



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
INFORMÁTICA

GRADO EN INGENIERÍA DEL SOFTWARE

Curso académico 2019/2020

TRABAJO FIN DE GRADO

**BÚSQUEDA DEL CLIQUE DE RATIO MÁXIMO
MEDIANTE EL ALGORITMO GRASP**

Autor:

José Miguel García Benayas

Tutores:

Dr. Jesús Sánchez-Oro Calvo

Dr. Alfonso Fernández Timón

Resumen

TODO

Agradecimientos

Todo este trabajo no habría sido posible sin mi familia, sin su apoyo y cariño, en especial a Rebeca Muñoz, la cual ha sido mi compañera de viaje y la luz que me ilumina.

Sin olvidar a mis amigos en la universidad, con quienes he disfrutado cada día en clase, y a mis tutores, Jesús y Alfonso, por darme la oportunidad de realizar este trabajo y de los que he aprendido mucho.

Índice general

Resumen	III
Agradecimientos	V
1. Introducción	1
1.1. Definición y motivación del problema	1
1.2. Estado del arte	1
2. Objetivos	3
2.1. Objetivos principales	3
2.2. Objetivos secundarios	3
3. Descripción algorítmica	5
3.1. Metaheurística	5
3.1.1. GRASP	5
3.1.1.1. Fase constructiva	5
3.1.1.2. Fase de búsqueda	8
4. Descripción informática	9
4.1. Diseño	9
4.2. Implementación	9
4.3. Metodología empleada	11
5. Resultados	13
5.1. Recursos utilizados	13
5.1.1. Descripción de la máquina utilizada	13
5.1.2. Instancias utilizadas	13
5.2. Análisis de los resultados	14
6. Conclusiones	15

VIII

6.1. Consecución de los objetivos	15
6.2. Conocimientos adquiridos	15
6.3. Líneas de desarrollo futuras	16
Bibliografía	17

Índice de figuras

Índice de cuadros

Lista de Abreviaciones

GRASP	G reedy R andomized A daptive S earch P rocedure
IDE	I ntegrated D evelopment E nvironment

1 Introducción

1.1. Definición y motivación del problema

Este problema trata de encontrar una solución aproximada al MRCP por sus siglas en inglés Maximum Ratio Clique Problem. La solución se obtiene como un subgrafo completo o clique de ratio máximo de un grafo, para ello se realiza una aproximación lo más cercana posible a la solución óptima mediante un algoritmo GRASP por sus siglas en inglés Greedy Randomized Adaptive Search Procedure, es cercana ya que este problema pertenece al conjunto de resolución en tiempo polinomial no determinista o NP-complejos (Nondeterministic Polynomial time).

1.2. Estado del arte

2 Objetivos

En este capítulo se indican los objetivos, tanto principales como secundarios, para la realización de este trabajo fin de grado:

2.1. Objetivos principales

- Estudio y comprensión del problema de la búsqueda del clique de ratio máximo.
- Implementación del algoritmo metaheurístico GRASP para la resolución del problema.

2.2. Objetivos secundarios

- Aprendizaje del lenguaje de programación Python para el desarrollo del algoritmo metaheurístico GRASP.
- Profundización y mejora en técnicas algorítmicas, de programación y de estructuras de datos para la realización de este trabajo fin de grado.

3 Descripción algorítmica

En este capítulo se describe el algoritmo utilizado para el desarrollo de este trabajo final de grado, exponiendo todas sus características.

3.1. Metaheurística

Que es la metaheurística cuales hay y por que se ha usado este algoritmo.

3.1.1. GRASP

Definición de la metaheurística utilizada.

El pseudocódigo del algoritmo GRASP que se ha empleado para el desarrollo en este proyecto es el siguiente:

Para los métodos con los que es posible obtener las soluciones en la fase constructiva se tiene el siguiente pseudocódigo:

El cuál es común a ambos, solo se diferencian en como obtiene cada uno su mejor nodo a escoger para incluir en su solución, los cuales se pueden definir de la siguiente manera:

1. El constructivo 1 el cual obtiene una solución mediante un algoritmo voraz buscando el mayor ratio de cada nodo adyacente al de partida es el siguiente:
2. El constructivo 2 el cual obtiene una solución mediante un algoritmo voraz buscando el mayor número de vecinos de cada nodo adyacente al de partida es el siguiente:

3.1.1.1. Fase constructiva

Definición de los constructivos.

