

Algoritmos y Estructuras de Datos III

TP3

1 de junio de 2015

Integrante	LU	Correo electrónico
Martin Baigorria	575/14	martinbaigorria@gmail.com
Federico Beuter	827/13	federicobeuter@gmail.com
Juan Rinaudo	864/13	jangamesdev@gmail.com
Mauro Cherubini	835/13	cheru.mf@gmail.com

Reservado para la cátedra

Instancia	Docente	Nota
Primera entrega		
Segunda entrega		

Índice

1. Introducción	3
1.1. Definiciones	3
1.2. Introducción	3
1.3. Maximalidad y dominancia	3
1.4. Modelado	4
1.4.1. Planificador Urbano	4
2. Algoritmo Exacto	5
2.1. Algoritmo	5
2.2. Podas y estrategias	5
2.3. Complejidad	5
2.4. Complejidad Espacial	5
2.5. Complejidad Temporal	5
2.6. Experimentacion	6
2.7. Código	7
3. Heurística Constructiva Golosa	9
3.1. Algoritmo	9
3.2. Complejidad	9
3.3. Efectividad de la heurística	9
4. Experimentacion	10
4.1. Código	11
5. Heurística de Búsqueda Local	13
6. Metaheurística GRASP	14

1. Introducción

1.1. Definiciones

Antes de enunciar el problema a resolver en este trabajo práctico, es necesario definir algunos conceptos.

Sea $G = (V, E)$ un grafo simple:

Definición Un conjunto $I \subseteq V$ es un *conjunto independiente* de G si no existe ningún eje de E entre los vértices de I . Es decir, los ejes de I no están conectados por las aristas de G .

Definición Un conjunto $D \subseteq V$ es un *conjunto dominante* de G si todo vértice de G está en D o bien tiene al menos un vecino que está en D .

Definición Un conjunto *conjunto independiente dominante* de G es un conjunto independiente que a su vez es dominante del grafo G . Desde un conjunto independiente dominante se puede acceder a cualquier vértice del grafo G con solo recorrer una arista desde uno de sus vértices.

Definición Un *Conjunto Independiente Dominante Mínimo* (CIDM) es el conjunto independiente dominante de G de mínima cardinalidad.

Mostrar como el CIDM y el CDM no son necesariamente los mismos, y de ser posible mostrar que hasta pueden tener diferente cardinalidad.

Cada definición debería ser acompañada con un gráfico. Por ejemplo, podemos mostrar dos conjuntos independientes y dominantes del mismo grafo, donde uno es el CIDM.

1.2. Introducción

En 1979, Garey y Johnson probaron que el problema de encontrar el CIDM de un grafo es un problema NP-Hard¹. El objetivo del trabajo es utilizar diferentes técnicas algorítmicas para resolver este problema. En un principio diseñaremos e implementaremos un algoritmo exacto para el mismo. Dada la complejidad del problema, luego propondremos diferentes algoritmos heurísticos para llegar a una solución que sea lo suficientemente buena a fines prácticos en un tiempo razonable.

Si recordamos el problema 3 del TP1, podemos ver claramente que el mismo es un caso particular del problema del conjunto dominante óptimo. Esto se debe a que el problema de los caballos imponía cierta estructura sobre el grafo en el que se efectuaba la búsqueda. El grafo en sí no era completo, dado que cada casilla era representada por un nodo, y un caballo no podía acceder a los nodos adyacentes. El movimiento de los caballos se modelaba con aristas entre nodos. En cambio, el problema de encontrar el CDM se aplica a cualquier tipo de grafo. Sin embargo, no es equivalente al problema del CIDM dado que la solución al problema no necesariamente era independiente. Dado que el problema de los caballos era computacionalmente costoso, considerando que la minimalidad no afecta mucho la estructura del problema podemos conjeturar, como ya lo confirma la literatura, que este problema se resolverá en tiempo no polinomial.

1.3. Maximalidad y dominancia

Las siguientes proposiciones serán útiles a lo largo del trabajo:

Proposición 1.1 Sea M un conjunto independiente maximal de G . $\forall v \in G.V$, si $v \notin M \implies \exists u \in M$ tal que u es adyacente a v .

