

Dibujos jerárquicos y clasificaciones

Ana Paulina Figueroa

Departamento Académico de Matemáticas, ITAM
apaulinafg@gmail.com y

Adjani Gama Dessavre

Departamento Académico de Matemáticas, ITAM
adjani.gd@gmail.com

1. Introducción

Los problemas de clasificación aparecen en diversos problemas cotidianos, por ejemplo, si pensamos en la asociación de tenistas profesionales (ATP), aunque no todos los jugadores llegan a jugar entre sí en los torneos del año, se busca tener una clasificación o *ranking*. Imaginemos que Murray le gana a todos; Djokovic le ha ganado a Raonic, Raonic le gana a Federer y Federer recuerda sus viejos tiempos ganándole a Djokovic. ¿Por qué Murray tiene que ser el número uno? Podemos modelar esta situación con una digráfica¹ como se muestra en la figura 1.

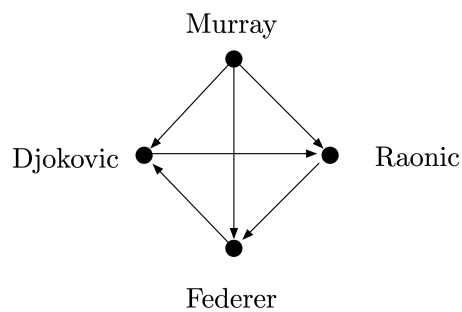


Figura 1.

¹Dado un conjunto V no vacío, podemos modelar a una relación entre los elementos de V con flechas que inician en un elemento u y terminan en un elemento v cuando u esté relacionado con v . A este objeto se le conoce como digráfica. Formalmente una digráfica D está formada por un conjunto V de elementos llamados vértices y un conjunto A de parejas ordenadas de vértices $(u, v) \in A$ a los que llamamos flechas (u está relacionado con v si $(u, v) \in A$).

Pensemos en una digráfica en la que los vértices son los jugadores, y como los jugadores juegan diversos partidos en el año para competir en los torneos inscritos en la ATP, habrá una flecha de un jugador A a otro B si A ha obtenido más victorias sobre B . Si analizamos la gráfica dirigida de la figura 1, lo que nos impide ordenar a los cuatro jugadores en una clasificación que muestre quién es el mejor jugador, es el ciclo dirigido formado por Djokovic, Federer y Raonic. Si quitáramos alguna de las tres flechas de ese triángulo dirigido, por ejemplo, la del partido entre Federer y Djokovic, los jugadores quedarían totalmente ordenados con el orden Murray, Djokovic, Raonic y Federer. Entonces, buscar una clasificación de los jugadores que refleje mejor los resultados del año es análogo a quitar algunas flechas de D , idealmente pocas, para encontrar una subdigráfica S de D que sea acíclica². Notemos que si encontramos una subdigráfica acíclica S de tamaño máximo³ de D , podemos pensar que los vértices de D están ordenados v_1, v_2, \dots, v_n de forma que si $(v_i, v_j) \in A(S)$ entonces $i < j$ y si $(v_i, v_j) \in A(D) \setminus A(S)$ entonces $j \leq i$ (figura 2).

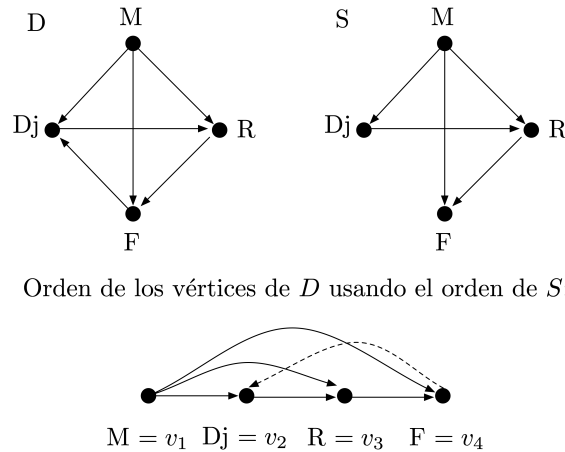


Figura 2. A la derecha se muestra la digráfica D que describe las victorias de los jugadores, donde cada jugador está representado por la primera o primeras letras de su nombre. A la izquierda, la subdigráfica acíclica S de D creada al quitar la flecha (F, Dj) . Abajo de estas se observa el orden de los vértices que surge de S , la flecha de regreso es la marcada en líneas punteadas.

Una forma ingenua de encontrar tal digráfica S , es ordenar a los vértices de todas las formas posibles y quitar las aristas que evitan que

² Una digráfica S es acíclica si no tiene ciclos dirigidos o equivalentemente si sus vértices se pueden ordenar de forma que cada que uv es una flecha de S implique que u está antes que v en el orden de los vértices.

³Con el máximo número de aristas posible.

la gráfica sea acíclica. Sin embargo, este método se vuelve menos viable entre más jugadores tengamos, simplemente con diez jugadores hay más de tres millones de formas de ordenarlos.