Algoritmo 1: Pseudocódigo algoritmo GRASP.

```

(1)  $v \leftarrow \text{rnd}(V)$ 
(2)  $S \leftarrow \{v\}$ 
(3)  $CL \leftarrow \{u \in V : (u, v) \in E\}$ 
(4) mientras  $|CL| \neq 0$  hacer
(5)    $g_{\min} \leftarrow \min_{c \in CL} g(c)$ 
(6)    $g_{\max} \leftarrow \max_{c \in CL} g(c)$ 
(7)    $\mu \leftarrow g_{\max} - \alpha(g_{\max} - g_{\min})$ 
(8)    $RCL \leftarrow \{c \in CL : g(c) \geq \mu\}$ 
(9)    $u \leftarrow \text{rnd}(RCL)$ 
(10)   $S \leftarrow S \cup \{u\}$ 
(11)   $CL \leftarrow CL \setminus \{u\} \setminus \{w : (u, w) \notin E\}$ 
(12) fin
(13) devolver  $S$ 

```

Algoritmo 2: Constructivo voráz

```

(1)  $S \leftarrow \emptyset$ 
(2)  $\text{Adyacentes} \leftarrow \text{SeleccionAdyacentes}(\text{nodo})$ 
(3) mientras  $|\text{Adyacentes}| \neq 0$  hacer
(4)    $\text{candidato} \leftarrow \text{buscarMejor}(\text{adyacente})$ 
(5)   si  $\text{formaClique}(\text{candidato})$  entonces
(6)      $\text{Adyacentes} \leftarrow \text{Adyacentes} \cap \text{SeleccionAdyacentes}(\text{candidato})$ 
(7)      $S \leftarrow S \cup \{\text{candidato}\}$ 
(8)   fin
(9)   en otro caso
(10)     $\text{Adyacentes} \leftarrow \text{Adyacentes} \setminus \{\text{candidato}\}$ 
(11)  fin
(12) fin
(13) devolver  $S$ 

```

Algoritmo 3: Pseudocódigo método buscarMejor de tipo ratio.

```
(1)  $ratio \leftarrow -1$ 
(2)  $nodoElegido \leftarrow NULO$ 
(3) para  $nodo$  hacer
(4)    $ratioNodo \leftarrow calcularRatio(nodo)$ 
(5)   si  $ratioNodo > ratio$  entonces
(6)      $ratio \leftarrow ratioNodo$ 
(7)      $nodoElegido \leftarrow nodo$ 
(8) devolver  $nodoElegido$ 
```

Algoritmo 4: Pseudocódigo método buscarMejor de tipo adyacentes.

```
(1)  $vecinos \leftarrow -1$ 
(2)  $nodoElegido \leftarrow NULO$ 
(3) para  $nodo$  hacer
(4)    $vecinosNodo \leftarrow SeleccionAdyacentes(nodo)$ 
(5)   si  $|vecinosNodo| > vecinos$  entonces
(6)      $vecinos \leftarrow |vecinosNodo|$ 
(7)      $nodoElegido \leftarrow nodo$ 
(8) devolver  $nodoElegido$ 
```

3.1.1.2. Fase de búsqueda

Definición de la fase de búsqueda.

4 Descripción informática

En este capítulo se describe el desarrollo completo del proyecto, desde el diseño hasta la implementación de este, así como la metodología utilizada para la correcta evolución del mismo.

4.1. Diseño

aaa

4.2. Implementación

La implementación de este proyecto se basa en la ejecución de un script¹ escrito en lenguaje Python el cual parte de una clase en la que se encuentra la siguiente sentencia de código:

```
if __name__ == "__main__"
```

la cual posibilita la ejecución del script como aplicación standalone² mediante el comando:

```
python grasp_main.py
```

Este script, en adelante Grasp Main, tiene la configuración necesaria para ajustar el programa:

- Número de iteraciones a realizar por cada fichero.
- Ruta donde se encuentran los ficheros de definición de los grafos a procesar.
- Ruta de los ficheros de resultados generados por el programa.

¹script

²standalone

Grasp Main se encarga de recorrer recursivamente los ficheros que se encuentran en la ruta de recursos definida y por cada uno de los ficheros encontrados crea un objeto de tipo Instance, añadiendo la información del grafo:

- Número de nodos.
- Número de aristas.
- Estructura de datos con los nodos del grafo.

Esta estructura de datos contiene tantos objetos de tipo Node como tengo el grafo, cada uno de ellos con la siguiente información:

- Identificador del nodo.
- Valor del peso p.
- Valor de peso q.
- Grado del nodo.
- Estructura de datos con las relaciones de este nodo con otros nodos del grafo.

