

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA TRIENNALE IN INGEGNERIA INFORMATICA

ESPERIMENTI DI PROGRAMMAZIONE LINEARE
INTERA PER PROBLEMI DI VERTEX COVER

Relatore

Prof. Domenico Salvagnin

Laureando

Giacomo Camposampiero

Anno Accademico 2020/2021

Padova, 22 luglio 2021

A mamma, papà e Sveva

Indice

Introduzione	1
1 Nozioni preliminari	3
1.1 Problema di ottimizzazione	3
1.2 Programmazione lineare intera	4
1.2.1 Algoritmo Branch and Bound	4
1.2.2 Algoritmo Branch and Cut	6
1.3 Vertex cover	7
2 Approccio sperimentale	9
2.1 Generazione dei grafi	9
2.1.1 Modello di Erdős–Rényi	10
2.1.2 Modello di Steger-Wormald	11
2.1.3 Modello di Watts-Strogatz	13
2.1.4 Modello di Barabási-Albert	15
2.2 Generazione delle istanze di problemi MIP	18
2.2.1 Modellazione algebrica	18
2.2.2 Modellazione nel software	20
2.3 Risoluzione delle istanze di problemi MIP	20
2.4 Estrazione ed analisi dei risultati	21
3 Risultati sperimentali	23
3.1 Risultati per grafi $G(n, p)$	24
3.2 Risultati per grafi regolari	26
3.3 Risultati per grafi di Watts-Strogatz	28
3.4 Risultati per grafi di Barabási-Albert	30
4 Conclusioni	33
Bibliografia	35

Introduzione

La ricerca operativa, dall'inglese *operational research*, è una disciplina scientifica relativamente giovane, nata con lo scopo di fornire strumenti matematici di supporto ad attività decisionali in cui occorre gestire e coordinare attività e risorse limitate. La ricerca operativa permette infatti di trovare, mediante la formalizzazione matematica di un problema, una soluzione ottima o ammissibile (quando possibile) al problema stesso e costituisce di fatto un approccio scientifico alla risoluzione di problemi complessi che ha trovato grande applicazione in moltissimi ambiti, non ultimo quello industriale.

Una delle diverse branche che compongono la ricerca operativa è l'ottimizzazione. Quest'ultima si occupa principalmente di problemi che comportano la minimizzazione o la massimizzazione di una funzione, detta funzione obiettivo, sottoposta ad un dato insieme di vincoli.

L'oggetto del presente studio è stata una particolare classe di problemi di ottimizzazione, il *vertex cover* minimo, anche detto problema di copertura dei vertici. In particolare, ne è stato analizzato l'andamento della complessità di risoluzione per alcune tipologie di grafi, generati casualmente, al variare dei parametri utilizzati nella generazione stessa.

La stesura di questo elaborato è stata articolata in quattro capitoli. Inizialmente, verranno brevemente esposti alcuni concetti teorici rilevanti nel contesto degli esperimenti svolti. La trattazione proseguirà poi con la presentazione dell'impostazione utilizzata nello svolgimento degli esperimenti e dei risultati ottenuti. Infine, saranno tratte alcune conclusioni riguardo il lavoro svolto.

Tutto il codice sviluppato in relazione a questo lavoro è stato scritto in Python [1], linguaggio di programmazione *general-purpose* di alto livello che ha permesso di gestire ognuna delle diverse fasi in cui si è articolato lo svolgimento degli esperimenti, dalla generazione dei grafi all'elaborazione grafica dei risultati. Il codice sorgente, così come tutti i risultati delle elaborazioni, sono liberamente consultabili online nella *repository* utilizzata per gestire il controllo di versione del progetto [2].

Capitolo 1

Nozioni preliminari

In questo capitolo sono stati raccolti alcuni concetti teorici di ottimizzazione legati al lavoro svolto. È stata come prima cosa data una definizione formale al concetto di problema di ottimizzazione. Si è proseguito poi con un approfondimento sulla programmazione lineare intera e su due algoritmi utilizzati nella risoluzione di problemi di ottimizzazione, *branch-and-bound* e *branch-and-cut*. Infine, è stata esposta una definizione formale al problema di ottimizzazione studiato in questo lavoro, il problema di copertura dei vertici minimo. Molte delle definizioni in questo capitolo sono ispirate alle dispense del corso di *Modelli e Software per l'Ottimizzazione Discreta*, tenuto dal professore Domenico Salvagnin [3][4].