En este trabajo escribiremos sobre los conjuntos de **flechas de retroalimentación de una digráfica** y los llamaremos **FAS** por sus siglas en inglés: *feedback arc set*. Un subconjunto $S \subseteq A(D)$ es un FAS si $D \setminus S$ es una digráfica acíclica. El problema de encontrar un FAS mínimo surgió en los años 60 a partir de un problema análogo, el problema de vértices de retroalimentación (*Feedback Vertex Set*) FVS, en el que se desea encontrar una digráfica acíclica a partir de la eliminación de vértices. El FVS se originó en el área de combinatoria [18], en particular, en el diseño de circuitos [11]. El FAS mínimo fue planteado por primera vez por Ruyon y generó gran interés en diversas áreas como la de investigación de operaciones cuando Tucker [23], en 1960, dio una formulación del problema para programación entera. En 1972, apareció como uno de los 21 problemas de Karp [17] NP-completos ⁴.

En las siguientes secciones hablaremos de las equivalencias entre las distintas definiciones para el FAS mínimo; expondremos algunos resultados y conjeturas interesantes de este invariante; finalmente, dado que desde sus orígenes el problema del FAS mínimo ha tenido diferentes aplicaciones, en la última sección de este trabajo describiremos dos de ellas: los dibujos de digráficas jerárquicas, y la comparación de sistemas de traducción en el campo de la lingüística computacional.

2. Definiciones equivalentes del FAS mínimo y una conjetura famosa

Tomaremos como definición básica de un FAS a un conjunto S de flechas de D que al borrarlo de D produce una digráfica acíclica. Un FAS mínimo (MFAS) es un FAS de cardinalidad mínima, esta cardinalidad será denotada como $\tau_1(D)$.

Existen distintas definiciones equivalentes para el FAS que están basadas en las siguientes definiciones. Si D es una digráfica y $\pi = v_1, v_2, \dots, v_n$ es un orden de sus vértices, entonces $(v_i, v_j) \in A(D)$ es una **flecha de regreso** de π si $i > j$ (véase la figura 3). Se dice que π es τ -óptimo si el número de flechas de regreso es el menor entre todos los órdenes de vértices de D . En el siguiente teorema se demuestra que hay una equivalencia entre el **orden τ -óptimo**, el FAS mínimo y la cardinalidad de la máxima subdigráfica acíclica de D .

⁴Intuitivamente un problema es NP completo si es un problema que no se puede resolver en un tiempo razonable con el poder de cálculo que se tiene hasta el momento, aunque es fácil verificar si un objeto es una solución del mismo. Para encontrar una definición formal del problema referimos al lector al capítulo 18 de Bang-Jensen [3].

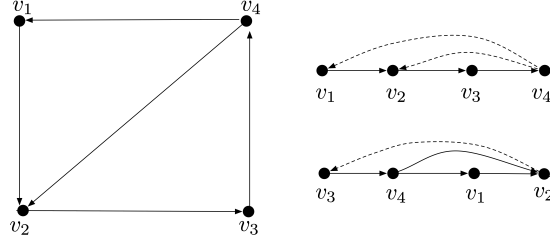


Figura 3. Los vértices de la digráfica de la izquierda pueden ser ordenados de diferentes maneras. Se muestran dos formas, donde la de abajo es τ -óptima. Las flechas de regreso se muestran en líneas punteadas.

Teorema 2.1. *Sea D una digráfica y $S \subseteq A(D)$, entonces las siguientes afirmaciones son equivalentes:*

- S es un FAS de tamaño mínimo.*
- $D \setminus S$ es una subdigráfica acíclica de D de tamaño máximo.*
- Existe un orden π de $V(D)$ que es τ -óptimo tal que S es el conjunto de flechas de regreso de π .*

Demostración. Sea S un FAS de tamaño mínimo. Supongamos, por contradicción, que existe D' subdigráfica acíclica de D con más flechas que $D \setminus S$, es decir, $|A(D \setminus S)| < |A(D')|$. Podemos escribir $D' = D \setminus S'$ con $S' \subseteq A(D)$. Notemos que S' es un FAS y que $|A(D \setminus S)| < |A(D \setminus S')|$, pero esto implica que $|S'| < |S|$, lo cual no es posible porque S es un FAS de tamaño mínimo. Y por lo tanto, sea S es un FAS de tamaño mínimo, $D \setminus S$ es una subdigráfica acíclica de D de tamaño máximo.

Por otra parte, si $D \setminus S$ una subdigráfica acíclica de D de tamaño máximo. Como $D \setminus S$ es acíclica existe π un orden v_1, v_2, \dots, v_n de los vértices de D , tal que si $(v_i, v_j) \in A(D \setminus S)$, entonces $i \leq j$. Sea P el conjunto de flechas de regreso de π en D , entonces $P \subseteq S$. Por otro lado, notemos que $D \setminus P$ es acíclica porque P son las flechas de regreso de π . Si existe $s \in S \setminus P$, entonces tenemos que $|S| > |P|$, lo que implica que $|A(D \setminus S)| < |A(D \setminus P)|$, lo que no es posible porque $D \setminus S$ es una subdigráfica acíclica de D con el máximo número de flechas. De esta manera, $S \subseteq P$ y por lo anterior $P = S$. Por lo tanto S es el conjunto de flechas de regreso de π . Ahora mostraremos que S es τ -óptimo, es decir, el conjunto de flechas de regreso de π es el de menor tamaño entre todos los órdenes de los vértices de $V(D)$. Supongamos, por contradicción que existe π' un orden de $V(D)$ donde su conjunto de flechas de regreso, S' , tiene menor tamaño que el de π ($|S'| < |S|$). Esto significa que $D \setminus S'$ es una subdigráfica acíclica y además, $|A(D \setminus S')| > |A(D \setminus S)|$, lo que contradice que $D \setminus S$ es una subdigráfica acíclica de D de tamaño máximo. Por lo tanto, $|S| \leq |S'|$ y S es τ -óptimo.