Demostración Por absurdo. Sea M un conjunto independiente maximal y $v \notin G.V$. $\nexists u \in M$ tal que u es adyacente a v . Por lo tanto, puedo agregar v a M y el conjunto va a seguir siendo independiente. Esto es absurdo, dado que el conjunto era maximal.

Proposición 1.2 Dado $G(V, E)$, todo conjunto independiente maximal es un conjunto independiente dominante.

Demostración Sea M un conjunto independiente maximal. Por la propiedad anterior, si $v \notin M \implies \exists u \in M$ tal que u es adyacente a v . Por lo tanto, si $v \notin M$ entonces tiene algún vecino que está en M . Esto significa que M es dominante.

¹M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman and Company, San Francisco (1979).

1.4. Modelado

Muchos problemas se pueden modelar con grafos y se pueden resolver mediante la búsqueda del conjunto independiente dominante mínimo.

1.4.1. Planificador Urbano

Supongamos que un planificador urbano está diseñando una ciudad con muchos barrios. Con el objetivo de proveer un buen sistema de salud para los habitantes, el planificador determina que cada barrio debe tener que cruzar a lo sumo un barrio para acceder a un hospital público. Aquí podemos modelar a cada barrio con un vértice, y representar la adyacencia entre barrios con una arista. Al obtener el CIDM, obtenemos la ubicación y la mínima cantidad de hospitales públicos necesarios para cumplir con los objetivos del planificador.

2. Algoritmo Exacto

2.1. Algoritmo

Utilizando backtracking, recorremos todas los conjuntos dominantes independientes y luego seleccionamos el de menor cardinalidad. Representamos al grafo con un arreglo $graph[n]$ de nodos. Cada nodo tiene los siguientes atributos:

1. adj: Lista de nodos adyacentes al nodo actual.
2. degree: Grado del nodo actual.
3. added: Bool que indica si el nodo ha sido agregado al conjunto que representa el cubrimiento.
4. reachable: Bool que indica si el nodo actual puede ser alcanzado desde un nodo perteneciente al cubrimiento.

Comenzamos definiendo la función *backtracking*, que lo que hace es tomar un nodo del grafo, y luego considera los casos en los que el nodo pertenece o no a un posible cubrimiento. En caso de agregar el nodo al cubrimiento, todos los nodos adyacentes al mismo son ignorados en futuras llamadas recursivas. Si consideráramos los nodos adyacentes, romperíamos la independencia de los cubrimientos y además no solo incrementaría la complejidad del código sino que también el tiempo de ejecución del mismo.

2.2. Podas y estrategias

Para poder resolver el problema lo mas rapido posible, en primer lugar buscamos una forma rapida de verificar si el conjunto actual al llamar la funcion recursiva es un cubrimiento. Esto lo resolvimos simplemente haciendo que la funcion backtracking lleve la cuenta del total de nodos alcanzables por el cubrimiento. Si ese numero es igual al numero total de nodos, significa que llegamos a un cubrimiento. De esta manera evitamos funciones auxiliares.

Ademas, antes de comenzar la búsqueda agregamos a todos los vértices de $d(v) = 0$ al conjunto solución final. Esto se debe a que estos vértices necesariamente estarán en la solución. Es muy simple probar esto, dado que si no lo estuvieran, algún vértice adyacente debería estar en el conjunto para que lo cubra. Sin embargo, tal vértice no existe.

Una poda muy común que también hemos implementado es la de la solución local actual. Dada una solución posible (que aun no sabemos si es la mínima), si en el estado actual del algoritmo se esta considerando un numero de vértices que no le puede ganar a esta solución, ignoramos esa rama del árbol de estados posibles.

Una estrategia natural también fue evitar los nodos adyacentes a los que ya agrego el algoritmo al potencial conjunto solución sean considerados. De esta forma mantenemos la independencia del conjunto y evitamos tener que agregar innecesariamente muchos nodos. Ademas, evitamos tener que programar una función que verifique la independencia del conjunto.

