



**DEPARTAMENTO
DE COMPUTACION**

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico Número 3

16 de Mayo de 2016

Algoritmos y Estructuras de Datos III

Grupo 8

Integrante	LU	Correo electrónico
Ciruelos Rodríguez, Gonzalo	063/14	gonzalo.ciruelos@gmail.com
Costa, Manuel José Joaquín	035/14	manucos94@gmail.com
Gatti, Mathias Nicolás	477/14	mathigatti@gmail.com
Maddonni, Axel Ezequiel	200/14	axel.maddonni@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

<http://www.fcen.uba.ar>

Índice

0. Introducción	3
0.1. Experimentación	3
1. Ejercicio 1	4
2. Ejercicio 2: Algoritmo exacto	5
2.1. Explicación detallada del algoritmo	5
2.1.1. Complejidad del algoritmo	7
2.2. Performance del algoritmo	8
3. Ejercicio 3	9
4. Ejercicio 4	10
5. Ejercicio 5	11
6. Ejercicio 6: Tabu Search	12
6.1. Explicación detallada del algoritmo	12
6.2. Performance del algoritmo	12
7. Ejercicio 7	13
8. Apéndice	14
8.1. Generación de grafos conexos aleatorios	14
8.2. Partes relevantes del código	15

0. Introducción

En este trabajo desarrollaremos varias soluciones para el problema de el subgrafo común máximo entre dos grafos G_1 y G_2 , con respecto a los vértices.

Más precisamente, dados $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$ dos grafos simples, el problema de máximo subgrafo común consiste en encontrar un grafo $H = (V_H, E_H)$ isomorfo tanto a un subgrafo de G_1 como a un subgrafo de G_2 que maximice $|E_H|$.

Nuestros acercamientos al problema van a ser dos. Primero, vamos a desarrollar una solución exacta, es decir, una solución que encuentra el subgrafo común que maximiza la cantidad de aristas. Sin embargo, como veremos, esta forma de resolverlo no es razonable dado que se desconoce una solución en tiempo polinomial, por lo que encontrar la mejor solución no es viable para entradas grandes.

Luego, veremos varios algoritmos aproximados para el problema del subgrafo común máximo. Estos consisten en sacrificar exactitud a cambio de tiempo de ejecución. Su complejidad será polinomial, pero como dijimos, las soluciones que generen serán *aproximadas*, o sea, no exactas.

0.1. Experimentación

La experimentación en general sigue los pasos sugeridos por las consignas del trabajo. Los métodos de generación de casos estarán explicados al final de cada sección de experimentación, o en su defecto, en el apéndice.

Sobre la experimentación de tiempos, como las complejidades en general dependen de muchos parámetros, los resultados se vuelven difíciles de representar. Es por eso que seguiremos el mismo método de representación que utilizamos en los TPs anteriores. Supongamos que la complejidad del algoritmo es $O(f(n, m))$, entonces nuestro gráfico tendrá $f(n, m)$ en el eje x y $T(n, m) = \text{“El tiempo que tarda el algoritmo para una entrada de tamaño (n,m)”}$ en el eje y , de esta manera, nos interesará ver que el gráfico es el de una constante.

Por supuesto, también haremos experimentos en los que fijamos parámetros y movemos otros, para corroborar que las performances se comportan como deben. En caso de que las complejidades dependan de un parámetro, haremos un gráfico clásico, en caso contrario, haremos lo mismo que explicamos anteriormente.

1. Ejercicio 1

2. Ejercicio 2: Algoritmo exacto

2.1. Explicación detallada del algoritmo

El problema del sugrafo común máximo (con respecto a aristas) es un problema perteneciente a la clase NP, por lo que hasta el momento no se conocen algoritmos que lo resuelvan de manera exacta en tiempo polinomial.

Por esa razón, el algoritmo que lo resuelve de manera exacta debe ser de la clase de algoritmos que exploran todo el espacio de soluciones y se quedan con la mejor. De entre esos algoritmos, elegiremos un algoritmo de backtracking, dado que proponiendo buenas podas, se pueden mejorar los tiempos de ejecución del algoritmo, evitando mirar el espacio de soluciones en su enteridad.

Nuestras soluciones, es decir, nuestros isomorfismos, van a estar representados como un vector de pares. Cada par (v_{1i}, v_{2i}) es un par de vértices que cumplen que $v_{1i} \in V(G_1)$ y $v_{2i} \in V(G_2)$ y nuestro isomorfismo los mapea.