Por último sea π un orden de $V(D)$ τ -óptimo con S el conjunto de flechas de regreso de π , demostraremos que S es un FAS mínimo. Como S es el conjunto de flechas de regreso de π , eso implica que $D \setminus S$ es acíclica y por lo tanto S es un FAS de D . Para demostrar que es mínimo, supongamos por contradicción, que existe S' un FAS de menor tamaño que S . Esto implica que $D \setminus S'$ es una subdigráfica acíclica, por lo tanto existe un orden π' de los vértices de D , v_1, v_2, \dots, v_n y como ya demostramos que a implica c , π' es un orden de los vértices de D en el que S' es el conjunto de flechas de regreso de π' , lo que implica que S no es τ -óptimo, lo que contradice la hipótesis. Por lo tanto, las afirmaciones a, b y c son equivalentes. \square

Una de las famosas conjeturas en digráficas relacionadas con el FAS mínimo fué propuesta por Chudnovsky, Seymour y Sullivan en 2008 [8]. Dicha conjetura está definida para gráficas orientadas⁵ y relaciona la cardinalidad de un FAS mínimo, $\tau_1(D)$, con la cantidad de aristas ausentes de la gráfica subyacente de D ⁶, denotadas como $\overline{m}(D)$ ⁷.

Conjetura 2.1. *Si una gráfica orientada D no tiene ciclos dirigidos de longitud 3, entonces $\tau_1(D) \leq \frac{1}{2}\overline{m}(D)$.*

Estos mismos autores probaron la conjetura para dos casos:

1. Cuando se puede hacer una partición de $V(D)$ en dos subconjuntos Y y Z tal que $D\langle Y \rangle$ ⁸ y $D\langle Z \rangle$ sean torneos⁹.
2. Cuando se pueden arreglar los vértices de D en un círculo tal que si u, v, w en $V(D)$ están ordenados según las manecillas del reloj y (u, w) está en $A(D)$, entonces (u, v) y (v, w) también están en $A(D)$.

Aunque la conjetura sigue abierta, pudieron probar que si una gráfica orientada D no tiene ciclos dirigidos de longitud 3, entonces $\tau_1(D) \leq \overline{m}(D)$. Expondremos la demostración de este resultado utilizando los siguientes lemas y definiciones. Sea $P_2(D)$ el conjunto de las trayectorias inducidas¹⁰ de longitud dos en D . Dado un vértice $v \in V(D)$ definimos los siguientes subconjuntos de $P_2(D)$: A_v está formado por

⁵Una gráfica orientada es una digráfica en la que $(u, v) \in A(D)$ implica que $(v, u) \notin A(D)$ para cualesquiera $u, v \in V(D)$.

⁶La gráfica subyacente de una digráfica D , \underline{D} , es la gráfica que se obtiene de D reemplazando cada flecha por una arista, es decir, cambiando las parejas ordenadas por parejas no ordenadas.

⁷Para una digráfica D , el número de aristas ausentes de su gráfica subyacente, $\overline{m}(D)$, es el número de parejas (no ordenadas) que no son aristas de la gráfica subyacente de D .

⁸ $D\langle X \rangle$ es la digráfica con vértices X y las flechas son exactamente las que existían en D entre los vértices de X .

⁹Un torneo es una digráfica que tiene exactamente una flecha por cada par de vértices.

¹⁰ u, v, w es una trayectoria inducida de D de longitud dos si $(u, v), (v, w) \in A(D)$ pero $(u, w), (w, u) \notin A(D)$

las que comienzan en v y B_v por las que tienen a v como vértice intermedio. Denotemos por $f(v)$ y $g(v)$ a las cardinalidades de A_v y B_v , respectivamente.

Lema 2.1. *Sea D una digráfica, entonces*

$$\sum_{v \in V(D)} f(v) = \sum_{v \in V(D)} g(v).$$

Demostración. Podemos construir dos particiones diferentes del conjunto $P_2(D)$. En la primera $P_2(D)$ la podemos ver como la unión disjunta de clases A_v para cada $v \in V(D)$. Pero también podemos partir a $P_2(D)$ con las clases B_v , $v \in V(D)$. Entonces el lema se sigue de que

$$\sum_{v \in V(D)} f(v) = |P_2(D)| = \sum_{v \in V(D)} g(v). \quad \square$$

Lema 2.2. *Sea D una digráfica, existe un vértice v tal que $f(v) \leq g(v)$.*

Demostración. Supongamos por contradicción que para todo vértice v en $V(D)$, $f(v) > g(v)$, esto implicaría que $\sum_{v \in V(D)} f(v) > \sum_{v \in V(D)} g(v)$ lo cual, por el lema anterior, es una contradicción. \square

Teorema 2.2. *Si una gráfica orientada D no tiene ciclos dirigidos de longitud 3, entonces $\tau_1(D) \leq \bar{m}(D)$.*

Demostración. Se probará por inducción sobre n , el número de vértices de D . Para $n \leq 2$ es claro que se cumple el resultado, pues $\tau_1(D) = 0$ y $\bar{m}(D) \geq 0$. Por el lema anterior, existe un vértice v tal que $f(v) \leq g(v)$. Ahora, dividiremos a los vértices de D en las siguientes subdigráficas $D_2 = D(N^+(v))$, $D_1 = D(V(D) \setminus N^+[v])$ y $\{v\}$ ¹¹, como en la figura 4. Mostraremos que por cada flecha que pertenece al FAS mínimo, existe al menos una arista que no pertenece a \underline{D} .