2.3. Complejidad

2.4. Complejidad Espacial

Para la representación del grafo, utilizamos un arreglo de nodos. Cada nodo tiene una lista de adyacencia. Por lo tanto, la complejidad espacial de nuestro algoritmo es de $\mathcal{O}(n + 2m)$, donde n es la cantidad total de vértices y m la cantidad total de aristas.

2.5. Complejidad Temporal

Nuestro algoritmo, sin considerar las podas, recorre cada conjunto independiente dominante una vez. Cada vez que encuentra uno, lo guarda en una estructura auxiliar en $\mathcal{O}(n)$. Si todos los nodos tienen grado 0, son agregados automaticamente, y el algoritmo resuelve el problema en n iteraciones. En el peor de los casos, el algoritmo recorre todos los conjuntos independientes y dominantes, comenzando con el de mayor cardinalidad. Cada vez que lo encuentra, actualiza la estructura donde guardamos la solución. Para que esto suceda, en realidad todos los conjuntos dominantes deben tener diferente cardinalidad, cosa que en general no sucede. Sin embargo, como base pensemos que si. Moon & Moser probaron que un grafo de n vértices tiene a lo sumo $3^{n/3}$ conjuntos independientes y dominantes². Por lo tanto, una cota no muy buena para la actualización de soluciones locales es $\mathcal{O}(n \times 3^{n/3})$.

Por otro lado, recorremos cada vértice y sus aristas adyacentes una vez por iteracion. Sin embargo hay muchas combinaciones en las que podemos hacer esto. En total, en el peor de los casos sin la poda del minimo, la cantidad de nodos recorridos esta acotada por 2^n .

Por lo tanto, el orden del algoritmo es de $\mathcal{O}(n \times 3^{n/3})$.

²Moon, J. W.; Moser, Leo (1965), 'On cliques in graphs', Israel Journal of Mathematics 3 (1): 23–28, doi:10.1007/BF02760024

2.6. Experimentacion

2.7. Código

```
1 #include <iostream>
2 #include <forward_list>
3
4 using namespace std;
5
6 struct Node {
7     forward_list<int> adj;
8     unsigned int degree;
9     bool added;
10    bool reachable;
11
12    Node() {
13        degree = 0;
14        added = false;
15        reachable = false;
16    }
17 };
18
19 void backtracking(int current, int& n, int coveredNodes, int usedNodes, Node graph[],
20     bool localSolution[], int& nodesUsedInSolution);
21
22 int main() {
23     int n, m; // n: vertices, m: edges
24     cin >> n >> m;
25
26     Node graph[n]; // graph container
27
28     bool localSolution[n];
29     int nodesUsedInSolution = n; // worst case scenario is n, that way I avoid
30         setting all the array as true.
31
32     int u, v;
33     for (int i = 1; i <= m; ++i) { // (u,v) edges
34         cin >> u >> v;
35         u--; // nodes are counted from 0 in array.
36         v--;
37         graph[u].adj.push_front(v);
38         graph[v].adj.push_front(u);
39         graph[u].degree++;
40         graph[v].degree++;
41     }
42
43     for (int i = 1; i <= m; ++i) { // add d(v)=0 nodes to cover.
44         if (graph[i].degree == 0) {
45             graph[i].added = true;
46             graph[i].reachable = true;
47         }
48     }
49
50     backtracking(0, n, 0, 0, graph, localSolution, nodesUsedInSolution);
51
52     // display solution
53     cout << nodesUsedInSolution;
54     for (int i = 0; i < n; ++i) {
55         if (localSolution[i] == true) cout << " " << i + 1;
```

```

55     }
56     cout << endl;
57
58     return 0;
59 }
60
61 void backtracking(int current, int& n, int coveredNodes, int usedNodes, Node graph[],
62     bool localSolution[], int& nodesUsedInSolution) {
63     // cout << "current: " << current << " n: " << n << " coveredNodes: " <<
64     // coveredNodes << " usedNodes: " << usedNodes << " nodesUsedInSolution: " <<
65     // nodesUsedInSolution << endl;
66     if (graph[current].reachable == true) return backtracking(current + 1, n,
67         coveredNodes, usedNodes, graph, localSolution, nodesUsedInSolution);
68     if (current == n) return; // no nodes left to add.
69     if (usedNodes + 1 == nodesUsedInSolution) return; // cant beat current solution
70
71     int pushed = 0;
72     forward_list<int> added; // save changes to graph to then restore
73     graph[current].added = true;
74     for (auto it = graph[current].adj.begin(); it != graph[current].adj.end(); ++it)
75     {
76         int adjNode = *it;
77         if (graph[adjNode].reachable == false) { // node reaches these new vertices
78             graph[adjNode].reachable = true;
79             added.push_front(adjNode);
80             ++pushed;
81         }
82     }
83
84     int tempcoveredNodes = coveredNodes + pushed + 1;
85     if (tempcoveredNodes == n) { // coverage found
86         for (int i = 0; i < n; ++i) {
87             localSolution[i] = graph[i].added;
88         }
89         nodesUsedInSolution = ++usedNodes;
90     } else {
91         backtracking(current + 1, n, tempcoveredNodes, usedNodes + 1, graph,
92             localSolution, nodesUsedInSolution); // adding current element to coverage
93     }
94
95     // restore graph state
96     graph[current].added = false;
97     for (auto it = added.begin(); it != added.end(); ++it) {
98         graph[*it].reachable = false;
99     }
100     backtracking(current + 1, n, coveredNodes, usedNodes, graph, localSolution,
101         nodesUsedInSolution); // skip current node
102 }

