



Universidad
Rey Juan Carlos

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

*«Victory is always possible for the person who refuses to stop fighting»,
Napoleon Hill*

Resumen

En la presente memoria de este Trabajo Fin de Grado se trata el problema del clique de ratio máximo también llamado Maximal Ratio Clique Problem (MRCP), el cual consiste en encontrar un subgrafo completo o clique, de ratio máximo.

A través de esta memoria se intenta explicar, de una manera clara, como se ha abordado el problema desde el diseño hasta la implementación, pasando por el estudio y comprensión de la metaheurística empleada, para la búsqueda óptima de una solución factible. La implementación se ha realizado mediante el algoritmo Greedy Randomized Adaptive Search Procedure (GRASP) o Procedimiento de Búsqueda Voráz Aleatoria y Adaptativa, el cual ofrece buenas soluciones en un tiempo de cómputo razonable.

Este problema es una derivación del clásico Problema del Clique Máximo o Maximal Clique Problem (MCP) que consiste en encontrar el clique máximo en un grafo.

El MRCP tiene aplicaciones tan actuales como el machine-learning o las redes sociales, ya que un clique es un grupo de personas con intereses afines y por ello se ha intentado implementar un algoritmo que trate los datos de la manera más eficiente posible, de forma que se agilice el tratamiento de estos posteriormente.

Agradecimientos

Todo este trabajo no habría sido posible sin mi familia, sin su apoyo y cariño, en especial a Rebeca, 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	v
Agradecimientos	vii
1. Introducción	1
1.1. Estructura de la memoria	1
1.2. Motivación del problema	2
1.3. Conceptos previos	4
1.3.1. Optimización combinatoria	4
1.3.2. Heurística y Metaheurística	4
1.3.3. Clique	6
1.4. Definición del problema	8
1.4.1. Problema del clique de ratio máximo	8
1.5. Estado del arte	10
2. Objetivos	13
2.1. Objetivos principales	13
2.2. Objetivos secundarios	13
3. Descripción algorítmica	15
3.1. Metaheurística GRASP	15
3.1.1. Fase constructiva	17
3.1.2. Fase de búsqueda	19
4. Descripción informática	21
4.1. Diseño	21
4.2. Implementación	24
4.3. Metodología empleada	26

5. Resultados	29
5.1. Recursos utilizados	29
5.1.1. Descripción de la máquina utilizada	29
5.1.2. Instancias utilizadas	30
5.2. Análisis de los resultados	30
5.2.1. Experimentos preliminares	31
5.2.2. Experimento final	33
6. Conclusiones	39
6.1. Consecución de los objetivos	39
6.2. Conocimientos adquiridos	40
6.3. Líneas de desarrollo futuras	40
Bibliografía	43

Índice de figuras

1.1. Grafo de relaciones en Facebook.	2
1.2. Grafo de telecomunicaciones.	3
1.3. Grafo sección del metro de Madrid.	3
1.4. Representación estándar y en grafo de la molécula del Tolueno.	3
1.5. Clasificación de metaheurísticas	5
1.6. Diagrama de un grafo.	6
1.7. Cliques de cardinalidad 2 del grafo.	7
1.8. Cliques de cardinalidad 3 del grafo.	7
1.9. Clique máximo del grafo.	8
1.10. Diagrama de problemas NP.	9
1.11. MRCP clasificado en los problemas NP-difícil.	10
4.1. Diagrama de clases del proyecto.	23
4.2. Ejemplo del sistema de tarjetas de Trello.	26

Índice de tablas

5.1. Resultados de los experimentos preliminares con constructivo adyacentes.	31
5.2. Resultados de los experimentos preliminares con constructivo ratio. . . .	32
5.3. Resultados del experimento final.	34

Listado de abreviaciones

API Application Programming Interface. 2

GIL Global Interpreter Lock. 41

GRASP Greedy Randomized Adaptive Search Procedure. 15, 21, 25, 31, 33, 39, 40

IDE Integrated Development Environment. 29

MCP Maximum Clique Problem. 8, 40

MRCP Maximum Ratio Clique Problem. 9

MSM Multi-Start Methods. 11

MWCP Maximum Weight Clique Problem. 9, 39

VNS Variable Neighborhood Search. 11

Capítulo 1

Introducción

En este capítulo se presenta la motivación del problema, los conceptos previos que ayudarán al lector a entender mejor el desarrollo, siguiendo con la definición del problema, y por último, se muestra una revisión del estado del arte relacionado con este problema.

1.1. Estructura de la memoria

La memoria de este trabajo final de grado se estructura de la siguiente manera:

- Capítulo 1, en este capítulo se introduce el tema a tratar, describiendo el problema y el estado del arte del mismo, así como conceptos previos a tener en cuenta.
- Capítulo 2, en este capítulo se muestran los objetivos que se esperan alcanzar con este trabajo final de grado.
- Capítulo 3, en este capítulo se describe de manera algorítmica la metaheurística empleada para abordar el problema tratado.
- Capítulo 4, en este capítulo se explica el desarrollo de las funciones para obtener una solución al problema.
- Capítulo 5, en este capítulo se exponen los resultados recopilados durante el procesamiento del problema, así como el análisis de los mismos.
- Capítulo 6, en este capítulo se muestran las conclusiones finales obtenidas durante todo este trabajo final de grado.

1.2. Motivación del problema

En la actualidad, los datos se han convertido en una pieza fundamental en la vida cotidiana de las personas. Tanto su obtención como su posterior tratamiento son grandes retos que deben ser estudiados y, por supuesto, realizar esto de una manera eficiente y en el menor tiempo posible, es un reto aún mayor. A partir de esto se obtiene el término grafo, el cual es fundamental en este ámbito y que se define como la relación entre nodos o vértices mediante uniones llamadas aristas, y esto, es lo realmente interesante a estudiar, las relaciones existentes entre los distintos nodos.

Existen muchos casos en los que se puede realizar la obtención y tratamiento de estos datos, como por ejemplo, la red social Facebook¹, que como se observa en la figura 1.1 define su grafo como las relaciones entre las personas y aficiones que les unen. Además ofrece una API² con la que poder interactuar con los datos que son recopilados en ella [8].

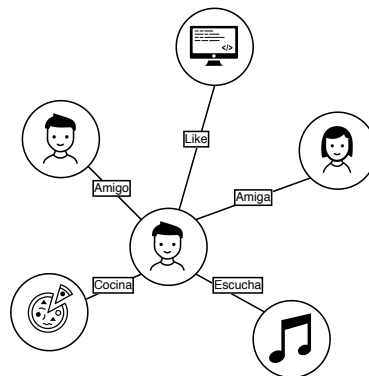


FIGURA 1.1: Grafo de relaciones en Facebook.

Por otro lado, también muy cercano a la vida cotidiana de las personas, se encuentran las redes de telecomunicaciones, figura 1.2, que interconectan todos los aparatos electrónicos, que de una manera u otra se comunican entre ellos, y las redes de metro, figura 1.3, que permiten el movimiento de las personas en la ciudad y pueden ser estudiadas por uso, localización, etcétera.

¹<https://es-es.facebook.com/>

²Conjunto de métodos que forman parte de una librería y que a modo de capa de abstracción son publicados para que puedan ser usados en otros desarrollos software.

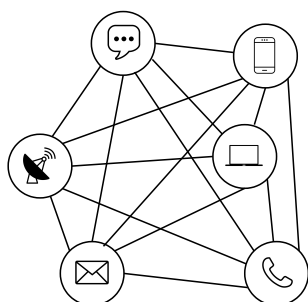


FIGURA 1.2: Grafo de telecomunicaciones.

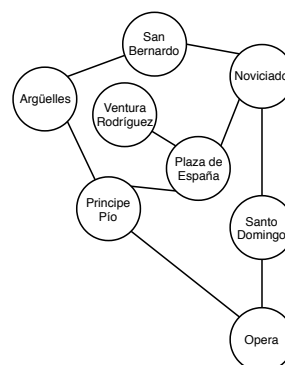


FIGURA 1.3: Grafo sección del metro de Madrid.

Y por último, otra representación de grafo, es el de las moléculas y los enlaces químicos [10], que se pueden representar como en la figura 1.4.

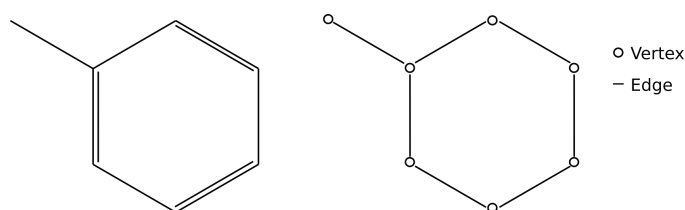


FIGURA 1.4: Representación estándar y en grafo de la molécula del Tolueno.

Fuente: <https://www.blopig.com/blog/2019/01/graph-based-methods-for-cheminformatics>

Estos ejemplos, tienen en común las conexiones entre todos los nodos que forman su estructura de grafo, y son muchas las aplicaciones e información que se obtiene de ellos como, nuevas terapias [2], visión computacional y reconocimiento de patrones [5], análisis del mercado bursátil para conocer nuevas formas de predicción de éxito en inversiones [17].

Por todo esto, es muy importante tratar los datos de forma adecuada y responsable, así como las relaciones entre ellos eficientemente. Para ello se han creado incluso base de datos orientadas a grafos, como es el caso de Neo4j³, con el fin de entender las relaciones y así poder hallar nuevas técnicas y más posibilidades para afrontar problemas y resolverlos de una manera mucho más ágil.

