{f(u,v) \in E} y_{uv} - \sum_{i=1}^k (n_i - 1) = \sum_{i=1}^k n_i - \sum_{i=1}^k 1 = n - k,$$

es decir, se necesitan al menos  $k$  aristas para obtener un equiparticionamiento conexo. Por otro lado, las desigualdades para los casos de maximización  $FG -1$  y  $FG -2$  se obtienen mediante las equivalencias del Teorema 3 y del Corolario 1 usando la relación  $y^0 = 1 - y$ . Así,

$$\begin{aligned} \sum_{f(u,v) \in E} y_{uv} &\geq n - k, \\ \sum_{f(u,v) \in E} (1 - y_{uv}^0) &\geq n - k, \\ m - \sum_{f(u,v) \in E} y_{uv}^0 &\geq n - k, \\ \sum_{f(u,v) \in E} y_{uv}^0 &\leq m - n + k. \end{aligned}$$

□

El cuarto tipo de desigualdades identifica los pares de nodos que no pueden ser incluidos en la misma componente conexa. Puede ser expresado por el siguiente resultado:

**Teorema 7.** Sean  $u$  y  $v$  dos nodos de  $G$  tal que  $(u, v) \in E$ , entonces las desigualdades:

$$x_{ui} + x_{vi} \leq 1, \quad \forall i \in [k],$$

son válidas para  $FG -1, FG -2$  y  $FG -3$ .

*Demostración.* Tomemos dos nodos  $u$  y  $v$ , se conoce que el tamaño máximo de una componente es  $\lceil n/k \rceil$ . Como se tiene que  $d(u, v) \geq \lceil n/k \rceil$ , entonces cualquier camino que permite conectar a los nodos  $u$  y  $v$  en el grafo utiliza al menos  $\lceil n/k \rceil$  aristas. Esto implica que, si alguna componente conexa contiene a los nodos  $u$  y  $v$ , entonces dicha componente debe tener al menos  $\lceil n/k \rceil + 1$  nodos violando de este modo la condición de equipartición.  $\square$

El quinto tipo se deriva de las desigualdades triangulares introducidas por Grötschel & Wakabayashi (1989), que indican que, si tres nodos se encuentran dentro de un mismo subconjunto, entonces las aristas que los conectan también deben ser usadas en la solución. Así, dados tres nodos  $u, v, w \in V$  tal que  $\{u, v\}, \{v, w\}, \{u, w\} \in E$ , entonces las desigualdades:

$$\begin{aligned} -y_{vw} + y_{wu} + y_{uv} &\leq 1, \\ y_{vw} - y_{wu} + y_{uv} &\leq 1, \\ +y_{vw} + y_{wu} - y_{uv} &\leq 1 \end{aligned}$$

son válidas para  $F G-3$ . Utilizando la equivalencia entre las formulaciones se tiene que:

$$\begin{aligned} y_{vw} - y_{wu} - y_{uv} &\leq 0, \\ -y_{vw} + y_{wu} - y_{uv} &\leq 0, \\ -y_{vw} - y_{wu} + y_{uv} &\leq 0, \end{aligned}$$

son desigualdades válidas para  $F G-1$  y  $F G-2$ .

El sexto tipo de desigualdades es útil en grafos densos. La idea consiste en identificar cliques con  $k + 1$  nodos y por el hecho de que el grafo debe ser particionado en  $k$  componentes, entonces al menos 2 nodos de la clique deben ser asignados al mismo subconjunto. El resultado es presentado en el siguiente teorema:

**Teorema 8.** Sea  $T = (T', E')$  una clique en  $G$  con  $|T'| = k + 1$  nodos, entonces la desigualdad:

$$\hat{a}_{\{u,v\} \in E'} y_{uv} \leq \frac{(k+1)k}{2} - 1,$$

es válida para  $F G-1$  y  $F G-2$ . Además, la desigualdad:

$$\hat{a}_{\{u,v\} \in E'} y_{uv} \geq 1,$$

es válida para  $F G-3$ .

*Demostración.* Sea  $T = (T', E')$  una clique con  $k + 1$  nodos contenida en  $G$ . Es fácil notar que jamás podría presentarse el caso de que los nodos de la clique sean asignados a componentes diferentes ya que se tiene  $k + 1$  nodos en  $T'$  y tan solo  $k$  subconjuntos en los que debe ser particionado el conjunto de nodos. Esto implica que al menos dos nodos en  $T'$  deben ir juntos en una misma componente conexa, es decir, al menos una arista perteneciente a  $E'$  debe ser elegida en la solución. Para la formulación  $F G-3$  se tiene:

$$\hat{a}_{\{u,v\} \in E'} y_{uv} \geq 1.$$

Las desigualdades para los casos de maximización ( $F G-1$  y  $F G-2$ ) se obtienen mediante las equivalencias del Teorema 3. Así,

$$\hat{a}_{\{u,v\} \in E'} y_{uv} \geq 1,$$

$$\begin{aligned} \hat{a}_{\{u,v\} \in E'} (1 - y'_{uv}) &\geq 1, \\ E' - \hat{a}_{\{u,v\} \in E'} y'_{uv} &\geq 1, \\ \frac{(k+1)k}{2} - \hat{a}_{\{u,v\} \in E'} y'_{uv} &\geq 1, \\ \hat{a}_{\{u,v\} \in E'} y'_{uv} &\leq \frac{(k+1)k}{2} - 1. \end{aligned}$$

$\square$

Las desigualdades del anterior tipo pueden ser separadas por un algoritmo de fuerza bruta en un tiempo  $O(|V|^{k+1})$ .