```

3. Heurística Constructiva Golosa

3.1. Algoritmo

Para poder armar un algoritmo para el problema del CIMD, en primer lugar hay que buscar un buen criterio para seleccionar que nodos pertenecerán al cubrimiento, dado los nodos que ya han sido agregados. Decidimos utilizar como criterio el numero de nodos adyacentes efectivos a los que cada nodo puede acceder. Definimos a un nodo adyacente efectivo (score) como un nodo que es adyacente y a su vez no puede ser accedido por otros nodos que ya pertenecen al cubrimiento. De esta forma, este criterio también nos garantiza la independencia del conjunto, dado que si tomamos dos nodos de la solución, por construcción no pueden ser adyacentes.

Cada nodo va a tener como atributos su score, un flag que indica si ha sido agregado y otro que indica si es alcanzable por el cubrimiento actual.

El algoritmo va a iterar un arreglo de nodos n^2 veces. Cada vez que busquemos un nodo para agregar al conjunto, los iteraremos todos para buscar el de maximo score. Luego, al identificarlo, actualizaremos los scores de los nodos adyacentes a los adyacentes del mismo.

3.2. Complejidad

El algoritmo recorre cada nodo del arreglo n veces. A su vez, actualizar los scores al identificar un máximo se hace m veces. Por lo tanto, el algoritmo tiene orden $\mathcal{O}(n^2 + m)$.

Notar que la forma en que buscamos el máximo es sumamente ineficiente. Esto se debe a que si utilizamos sort, luego es bastante difícil encontrar el nodo al que le debemos actualizar su respectivo score. A su vez, dado que en cada iteracion actualizamos el score, mantener el orden es sumamente costoso. Es muy posible que exista una estructura de datos mucho mas eficiente para resolver este problema.

Podríamos resolver el problema en $\mathcal{O}(n \times \log(n))$ simplemente ignorando la actualización de los scores, desencolando de un heap n veces. Sin embargo, este criterio es a simple vista inferior que el de actualización de scores. Aquí hay un tradeoff entre hacer el mejor pick y la complejidad temporal del algoritmo.

3.3. Efectividad de la heurística

Nuestra heurística no siempre devuelve la solución óptima. Considerar los siguientes ejemplos:

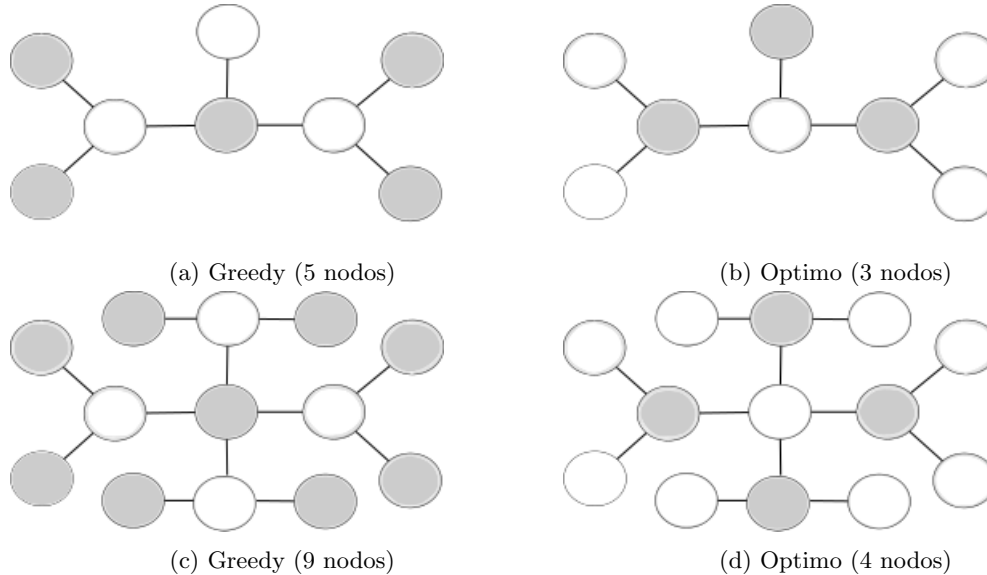


Figura 1: Ejemplos de nuestra heurística comparado con el óptimo.

El peor caso es claramente el de la figura (c) y (d). Tenemos un nodo v de grado $d(v) = n$, con sus nodos adyacentes de grado $d(u) = n - 1$. Si tenemos c componentes conexas de ese tipo, utilizaremos $c \times (n \times (n - 2) + 1)$ nodos, cuando en realidad el óptimo tiene $c \times n$ nodos.

4. Experimentacion

4.1. Codigo

```
1 #include <iostream>
2 #include <forward_list>
3 #include <algorithm>
4
5 using namespace std;
6
7 struct Node {
8     forward_list<int> adj;
9     unsigned int score;
10    bool added;
11    bool reachable;
12
13    Node() {
14        score = 0;
15        added = false;
16        reachable = false;
17    }
18
19 };
20
21 int main() {
22
23     int n, m; // n: vertices , m: edges
24     cin >> n >> m;
25
26     Node graph[n]; // graph container
27     int nodesUsedInSolution = 0;
28
29     int u, v;
30     for (int i = 1; i <= m; ++i) { // (u,v) edges
31         cin >> u >> v;
32         u--; // nodes are counted from 0 in array.
33         v--;
34         graph[u].adj.push_front(v);
35         graph[v].adj.push_front(u);
36         graph[u].score++;
37         graph[v].score++;
38     }
39
40     for (int i = 0; i < n; ++i) {
41         if (graph[i].reachable == true) continue;
42
43         int greatest = i;
44         unsigned int score = graph[i].score;
45
46         if (score == 0) {
47             graph[i].reachable = true;
48             graph[i].added = true;
49         }
50
51         for (int j = 0; j < n; ++j) {
52             if (graph[j].reachable == true) continue;
53             if (graph[j].score > score) {
54                 greatest = j;
55                 score = graph[j].score;
56             }
57         }
58     }
59 }
```

```

57     }
58
59     graph[i].added = true;
60
61     // update adjacent nodes of reachable nodes scores.
62     for (auto it = graph[greatest].adj.begin(); it != graph[greatest].adj.end();
        ++it) {
63         int adjNode = *it;
64         graph[adjNode].reachable = true;
65         for (auto it2 = graph[adjNode].adj.begin(); it2 != graph[adjNode].adj.end
            ()); ++it2) {
66             graph[*it2].score--;
67         }
68     }
69
70     nodesUsedInSolution++;
71 }
72
73 // display solution
74 cout << nodesUsedInSolution;
75 for (int i = 0; i < n; ++i) {
76     if (graph[i].added == true) cout << " " << i + 1;
77 }
78 cout << endl;
79
80 return 0;
81 }

```

5. Heurística de Búsqueda Local

6. Metaheurística GRASP