Recordemos que $g(v)$ es el número de trayectorias inducidas de longitud 2 que tienen v como vértice intermedio. Observemos que toda pareja que no es flecha de D_1 tampoco es arista de \underline{D} , lo mismo sucede con D_2 y finalmente, por cada trayectoria inducida $p = a, v, b$ en D , (a, b) no es una flecha de \underline{D} ya que p es inducida. Entonces,

$$\bar{m}(D) \geq \bar{m}(D_1) + \bar{m}(D_2) + g(v).$$

Por otro lado, tenemos que por la hipótesis de inducción $\tau_1(D_1) \leq \bar{m}(D_1)$ y $\tau_1(D_2) \leq \bar{m}(D_2)$. Llamaremos X_i a el FAS mínimo de D_i . Lo siguiente será fijarnos en el conjunto $X_3 = \{ab \in A(D) \mid a \in N^+(v), b \notin N^+(v)\}$ que son todas las flechas que van de D_2 a D_1 . Como no tenemos

¹¹ La exvecindad de v es: $N^+(v) = \{x \in V(D) \mid (v, x) \in A(D)\}$. La invecindad de v es: $N^-(v) = \{x \in V(D) \mid (x, v) \in A(D)\}$. Entonces, la exvecindad cerrada y la invecindad cerrada de v se definen como $N^+[v] = N^+(v) \cup \{v\}$ y $N^-[v] = N^-(v) \cup \{v\}$ respectivamente.

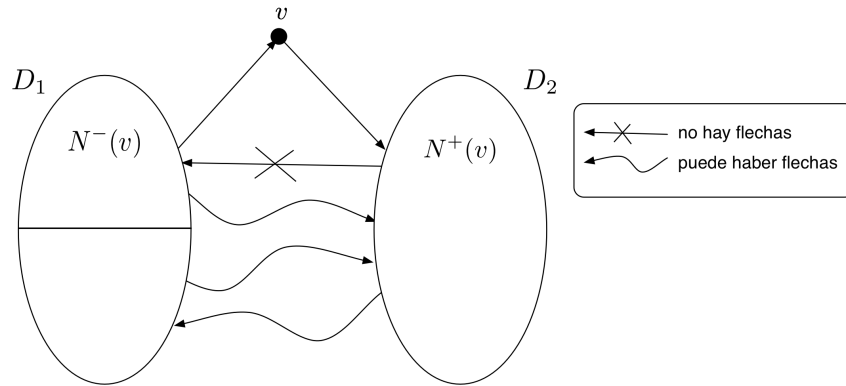


Figura 4.

ciclos dirigidos de tamaño 3, ninguna flecha en X_3 es incidente¹² a ningún vértice en $N^-(v)$ (esto se puede visualizar en la figura 4), por lo tanto, $|X_3| = f(v)$. Además, si definimos $X = X_1 \cup X_2 \cup X_3$, $D - X$ es acíclica, por lo que X es un FAS de D . Con todo lo anterior tenemos $\tau_1 \leq |X| = |X_1| + |X_2| + |X_3| = \tau_1(D_1) + \tau_1(D_2) + f(v) \leq \bar{m}(D_1) + \bar{m}(D_2) + f(v) \leq \bar{m}(D)$. \square

Es importante no solo encontrar el tamaño del MFAS, sino también estudiar la estructura que tiene este en diversos tipos de digráficas. De hecho, no se ha resuelto el problema de caracterizar cuáles son las digráficas que cumplen que su MFAS es isomorfo a una digráfica dada. En particular, esto sería muy útil para los torneos, ya que para estos Isaak y Narayan [16] dan condiciones necesarias y suficientes para que un FAS mínimo sea un torneo acíclico.

3. ¿Cómo encontrar un MFAS?

Como se mencionó antes, Karp, en 1971, fue el primero en demostrar que este problema es NP-completo. Aún más, se ha encontrado que el problema es NP-completo para familias particulares de digráficas como: las digráficas de líneas, lo cual fue demostrado por Gavril en 1977; los torneos, lo cual fue conjeturado por Bang-Jensen y Thomassen [2] en 1992, y posteriormente, alrededor del 2006, probado de manera independiente por distintos investigadores como Alon [1], Thomassé y Yeó [7] y Conitzer [9].

¹²Una flecha es incidente a un vértice v si v es su vértice inicial o final.