Tras terminar esta operación, realiza el procesado un número N de veces, según se haya definido previamente en la configuración de la aplicación en Grasp Main, y por cada tipo de constructivo de los que dispone la aplicación, en este caso, constructivo ratio y constructivo adyacente, los cuales serán explicados más adelante. La implementación del algoritmo GRASP se encuentra en la clase SolutionGrasp y contiene los métodos necesarios para la obtención mediante el algoritmo de una solución, partiendo de la función *find_grasp_solution*, la cual inicializa los siguientes datos:

- vertex, el cuál es obtenido de manera aleatoria entre todos los nodos del grafo.
- solution, conjunto inicializado con el vértice obtenido anteriormente.
- cl, lista de candidatos posibles para encontrar una solución.

//TODO ¿se debe explicar cada método que forma parte de esta funcion?

Esta función se apoya en otra auxiliar, nombrada como *find_solution_aux*, la cuál implementa el algoritmo utilizado, de esta manera se desacopla el tipo de algoritmo de la búsqueda de una solución, dando la posibilidad a un cambio posterior si fuera necesario. Dicha función, procesará los nodos del grafo tal como se describió en el capítulo

anterior. Para la fase constructiva del algoritmo, según el tipo elegido en la configuración inicial, se creará un objeto del constructivo específico, `SolutionGreedyRatio` o `SolutionGreedyAdjacent`. Estos heredan de la clase de la clase abstracta³ `SolutionGreedy`, la cuál tiene la información compartida por ambos tipos, y delega la implementación de la función *find_better* en cada una de las clases específicas, quienes mediante un algoritmo de tipo voráz buscan una solución factible en un tiempo muy reducido.

Para mantener cierta información en un único lugar se ha implementado la clase `GraphUtils`, la cuál contiene información necesaria para los grafos y métodos útiles para usarse durante el procesado de los ficheros.

Para probar las diferentes funciones del proyecto se han escrito casos de prueba mediante la librería de Python `unittest`⁴, para conseguir un código tolerable a posibles fallos y mantenible.

//TODO Explicación del proceso, partiendo de la lectura de fichero, luego crear solución voraz

4.3. Metodología empleada

Para el desarrollo de este proyecto se ha elegido por una metodología ágil ya que como afirma Joaquín Alviz (2016) [1] "para poder contar con versiones del producto funcional que se puedan mostrar al cliente conforme se va mejorando y completando, es necesario contar con un enfoque ágil, flexible y con una retroalimentación constante", por esto para la elaboración de este trabajo final de grado se ha optado por seguir una metodología ágil de tipo iterativa e incremental, lo que permite evolucionar el proyecto progresivamente e ir adaptando los requisitos del cliente, en este caso los tutores del proyecto, en el menor tiempo posible, mejorando así la calidad del producto final con el menor esfuerzo.

Estas iteraciones e incrementos de funcionalidad se han realizado durante todo el desarrollo del proyecto, mediante reuniones, con un lapso de aproximadamente 3 a 4 semanas entre ellas, corrigiendo errores de la iteración anterior si los hubiera y aumentando la funcionalidad del producto a entregar tanto en funcionalidad como en calidad, de

³clase abstracta

⁴unittest

esta manera se consigue una evolución progresiva y segura sobre el producto y los requerimientos que el proyecto exige.

Para mantener el control y administrar lo correspondiente sobre las tareas a realizar, las que se han realizado y las realizadas del proyecto, se ha usado la utilidad Trello⁵, la cuál mediante tarjetas sobre el tablero del proyecto permite conocer las tareas del proyecto, así como conocer su estado, y añadir nuevas si así fuera necesario.

En cuanto al mantenimiento de versiones del proyecto se ha usado el sistema de control de versiones Git⁶ a través del portal de alojamiento de repositorios GitHub⁷, para interactuar entre el repositorio local y el repositorio remoto se ha optado por hacer uso tanto de la terminal mediante los comandos del propio sistema de control de versiones Git como del cliente para tal propósito GitKraken⁸, el cuál permite mediante su sencilla e intuitiva interfaz mantener un control exhaustivo sobre las ramas y versionado de las distintas piezas de código del proyecto en el que se trabaja, así como revisar posibles conflictos que se produzcan.