³<https://neo4j.com/>

1.3. Conceptos previos

Para comprender mejor todas las explicaciones que se van a exponer a lo largo del documento, se describen a continuación los conceptos más importantes.

1.3.1. Optimización combinatoria

La optimización combinatoria es el área, dentro de las matemáticas aplicadas, que se encarga de maximizar o minimizar una función en un espacio de soluciones, el cuál debe ser finito. Los problemas que pertenecen a este área tienen en común la dificultad de encontrar soluciones factibles, puesto que existen muchas posibles y alguna de estas es óptima.

Algunos casos conocidos son el problema de la mochila⁴ o el problema del vendedor viajero⁵. Estos problemas tienen un planteamiento sencillo pero son difíciles de solucionar, y mediante ciertos algoritmos, como se explicará en la sección 1.3.2, se pueden obtener soluciones factibles en tiempos de cómputo reducidos.

En la actualidad es un campo con un gran crecimiento ya que “muchos de los problemas de la vida real pueden ser formulados mediante optimización combinatoria” [13].

1.3.2. Heurística y Metaheurística

Se define heurística, según el Diccionario de la Real Academia de la Lengua Española [7], como “técnica de la indagación y del descubrimiento” y “en algunas ciencias, manera de buscar la solución de un problema mediante métodos no rigurosos, como por tanteo, reglas empíricas, etc.”. Aplicándolo a temas científicos se define como el proceso de creación de medios, estrategias y principios para alcanzar un objetivo eficaz al problema dado [6]. El término heurística fue acuñado por George Polya en su libro “How to Solve It” [21], más tarde traducido a “Cómo plantear y resolver problemas” [20].

Añadiendo al término heurística el prefijo “meta”, procedente del griego que significa “más allá” o “nivel superior”, se puede definir metaheurística como el conjunto de procedimientos heurísticos combinados para obtener una solución a un problema

⁴<https://www.sciencedirect.com/topics/computer-science/knapsack-problem>

⁵<https://www.sciencedirect.com/topics/computer-science/travelling-salesman-problem>

que no tiene un algoritmo heurístico específico o su aplicación es ineficiente [15]. Este término lo acuñó Fred Glover en sus trabajos sobre búsqueda tabú en 1986 [11].

Existen una gran variedad de algoritmo metaheurísticos que se pueden clasificar de diferentes maneras según el enfoque, en la figura 1.5 se muestra una posible clasificación.

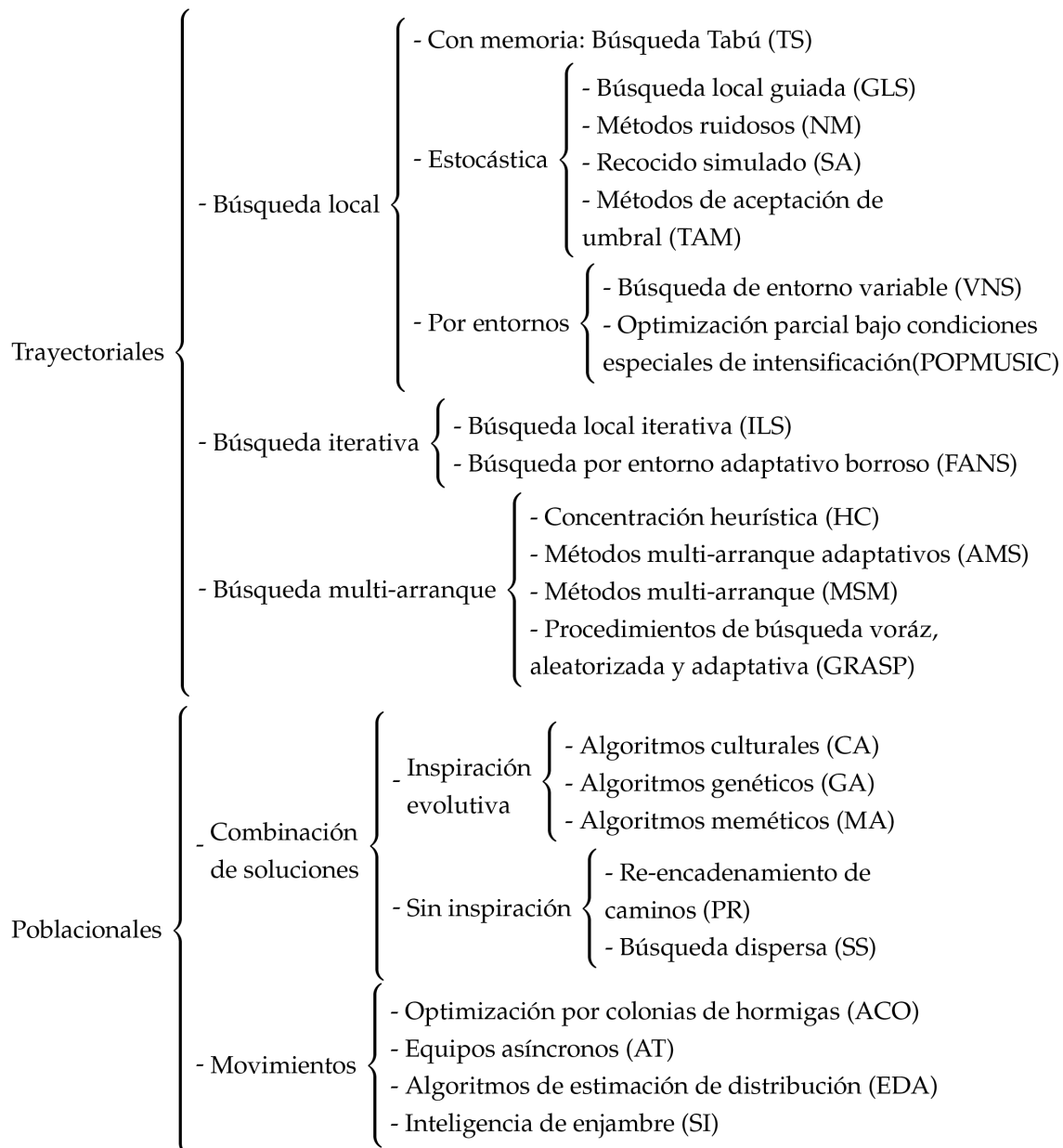


FIGURA 1.5: Clasificación de metaheurísticas

Dos conceptos a tener en cuenta en un algoritmo metaheurístico son la intensificación, que indica la exhaustividad con la que el algoritmo explora una región, y la diversificación, que indica como el algoritmo es capaz de explorar nuevas regiones del espacio de soluciones [1].

1.3.3. Clique

El término clique se puede asemejar a un grupo de personas, las cuales serían los nodos o vértices de un grafo, a las que les unen los mismos intereses por algún tema en concreto, estas uniones serían las aristas que unen los nodos del grafo [14].

En teoría de grafos, un clique, es el subgrafo perteneciente a un grafo, en el cuál todos sus nodos o vértices son adyacentes entre sí, es decir, todo par de nodos o vértices están conectados mediante una arista. En términos matemáticos se describe como, dado un grafo $G = (V, E)$ donde V indica el conjunto de vértices del grafo y E indica el conjunto de aristas del grafo [30], un clique C se define como:

$$C \subseteq V(G) \wedge u, v \in C \wedge u \neq v \Rightarrow u, v \in E(G)$$

Mostrandolo de una manera gráfica se indica en la figura 1.6 un grafo compuesto por el conjunto de nodos $V = \{A, B, C, D, E\}$.

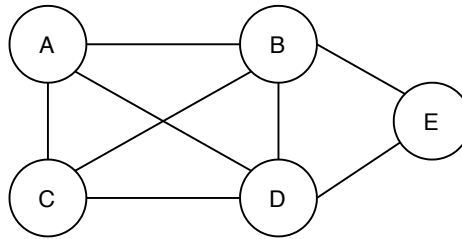


FIGURA 1.6: Diagrama de un grafo.

En este grafo de ejemplo se encuentran los cliques con cardinalidad igual a 2 como se indican resaltado en negro en la figura 1.7.

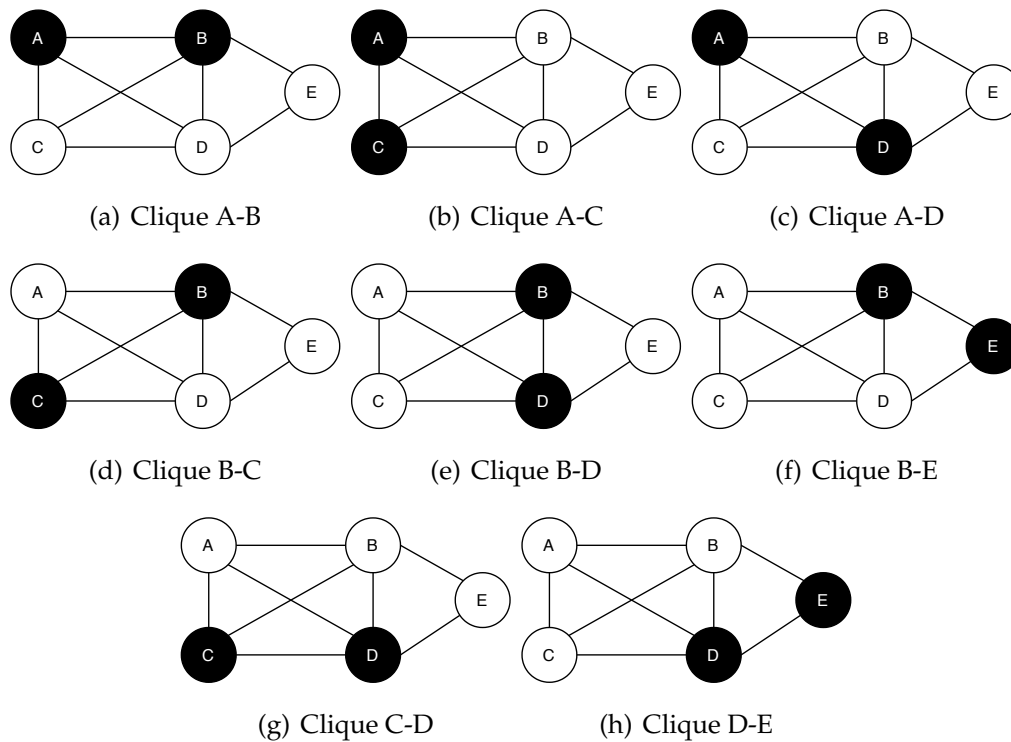


FIGURA 1.7: Cliques de cardinalidad 2 del grafo.