El último tipo de desigualdades válidas se centran en eliminar subconjuntos de nodos que no satisfacen las condiciones de equiparticionamiento, es decir, realizar una búsqueda de aquellos subconjuntos que disponen de máximo  $\lceil n/k \rceil - 1$  nodos y asegurar que exista una arista en el corte para garantizar la conexidad de la solución. El resultado es formalizado de la siguiente forma:

**Teorema 9.** Sea  $Q \subset V$  tal que  $1 \leq |Q| \leq \lceil n/k \rceil - 1$ , entonces la desigualdad:

$$\hat{a}_{\{u,v\} \in d(Q)} y_{uv} \leq |d(Q)| - 1,$$

es válida para  $F G-1$  y  $F G-2$ . Además, la desigualdad:

$$\hat{a}_{\{u,v\} \in d(Q)} y_{uv} \geq 1,$$

es válida para  $F G-3$ .

*Demostración.* Sea  $Q \subset V$  tal que  $1 \leq |Q| \leq \lceil n/k \rceil - 1$ . Dado que cada componente debe tener al menos  $\lceil n/k \rceil$  nodos, entonces al menos un nodo en la vecindad debe ser incluido en el conjunto  $Q$  para alcanzar el número mínimo de nodos en la componente conexa y satisfacer la condición de equipartición. Esto implica que para la formulación  $F G-3$  debe existir al menos una arista en el corte de  $Q$ . Por otro lado, para las formulaciones  $F G-1$  y  $F G-2$  se puede usar el mismo argumento, es decir, para asegurar que el conjunto  $Q$  se transforme en una componente conexa, a lo más  $|d(Q)| - 1$  aristas deben ser incluidas en el corte. Si  $Q$  contiene nodos de varios subconjuntos de la partición, entonces más de una arista en  $d(Q)$  será incluida en la solución.  $\square$

#### 4. RESULTADOS COMPUTACIONALES

En esta sección, se reportan los resultados computacionales obtenidos con las formulaciones  $F G-1$ ,  $F G-2$ , y  $F G-3$  junto con las desigualdades válidas reportadas en el presente trabajo. Todos los experimentos computacionales fueron realizados en un computador Intel Core i7-9700K CPU 3.60 Ghz con 32 GB de memoria RAM bajo el sistema operativo Windows 10 Education. Además, se utilizó la versión 9.1.1 del solver de optimización Gurobi (Gurobi Optimization, LLC, 2021) a través de la distribución Anaconda. Adicionalmente, se programó en la aplicación Jupyter

Notebook mediante el lenguaje de programación Python.

Las instancias utilizadas fueron simuladas de la siguiente forma: se generaron aleatoriamente  $n$  puntos  $(x,y)$  en el plano, donde  $x \in [0;200]$  e  $y \in [0;200]$ . La posición de cada nodo fue asociada a cada punto  $(x,y)$ . Para cada par de puntos  $u,v \in V$  con  $u \neq v$ , se generó una arista con costo  $c_{uv}$  igual a la distancia euclidiana entre los puntos extremos de dicha arista. Además, un número real en el intervalo  $[0, 1]$  fue considerado, de forma que la arista existe basada en una probabilidad igual a dicho valor. Dependiendo del número de aristas y nodos, diremos que el grafo generado tiene densidad igual a  $d = 2|E|/(|V|(|V| - 1))$ .

Iniciamos comparando los tamaños de los modelos, donde se puede notar que las formulaciones  $FG-1$  y  $FG-3$  dependen únicamente de  $n = |V|$  y  $m = |E|$  por lo que su tamaño crece polinomialmente en función de la instancia de entrada. Por otra parte, el modelo  $FG-2$  precisa de la cardinalidad del conjunto  $G_1$  lo que produce que el número de restricciones crezcan exponencialmente. La Tabla 1 muestra los tamaños de los modelos, donde  $a = \frac{n(n-1)}{2} - m$  y  $b = \sum_{\{u,v\} \notin E} |G_1(u,v)|$ .

**Tabla 1.** Tamaño de los modelos

Modelo	# de variables	# de restricciones
$FG-1$	$3m + 2nk$	$m + 2n + 2k + k(3m + n)$
$FG-2$	$m + nk$	$n + k + \frac{3}{2}mk + \frac{1}{2}mk^2 + abk$
$FG-3$	$3m + 2nk$	$m + 2n + k + \frac{1}{2}mk + \frac{1}{2}mk^2 + nk$

Se plantea resolver los modelos enteros incluyendo diferentes configuraciones de desigualdades válidas y comparar sus comportamientos. Para las diferentes formulaciones se incluyeron las desigualdades de tipo: Separador (Teorema 5), Cota (Teorema 6), Camino (Teorema 7), Clique (Teorema 8), Corte (Teorema 9) y Triángulo (desigualdades triangulares). Para todas las instancias, los algoritmos de separación utilizados son del tipo fuerza bruta y los cortes del solver Gurobi fueron deshabilitados. Es importante notar que el número de desigualdades de tipo Cota y Camino tienen un orden polinomial por lo que fueron añadidas en el nodo raíz. En la formulación  $FG-2$ , las restricciones relacionadas con los conjuntos separadores fueron incluidas como desigualdades tipo lazy.