Por ejemplo, si nuestro isomorfismo es $f : V(G_1) \rightarrow V(G_2)$ tal que

$$f(0) = f(1)$$

$$f(1) = f(2)$$

$$f(3) = f(5)$$

$$f(6) = f(0)$$

Lo vamos a almacenar de la siguiente manera: $(0, 1), (1, 2), (3, 5), (6, 0)$. Nótese que no importa que el “isomorfismo” sea parcial, dado que en realidad estamos describiendo subgrafos isomorfos.

Por lo tanto, nuestro algoritmo de backtracking se va a basar en probar todas las posibles combinaciones de listas de pares, y buscar cual representa el isomorfismo con la mayor cantidad de aristas.

Una primera poda que proponemos (bastante fácil de explicar y muy poderosa) es la que sigue: podemos notar que si ciegamente consideramos todas las combinaciones de pares, estaremos considerando todos los isomorfismos varias veces. Por ejemplo, el vector $(0, 1), (1, 2)$ y el vector $(1, 2), (0, 1)$ representan el mismo isomorfismo, pero nuestro algoritmo naif los analizará 2 veces.

Por esta razón, diseñamos la siguiente poda: solo considerar vectores cuyo vector de primera coordenadas esté ordenado ascendentemente. En el ejemplo anterior, cuando le llegue al turno a $(1, 2), (0, 1)$, no lo analizaremos ni a él, ni a ninguno de sus descendientes, dado que todos ellos serán analizados como descendientes de $(0, 1), (1, 2)$.

Otra poda que realizamos es que, si encontramos la solución máxima posible teóricamente (es decir, la solución cuya cantidad de aristas coincide con el mínimo de las aristas de G_1 y G_2), terminar con el algoritmo inmediatamente.

Una última poda que realizamos es considerar solamente aquellos vectores cuyo largo sea exactamente el del mínimo de los vértices de G_1 y G_2 . Esto se debe a que, si un vector es más chico, vamos a poder considerar a un vector extendido, cuyo isomorfismo potencialmente tendrá más aristas.

El algoritmo utiliza una variable global llamada *solucin*, que tiene 2 campos, *aristas* e *isomorfismo*. En esta variable global irá guardando la mejor solución encontrada hasta el momento. *solucin.aristas* debe inicializarse en 0.

Sin más que analizar, pasemos a ver el pseudocódigo del algoritmo. Vale la pena aclarar que asumimos como precondición que G_1 tiene menos nodos que G_2 (en tal caso de que así no fuere, el llamador debe ocuparse de dar vuelta los parámetros).

Algorithm 1 Pseudocódigo del procedimiento Backtracking

```

1: procedure BT(Grafo  $g1$ , Grafo  $g2$ , vector<int>  $vertices1$ , vector<int>  $vertices2$ ,
   Isomorfismo  $iso$ )
2:   if  $solucion.aristas == MIN(g1.m, g2.m)$  then ▷  $O(1)$ 
3:     return
4:   if  $!ordenado\_asc(iso)$  then ▷  $O(n_1)$ 
5:     return
6:   if  $iso.size == g1.n$  then ▷  $O(1)$ 
7:      $aristas \leftarrow contar\_aristas\_isomorfismo(g1, g2, iso)$  ▷  $O(n_1^2)$ 
8:     if  $aristas > solucion.aristas$  then ▷  $O(1)$ 
9:        $solucion.isomorfismo \leftarrow iso$  ▷  $O(n_1)$ 
10:       $solucion.aristas \leftarrow aristas$  ▷  $O(1)$ 
11:   return
12:   for  $u \in vertices1$  do ▷  $v_1$  veces  $O(1)$ 
13:     for  $v \in vertices2$  do ▷  $v_1 v_2$  veces  $O(1)$ 
14:        $nuevo\_iso = iso$  ▷  $v_1 v_2$  veces  $O(n_1)$ 
15:        $nuevo\_iso.push\_back(make\_pair(u, v))$  ▷  $v_1 v_2$  veces  $O(1)$ 
16:        $bt(g1, g2, copiar\_sin(vertices1, u), copiar\_sin(vertices2, v), nuevo\_iso)$  ▷
        $v_1 v_2$  veces  $T(n_1, n_2, v_1 - 1, v_2 - 1)$ 