También se encuentran, como indica la figura 1.8, los cliques con cardinalidad igual a 3.

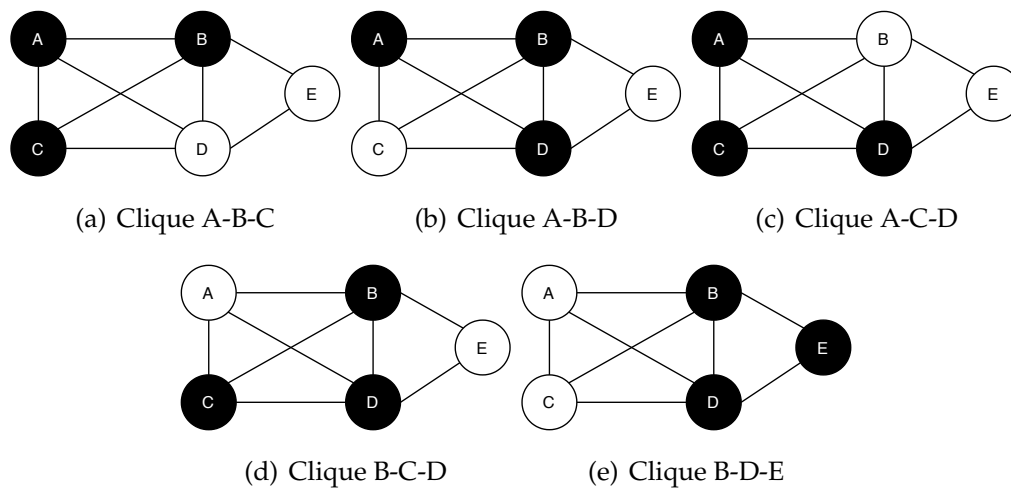


FIGURA 1.8: Cliques de cardinalidad 3 del grafo.

Cabe destacar que los nodos, por si solos, también forman clique.

Una característica importante de un clique es que este puede ser máximo [32]. Esto quiere decir que no se pueden añadir más nodos adyacentes que cumplan las restricciones necesarias para formar un clique, y a su vez, es el de mayor tamaño del grafo, como se muestra en la figura 1.9, a diferencia de un clique que se podría denominar simple [31].

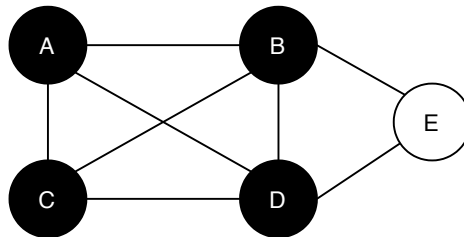


FIGURA 1.9: Clique máximo del grafo.

En este caso, se obtiene $\omega(G) = 4$, donde ω denota el número de vértices del clique, su cardinalidad.

1.4. Definición del problema

1.4.1. Problema del clique de ratio máximo

Como se introdujo en la sección 1.3.3, el concepto clique es fundamental y muy estudiado, en concreto, la búsqueda del clique máximo dentro de un grafo. A este problema se le conoce como el problema del clique máximo o MCP (Maximum clique problem), y es catalogado como un NP-completo.

En teoría de complejidad computacional, a los problemas denominados como NP, acrónimo de non-deterministic polynomial time o tiempo polinomial no determinista, se les conoce como el conjunto de problemas que se pueden resolver en un tiempo polinómico por una máquina de Turing no determinista. Esta clasificación, como se muestra en la figura 1.10, también contiene todos los problemas de tipo P y de tipo NP-completos, como es el caso del problema del clique máximo.

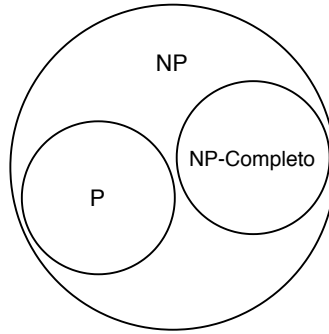


FIGURA 1.10: Diagrama de problemas NP.

Una variante de este problema es el llamado problema del clique de peso máximo o MWCP (Maximum Weight Clique Problem), en el que se asocia un peso no negativo a cada vértice, y cuyo objetivo es encontrar un clique con el máximo valor en la suma de los pesos de sus vértices. Este problema está estrechamente ligado al problema tratado en este trabajo.

Otra variante, es la tratada en este trabajo y que se corresponde con el problema del clique de ratio máximo o MRCP (Maximum Ratio Clique Problem) que trata de encontrar el subgrafo completo de ratio máximo. Este ratio se define como la proporción de la suma de los pesos no negativos asociados a los vértices del grafo como se indica a continuación:

$$\frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n q_i x_i}$$

donde p y q son pesos no negativos asociados a cada vértice i , y x se determina como:

$$x_i = \begin{cases} 1 & : \text{si el vértice } i \text{ forma parte de la solución.} \\ 0 & : \text{en otro caso} \end{cases}$$

Como se expone en el modelo de ecuación 1.1, partiendo de un grafo simple no dirigido $G = (V, E)$, donde V se asocia al conjunto de vértices pertenecientes al grafo, $\{v_1, \dots, v_n\}$, y E es el conjunto de aristas que conectan los vértices del grafo, $\{(v_i, v_j)\}$ tal que $i \neq j$ y $v_i, v_j \in V$, y suponiendo que los pesos asociados a cada vértice son positivos, se obtiene un clique máximo \hat{S} , siempre y cuando se cumplan las restricciones 1.2 - 1.4.

$$\begin{array}{ll} \text{maximizar} & f = \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n q_i x_i} \end{array} \quad (1.1)$$

sujeto a :

$$x_i + x_j \leq 1 : \forall (v_i, v_j) \notin E, i \neq j, \quad (1.2)$$

$$\sum_{i=1}^n (a_{ij}) x_i \geq 1 : \forall v_j \in V, \quad (1.3)$$

$$x_i \in 0, 1 : \forall v_i \in V. \quad (1.4)$$

El problema del clique de ratio máximo ha sido catalogado como un problema NP-difícil o NP-complejo como se indica en la figura 1.11, por lo que no es posible obtener una solución factible por métodos heurísticos o exactos.

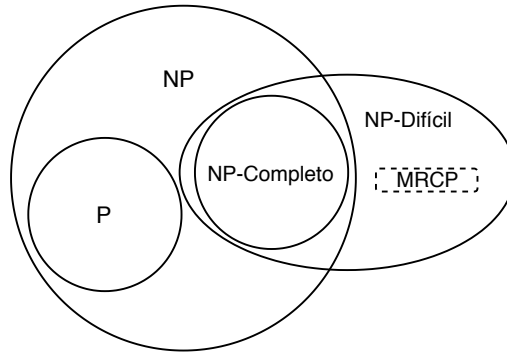


FIGURA 1.11: MRCP clasificado en los problemas NP-difícil.

1.5. Estado del arte

Si bien es cierto que no existe demasiada literatura al respecto, ya que se trata de un problema poco estudiado en comparación con el clásico problema de búsqueda de la clique máxima, [3], [18], [33], o [22], este análisis, y todo lo que rodea al mismo, ha sido documentado mediante la literatura que se explica a continuación.

Un estudio muy cercano al tratado en este trabajo es el problema de la búsqueda del clique de peso máximo, en [29] y [27] se implementan algoritmos de búsqueda local, llamados SCCWalk y SCCWalk4L, en los que se hace uso del muestreo no determinista y el machine-learning para abordar el problema, eliminando variables de decisión que no formarían parte de una solución óptima.

En el documento desarrollado por Samyukta Sethuraman y Sergiy Butenko [25] se trata el problema desde tres puntos de vista, mediante el IBM CPLEX, para resolver el problema de manera lineal, la aplicación de la búsqueda binaria y el método de Newton.

En [16] se trata el problema basándose en el enfoque eficiente que dan las funciones de diferencia de convexos (DC) y sus algoritmos (DCA), que proveen resultados competitivos y a su vez introducen las desigualdades válidas, que ayudan a mejorar el tiempo computacional en la obtención de resultados de calidad al problema.

Cabe destacar entre todos los trabajos el de Dominik Goeke, Mahdi Moeini y David Poganiuch [12] el cuál se ha tomado como referencia para realizar este trabajo. En este artículo, los autores parten de un punto de vista multi-arranque MSM, añadiendo una búsqueda de vecindario variable VNS, lo que permite obtener soluciones de gran calidad con un tiempo de computo menor.