En un primer experimento, se resolvieron los modelos  $FG-1$ ,  $FG-2$  y  $FG-3$  sobre instancias pequeñas ( $n \leq 20$ ) con la finalidad de identificar la validez de las desigualdades reportadas y determinar el mejor comportamiento del algoritmo exacto de solución respecto al tiempo de ejecución, la cantidad de nodos explorados y el valor de la brecha de optimalidad.

Para cada instancia  $(n,k,d)$  se reportan los experimentos utilizando varias configuraciones de desigualdades, con todas las desigualdades y sin usar ninguna de ellas. Las Tablas 2 – 4 reportan los resultados de cada caso y se encuentran organizadas de la siguiente forma: las dos primeras columnas reportan el valor de la función objetivo considerando el sentido de minimización y su valor equivalente en el sentido de maximización; las tres columnas siguientes indican el número de nodos explorados

en el proceso de optimización, el tiempo de ejecución en segundos y la brecha de optimalidad en porcentaje. La última columna representa el número total de apariciones de todas las desigualdades válidas en cada experimento. El tiempo máximo de ejecución para los experimentos se fijó en 1800 segundos.

La diferencia en la densidad de los grafos de entrada permite evidenciar que las desigualdades tipo Camino aparecen cuando el grafo es disperso, mientras que las desigualdades tipo Clique surgen cuando el grafo es denso. Respecto a la desigualdad tipo Cota se puede ver que su inclusión es favorable, ya que genera una importante disminución en el tiempo de ejecución respecto al caso sin desigualdades. Usando las desigualdades tipo Separador, se evidencia la gran cantidad de desigualdades que son incluidas en una instancia, pero no tienen un efecto significativo en la solución. Las desigualdades tipo Triángulo no tienen un gran impacto en el tiempo de ejecución, pero en algunas instancias ayudan a encontrar la solución óptima explorando una menor cantidad de nodos. Además, es importante notar que en la gran mayoría de los modelos existe al menos una estrategia que mejora el tiempo de ejecución respecto al caso en el que no se incluyen desigualdades válidas, verificando la validez del método de solución diseñado en el presente trabajo. Para el modelo  $FG-1$ , la Figura 1 muestra la distribución del número total de apariciones de los tipos de desigualdades válidas en las instancias consideradas en este experimento.

Por otro lado, podríamos afirmar que el modelo que mejor se comporta respecto al tiempo de ejecución en estas instancias es el modelo  $FG-1$ , ya que independientemente de la desigualdad aplicada y la densidad de la instancia, el algoritmo asociado termina en un menor tiempo comparado con los otros dos modelos. Además, podemos descartar al modelo  $FG-2$  para el resto de experimentos, pues al depender de los conjuntos separadores, el tamaño del modelo crece al punto en que se vuelve computacionalmente intratable. Adicionalmente, debido al número de elementos en el conjunto  $G_1$  en instancias de mayor tamaño, las desigualdades válidas tipo Separador también las podemos excluir para los siguientes análisis.

En un segundo experimento, se procede a resolver los modelos  $FG-1$  y  $FG-3$  sobre instancias de mayor tamaño, cada uno usando las estrategias que mejoran su eficiencia según los análisis anteriores. Las nuevas instancias consideran grafos con  $n \in \{22,24,26,28,30\}$  y la cantidad de subconjuntos en la que debe ser particionado el grafo varía entre 5 y 10. De la misma forma que en las instancias anteriores, se consideran experimentos con grafos usando densidad alta y baja, pues de esto depende la existencia de varios tipos de desigualdades. La Tabla 5 reporta los resultados obtenidos para este experimento y se encuentra organizada de la siguiente forma: las tres primeras columnas muestran las características de la instancia  $(n,k,d)$ ; las siguientes seis columnas son asociadas al modelo  $FG-1$  y reportan el valor de la función objetivo transformada al caso de minimización, el valor de la función objetivo, la cantidad de nodos explorados en el proceso de optimización, el tiempo de ejecución en segundos, la brecha de optimalidad y la cantidad de desigualdades tipo Clique incluidas en el proceso de solución; las subsiguientes seis