1.1 Problema di ottimizzazione

Un problema di ottimizzazione può essere formulato come

$$\begin{array}{ll} \min(or \max) f(x) & \\ S & \\ x \in D & \end{array} \quad (1.1)$$

dove $f(x)$ è una funzione a valori reali nelle variabili x , D è il dominio di x e S un insieme finito di vincoli. In generale, x è una tupla (x_1, \dots, x_n) e D è un prodotto cartesiano $D_1 \times \dots \times D_n$, e vale $x_j \in D_j$.

Un problema nella forma (1.1) è intrattabile, ovvero non esistono algoritmi efficienti (o non esiste proprio alcun algoritmo) per la sua risoluzione. Si rende quindi necessario considerare dei casi particolari di questa formulazione, i quali possiedono determinate proprietà che possono essere sfruttate nella definizione di algoritmi risolutivi ad-hoc.

1.2 Programmazione lineare intera

Uno dei casi particolari della formulazione generale (1.1) a cui si è precedentemente fatto riferimento viene trattato dalla programmazione lineare intera. Un problema di programmazione lineare intera consiste nella minimizzazione (o massimizzazione) di una funzione lineare soggetta ad un numero finito di vincoli lineari, con in aggiunta il vincolo che alcune delle variabili del problema debbano assumere valori interi. In generale, il problema può quindi essere riformulato come

$$\begin{aligned} \min & cx \\ a_i x & \sim b_i & i = 1, \dots, m \\ l_j & \leq x_j \leq u_j & j = 1, \dots, n = N \\ x_j & \in \mathbb{Z} & \forall j \in J \subseteq N = 1, \dots, n \end{aligned}$$

Se $J = N$ si parla di programmazione lineare intera pura, altrimenti di programmazione lineare intera mista (o MIP, dall'inglese *Mixed Integer Programming*).

La programmazione lineare intera restringe quindi notevolmente il tipo di vincoli a disposizione nel processo di formalizzazione matematica del problema, determinando una maggior difficoltà in fase di modellazione. Tuttavia, questo non comporta eccessive restrizioni, almeno per la MIP, sulle tipologie di problemi che possono essere formulati secondo questo paradigma. Alcuni esempi classici di problemi risolvibili mediante MIP sono *knapsack*, problemi di *scheduling*, *facility location* e, naturalmente, *minimum vertex cover*. Inoltre, l'introduzione dei vincoli di linearità ed interezza comporta notevoli vantaggi nella definizione ed implementazione di algoritmi risolutivi, alcuni dei quali verranno descritti più nel dettaglio nelle seguenti sezioni.

1.2.1 Algoritmo Branch and Bound

L'algoritmo branch-and-bound (B&B) è un algoritmo di ottimizzazione generica basato sull'enumerazione dell'insieme delle soluzioni ammissibili di un problema di ottimizzazione combinatoria, introdotto nel 1960 da A. H. Land e A. G. Doig [5].

Questo algoritmo permette di gestire il problema dell'esplosione combinatoria mediante il *pruning* di intere porzioni dello spazio delle soluzioni, che può essere effettuato quando si riesce a dimostrare che queste ultime non contengono soluzioni migliori di quelle note. Branch-and-bound implementa inoltre una strategia *divide and conquer*, che permette di partizionare lo spazio di ricerca e di risolvere ognuna delle partizioni ottenute separatamente. Viene riportata di seguito una breve descrizione dell'algoritmo.

Sia F l'insieme delle soluzioni ammissibili di un problema di minimizzazione (simmetrico nel caso della massimizzazione, a meno di un cambio di segno della funzione obiettivo), $c : F \rightarrow \mathbb{R}$ la funzione obiettivo e $\bar{x} \in F$ una soluzione ammissibile nota, generata mediante euristiche o mediante assegnazioni casuali. Sia $z = f(\bar{x})$ il costo di tale soluzione nota, anche detto *incumbent*, che rappresenta per sua natura un limite superiore al valore della soluzione ottima.