Capítulo 2

Objetivos

En este capítulo se indican los objetivos, tanto principales como secundarios, establecidos al inicio de este trabajo fin de grado.

2.1. Objetivos principales

- Estudio y comprensión del problema de la búsqueda del clique de ratio máximo.
- Estudio, diseño e implementación del algoritmo metaheurístico GRASP para la resolución del problema.

2.2. Objetivos secundarios

- Conocer los distintos algoritmos metaheurísticos existentes.
- Comprender la complejidad computacional relacionada con la búsqueda del clique de ratio máximo.
- 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.

Capítulo 3

Descripción algorítmica

En este capítulo se describe el algoritmo metaheurístico utilizado, exponiendo todas sus características para la obtención de una solución al problema.

3.1. Metaheurística GRASP

El acrónimo GRASP (Greedy Randomized Adaptative Search Procedure) o en castellano procedimiento de búsqueda voráz aleatorizado y adaptativo, fue introducido por primera vez por Feo y Resende en 1995 en su artículo con el mismo nombre [9].

Este algoritmo se basa en el multi-arranque, dónde cada uno de ellos es una iteración de un procedimiento que está constituido por dos partes bien diferenciadas, la fase constructiva, en la que se obtiene una solución de buena calidad, y una fase de mejora, en la que partiendo de la solución obtenida en la fase anterior, se intenta mejorar localmente [1]. En [26] [24] [19] [4] [28] [23] se pueden encontrar diversos documentos en los que se tratan problemas aplicando la metaheurística GRASP.

En el algoritmo 1 se muestra el pseudocódigo de la metaheurística GRASP que se ha empleado para el desarrollo y obtención de una solución preliminar para este problema, y posteriormente, se muestra el algoritmo 5, con el cual se ha refinado esta solución para obtener una mejor.

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  $S$ 

```

Dichos algoritmos se puede explicar, aplicándolo al problema tratado, de la siguiente manera:

Partiendo de un grafo $G = (V, E)$ donde V son los vértices o nodos del grafo, y E las aristas que unen estos nodos, se toma un vértice v aleatorio de entre los vértices del grafo. El vértice v se incluye en la solución S ya que cumple con las restricciones del problema, descritas en la sección 1.4.1. A partir de v se construye la lista de candidatos CL , definida como todos los nodos adyacentes a v que forman parte de la lista de nodos del grafo de partida. A continuación se toma un elemento de la lista de candidatos, con el que se obtiene mediante una función voraz un listado de nodos. Esta función será determinada antes de iniciar el proceso e indicará si se deben obtener los nodos según su ratio o según el número de adyacentes al nodo tratado, será explicada con mayor detalle en la sección 3.1.1. De este listado de nodos se escogen los valores máximo y mínimo, que junto a α se obtendrá el valor de μ .

El valor de α será configurado antes de este proceso, el cual oscilará entre 0 y 1, marcando la aleatoriedad de la función de manera que, si es igual a 1 la función será menos

aleatoria obteniéndose solo el mejor candidato, y si es igual 0 la función es más aleatoria y se obtendrán todos los candidatos posibles.

El valor μ permite obtener la lista de candidatos restringida *RCL*. Esta lista se genera mediante la lista de candidatos *CL* ordenada de mayor a menor valor, tomando los primeros valores hasta alcanzar el valor de μ .

Tomando esta lista, se elegirá de manera aleatoria un nodo u , como se muestra en la línea 9 del algoritmo 1, el cuál se añadirá a la lista solución *S*, eliminándolo de la lista de candidatos junto con los nodos que no sean adyacentes a este.

Este procedimiento será repetido hasta que la lista de candidatos este vacía, obteniéndose en ese momento la lista *S* final, que conformará la solución preliminar.

A partir de este momento, se obtendrán todos los nodos adyacentes a los nodos que conforman la solución, de una manera ordenada, mediante el valor de su ratio calculado. Estos se irán incluyendo en la solución, comprobando si forman un clique, como indican las restricciones del problema, si no es así, se eliminarán los nodos que causen el conflicto para obtener un clique. Posteriormente se añadirán todos los nodos adyacentes a este, obteniendo de esta manera un clique máximo y solución factible al problema.

3.1.1. Fase constructiva

En esta fase se ha usado el algoritmo 2 para definir la función de ambos constructivos, ya que esta es común, y sólo se diferencian en como obtiene cada uno el mejor nodo a incluir en su resultado.

Algoritmo 2: Constructivo voráz.

```

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

```

En primer lugar se define el algoritmo 3 para el constructivo que obtiene una solución mediante un algoritmo voraz, buscando el mayor ratio de cada nodo adyacente al de partida.

Algoritmo 3: Pseudocódigo método buscarMejor de tipo ratio.

```

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

```

Y en segundo lugar, se define el algoritmo 4 para el constructivo que obtiene una solución mediante un algoritmo voraz buscando el mayor número de vecinos de cada nodo adyacente inicial.

Algoritmo 4: Pseudocódigo método buscarMejor de tipo adyacentes.

```

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

```

3.1.2. Fase de búsqueda

Para esta segunda fase, se define el siguiente algoritmo 5, el cuál parte de la solución obtenida previamente en la fase constructiva.

Algoritmo 5: Pseudocódigo algoritmo búsqueda local.

```

(1)  $vecinos \leftarrow \emptyset$ 
(2)  $vecinos \leftarrow obtenerVecinos(solucion)$ 
(3)  $vecinosOrdenados \leftarrow ordenarVecinos(vecinos)$ 
(4) para  $nodo \in vecinosOrdenados$  hacer
(5)    $solucion \leftarrow incluirNodo(solucion, nodo)$ 
(6)   si  $no esClique(solucion)$  entonces
(7)      $solucion \leftarrow excluirNodos(solucion, nodo)$ 
(8)    $solucion \leftarrow incluirAdyacentes(solucion, nodo)$ 
(9) devolver  $solucion$ 

```

Con esta solución se obtienen todos los vecinos de cada nodo que forma parte de ella, ordenados de mayor a menor ratio. Esta ordenación se realiza con el fin de aumentar las posibilidades de obtener un mejor valor de ratio. Para cada uno de estos nodos se

comprueba si forma o no una solución, es decir, un clique. En el caso de que no cumpliera las restricciones, se descartan todos aquellos nodos que impiden que se forme una solución

Un vez se tiene un clique, se añaden todos los nodos posibles, es decir, que sigan formando clique, con el fin de obtener el máximo clique, como marca la restricción del problema.

Capítulo 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

Para la realización del desarrollo del proyecto se ha optado por seguir un paradigma de programación orientada a objetos y se ha planteado la arquitectura que se muestra en el diagrama de clases de la figura 4.1. La implementación parte de un documento principal, en este caso se trata de un script ejecutable, `grasp_main.py`, que contiene la información inicial y las llamadas a las clases y métodos de estas necesarios para obtener los resultados al problema.

A continuación de este parten las clases:

- **Instance:** Esta clase contendrá toda la información relacionada con una instancia o definición de grafo del problema, la cuál se compone de la clase **Node**, que contiene toda la información relativa a un nodo, así como operaciones relacionadas con el mismo.
- **SolutionGrasp:** Esta clase contendrá los métodos necesarios para la implementación del algoritmo GRASP y de la búsqueda local.
- **GraphUtils:** Esta clase está formada por métodos estáticos en su mayoría, y contendrá todos los métodos de utilidad relacionadas con los grafos, así como la posterior exportación a fichero de los resultados obtenidos durante el proceso. También se compone de la clase **Logger**, la cuál ayuda a configurar el sistema de

registros del proyecto, con el que trazar errores y ver el estado del progreso más fácilmente.

También se encuentran las clases:

- **SolutionGreedy**: Esta clase abstracta contiene lo necesario para construir una solución completa mediante un algoritmo voraz. La definición completa se refiere a los atributos `density`, el cuál indica la densidad del grafo del que parte el proceso; `clique`, que contiene el listado de nodos que forman parte de la solución; `sol_value`, el cuál indica el valor de la solución o función objetivo; `cardinality`, que contiene el valor del número de nodos de la solución y `compute_time`, que indica el tiempo total de computo en hallar esa solución.
- **SolutionGreedyRatio**: Clase especializada en la obtención del mejor candidato posible por la condición de ratio.
- **SolutionGreedyNeighbors**: Al igual que **SolutionGreedyRatio**, esta clase se especializa en encontrar el mejor nodo candidato, en este caso, con mayor número de nodos vecinos.

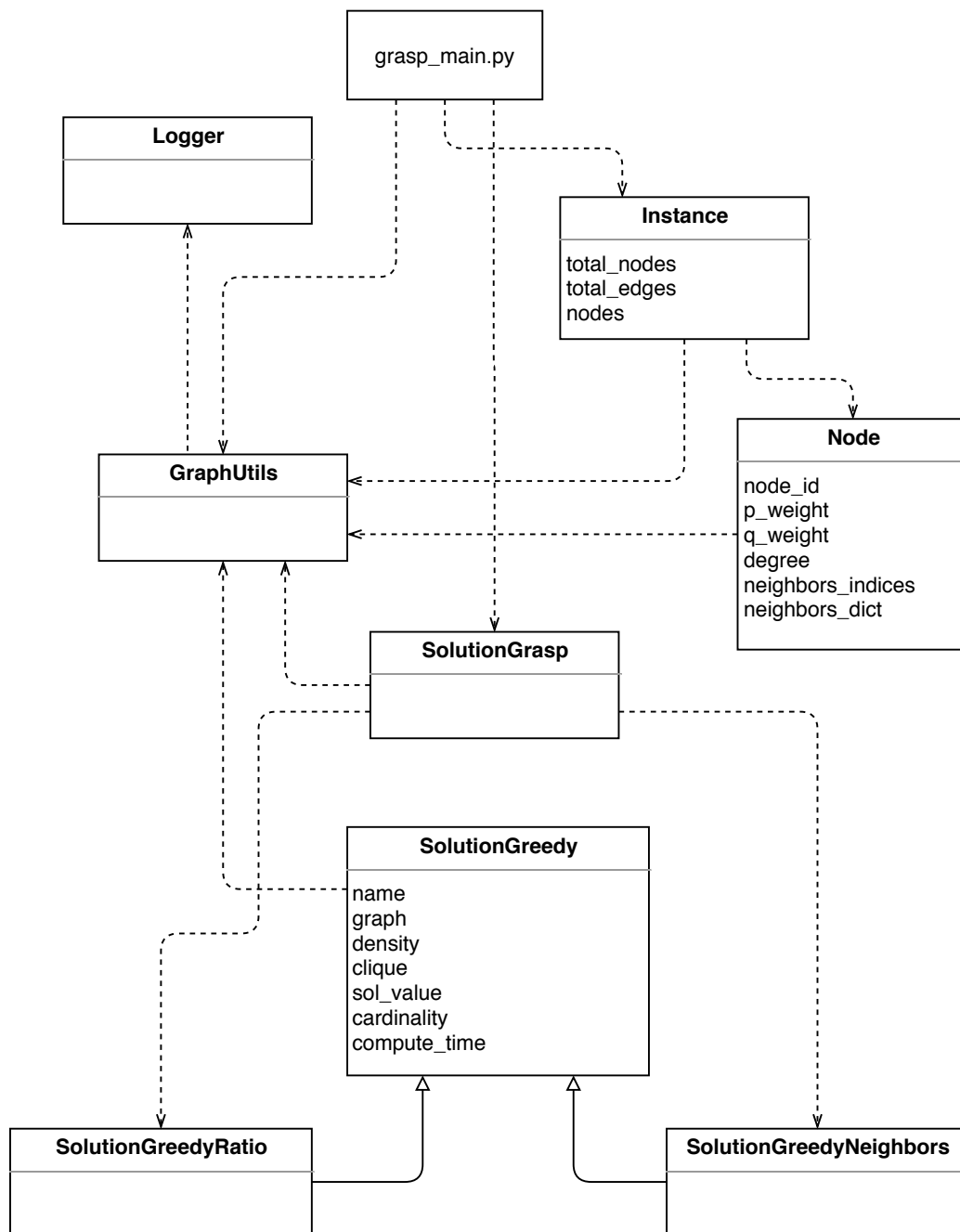


FIGURA 4.1: Diagrama de clases del proyecto.

4.2. Implementación

La implementación de este proyecto se basa en la ejecución de un script¹, escrito en lenguaje Python, que parte de una clase principal caracterizada por contener la siguiente sentencia de código:

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

Dicha sentencia posibilita la ejecución del script mediante el comando:

```
python grasp_main.py
```

Este script, en adelante Grasp Main, tiene la configuración necesaria para ajustar los parámetros del programa:

- Número de iteraciones a realizar por cada fichero.
- Ruta donde se encuentran los ficheros de definición de los grafos a procesar o instancias.
- Ruta de los ficheros de resultados generados por el programa.
- Valores de α para ajustar la aleatoriedad del algoritmo.

Grasp Main se encarga de recorrer recursivamente los ficheros que se encuentren en la ruta de recursos definida y, por cada uno de los ficheros encontrados, crea un objeto de tipo Instance, añadiendo la siguiente 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 tenga el grafo, cada uno de ellos con la siguiente información:

- Identificador del nodo.
- Valor del peso p.
- Valor de peso q.
- Grado del nodo.

¹Secuencia de ordenes o instrucciones que serán interpretadas para su ejecución.

- Estructura de datos con las relaciones de este nodo con otros nodos del grafo.

Tras terminar esta operación se realiza el procesamiento un número N de veces, según se haya definido previamente en la configuración de la aplicación, y por cada tipo de constructivo, los cuales serán explicados más adelante.

A partir de este punto se invoca la función *find_grasp_solution* disponible en la clase *SolutionGrasp* y que implementa al algoritmo GRASP anteriormente descrito. Esta función 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.

Para la fase constructiva del algoritmo, la implementación se apoya en la función *get_g* que obtiene un listado de posibles candidatos. Según el tipo elegido en la configuración inicial, se creará un objeto del constructivo específico, *SolutionGreedyRatio* o *SolutionGreedyAdjacent*. Estos heredan de la clase abstracta² *SolutionGreedy*, la cual 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 voraz buscan una solución factible en un tiempo muy reducido.

Una vez se obtiene el listado de candidatos, es procesado mediante la función *get_rcl*, la cual haciendo uso del valor de μ , calculado como se mostró en el algoritmo 1 con el valor de α , permite obtener la lista de candidatos restringida denominada en el algoritmo como *RCL*. A partir de este momento el siguiente paso es escoger de manera aleatoria un nodo de esta lista e incluirlo en el conjunto solución, eliminando de la lista *CL* los nodos que no son adyacentes a este, ya que no formarían una solución factible. Esta operación es realizada hasta que la lista de candidatos esté vacía.

Para mantener el código ordenado se ha implementado la clase *GraphUtils*, la cuál contiene información necesaria y métodos útiles para el procesamiento de las instancias, así como la exportación a ficheros de tipo CSV³ de los resultados obtenidos.

²En programación orientada a objetos, es un tipo de clase que no puede ser instanciada, y por lo general sirve para definir otras clases de este tipo.

³Es un tipo de ficheros de texto simple en el que se almacenan datos separados en columnas por comas o por punto y coma, y las filas por salto de línea.

Con el fin de probar las diferentes funciones del proyecto se han escrito casos de prueba mediante la librería de Python unittest⁴.

4.3. Metodología empleada

Para el desarrollo de este proyecto se ha optado por seguir una metodología de tipo iterativa e incremental, la cual permite evolucionar el proyecto progresivamente e ir adaptándose a 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 o 4 semanas entre ellas, corrigiendo errores de la iteración anterior si los hubiera y aumentando la funcionalidad del producto a entregar, de esta manera se consigue una evolución progresiva y segura de los requerimientos que el proyecto exige.

Para mantener el control y administrar las tareas a realizar, las que se están desarrollando y las que han finalizado, se ha usado el administrador de proyectos Trello⁵ el cuál mediante tarjetas permite conocer las tareas del proyecto, así como su estado, detalles de la misma y añadir nuevas si fuera necesario. En la figura 4.2 se muestra un ejemplo de uso del sistema de tarjetas que ofrece la herramienta.

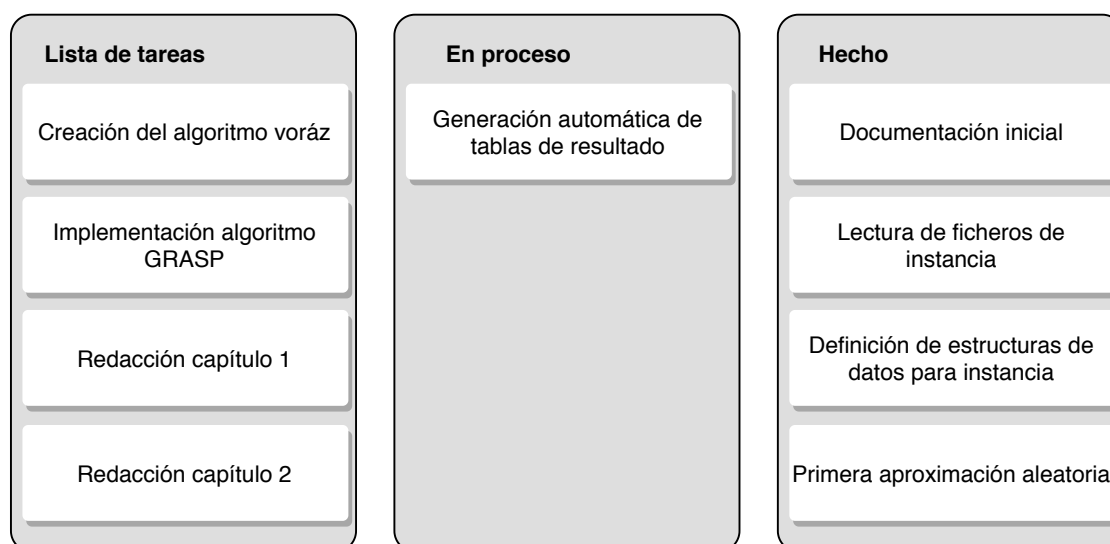


FIGURA 4.2: Ejemplo del sistema de tarjetas de Trello.

⁴<https://docs.python.org/3/library/unittest.html>

⁵<https://trello.com/es>

En cuanto al mantenimiento de versiones del proyecto, se ha usado el sistema de control de versiones Git⁶ y gestionado a su vez a través del portal de alojamiento remoto 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 cual permite, mediante su sencilla e intuitiva interfaz, 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.

⁶<https://git-scm.com/>

⁷<https://github.com/>

⁸<https://www.gitkraken.com/>

Capítulo 5

Resultados

En este capítulo se exponen los diferentes recursos, tanto hardware como software, para la realización de este trabajo fin de grado y cómo, 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 empleados para el desarrollo del código, así como las instancias utilizadas para comprobar la calidad del algoritmo .

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(R) Core(TM) i5-5257U CPU 2.70 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 entorno de desarrollo integrado o IDE (Integrated Development Environment) PyCharm² de JetBrains en su versión 2019.3.1.

¹<https://www.python.org/>

²<https://www.jetbrains.com/es-es/pycharm/>

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 conjuntos de grafos, tanto generados de manera aleatoria como obtenidos de diferentes fuentes de datos, como precios del mercado de valores o turbinas de viento.

A continuación se detallan los diferentes conjuntos en los que se dividen las instancias:

- **Conjuntos de tipo A y B:** Estos conjuntos son instancias de grafos generadas mediante una distribución de probabilidad uniforme, variando su número de nodos entre 100 y 500, la densidad de estos grafos oscila entre el 45,78 % y el 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 Challenge³, 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

En esta sección se muestran los resultados obtenidos tras el procesamiento de las instancias descritas en la sección 5.1.2 y empleando la máquina descrita en la sección 5.1.1. Estos resultados se han obtenido en dos fases, una primera fase preliminar, con

³<http://dimacs.rutgers.edu/programs/challenge/>

un conjunto reducido de las instancias y una segunda fase, con todas las instancias del problema para su posterior comparación con los resultados de estudios previos.

5.2.1. Experimentos preliminares

En esta fase ha sido seleccionado un subconjunto de 22 instancias del problema, con el fin de estudiar la calidad del algoritmo y comprobar los resultados obtenidos mediante distintos valores para α en varias iteraciones, estas se han elegido de todos los conjuntos para ampliar el rango de las pruebas. El número de nodos en estas instancias varía desde los 30 hasta los 1000 nodos, con densidades del 1 al 50 por ciento aproximadamente.