**Tabla 2.** Comparación de los modelos  $FG-1$ ,  $FG-2$  y  $FG-3$  con instancias de densidad baja

<b>Instancia (15, 6, 0.31)</b>						
	<b>F. Min.</b>	<b>F. Max.</b>	<b>Nodos</b>	<b>Tiempo</b>	<b>Gap</b>	<b># de apariciones</b>
$FG-1$						
Clique	594,68	2 733	57 062	26,67	0	0
Corte	594,68	2 733	57 062	27,41	0	0
Cota y Camino	594,68	2 733	441	0,39	0	115
Separador	594,68	2 733	24 804	21,16	0	30
Triángulo	594,68	2 733	42 933	24,85	0	87
Sin desigualdades	594,68	2 733	57 062	24,05	0	0
Todas las desigualdades	594,68	2 733	556	0,58	0	142
$FG-2$						
Clique	594,68	2 733	4 962	1 801,21	16,29	0
Corte	594,68	2 733	4 907	1 801,20	16,29	0
Cota y Camino	594,68	2 733	80 733	149,21	0	115
Separador	594,68	2 733	4 487	1 801,18	16,29	70
Triángulo	594,68	2 733	4 841	1801,11	16,29	164
Sin desigualdades	594,68	2 733	4 603	1 801,05	16,29	0
Todas las desigualdades	594,68	2 733	78 891	120,47	0	1 400
$FG-3$						
Clique	594,68	2 733	200 681	64,89	0	0
Corte	594,68	2 733	200 681	65,55	0	0
Cota y Camino	594,68	2 733	23 167	11,12	0	115
Separador	594,68	2 733	191 641	124,6	0	180
Triángulo	594,68	2 733	184 329	79,77	0	6 865
Sin desigualdades	594,68	2 733	200 681	48,46	0	0
Todas las desigualdades	594,68	2 733	18 915	17,58	0	2 308
<b>Instancia (20, 5, 0.36)</b>						
	<b>F. Min.</b>	<b>F. Max.</b>	<b>Nodos</b>	<b>Tiempo</b>	<b>Gap</b>	<b># de apariciones</b>
$FG-1$						
Clique	765,99	5 744	3 259 549	1 800,10	7,60	0
Corte	765,99	5 744	95 522	856,02	0	726
Cota y Camino	765,99	5 744	12 661	8,71	1	
Separador	765,99	5 744	1 915 191	1800,28	8,12	360
Triángulo	765,99	5 744	2 376 724	1800,32	7,45	1 587
Sin desigualdades	765,99	5 744	4 152 452	1800,02	7,21	0
Todas las desigualdades	765,99	5 744	6 243	134,34	0	861
$FG-2$						
Clique	1 071,83	5 438,16	1 460	1 802,27	19,71	0
Corte	1 036,71	5 473,28	1 201	1 861,65	18,86	10 316
Cota y Camino	765,99	5 744,00	3 153	1 801,40	2,74	1
Separador	789,8	5 720,19	1 477	1 802,21	13,81	10
Triángulo	1 071,83	5 438,16	1 459	1 804,01	19,71	17
Sin desigualdades	1 071,83	5 438,16	1 458	1 804,32	19,71	0
Todas las desigualdades	775,96	5734,03	1 952	1802,70	2,95	19 385
$FG-3$						
Clique	765,99	5 744,00	2 483 875	1 800,14	61,69	0
Corte	775,96	5 734,03	202 218	1 803,86	46,48	1 206
Cota y Camino	765,99	5 744,00	3 276 018	1 569,98	0	1
Separador	817,49	5 692,50	1 520 811	1 800,15	67,69	200
Triángulo	815,90	5 694,09	1 740 087	1 800,04	58,00	954
Sin desigualdades	765,99	5 744,00	2 375 132	1 800,04	62,05	0
Todas las desigualdades	765,99	5 744,00	192 932	18 06,02	12,06	2 531

columnas están asociados al modelo  $FG-3$  e indican el valor de la función objetivo, la cantidad de nodos explorados en el proceso de optimización, el tiempo de ejecución en segundos, la brecha de optimalidad, el número de desigualdades tipo Clique y el número de desigualdades tipo Corte incluidas en el proceso

de solución. Se aplicaron el mismo número de desigualdades válidas tipo Cota y Camino en ambos modelos en el nodo raíz y los valores se indican en la última columna. El tiempo máximo de ejecución para los experimentos mencionados se ha extendido a 3600 segundos.