Esto lleva a la pregunta que ocurre en muchos problemas difíciles de resolver que tienen aplicaciones ¿Cómo obtener una solución buena que no nos lleve una eternidad calcularla? Una acercamiento deseable es utilizar algoritmos de aproximación. Estos algoritmos no necesariamente encuentran la solución óptima, pero garantizan cierta distancia entre la solución encontrada y la óptima. En 1995, Seymour [20] demostró que para el problema del FAS existe un algoritmo de aproximación de orden $O(\log n \log \log n)$. Algunos de estos algoritmos utilizan formulación de programación lineal entera del problema:

$$\begin{aligned} \text{Min} \quad & \sum_{e \in A} c(e) \cdot d(e) \\ \text{sujeto a} \quad & \sum_{e \in C} d(e) \geq 1 \text{ para todo ciclo } C \\ & d(e) \in \{0, 1\} \text{ para toda flecha } e \in A \end{aligned}$$

En esta formulación $c(\cdot)$ es la función de costos, con la cual se asigna un costo o peso a cada flecha. Cuando se estudia el problema sin pesos, tenemos que $c(e) = 1$ para toda flecha. La variable $d(e)$ es la que se asigna a cada flecha y podrá tomar valor 1 o 0 si la flecha está o no en el FAS. Al resolver el problema, para aquellas flechas que se encuentren en el FAS, $d(e)$ toma el valor 1. La primera restricción implica que el FAS debe tener al menos una flecha de todos los ciclos y minimizar la función objetivo garantiza que el FAS es mínimo.

Even, Naor, Schieber y Sudan [10] describen un algoritmo de aproximación para una generalización del problema del FAS con pesos en las flechas. Este algoritmo utiliza una reducción al problema del multicorte dirigido¹³ (*directed multicut*) y tiene orden de aproximación $O(\min\{\log \tau^* \log \log \tau^*, \log n \log \log n\})$ donde τ^* es el costo de una solución a la relajación del problema de programación entera descrito anteriormente.

A pesar de que los algoritmos de aproximación son muy importantes y útiles debido a su proximidad con la solución óptima, en algunos casos son muy complejos o no son lo suficientemente eficientes. Por lo anterior, se han buscado distintos enfoques para abordar este problema como los métodos heurísticos. Estos son métodos que tratan de descubrir una solución factible cercana a la óptima de forma eficiente, aunque, en realidad, no garantizan la calidad de esta solución [15].

Para el problema del FAS mínimo se han desarrollado diversas heurísticas como heurísticas de ordenamiento y heurísticas híbridas, algunas

¹³Si se tiene una red $N = (V, A, c(\cdot), \{s_i, t_i\}_{i=1 \dots k})$, donde (V, A) es una digráfica, $c(\cdot)$ es una función de capacidades definida en las flechas y $\{s_i, t_i\}_{i=1 \dots k}$ son parejas de vértices, un multicorte es un subconjunto $F \subseteq A$ tal que para cada pareja (s_i, t_i) , toda trayectoria de s_i a t_i contiene al menos una flecha de F .

de las cuales exploramos en [13]. Un ejemplo es *Sifting*, el cual es un método iterativo que comienza con un orden aleatorio de los vértices $\pi = v_1, v_2, \dots, v_n$. En la iteración i remueve el vértice v_i y lo reinserta en la mejor posición posible, donde la mejor posición es aquella en la que v_i induce el menor número de flechas de regreso¹⁴. En caso de que haya más de una posición óptima, inserta v_i en la que se encuentre más a la izquierda. A continuación mostramos un ejemplo de este método utilizando la digráfica de la figura 5 y con todas sus iteraciones mostradas en el cuadro 1. Además, en la figura 6 se muestra el proceso de una iteración del método.

Ejemplo 3.1. *Supongamos que el orden inicial es $\pi = v_1, v_2, v_3, v_4$.*

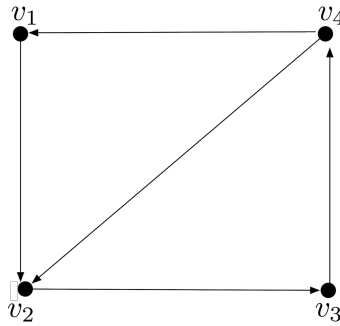


Figura 5.

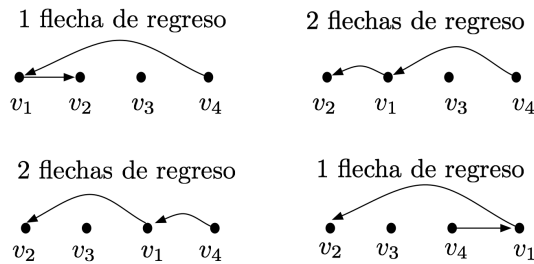


Figura 6. Ejemplo de una iteración del método heurístico Sifting.

No siempre resolver el problema del FAS mínimo es difícil. Existen subfamilias de digráficas, como las digráficas planas¹⁵, en las que el

¹⁴ Las flechas de regreso inducidas por v_i son aquellas flechas de regreso que tienen a v_i como vértice inicial o final.

¹⁵ Una digráfica plana es una digráfica que puede ser dibujada en el plano sin que sus flechas se crucen.