A esta selección de instancias se les ha aplicado los dos constructivos desarrollados, descritos en la sección 4.2, y a su vez se han aplicado los valores de $\alpha \in \{0.25, 0.5, 0.75\}$ y adicionalmente un valor de α aleatorio. Este experimento se ha realizado procesando cada instancia 100 veces, que unido a lo anteriormente descrito, hace un total de 800 iteraciones por fichero.

Durante los experimentos preliminares se han recopilado los datos mostrados en la tabla 5.1 para el constructivo por número de adyacentes y la tabla 5.2 para el constructivo por el valor del ratio. Las tablas están divididas en el uso del algoritmo GRASP y su posterior mejora mediante la búsqueda local y donde:

- f : Indica el valor de la función objetivo calculado.
- c : Indica la cardinalidad del clique encontrado.
- t : Indica el tiempo total empleado en encontrar la solución en segundos.
- α : Indica el valor de α con el que se ha obtenido esa solución.
- Mejor: Indica si la solución obtenida mediante este constructivo es mejor respecto al otro.

TABLA 5.1: Resultados de los experimentos preliminares con constructivo adyacentes.

Instancias	GRASP				GRASP y búsqueda local				Mejor
	f	c	t (sec)	α	f	c	t (sec)	α	
random-1	1.2421	15	0.0620	0.8377	1.2421	15	0.1430	0.5000	No
random-4	1.0168	26	1.6731	0.7500	1.4650	27	5.7613	0.5000	No

Continúa en la siguiente página

Tabla 5.1 Continuación de la página anterior

Instancias	GRASP				GRASP y búsqueda local				Mejor
	f	c	t (sec)	α	f	c	t (sec)	α	
random-5	0.9130	31	3.6512	0.7500	1.4592	32	3.4072	0.5000	No
random-6	0.9814	19	0.0990	0.0631	0.9672	18	0.2590	0.5000	No
random-8	0.8551	22	0.7140	0.7500	0.9229	23	1.9691	0.7500	No
random-10	0.8989	31	3.3232	0.7500	1.0688	30	2.5841	0.4817	No
market-1	1.5434	8	0.0300	0.0111	2.1429	8	0.2550	0.0111	Si
market-2	1.2310	8	0.0510	0.7500	1.3260	8	0.3060	0.5000	Si
market-5	1.4725	8	0.0230	0.2500	2.8116	8	0.3880	0.2053	Si
wind-2005	90 391.7500	4	0.0030	0.2500	90 391.7500	4	0.0160	0.5000	No
wind-2006	87 219.5122	5	0.0050	0.5000	90 034.2000	5	0.0520	0.5000	No
brock200-2	0.5455	12	0.0620	0.7500	0.5455	12	0.8240	0.7500	Si
football	0.5263	10	0.0040	0.2500	0.5263	10	0.1650	0.7500	Si
johnson16-2-4	0.5556	10	0.0640	0.2500	0.5556	10	0.4110	0.5000	Si
karate	0.5455	6	0.0010	0.2500	0.5455	6	0.0100	0.2500	No
keller4	0.5417	13	0.0810	0.7500	0.5417	13	0.5060	0.5000	Si
c-fat500-10	1.0080	126	84.0878	0.5000	1.0080	126	98.4496	0.2500	Si
email	1.4532	12	0.0040	0.7500	1.4532	12	0.0040	0.0981	No
jazz	0.8886	30	0.2160	0.8065	0.8886	30	0.0630	0.5000	No
p-hat1000-1	0.8058	9	0.1470	0.7500	2.0753	10	0.6540	0.3496	Si
polbooks	2.9750	6	0.0010	0.7500	2.9750	6	0.0040	0.5000	No
sanr400-0-7	1.3352	18	0.5860	0.2500	2.0469	18	3.1032	0.5000	No

Concluido

TABLA 5.2: Resultados de los experimentos preliminares con constructivo ratio.

Instancias	GRASP				GRASP y búsqueda local				Mejor
	f	c	t (sec)	α	f	c	t (sec)	α	
random-1	1.2421	15	0.0460	0.8377	1.2421	15	0.1170	0.5000	Si
random-4	0.9821	28	1.3561	0.7500	1.1858	28	2.5681	0.7500	Si
random-5	0.8088	31	1.3531	0.7500	1.4592	32	2.9242	0.7500	Si
random-6	0.9514	18	0.0790	0.5000	0.9672	18	0.2570	0.0631	Si
random-8	0.9073	22	0.2110	0.7500	0.9229	23	0.8310	0.7500	Si
random-10	0.9014	30	1.3901	0.7500	1.0688	30	2.5391	0.7500	Si
market-1	1.8981	8	0.0260	0.7500	2.1429	8	0.2730	0.0111	No
market-2	0.8790	8	0.0470	0.4401	1.3260	8	0.3230	0.7500	No
market-5	1.2501	8	0.0580	0.5000	3.8616	7	0.3630	0.7500	No
wind-2005	90 391.7500	4	0.0030	0.2500	90 391.7500	4	0.0150	0.7500	Si
wind-2006	88 663.4000	5	0.0040	0.6243	90 034.2000	5	0.0220	0.5000	Si
brock200-2	0.5238	11	0.0540	0.1726	0.5238	11	0.4620	0.5000	No
football	0.5263	10	0.0040	0.2500	0.5263	10	0.1720	0.7500	No
johnson16-2-4	0.5294	9	0.0730	0.2500	0.5294	9	0.2820	0.7500	No
karate	0.5455	6	0.0010	0.2500	0.5455	6	0.0100	0.7500	Si
keller4	0.5217	12	0.0760	0.2500	0.5217	12	0.3130	0.7500	No
c-fat500-10	1.0080	126	99.9827	0.2500	1.0080	126	113.8295	0.7500	No
email	1.4532	12	0.0080	0.5000	1.4532	12	0.0020	0.7500	Si
jazz	0.8886	30	0.1700	0.7500	0.8886	30	0.0240	0.7500	Si

Continúa en la siguiente página

Tabla 5.2 Continuación de la página anterior

Instancias	GRASP				GRASP y búsqueda local				Mejor
	f	c	t (sec)	α	f	c	t (sec)	α	
p-hat1000-1	2.9810	9	0.1580	0.3496	2.0753	10	0.7830	0.7500	No
polbooks	2.9750	6	0.0010	0.5000	2.9750	6	0.0030	0.7500	Si
sanr400-0-7	2.9186	18	0.7960	0.2500	2.3953	19	3.5592	0.5000	Si

En estas tablas se encuentran los mejores resultados tras todas las iteraciones de ambos constructivos y la posterior mejora del algoritmo con la búsqueda local.

Aunque en algunos casos, añadir al algoritmo la mejora mediante la búsqueda local no parece aumentar la calidad de la solución, en la mayoría de casos si, por lo que finalmente se añadirá para la experimentación final.

Además, tras estas iteraciones se ha comprobado que el constructivo mediante el cálculo por ratio ha obtenido mejores resultados en relación ratio-cardinalidad, alcanzado 13 soluciones mejores respecto a las 9 soluciones mejores obtenidos mediante el constructivo por adyacentes.

Por lo tanto el constructivo basado en el ratio, permite obtener valores para la función objetivo mayores y cliques con mayor número de nodos en tiempos de computo razonables. A esto hay que añadir que el valor de α con el que se han obtenido mejores resultados ha sido 0.75, llegando a 16 soluciones mejores de las 22 instancias procesadas en este caso.

5.2.2. Experimento final

En esta sección se han recopilado todos los datos obtenidos en las pruebas preliminares para el procesamiento final y se ha configurado de la siguiente manera:

- Algoritmo GRASP.
- Búsqueda local.
- Constructivo calculado por ratio.
- Valor de $\alpha = 0,75$.
- 100 iteraciones por instancia.

Con los que tras el procesamiento de todas las instancias, se han obtenido los datos mostrados en la tabla 5.3, donde:

- f : Indica el valor de la función objetivo calculado.
- c : Indica la cardinalidad del clique encontrado.
- t : Indica el tiempo total empleado en encontrar la solución en segundos.
- $\Delta f(\%)$: Indica la variación respecto al valor de la función objetivo en la que se ha basado este trabajo.
- Mejor: Indica si la solución obtenida mediante el algoritmo es mejor o no que la del trabajo anterior.

Los datos recopilados durante el experimento final han sido comparados con los obtenidos previamente en el trabajo realizado por Dominik Goeke, Mahdi Moeini y David Poganiuch [12], los cuales se pueden observar en la tabla 2 del mismo.

TABLA 5.3: Resultados del experimento final.

Instancias	f	c	t (sec)	$\Delta f(\%)$	Mejor
random 1	1.2421	15	0.1320	62,13	No
random 2	1.0816	20	0.3720	76,94	No
random 3	0.9475	21	0.4400	77,49	No
random-4	1.5226	27	2.4531	68,48	No
random-5	1.4592	32	3.4452	59,91	No
random-6	0.9672	18	0.2740	15,89	No
random-7	1.0473	19	0.3310	12,73	No
random-8	0.9504	23	0.7030	20,14	No
random-9	0.9670	28	1.4321	26,74	No
random-10	0.9905	31	3.6612	27,70	No
market-1	2.1429	8	0.2810	89,90	No
market-2	1.3260	8	0.3240	94,40	No