**Tabla 3.** Comparación de los modelos  $FG-1$ ,  $FG-2$  y  $FG-3$  con instancias de densidad media

Instancia (15, 6, 0.64)						
	F. Min.	F. Max.	Nodos	Tiempo	Gap	# de apariciones
$FG-1$						
Clique	439,26	6 308,64	226 360	68,36	0	0
Corte	439,26	6 308,64	226 360	68,92	0	0
Cota y Camino	439,26	6 308,64	11 099	5,44	0	1
Separador	439,26	6 308,64	206 441	74,27	0	9 462
Triángulo	439,26	63 08,64	191 284	78,39	0	12 455
Sin desigualdades	439,26	6 308,64	226 360	86,42	0	0
Todas las desigualdades	439,26	6 308,64	11 970	9,22	0	2 475
$FG-2$						
Clique	439,26	6 308,64	869 596	268,18	0	0
Corte	439,26	6 308,64	869 596	268,34	0	0
Cota y Camino	439,26	6 308,64	358 990	102,53	0	1
Separador	439,26	6 308,64	839 947	297,94	0	0
Triángulo	439,26	6 308,64	827 249	358,74	0	140 421
Sin desigualdades	439,26	6 308,64	869 596	277,48	0	0
Todas las desigualdades	439,26	6 308,64	448 805	261,22	0	45 038
$FG-3$						
Clique	439,26	6 308,64	4 234 283	1 800,02	24,19	0
Corte	439,26	6 308,64	4 104 240	1 800,02	25,38	0
Cota y Camino	439,26	6 308,64	1 165 916	439,19	0	1
Separador	439,26	6 308,64	2 477 993	1 367,10	0	710 758
Triángulo	439,26	6 308,64	2 957 095	1 800,05	26,06	1 258 109
Sin desigualdades	439,26	6 308,64	5 615 784	1 685,66	0	0
Todas las desigualdades	439,26	6 308,64	1 080 332	743,08	0	996 635
Instancia (20, 5, 0.61)						
	F. Min.	F. Max.	Nodos	Tiempo	Gap	# de apariciones
$FG-1$						
Clique	790,31	10 399,06	2 226 246	825,24	0	68 602
Corte	810,47	16 433,74	60 617	1 830,97	1,41	12 600
Cota y Camino	790,31	10 399,06	91 716	49,47	0	1
Separador	790,31	10 399,06	1 583 107	1 033,24	0	3 673
Triángulo	790,31	10 399,06	1 776 864	1 010,6	0	81 866
Sin desigualdades	790,31	10 399,06	2 544 456	481,96	0	0
Todas las desigualdades	790,31	10 399,06	25 730	837,05	0	17 174
$FG-2$						
Clique	790,31	10 399,06	1 901 739	1 800,18	2,45	47 129
Corte	790,31	10 399,06	217 834	1 803,37	3,04	22 102
Cota y Camino	790,31	10 399,06	518 569	647,68	0	1
Separador	790,31	10 399,06	1 117 591	1 800,02	2,39	74 195
Triángulo	790,31	10 399,06	1 608 429	1 800,03	1,9	216 091
Sin desigualdades	790,31	10 399,06	2 469 429	1 800,01	2,23	0
Todas las desigualdades	790,31	10 399,06	124 675	1 808,95	1,63	101 913
$FG-3$						
Clique	897,15	10 292,22	2 381 897	1 800,2	42,75	2 772 055
Corte	793,17	10 396,2	318 527	1 802,66	31,24	347 662
Cota y Camino	790,31	10 399,06	4 174 153	1 800,01	10,82	1
Separador	790,31	10 399,06	1876 661	1 800,22	29,45	6 216
Triángulo	793,17	10 396,2	1 919 882	1 800,14	40,25	149 537
Sin desigualdades	790,31	10 399,06	3 970 298	1 800,01	32,48	0
Todas las desigualdades	844,12	10 345,25	257 777	1 802,28	24,12	658 621

Al comparar los modelos  $FG-1$  y  $FG-3$  se puede observar un mejor comportamiento del modelo  $FG-1$ , ya que en muchas instancias se obtiene la solución óptima y en el resto de ellas el valor del gap se aproxima a cero. Por otro lado, la tercera formulación reporta altos valores del gap en todas las instancias, con

un valor promedio del 52.52%. Lo anterior nos permite concluir que el modelo  $FG-3$  no presenta un desempeño satisfactorio para resolver el problema sobre instancias de mayor tamaño y es descartado para los siguientes experimentos.