```

Algorithm 2 Pseudocódigo del procedimiento contar aristas isomorfismo

```

1: procedure CONTAR_ARISTAS_ISOMORFISMO(Grafo  $g1$ , Grafo  $g2$ , Isomorfismo  $iso$ )
    $\rightarrow$  Int
2:    $aristas \leftarrow 0$  ▷  $O(1)$ 
3:   for  $p \in iso$  do ▷  $n_1$  veces  $O(1)$ 
4:      $vg1 = p.first$  ▷  $n_1$  veces  $O(1)$ 
5:      $vg2 = p.second$  ▷  $n_1$  veces  $O(1)$ 
6:     for  $q \in iso$  do ▷  $n_1$  veces  $O(1)$ 
7:        $ug1 = q.first$  ▷  $n_1^2$  veces  $O(1)$ 
8:        $ug2 = q.second$  ▷  $n_1^2$  veces  $O(1)$ 
9:       if  $g1.adj\_matrix[vg1][ug1] \wedge g2.adj\_matrix[vg2][ug2]$  then ▷  $n_1^2$  veces  $O(1)$ 
10:       $aristas++$  ▷  $n_1^2$  veces  $O(1)$ 
   return  $aristas$ 

```

Donde $n_i = |V(G_i)|$ y $m_i = |E(G_i)|$, para $i = 1, 2$. Nótese que el largo del vector isomorfismo está acotado superiormente por n_1 , pues $n_1 < n_2$ y el isomorfismo mapea a lo sumo a todos los vértices de n_1 y no puede mapear más cosas.

Además, v_i para $i = 1, 2$ es el tamaño del vector *vertices*.

El algoritmo debe ser llamado de la siguiente manera:

$$bt(grafo1, \{1, \dots, |V(grafo1)|\}, grafo2, \{1, \dots, |V(grafo2)|\}, \{\})$$

.

Dado que inicialmente todos los vértices están sin ser utilizados, y el isomorfismo es el isomorfismo vacío.

2.1.1. Complejidad del algoritmo

Calculemos la complejidad del algoritmo. Primero, como vimos, la complejidad del algoritmo `contar_aristas_isomorfismo` es $O(n_1^2)$ y es bastante fácil de calcular.

Ahora, calculemos la complejidad del algoritmo de backtracking. Sea $T(n_1, n_2, v_1, v_2)$ el tiempo que el algoritmo tarda para una entrada de ese tamaño. El tamaño del vector *iso* puede acotarse por n_1 , como vimos antes.

Como se ve en el análisis de complejidad, $T(n_1, n_2, v_1, v_2) = n_1 + n_1^2 + v_1 v_2 n_1 + v_1 v_2 + v_1 v_2 T(n_1, n_2, v_1 - 1, v_2 - 1)$.

Además, notemos que siempre pasa que $v_i < n_i$, luego, la complejidad puede acotarse por $T(n_1, n_2, v_1, v_2) < n_1 + n_1^2 + v_1 v_2 n_1 + v_1 v_2 + v_1 v_2 T(n_1, n_2, v_1 - 1, v_2 - 1)$. Nos tomamos la libertad de escribir $T(v_1, v_2)$, dado que son los únicos parámetros variables (n_1 y n_2 están fijos).

Como $n_i > v_i$, $T(v_1, v_2) < n_1 + n_1^2 + n_1 n_2 n_1 + n_1 n_2 + v_1 v_2 T(v_1 - 1, v_2 - 1)$.

Luego, $T(v_1, v_2) < 4n_1^2 n_2 + v_1 v_2 T(v_1 - 1, v_2 - 1)$.

Además, $T(n_1, n_2, 0, v_2) = n_1 + n_1^2$ (recordemos que siempre va a pasar que $n_1 < n_2$, por lo tanto $v_1 < v_2$). Por lo que la profundidad de la fórmula recursiva depende solo de v_1 . Luego,

$$\begin{aligned} T(v_1, v_2) &< 4n_1^2 n_2 + v_1 v_2 T(v_1 - 1, v_2 - 1) \\ &< 4n_1^2 n_2 + v_1 v_2 (4n_1^2 n_2 + (v_1 - 1)(v_2 - 1)T(v_1 - 2, v_2 - 2)) \\ &= 4n_1^2 n_2 + v_1 v_2 4n_1^2 n_2 + v_1 v_2 (v_1 - 1)(v_2 - 1)T(v_1 - 2, v_2 - 2) \\ &< 4n_1^2 n_2 + 4n_1^3 n_2^2 + v_1 v_2 (v_1 - 1)(v_2 - 1)T(v_1 - 2, v_2 - 2) \\ &< \dots \\ &< 4n_1^2 n_2 + \dots + 4n_1^{v_1} n_2^{v_1-1} + v_1 (v_1 - 1) \dots (v_1 - v_1 + 1) v_2 (v_2 - 1) \dots (v_2 - v_1 + 1) T(0, v_2 - v_1) \\ &< 4n_1^2 n_2 + 4n_1^3 n_2^2 + \dots + 4n_1^{v_1} n_2^{v_1-1} + 4n_1^{v_1+1} n_2^{v_1} \\ &< 4n_1^2 n_2 + 4n_1^3 n_2^2 + \dots + 4n_1^{v_1+1} n_2^{v_1} \\ &< 4^{v_1+1} n_1^{v_1+1} n_2^{v_1+1} \\ &= (4n_1 n_2)^{v_1+1} \end{aligned}$$