⁵<https://trello.com/es>

⁶<https://git-scm.com/>

⁷<https://github.com/>

⁸<https://www.gitkraken.com/>

5 Resultados

En este capítulo se describen los diferentes recursos tanto hardware como software para la realización de este trabajo fin de grado, así como a partir de estos, se han obtenido los resultados para su posterior análisis.

5.1. Recursos utilizados

A continuación se detallará la máquina y software para el desarrollo y procesado de las instancias para la comprobación del algoritmo, y las instancias utilizadas que describen diferentes casos y situaciones reales.

5.1.1. Descripción de la máquina utilizada

Para la realización de las diversas pruebas y procesado de las instancias de este problema, se ha utilizado una máquina con las siguientes características:

- **Procesador:** Intel Core i5 2,7 GHz
- **Memoria RAM:** 8 GB 1867 Mhz DDR3

El desarrollo del código se ha realizado mediante el lenguaje de programación Python¹, en su versión 3.7.4, a través del IDE², PyCharm³ de JetBrains en su versión 2019.3.1.

5.1.2. Instancias utilizadas

Las instancias con las que se ha contado para comprobar la eficiencia del algoritmo desarrollado han sido proporcionadas por los estudios previos en los que se basa y compara este trabajo final de grado, estas, hacen referencia a diferentes conjuntos de

¹<https://www.python.org/>

²Integrated Development Environment o entorno de desarrollo integrado.

³<https://www.jetbrains.com/es-es/pycharm/>

grafos, tanto generados de manera aleatoria como obtenidos de diferentes fuentes de datos como precios del mercado de valores y turbinas de viento. A continuación se detallan los diferentes conjuntos:

- **Conjuntos de tipo A y B:** Estos conjuntos son instancias de grafos generadas mediante una distribución de probabilidad uniforme, variando entre los 100 y los 500 nodos, así como la densidad del mismo que varía entre 45,78 % y 53,64 %.
- **Conjunto de tipo C:** Los datos pertenecientes a las instancias de este tipo hacen referencia a datos de los precios del mercado de valores.
- **Conjunto de tipo D:** Las instancias de este conjunto son datos para la construcción de turbinas de viento, donde cada nodo representa una localización de estas turbinas y sus pesos son la media de la velocidad del viento y el coste de construcción de una turbina en ese punto.
- **Conjunto de tipo E:** Estas instancias están extraídas del segundo y décimo DIMACS Implementation Challenge ⁴, donde cada nodo tiene un peso $p_i = 1$ y un peso $q_i = 2$. Adicionalmente se ha añadido un nodo más a cada instancia de este conjunto, el cual está conectado al resto de nodos de la misma, con un peso $p_i = 1$ y un peso $q_i = 1$, donde i es el número del nodo dentro de esa instancia.
- **Conjunto de tipo F:** Estas instancias son las mismas que en el conjunto E pero en este caso los pesos de cada nodo son $p_i = i$ y $q_i = |V| - i + 1$, donde i es el número del nodo dentro de esa instancia.

5.2. Análisis de los resultados

Mostrar y comentar los resultados obtenidos.

⁴<http://dimacs.rutgers.edu/programs/challenge/>

6 Conclusiones

En este capítulo se describen las conclusiones finales alcanzadas tras el desarrollo del proyecto, así como las lecciones aprendidas durante el mismo.

6.1. Consecución de los objetivos

Los objetivos establecidos al comienzo del proyecto son:

- aaaa

Estos objetivos se han cumplido de manera satisfactoria:

6.2. Conocimientos adquiridos

Durante el proceso de realización de este trabajo fin de grado me he encontrado con diferentes retos los cuales me han permitido adquirir muchos conocimientos sobre el desarrollo de software y la gestión de un proyecto. También he adquirido y profundizado en conceptos sobre algoritmia y estructuras de datos para obtener soluciones mejores y más eficientes al planteamiento de problemas. Por lo tanto destaco:

- El aumento y mejorar mis conocimientos en el lenguaje de programación Python, usado para la implementación de este proyecto.
- La comprensión sobre conceptos de algoritmia y estructuras de datos, así como la mejora continua del código, enfocados en la resolución de problemas de optimización.
- El aprendizaje sobre \LaTeX , al utilizarlo para documentar el trabajo fin de grado.

6.3. Líneas de desarrollo futuras

ccc

Bibliografía

- [1] *Principios del desarrollo ágil. Cómo aplicar metodologías ágiles*. URL: <https://www.renacen.com/blog/principios-del-desarrollo-agil-metodologias-agiles/>.