market-3	2.0168	8	0.3500	85,51	No
market-4	2.3678	8	0.3520	87,36	No
market-5	2.8116	8	0.3140	80,19	No
wind-2004	91 286.0000	4	0.0260	0	Si
wind-2005	90 391.7500	4	0.0170	4,54	No
wind-2006	90 034.2000	5	0.0240	8,57	No
Set E					
brock200-1	0.5135	19	1.7311	3,11	No
brock200-2	0.5263	10	0.4600	4,31	No
brock200-3	0.5185	14	0.7350	3,98	No
brock200-4	0.5172	15	1.0301	2,41	No

Continúa en la siguiente página

Tabla 5.3 Continuación de la página anterior

Instancias	f	c	t (sec)	Δf (%)	Mejor
c-fat200-1	0.5200	13	0.5160	0	Si
c-fat200-2	0.5102	25	0.8730	0,04	Si
c-fat200-5	0.5043	59	8.4095	0,85	Si
C-fat500-1	0.5172	15	5.5273	0,53	No
c-fat500-2	0.5094	27	4.5143	0,11	No
c-fat500-5	0.5039	65	19.1571	0,78	Si
c-fat500-10	0.5020	127	146.8904	0,4	Si
hamming6-2	0.5077	33	1.1881	2,37	No
hamming6-4	0.5556	5	0.0280	7,41	No
hamming8-4	0.5152	17	0.9801	8,01	No
johnson8-2-4	0.5556	5	0.0080	0,79	No
johnson8-4-4	0.5172	15	0.1380	2,41	No
johnson16-2-4	0.5294	9	0.2810	0,11	No
johnson32-2-4	0.5152	17	15.6079	0,93	No
keller4	0.5217	12	0.2870	5,14	No
p-hat300-1	0.5333	8	0.6900	6,43	No
p-hat300-2	0.5111	23	1.3121	5,35	No
p-hat300-3	0.5082	31	5.5273	2,27	No
p-hat500-1	0.5294	9	2.2275	7,12	No
p-hat700-1	0.5263	10	3.3702	7,66	No
p-hat1000-1	0.5263	10	7.5684	7,66	No
san200-0-7-1	0.5111	23	1.5341	3,56	No
san200-0-7-2	0.5161	16	1.4451	4,42	No
san200-0-9-1	0.5039	64	16.6310	1,19	No
san200-0-9-2	0.5057	44	11.3847	0,83	No
san200-0-9-3	0.5075	34	9.2055	0,50	No
san400-0-5-1	0.5263	10	2.6161	6,02	No
san400-0-7-1	0.5111	23	7.3954	0	Si
san400-0-7-2	0.5135	19	6.3104	3,11	No
san400-0-7-3	0.5152	17	5.1413	0	Si
san400-0-9-1	0.5043	59	49.7058	0	SI
san1000	0.5238	11	23.8044	4,76	No
sanr200-0-7	0.5143	18	1.3771	2,96	No
sanr200-0-9	0.5070	36	11.1576	0,58	No
sanr400-0-5	0.5200	13	1.7641	5,45	No
sanr400-0-7	0.5135	19	4.8443	3,11	No
karate	0.5455	6	0.0100	9,09	No
dolphins	0.5455	6	0.0290	9,09	No

Continúa en la siguiente página

Tabla 5.3 Continuación de la página anterior

Instancias	f	c	t (sec)	Δf (%)	Mejor
polbooks	0.5385	7	0.1730	10,26	No
adjnoun	0.5455	6	0.1370	9,09	No
football	0.5263	10	0.1810	12,28	No
jazz	0.5082	31	0.5830	15,30	No
celegans-metabolic	0.5263	10	9.5435	12,28	No
email	0.5200	13	94.5984	13,33	No
Set F					
brock200-1	1.5021	18	1.1171	78,17	No
brock200-2	1.6833	11	0.2330	89,00	No
brock200-3	3.3226	12	0.3790	76,08	No
brock200-4	2.1094	14	0.6250	75,56	No
c-fat200-1	1.1383	12	0.0210	6,70	No
c-fat200-2	1.0000	24	0.2150	13,04	No
c-fat200-5	1.0201	58	6.1404	0,01	Si
c-fat500-1	1.0745	14	0.0370	14,72	No
c-fat500-2	1.0745	26	0.4100	0,42	Si
c-fat500-5	1.0080	64	9.2605	3,08	No
c-fat500-10	1.0080	126	116.2877	0,20	No
hamming6-2	1.0000	32	0.6660	39,39	No
hamming6-4	1.0000	4	0.0040	65,99	No
hamming8-4	1.0000	16	0.4560	49,75	No
johnson8-2-4	1.0714	4	0.0030	6,83	No
johnson8-4-4	1.0000	14	0.0990	29,08	No
johnson16-2-4	1.0209	8	0.1720	18,98	No
johnson32-2-4	1.1268	16	9.2425	0	Si
keller4	1.5126	11	0.1570	63,55	No
p-hat300-1	6.2907	7	0.0590	75,58	No
p-hat300-2	1.0048	21	1.0941	86,62	No
p-hat300-3	1.6018	32	4.7473	77,84	No
p-hat500-1	4.3871	8	0.2310	77,69	No
p-hat700-1	4.5360	8	0.4420	78,48	No
p-hat1000-1	2.0753	10	0.8270	94,07	No
san200-0-7-1	1.1401	25	0.9961	89,60	No
san200-0-7-2	2.7520	14	1.2351	84,83	No
san200-0-9-1	1.1930	55	11.8537	75,85	No
san200-0-9-2	1.3278	43	10.2146	73,44	No
san200-0-9-3	1.1145	35	9.7036	74,95	No
san400-0-5-1	1.7034	9	1.2581	97,05	No

Continúa en la siguiente página

Tabla 5.3 Continuación de la página anterior

Instancias	f	c	t (sec)	Δf (%)	Mejor
san400-0-7-1	3.8660	22	5.2073	69,15	No
san400-0-7-2	3.3273	18	4.6083	79,45	No
san400-0-7-3	3.4127	16	3.1572	80,98	No
san400-0-9-1	1.7565	55	50.8769	0	Si
san1000	2.9117	10	9.4345	96,51	No
sanr200-0-7	1.3582	17	0.7650	86,50	No
sanr200-0-9	1.4446	36	10.0596	70,28	No
sanr400-0-5	1.0058	12	0.7040	93,40	No
sanr400-0-7	4.0658	19	3.4942	71,35	No
karate	0.1589	5	0.0010	99,04	No
dolphins	1.2028	5	0.0010	89,63	No
polbooks	2.9750	6	0.0040	75,71	No
adjnoun	0.4637	5	0.0200	97,63	No
football	1.9659	9	0.0020	79,41	No
jazz	0.8886	30	0.0510	87,81	No
celegans-metabolic	1.2098	9	0.0260	80,86	No
email	1.4532	12	0.0020	97,46	No

Como indica el porcentaje de variación de la función objetivo, Δf (%), se puede observar que en todas las instancias se han obtenido valores, en algunos casos, muy cercanos al trabajo anterior, ya que son valores cercanos a 0, y por lo tanto el algoritmo cumple cierta calidad. Aún así, sólo en 10 casos se han obtenido valores mejores.

También cabe destacar que los resultados obtenidos se han realizado en un tiempo de computo significativamente bajo, por lo que aunque no supere al trabajo de referencia, si mejora los resultados obtenidos mediante la linealización y el método de Newton con el solucionador Gurobi⁴ con los que se compara dicho trabajo.

⁴<https://www.gurobi.com/es/>

Capítulo 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:

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

Y por otro lado los objetivos secundarios:

- Estudio de la importancia de los grafos en la vida real.
- Estudio de los distintos algoritmos metaheurísticos existentes.
- Estudio y comprensión de la complejidad computacional relacionada con la búsqueda de clique de ratio máximo.
- 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.

Estos objetivos se han cumplido de manera satisfactoria, puesto que se ha estudiado el estado del arte del problema del clique de ratio máximo para poder abordarlo, así como problemas relacionados como son el problema del clique de peso máximo (MWCP)

y el clásico problema del clique máximo (MCP). También se han estudiado las distintas familias en las que se categorizan los algoritmos metaheurísticos y en especial el algoritmo GRASP. Todo esto ha ayudado en la comprensión del problema y su posterior diseño e implementación.

6.2. Conocimientos adquiridos

El proceso de realización de este trabajo fin de grado ha supuesto la superación de diferentes retos los cuales han permitido ampliar conocimientos sobre el desarrollo de software y la gestión de un proyecto. También se ha profundizado en conceptos sobre algoritmia y estructuras de datos para obtener soluciones mejores y más eficientes al planteamiento de problemas de optimización. Destacando:

- El aumento y mejora de 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, enfocado en la resolución de problemas de optimización.
- Adquisición de conocimientos sobre heurísticas y metaheurísticas aplicadas a la resolución de problemas.
- Profundización en el uso de grafos en programación, así como la relación con el mundo real.
- El aprendizaje sobre \LaTeX , al utilizarlo para documentar el trabajo fin de grado.

6.3. Líneas de desarrollo futuras

En cuanto a las líneas de desarrollo futuras para este problema se describen algunas ideas como:

En primer lugar, un caso que ofrece un considerable aumento en el rendimiento del código, es la utilización de Cython¹ mediante pequeñas modificaciones en el código para ajustarlo a su lenguaje propio. Este compilador optimizado permite aumentar el rendimiento de las funciones escritas en Python. Algunas de sus características son la