Iteración	órdenes posibles	Flechas de regreso de v_i	orden elegido
1	v_1, v_2, v_3, v_4	1	$\sigma = v_1, v_2, v_3, v_4$
	v_2, v_1, v_3, v_4	2	
	v_2, v_3, v_1, v_4	2	
	v_2, v_3, v_4, v_1	1	
2	v_2, v_1, v_3, v_4	2	$\sigma = v_1, v_2, v_3, v_4$
	v_1, v_2, v_3, v_4	1	
	v_1, v_3, v_2, v_4	2	
	v_1, v_3, v_4, v_2	1	
3	v_3, v_1, v_2, v_4	1	$\sigma = v_1, v_2, v_3, v_4$
	v_1, v_3, v_2, v_4	1	
	v_1, v_2, v_3, v_4	0	
	v_1, v_2, v_4, v_3	0	
4	v_4, v_1, v_2, v_3	1	$\sigma = v_4, v_1, v_2, v_3$
	v_1, v_4, v_2, v_3	2	
	v_1, v_2, v_4, v_3	3	
	v_1, v_2, v_3, v_4	2	

Cuadro 1. Con lo anterior se obtiene como resultado el orden v_4, v_1, v_2, v_3 , y un FAS de tamaño 1, $\{(v_3, v_4)\}$. En la figura 6 podemos observar cómo funciona la primera iteración del ejemplo 3.1 utilizando el método Sifting.

problema del FAS mínimo puede resolverse en tiempo polinomial. En particular, en el mundo de digráficas planas, se puede transformar cada instancia del problema del MFAS a una en la que se debe de encontrar el cubrimiento de dicortes mínimo del dual de la digráfica, para el cual existe un algoritmo que lo resuelve en tiempo polinomial [3].

4. Aplicaciones

El problema del FAS es un problema que se ha estudiado en diversos campos por más de 50 años y se ha aplicado en distintas áreas. A continuación describiremos dos de sus aplicaciones: la primera, en problemas de clasificación, donde se muestra un ejemplo en la lingüística computacional; la segunda, en el dibujo de digráficas en capas.

Una de las principales aplicaciones del problema del FAS mínimo es en los problemas de clasificación (*ranking problems*). En un problema de clasificación, se tiene un conjunto S con n elementos, los cuales se quiere clasificar según una aptitud $a(v)$. El problema surge cuando, al ordenar los objetos según su aptitud, la clasificación obtenida no es consistente. Un ejemplo, es el problema de los tenista expuesto anteriormente.

Para modelar estos problemas, se construye una digráfica D , donde los vértices son los objetos del conjunto S y cada vértice v tiene una aptitud $a(v) \in \mathbb{R}$, se tiene una flecha uv si $a(u) > a(v)$. Tener una

clasificación consistente es equivalente a que D sea una digráfica acíclica (Véase el ejemplo en la figura 7).

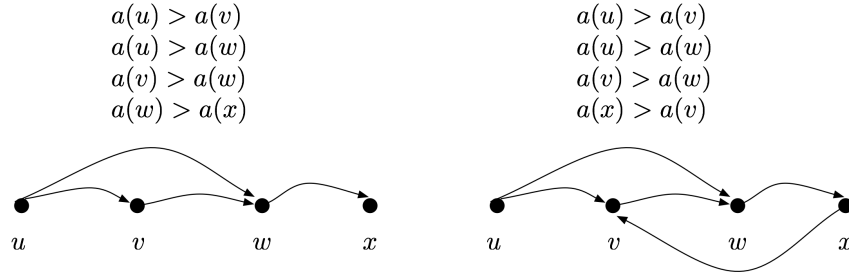


Figura 7. A la izquierda se muestra una situación donde existe una clasificación consistente como se verifica en la digráfica inferior que es acíclica. Por otro lado, en el lado derecho tenemos un situación donde no existe un ranking consistente, lo que se evidencia en la digráfica inferior que contiene un ciclo.

Un ejemplo de problema de clasificación muy peculiar se encuentra en el campo de la lingüística computacional y lo expone Adam Lopez [19] para el 7th *Workshop on Machine Translation (WMT)* [21] que se llevó a cabo en Montreal, Quebec, Canadá en el 2012. En este se utiliza evaluación humana para clasificar diferentes sistemas de traducción automática.

Para el estudio utiliza los datos publicados por la WMT 2010 [5] y la WMT 2011 [6]. En estas se tienen los datos de sistemas de traducción Inglés-Checo, Checo-Inglés, Inglés-Francés, Francés-Inglés, Inglés-Alemán, Alemán-Inglés, Inglés-Español, Español-Inglés. En cada una se comparan entre 4 y 22 sistemas de traducción, para los que se tienen entre 4000 y 13000 comparaciones por pareja entre los distintos sistemas.

La forma en la que se realizan las evaluaciones es dando a especialistas una oración seguida de cinco traducciones, cada una realizada por un sistema de traducción distinto, y se les pide que las clasifiquen de la mejor a la peor con empates permitidos. De esta forma se tiene que cada evaluación consiste de un conjunto de cinco sistemas y un orden de los sistemas, y de este se obtienen las comparaciones por parejas.

Para modelar el problema se genera una digráfica D en la que los vértices son los sistemas de traducción que se están comparando. Dados dos sistemas de traducción A, B , se tiene el peso $w = \text{contar}(A, B) - \text{contar}(B, A)$, donde $\text{contar}(A, B)$ es el número de veces que un experto prefirió una traducción del sistema A sobre una de B . Si w es positiva, se tiene la flecha (A, B) , si es negativa, la (B, A) , y el peso de dicha flecha es $|w|$. Adam Lopez [19] afirma que un enfoque adecuado para encontrar un buen clasificación de los sistemas de traducción es resolver

el problema del FAS con pesos para la digráfica D . En este problema el FAS mínimo es aquel FAS cuya suma de pesos sea la menor entre todos los FAS posibles.