En las Tablas 6 – 7, se reportan experimentos sobre instancias

**Tabla 4.** Comparación de los modelos  $FG-1$ ,  $FG-2$  y  $FG-3$  con instancias de densidad alta

Instancia (15, 6, 0.92)						
	F. Min.	F. Max.	Nodos	Tiempo	Gap	# de apariciones
$FG-1$						
Clique	468,65	10 778,19	1 860	59,70	0	805
Corte	468,65	10 778,19	5 993	5,56	0	0
Cota y Camino	468,65	10 778,19	3 256	4,00	0	1
Separador	468,65	10 778,19	5 993	5,88	0	0
Triángulo	468,65	10 778,19	6 151	10,92	0	50
Sin desigualdades	468,65	10 778,19	5 993	5,87	0	0
Todas las desigualdades	468,65	10 778,19	4 726	159,65	0	1 073
$FG-2$						
Clique	468,65	10 778,19	59 861	1 820,31	0,71	1 970
Corte	468,65	10 778,19	678 187	418,35	0	0
Cota y Camino	468,65	10 778,19	387 297	168,25	0	1
Separador	468,65	10 778,19	687 831	567,73	0	1 227 229
Triángulo	468,65	10 778,19	689 214	961,57	0	125 375
Sin desigualdades	468,65	10 778,19	678 187	358,44	0	0
Todas las desigualdades	468,65	10 778,19	56 157	1 814,98	0,61	174 075
$FG-3$						
Clique	468,65	10 778,19	33 324	1 167,73	0	4 109
Corte	468,65	10 778,19	10 4321	128,51	0	0
Cota y Camino	468,65	10 778,19	102 014	67,04	0	1
Separador	468,65	10 778,19	103 680	138,99	0	7
Triángulo	468,65	10 778,19	125 236	297,90	0	698
Sin desigualdades	468,65	10 778,19	104 321	128,58	0	0
Todas las desigualdades	468,65	10 778,19	49 469	1 824,38	17,35	8 182
Instancia (20, 5, 0.91)						
	F. Min.	F. Max.	Nodos	Tiempo	Gap	# de apariciones
$FG-1$						
Clique	1 144,51	16 433,74	15 457	1 053,04	0	3 857
Corte	1 144,51	16 433,74	22 020	471,30	0	624
Cota y Camino	1 144,51	16 433,74	13 124	30,36	0	1
Separador	1 144,51	16 433,74	25 403	36,21	0	0
Triángulo	1 144,51	16 433,74	26 383	81,61	0	23
Sin desigualdades	1 144,51	16 433,74	25 403	28,67	0	0
Todas las desigualdades	1 144,51	16 433,74	17 857	1 771,98	0	5 562
$FG-2$						
Clique	1 144,51	16 433,74	25 038	1 800,75	2,8	3 976
Corte	1 144,51	16 433,74	83 268	1 813,53	3,2	804
Cota y Camino	1 144,51	16 433,74	1 125 214	1 079,27	0	1
Separador	1 144,51	16 433,74	914 992	1 065,71	0	113
Triángulo	1 144,51	16 433,74	638 454	1 800,08	1,64	1 075
Sin desigualdades	1 144,51	16 433,74	914 367	814,76	0	0
Todas las desigualdades	1 144,51	16 433,74	18 263	1 808,23	3,18	6 740
$FG-3$						
Clique	1 231,77	16 433,74	24 410	1 828,50	40,91	34 267
Corte	1 194,18	16 384,07	95 524	1 803,78	37,13	160 378
Cota y Camino	1 144,51	16 433,74	84 285	109,17	0	1
Separador	1 144,51	16 433,74	141 732	265,34	0	6
Triángulo	1 144,51	16 433,74	137 594	569,98	0	1 283
Sin desigualdades	1 144,51	16 433,74	150 453	236,30	0	0
Todas las desigualdades	1 179,13	16 399,12	18 824	1 815,76	27,22	52 525

con grafos de  $40 \leq n \leq 60$  nodos y con  $6 \leq k \leq 21$  subconjuntos usando exclusivamente el modelo  $FG-1$  bajo dos tipos de estrategias notadas por  $D_1$  y  $D_2$ . En ambas estrategias, se incluyen las desigualdades tipo Cota, Camino y Clique, ya que en los experimentos anteriores disminuyeron la cantidad de nodos explorados y el tiempo de ejecución. Además, la estrategia  $D_1$

admite la desigualdad tipo Triángulo, mientras que  $D_2$  admite la desigualdad tipo Corte. Estas dos estrategias son comparadas con una tercera ( $D_3$ ) en la que no se incluyen desigualdades válidas. Para cada instancia y estrategia, el valor de la función objetivo, el número de nodos explorados, el tiempo de ejecución en segundos, la brecha de optimalidad en porcentaje y el número

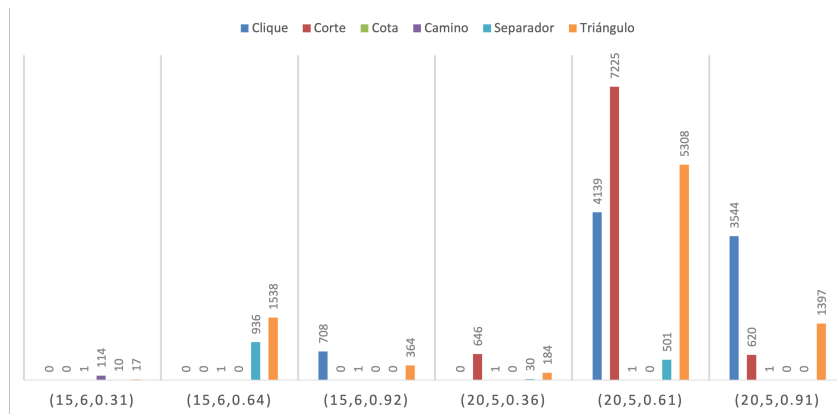


Figura 1. Distribución de apariciones de las desigualdades válidas en el modelo  $FG-1$

Tabla 5. Comparación de los modelos  $FG-1$  y  $FG-3$

Instancia			$FG-1$						$FG-3$						
$n$	$k$	$d$	F. Min.	F. Max.	Nodos	Tiempo	Gap	Clique	F. Min.	Nodos	Tiempo	Gap	Clique	Corte	Cota+Camino
22	5	0,93	1 438,39	20 404,13	20 498	1 940,86	0	4 569	2 247,68	28 960	3 712,11	66,68	10 299	2 710	1
		0,33	1 037,18	6 801,79	49 064	42,37	0	0	1 037,18	310 550	3 608,74	15,95	0	107 584	1
24	8	0,93	685,08	24 503,21	21 560	3 600,43	0,26	4 994	1 421,02	15 376	3 676,21	74,36	10 381	6 496	1
		0,36	872,23	8 771,38	760 954	1 225,68	0	0	876,50	178 079	3 608,90	33,14	0	67 676	97
26	8	0,9	749,01	24 709,58	22 086	3 645,93	0,77	6 113	1 986,66	15 327	3 667,67	81,21	4 177	2 756	1
		0,37	691,64	10 634,22	27 546	75,12	0	0	873,40	117 183	3 603,64	34,48	0	45 222	1
28	10	0,87	628,71	33 282,73	14 960	3654,93	0,38	3 628	903,44	11 786	3 623,22	57,44	5 924	0	1
		0,35	802,78	11 968,42	900 187	3600,11	0,11	0	843,35	333 383	3600,44	31,96	0	0	101
30	10	0,89	806,05	37 365,48	14 610	3 683,62	0,35	4 067	2 466,39	8 814	3 608,25	80,48	3 704	2 180	1
		0,31	1 075,49	13 385,55	215 847	715,35	0	0	1 507,00	79 030	3 614,71	49,54	0	15 030	101

de desigualdades de cada tipo incluidas en el proceso de solución son reportadas.