¹<https://cython.org/>

unificación de la legibilidad del código escrito en Python con el rendimiento de C, y la interacción eficiente con grandes conjuntos de datos mediante NumPy².

Otra opción podría ser la implementación del procesado paralelo aprovechando el módulo *concurrent.futures* que ofrece Python, este hace uso del módulo *multiprocessing*, por lo tanto no se ve afectado por el GIL³ a diferencia del módulo *multithreading*, y permite crear un conjunto de procesos, mediante la función *ProcessPoolExecutor*, para ejecutar llamadas asíncronas.

Esta opción se puede añadir para el cálculo de los valores a partir de la lista de todos los nodos candidatos, como se describió en la sección 3.1 en el algoritmo 1, con el fin de obtener los resultados de manera paralela, reduciendo considerablemente el tiempo de cómputo, ya que para instancias con una gran cantidad de nodos y una alta densidad este tiempo puede ser excesivo.

²<https://numpy.org/>

³<https://docs.python.org/3/glossary.html?highlight=gil#term-global-interpretor-lock>

Bibliografía

- [1] Micael Gallego Carrillo Abraham Duarte Muñoz Juan José Pantrigo Fernández. *Metaheurísticas*. Ed. por Servicio de Publicaciones Universidad Rey Juan Carlos. 2010.
- [2] Vicente. Falcó Montesinos Antonio. Amigó J. M. Villar Amigó. «Topología molecular / J. M. Amigó ... [et al.]» En: *Sociedad Española de Matemática Aplicada (SeMA)* 39 (2007), págs. 135-149.
- [3] Mikhail Batsyn y col. «Improvements to MCS algorithm for the maximum clique problem». eng. En: *Journal of Combinatorial Optimization* 27.2 (2014), págs. 397-416. ISSN: 1382-6905.
- [4] J. Beltran y col. «Procedimientos constructivos adaptativos (GRASP) para el problema del empaquetado bidimensional». En: *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial* 6 (ene. de 2002), págs. 26-33. DOI: 10.4114/ia.v6i15.755.
- [5] Pardalos P.M. Pelillo M. Bomze I.M. Budinich M. «The Maximum Clique Problem». En: *Pardalos P.M. (eds) Handbook of Combinatorial Optimization. Springer, Boston, MA* (1999).
- [6] Conceptodefinicion.de. *Heurística*. Jul. de 2019. URL: <https://conceptodefinicion.de/heuristica/>.
- [7] Real Academia Española. *Heurística*. URL: <https://dle.rae.es/heur%C3%ADstico>.
- [8] Facebook. *API Graph*. Mar. de 2020. URL: <https://developers.facebook.com/docs/graph-api>.
- [9] Thomas Feo y Mauricio Resende. «Greedy Randomized Adaptive Search Procedures». En: *Journal of Global Optimization* 6 (mar. de 1995), págs. 109-133. DOI: 10.1007/BF01096763.
- [10] L.R. Foulds. *Graph Theory Applications*. Universitext 7. Springer New York, 2012. ISBN: 9781461209331. URL: <https://books.google.es/books?id=5G4QBwAAQBAJ>.

- [11] Fred Glover. «Future paths for integer programming and links to artificial intelligence». En: *Computers operations research* 13 (1986), págs. 533-549.
- [12] Dominik Goeke, Mahdi Moeini y David Poganiuch. «A Variable Neighborhood Search heuristic for the maximum ratio clique problem». En: *Computers y Operations Research* 87 (2017), págs. 283 -291. ISSN: 0305-0548. DOI: <https://doi.org/10.1016/j.cor.2017.01.010>. URL: <http://www.sciencedirect.com/science/article/pii/S0305054817300102>.
- [13] Mauricio Granada-Echeverri y Jhon Santa. *Optimización combinatoria - de la teoría a la práctica*. Mayo de 2018. ISBN: 978-958-8859-43-9.
- [14] R D LUCE y A D PERRY. «A method of matrix analysis of group structure». En: *Psychometrika* 14.2 (jun. de 1949), págs. 95-116. DOI: 10.1007/bf02289146. URL: <https://pubmed.ncbi.nlm.nih.gov/18152948>.
- [15] *Metaheurística*. Oct. de 2019. URL: <https://es.wikipedia.org/wiki/Metaheur%C3%ADstica>.
- [16] Mahdi Moeini. «The Maximum Ratio Clique Problem: A Continuous Optimization Approach and Some New Results». En: *Modelling, Computation and Optimization in Information Systems and Management Sciences*. Ed. por Hoai An Le Thi, Tao Pham Dinh y Ngoc Thanh Nguyen. Cham: Springer International Publishing, 2015, págs. 215-227. ISBN: 978-3-319-18161-5.
- [17] Neo4j Neo4j Neo4j Neo4j. *Nuevas Formas de Predecir el Éxito de Inversiones Tableros Dinámicos 360 de Clientes. Detección de Fraude en Tiempo Real y Agilizando Procesos de AML*. Dic. de 2019. URL: <https://www.youtube.com/watch?v=KUtEee0et-w>.
- [18] A. Nikolaev, M. Batsyn y P. San Segundo. «Reusing the same coloring in the child nodes of the search tree for the maximum clique problem». En: vol. 8994. Springer Verlag, 2015, págs. 275-280. ISBN: 9783319190839.
- [19] Kuk-Kwon Park y col. «GRASP Algorithm for Dynamic Weapon-Target Assignment Problem». En: *Journal of the Korean Society for Aeronautical and Space Sciences* 47 (dic. de 2019), págs. 856-864. DOI: 10.5139/JKSAS.2019.47.12.856.
- [20] George Pólya. *Cómo Plantear y Resolver Problemas*. Ed. por Editorial Trillas. 1965.
- [21] George Pólya. *How to Solve It*. Ed. por Princeton. 1945.
- [22] Julio Ponce y col. «Algoritmo de Colonia de Hormigas para el Problema del Clique Máximo con un Optimizador Local K-opt». En: (ene. de 2008).
- [23] Jorge Ramírez. «Metaheurística GRASP para el problema Vertex Bisection Minimization.» En: 12 (ene. de 2018), págs. 28-41.

- [24] Joao Santos y col. «A parallel hybrid implementation using genetic algorithm, GRASP and reinforcement learning». En: jul. de 2009, págs. 2798 -2803. DOI: 10.1109/IJCNN.2009.5178938.
- [25] Samyukta Sethuraman y Sergiy Butenko. «The maximum ratio clique problem». En: *Computational Management Science* 12.1 (2015), págs. 197-218. DOI: 10.1007/s10287-013-0197-z. URL: <https://doi.org/10.1007/s10287-013-0197-z>.
- [26] Wang Shaochang y col. «Application of Greedy Random Adaptive Search Algorithm (GRASP) in Flight Recovery Problem». En: ene. de 2018.
- [27] Y. Sun, X. Li y A. Ernst. «Using Statistical Measures and Machine Learning for Graph Reduction to Solve Maximum Weight Clique Problems». En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019), págs. 1-1. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2019.2954827.
- [28] Víctor Villavicencio. «GRASP para el problema de ruta de vehículos». En: (mar. de 2020).
- [29] Yiyuan Wang y col. «SCCWalk: An efficient local search algorithm and its improvements for maximum weight clique problem». En: *Artificial Intelligence* 280 (2020), pág. 103230. ISSN: 0004-3702. DOI: <https://doi.org/10.1016/j.artint.2019.103230>. URL: <http://www.sciencedirect.com/science/article/pii/S0004370219302164>.
- [30] Eric W. Weisstein. *Clique*. From MathWorld—A Wolfram Web Resource. URL: <https://mathworld.wolfram.com/Clique.html>.
- [31] Eric W. Weisstein. *Maximal Clique*. From MathWorld—A Wolfram Web Resource.
- [32] Eric W. Weisstein. *Maximum Clique*. From MathWorld—A Wolfram Web Resource.
- [33] Gang Yang y col. «An improved competitive Hopfield network with inhibitive competitive activation mechanism for maximum clique problem». eng. En: *Neurocomputing* 130 (2014), págs. 28-35. ISSN: 0925-2312.