Para resolver el problema del FAS Lopez utiliza una heurística en la que en cada paso se tiene una clasificación de algún subconjunto de los vértices. Después, esta clasificación se extiende eligiendo un vértice que no ha sido ordenado aún, donde el costo de dicho vértice será la suma de los pesos de los vértices que no han sido ordenados aún. Los vértices los explora utilizando el algoritmo de Dijkstra. Además, en este artículo afirma que utilizar este proceso para el problema planteado, no toma más que unos minutos.

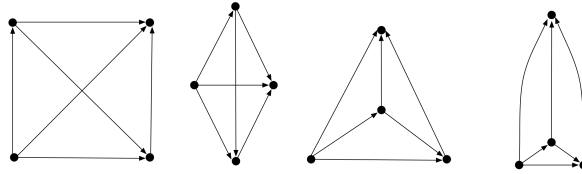


Figura 8. Distintas representaciones de TT_4 .

Otra aplicación del problema del FAS se encuentra en el dibujo de digráficas. Una digráfica se puede representar de infinitas formas en un plano, por ejemplo, en la figura 8 podemos observar cuatro representaciones distintas de TT_4 . Cada representación nos hace organizar la información que representa la digráfica de manera distinta. Un buen dibujo podría, por ejemplo, disminuir el costo de producción de un circuito o ayudarnos a visualizar grandes cantidades de información de manera efectiva.

Como es de esperarse, no hay un consenso de lo que significa un buen dibujo. Sin embargo, excepto por algunos casos especiales (como [24]), la estética no es aplicada a una digráfica en un sentido puramente artístico, sino que se busca revelar el significado y estructura que representa dicha digráfica. Por esto, en general los investigadores asocian la estética con la legibilidad y con la comprensión [4]. Por lo anterior, en el dibujo de digráficas se busca que los vértices estén distribuidos uniformemente y que las flechas tengan una longitud homogénea, asimismo, se trata de minimizar el número de cruces entre flechas [12].

Sugiyama, Tagawa y Toda [22] desarrollaron, en 1981, un algoritmo para dibujar digráficas en capas que consta de cuatro fases, donde la primera de ellas utiliza el problema del FAS mínimo. Es llamado el algoritmo de Sugiyama en honor a Kozo Sugiyama quien fue el primero en desarrollar este estilo de dibujo de digráficas. En el dibujo de digráficas en capas, además de los parámetros antes descritos, los vértices son

organizados en capas horizontales como en un organigrama y se trata de minimizar el número de flechas que apuntan hacia arriba. Por esto, es común utilizarlo para la representación de estructuras jerárquicas.

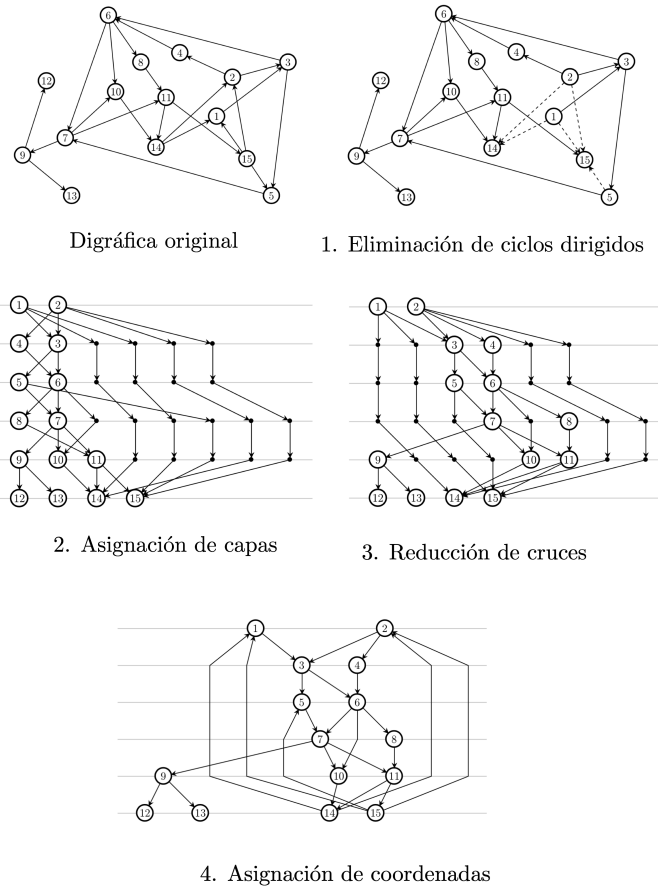


Figura 9. Ejemplo de las cuatro fases del algoritmo de Sugiyama. Obtenido de [14].

Las fases del algoritmo de Sugiyama son las siguientes:

1. **Eliminación de ciclos dirigidos.** La digráfica se hace acíclica invirtiendo la dirección de algunas flechas o eliminándolas, para esto se resuelve el problema del FAS mínimo.
2. **Asignación de capas.** Se asignan capas horizontales, también llamadas niveles, a los vértices. Se busca minimizar la altura y el ancho de la digráfica. Este problema es conocido como *layer assignment problem*.
3. **Reducción de cruces.** Los vértices son ordenados de forma que se minimice el número de intersecciones entre flechas.