De las 20 instancias presentadas, dos de ellas no se resolvieron mediante ninguna estrategia. En 15 instancias sucedió que alguna de las estrategias  $D_1$  y  $D_2$  reportó un mejor comportamiento que el caso sin desigualdades, ya que encontró una solución con menor valor en la función objetivo. En las tres instancias faltantes, el caso sin desigualdades obtuvo una mejor solución, pero en dos de ellos el valor del gap es mayor. Cabe notar que en la mayoría de las instancias, las estrategias  $D_1$  y  $D_2$  obtuvieron sus soluciones con menor cantidad de nodos explorados respecto al caso sin desigualdades. En los reportes anteriores, se evidencia que en instancias grandes las desigualdades válidas propuestas en el presente trabajo reducen el valor del gap y la 18/01 de nodos explorados. Comparando las estrategias  $D_1$  y  $D_2$ , se muestra que  $D_1$  en 12 ocasiones encontró una mejor solución con un menor valor del gap.

### 5. CONCLUSIONES

En este artículo se estudia el problema de equiparticionamiento de grafos en un número fijo de componentes conexas. El problema consiste en particionar un grafo no dirigido con costos sobre las aristas en un número fijo de componentes conexas, tal que el número de nodos en cada componente difiera en a lo más una unidad. El objetivo busca minimizar el costo total de las aristas con nodos finales en la misma componente. Tres modelos de

programación lineal entera junto con varias familias de desigualdades válidas e incluidas en un algoritmo exacto tipo Branch & Cut fueron propuestos para resolver el problema. La efectividad de las desigualdades válidas fueron probadas computacionalmente al obtener una significativa reducción en el número de nodos explorados, tiempo de ejecución y brecha de optimalidad. Al comparar las tres formulaciones se pudo concluir que la formulación  $FG-1$  es la que se comporta de mejor forma. Al igual que otros algoritmos exactos para el problema de particionamiento de grafos reportados previamente en la literatura, nuestro método permite resolver instancias de hasta 60 nodos.

Los resultados obtenidos muestran la funcionalidad del modelo  $FG-1$  ya que se alcanzan valores cercanos al óptimo usando las distintas estrategias de solución  $D_1$  y  $D_2$ . La brecha de optimalidad en promedio es igual al 3% y en algunos casos menor que 1%. Esto implica que se dispone de una formulación estable y un algoritmo exacto tipo Branch & Cut adecuado para resolver el problema de equiparticionamiento de grafos generales en componentes conexas.

Como trabajo futuro se podría pensar en mejorar las rutinas de separación y explorar otros tipos de desigualdades válidas para las formulaciones  $FG-2$  y  $FG-3$ . Finalmente, este trabajo puede ser extendido a problemas de equiparticionamiento considerando pesos sobre los nodos, de modo que no solo se busque un equiparticionamiento, si no también se obtenga un balance respecto al peso impuesto sobre los nodos.