Luego, $T(n_1, n_2) \in O((4n_1 n_2)^{n_1+1})$.

Nótese que esta es una cota bastante poco ajustada, pero es suficientemente exacta para el análisis que queremos hacer (una cota más ajustada, como se ve en los cálculos anteriores, involucra fórmulas con factoriales y combinatorios).

2.2. Performance del algoritmo

Asdf

3. Ejercicio 3

4. Ejercicio 4

5. Ejercicio 5

6. Ejercicio 6: Tabu Search

6.1. Explicación detallada del algoritmo

En la sección anterior analizamos varios algoritmos de búsqueda local. La búsqueda local es una heurística para algoritmos aproximados. Sin embargo, las heurísticas de búsqueda local pueden verse como un algoritmo de *gradient descent*. Estos algoritmos tienen el problema de los mínimos locales: cuando encuentran un mínimo local se quedan trabados y no pueden mejorar esa solución no óptima.

Por esa razón surgen las metaheurísticas. Una metaheurística es un método heurístico para resolver un problema computacional general. Una metaheurística usa los parámetros dados por el usuario sobre unos procedimientos genéricos y abstractos. Normalmente, estos procedimientos son heurísticos.

En particular, nosotros utilizaremos la metaheurística del tabu search, que a su vez se basa en la heurística de búsqueda local.

6.2. Performance del algoritmo

7. Ejercicio 7

8. Apéndice

8.1. Generación de grafos conexos aleatorios

Algorithm 3 Pseudocódigo del procedimiento para generar grafos conexos al azar

```

1: procedure GRAFO_RANDOM(int  $n$ , int  $m$ )  $\rightarrow$  Grafo
2:    $k_n \leftarrow \{(0, 1), (0, 2), \dots, (0, n), (1, 2), (1, 3), \dots, (n - 2, n - 1)\}$ 
3:    $vertices \leftarrow \{random.range(0, n)\}$   $\triangleright$  Empiezo con un vértice al azar
4:    $agm \leftarrow \{\}$ 
5:   while  $vertices.size() < n$  do
6:      $aristas \leftarrow$  “aristas  $(u, v)$  de  $k_n$  tal que  $u \in vertices$  y  $v \notin vertices$  o viceversa”
7:      $arista\_nueva \leftarrow random.choice(aristas)$ 
8:      $agm.add(arista\_nueva)$ 
9:      $k_n.remove(arista\_nueva)$ 
10:     $vertices.add(\text{“extremo de } arista\_nueva \text{ que no estaba en vertices”})$ 
11:     $\triangleright$  Cuando termina este ciclo tenemos un árbol de  $n$  vertices y  $n - 1$  aristas
12:     $grafo \leftarrow agm$ 
13:    while  $grafo.size() < m$  do
14:       $arista \leftarrow random.choice(k_n)$ 
15:       $grafo.add(arista)$ 
16:       $k_n.remove(arista)$ 
17:    for  $arista \in grafo$  do
18:       $peso(arista) \leftarrow random.random()$ 
return  $grafo$ 

```

El algoritmo, se basa en generar un grafo conexo minimal (es decir, un árbol) de n vértices. Para lograr esto, técnicamente lo que hacemos es empezar con K_n , es decir, el grafo completo de n vértices, con todos sus aristas de igual peso, y le encontramos un árbol generador mínimo utilizando Prim. Todo esto es obviamente trivial en este caso, dado que todas las aristas tienen igual peso, así que básicamente lo que hacemos es elegir una arista al azar en cada paso.

Luego, una vez que tenemos el árbol terminado, lo completamos con aristas al azar, hasta llegar al objetivo de m aristas.

Finalmente, se eligen pesos al azar para cada arista.

8.2. Partes relevantes del código