4. **Asignación de coordenadas.** Se asigna una coordenada en el eje x para cada vértice, se trata de minimizar los dobleces de las flechas y el ángulo de estos.

En la figura 9 se observa un ejemplo del algoritmo expuesto anteriormente en cada una de sus fases.

Como hemos visto en este artículo, el problema de FAS es de interés tanto desde el punto de vista teórico como por sus aplicaciones. Además, estos dos factores, teoría y práctica, se han ido desarrollando a la par, en ocasiones una aplicación inspiró el estudio teórico de alguna propiedad y en otras ocasiones sucedió lo contrario. Hay que notar que las aplicaciones expuestas anteriormente son solo algunas de las que se han estudiado hasta el momento, sin embargo, muestran lo diversa y basta que es la influencia de este problema en distintas áreas.

Bibliografía

- [1] N. Alon, «Ranking tournaments», *SIAM Journal on Discrete Mathematics*, vol. 20, núm. 1, 2006, 137–142.
- [2] J. Bang-Jensen y C. Thomassen, «A polynomial algorithm for the 2-path problem for semicomplete digraphs», *SIAM Journal on Discrete Mathematics*, vol. 5, núm. 3, 1992, 366–376.
- [3] J. Bang-Jensen y G. Gutin, *Digraphs: Theory, algorithms and applications*, Springer, 2009.
- [4] C. Bennett, J. Ryall, L. Spalteholz y A. Gooch, «The Aesthetics of Graph Visualization», en *Computational Aesthetics in Graphics, Visualization, and Imaging*, eds. Douglas W. Cunningham *et al.*, The Eurographics Association, 2007.
- [5] C. Callison-Burch, P. Koehn, C. Monz, K. Peterson, M. Przybocki y O. F. Zaidan, «Findings of the 2010 joint workshop on statistical machine translation and metrics for machine translation», en *2010 Proceedings of Workshop on Statistical Machine Translation (WMT)*, Association for Computational Linguistics, 2010.
- [6] ———, «Findings of the 2011 joint workshop on statistical machine translation and metrics for machine translation», en *2011 Proceedings of Workshop on Statistical Machine Translation (WMT)*, Association for Computational Linguistics, 2011.
- [7] P. Charbit, S. Thomassé y A. Yeo, «The minimum feedback arc set problem is np-hard for tournaments», *Combinatorics, Probability and Computing*, vol. 16, núm. 01, 2006, 1.
- [8] M. Chudnovsky, P. Seymour y B. Sullivan, «Cycles in dense digraphs», *Combinatorica*, vol. 28, núm. 1, 2008, 1–18.
- [9] V. Conitzer, «Computing Slater rankings using similarities among candidates», en *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1, AAAI'06*, AAAI Press, 2006, 613–619.
- [10] G. Even, J. (Seffi) Naor, B. Schieber y M. Sudan, «Approximating minimum feedback sets and multicuts in directed graphs», *Algorithmica*, vol. 20, núm. 2, 1998, 151–174.
- [11] P. Festa, P. M. Pardalos y M. G. C. Resende, «Feedback set problems», *Handbook of Combinatorial Optimization*, 1999, 209–258.
- [12] T. M. J. Fruchterman y E. M. Reingold, «Graph drawing by force-directed placement», *Software: Practice and Experience*, vol. 21, núm. 11, 1991, 1129–1164.

- [13] A. Gama Dessavre, «Algoritmos genéticos y heurísticas de ordenamiento para el conjunto de flechas de retroalimentación de una digráfica», tesis de doctorado, Instituto Tecnológico Autónomo de México, 2016.
- [14] P. Healy y N. S. Nikolov, «Hierarchical drawing algorithms», en *Handbook of Graph Drawing and Visualization*, Chapman and Hall/CRC, 2013, 409–413.
- [15] F. Hillier y G. Lieberman, *Introducción a la investigación de operaciones*, McGraw-Hill, 2010.
- [16] G. Isaak y D. A. Narayan, «A classification of tournaments having an acyclic tournament as a minimum feedback arc set», *Information Processing Letters*, vol. 92, núm. 3, 2004, 107–111.
- [17] R. M. Karp, «Reducibility among combinatorial problems», *Complexity of Computer Computations*, 1972, 85–103.
- [18] C. Kenyon-Mathieu y W. Schudy, «How to rank with few errors», *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing - STOC '07*, 2007, .
- [19] A. Lopez, «Putting human assessments of machine translation systems in order», en *Proceedings of the Seventh Workshop on Statistical Machine Translation*, WMT '12, Association for Computational Linguistics, 2012, 1–9.
- [20] P. D. Seymour, «Packing directed circuits fractionally», *Combinatorica*, vol. 15, núm. 2, 1995, 281–288.
- [21] Statmt.org, «Naacl 2012 seventh workshop on statistical machine translation», 2015.
- [22] K. Sugiyama, S. Tagawa y M. Toda, «Methods for visual understanding of hierarchical system structures», *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 11, núm. 2, 1981, 109–125.
- [23] A. W. Tucker, «On directed graphs and integer programs», *In Symposium on Combinatorial Problems*, 1960, 562–570.
- [24] W. Vogels, «Graphs as art - all things distributed», 2018.