Tabla 6. Modelo  $FG-1$  para grafos con  $n \in \{40, 45, 50\}$ 

Instancia	Estrategia	F. Obj.	Nodos	Tiempo	Gap	Cota	Camino	Clique	Corte	Triángulo
(40, 6, 0,77)	$D_1$	3 037,34	22 969	3 687,50	3,20	1	0	4 124	0	1 446
	$D_2$	4 584,80	2 986	3 661,15	6,00	1	0	1 909	569	0
	$D_3$	3 405,90	468 005	3 600,26	3,73	0	0	0	0	0
(40, 6, 0,15)	$D_1$	2 064,95	1 212 407	3 600,09	5,20	1	0	0	0	280
	$D_2$	2 328,20	17 800	3 730,05	7,98	1	0	0	1 581	0
	$D_3$	2 057,14	3 184 309	3 600,07	19,95	0	0	0	0	0
(40, 14, 0,8)	$D_1$	1 075,73	4 402	3 601,80	0,86	1	0	0	0	490
	$D_2$	825,95	4 776	3 600,42	0,44	1	0	0	0	0
	$D_3$	1 171,13	5 300	3 600,35	1,90	0	0	0	0	0
(40, 14, 0,16)	$D_1$	1 593,43	434 658	3 600,26	2,65	1	2 996	0	0	382
	$D_2$	1 593,43	694 265	3 600,49	2,46	1	2 996	0	0	0
	$D_3$	1 720,69	531 709	3 600,09	13,54	0	0	0	0	0
(45, 7, 0,78)	$D_1$	4 134,76	3 967	3 608,34	4,23	1	0	3 016	0	896
	$D_2$	4 942,46	1 327	3 612,59	5,52	1	0	1 111	252	0
	$D_3$	3 581,96	74 674	3 600,38	3,88	0	0	0	0	0
(45, 7, 0,16)	$D_1$	2 237,83	702 371	3 600,23	2,40	1	0	0	0	308
	$D_2$	-	13 286	3 600,00	-	1	0	0	-	0
	$D_3$	2 504,33	1 103 065	3 600,20	15,68	0	0	0	0	0
(45, 16, 0,77)	$D_1$	1 139,45	1 396	3 600,79	0,77	1	0	0	0	135
	$D_2$	1 159,47	1 491	3 601,02	0,78	1	0	0	0	0
	$D_3$	1 161,23	7 675	3 600,43	1,43	0	0	0	0	0
(45, 16, 0,15)	$D_1$	1 291,68	194 871	3 600,43	0,57	1	4 832	0	0	260
	$D_2$	1 291,68	168 869	3 600,52	0,64	1	4 832	0	0	0
	$D_3$	1 626,83	63 712	3 600,08	12,2	0	0	0	0	0
(50, 8, 0,78)	$D_1$	3 834,48	10 388	3 603,67	3,01	1	0	4 365	0	579
	$D_2$	4 968,80	1 858	3 723,45	4,37	1	0	695	174	0
	$D_3$	3 689,61	66 473	3 600,30	3,42	0	0	0	0	0
(50, 8, 0,15)	$D_1$	2 647,99	427 477	3 600,33	3,41	1	0	0	0	454
	$D_2$	-	11 645	3 600,00	-	1	0	0	-	0
	$D_3$	2 863,69	1 302 728	3 600,09	16,78	0	0	0	0	0
(50, 17, 0,79)	$D_1$	1 028,30	1 522	3 600,93	0,49	1	0	0	0	133
	$D_2$	1 068,79	1 387	3 600,50	0,54	1	0	0	0	0
	$D_3$	2 927,52	4 158	3 600,49	2,99	0	0	0	0	0
(50, 17, 0,28)	$D_1$	2 922,00	4 013	3 600,34	5,62	1	102	0	0	256
	$D_2$	1 316,06	5 581	3 600,29	0,80	1	102	0	0	0
	$D_3$	-	4 116	3 600,00	-	0	0	0	0	0

El uso del caracter - implica que en la optimización no se encontró ninguna solución factible.

Tabla 7. Modelo  $FG-1$  para grafos con  $n \in \{55, 60\}$ 

$FG-1$										
Instancia	Estrategia	F. Obj.	Nodos	Tiempo	Gap	Cota	Camino	Clique	Corte	Triángulo
(55, 9, 0,8)	$D_1$	4 595,42	4 956	3 601,55	2,86	1	0	2 500	0	331
	$D_2$	5 965,41	1 453	3 628,28	4,03	1	0	670	162	0
	$D_3$	5 741,28	10 130	3 600,14	4,49	0	0	0	0	0
(55, 9, 0,15)	$D_1$	2 912,69	241 584	3 600,24	5,13	1	0	0	0	768
	$D_2$	-	9 500	3 600,00	-	1	0	0	-	0
	$D_3$	-	640 197	3 600,00	-	0	0	0	0	0
(55, 19, 0,81)	$D_1$	1 650,44	466	3 600,77	0,86	1	0	0	0	74
	$D_2$	1 609,95	479	3 600,81	0,84	1	0	0	0	0
	$D_3$	1 870,52	4 772	3 600,49	1,47	0	0	0	0	0
(55, 19, 0,27)	$D_1$	-	2 177	3 600,00	-	1	8	0	0	-
	$D_2$	-	2 456	3 600,00	-	1	8	0	-	0
	$D_3$	-	3 430	3 600,00	-	0	0	0	0	0
(60, 10, 0,8)	$D_1$	6 239,52	2 960	3 608,13	3,49	1	0	1 697	0	161
	$D_2$	6 892,72	1 161	3 701,93	3,99	1	0	462	90	0
	$D_3$	7 427,40	4 838	3 600,70	4,82	0	0	0	0	0
(60, 10, 0,16)	$D_1$	3 622,15	89 836	3 601,04	7,11	1	0	0	0	1 123
	$D_2$	3 189,67	6 695	3 668,83	5,18	1	0	0	269	0
	$D_3$	4 190,93	151 338	3 600,09	18,26	0	0	0	0	0
(60, 21, 0,81)	$D_1$	1 801,44	181	3 602,53	0,94	1	0	0	0	22
	$D_2$	2 115,23	167	3 601,44	1,18	1	0	0	0	0
	$D_3$	2 029,76	2824	3 600,42	1,47	0	0	0	0	0
(60, 21, 0,28)	$D_1$	-	1 800	3 600,00	-	1	13	0	0	-
	$D_2$	-	2 382	3 600,00	-	1	13	0	-	0
	$D_3$	-	3 338	3 600,00	-	0	0	0	0	0

El uso del caracter - implica que en la optimización no se encontró ninguna solución factible.