L'algoritmo branch-and-bound esegue un'iniziale fase di *bounding* in cui uno o più vincoli del problema vengono rilassati, allargando di conseguenza l'insieme delle possibili soluzioni $G \supseteq F$. La soluzione di questo rilassamento, se esiste, rappresenta un *lower bound* alla soluzione ottima del problema iniziale. Se la soluzione di tale rilassamento appartiene a F o ha costo uguale all'attuale *incumbent*, allora l'algoritmo termina, in quanto si è trovata una soluzione ottima del problema. Se il rilassamento risulta essere impossibile, possiamo anche in questo caso terminare la ricerca di una soluzione e concludere che anche il problema di partenza è impossibile.

Nel caso in cui invece una soluzione al rilassamento esiste ma non è contenuta nell'insieme delle soluzioni ammissibili F , l'algoritmo procede con l'identificare una separazione F^* di F , ossia un insieme finito di sottoinsiemi tali che

$$\bigcup_{F_i \in F^*} F_i = F$$

Questa fase, detta di *branching*, è giustificata dal fatto che la soluzione ottima dell'intero problema è data dalla minima tra le soluzioni delle varie separazioni $F_i \in F^*$. F^* è spesso, anche se non necessariamente, una partizione dell'insieme iniziale F . A questo punto, tutti i figli di F vengono aggiunti alla coda dei sotto-problemi da processare.

L'algoritmo procede quindi con il selezionare un sotto-problema P_i dalla coda un rilassamento. Quattro sono i possibili casi:

- Se si trova una soluzione $\in F$ migliore dell'attuale *incumbent*, quest'ultimo viene sostituito dalla soluzione trovata e si procede con lo studio di un altro sotto-problema.
- Se il rilassamento del sotto-problema non ammette soluzione, allora si smette di esplorare l'intero sotto-albero a lui associato nello spazio di ricerca (*pruning by infeasibility*).
- Altrimenti, si confronta la soluzione trovata con il valore corrente dell'*upper-bound* dato dall'*incumbent*; se quest'ultimo è minore della soluzione del rilassamento trovata, è possibile anche in questo caso smettere di esplorare il sotto-albero associato

al sotto-problema corrente, in quanto non può portare ad una soluzione migliore di quella che già si conosce (*pruning by optimality*).

- Infine, se non è stato possibile scartare o concludere la visita del sotto-albero associato a P_i in alcun modo, è necessario eseguire nuovamente il *branching*, aggiungendo i nuovi sotto-problemi alla lista dei sotto-problemi da processare.

L'algoritmo prosegue nella selezione di sotto-problemi finché la lista di questi ultimi non si svuota. Quando ciò avviene, la soluzione rappresentata dall'attuale *incumbent* è la soluzione ottima al problema iniziale.

Quella appena descritta rappresenta una formulazione generica dell'algoritmo B&B. Questa formulazione può essere tuttavia specializzata nella risoluzione di problemi MIP con relativa semplicità, agendo sulle condizioni che regolano *bounding* e *branching*. Per quanto riguarda il primo, la scelta più diffusa consiste nel rilassamento del vincolo di interezza. Se la soluzione del rilassamento non è intera, una possibile separazione in sotto-problemi può essere fatta considerando la partizione

$$x_j \leq \lfloor x_j^* \rfloor \vee x_j \geq \lceil x_j^* \rceil$$

Infine, è importante notare come, per costruzione, ogni soluzione trovata dall'algoritmo è migliore dell'*incumbent* e, di conseguenza, l'andamento dell'upper bound del problema è strettamente decrescente. I lower bound dei singoli sotto-problemi non hanno invece, al contrario degli upper bound, valenza globale. Derivare un lower bound globale è comunque possibile, considerando il minimo tra tutti i lower bound dei sotto-problemi ancora aperti. Avendo a disposizione in ogni momento entrambi i bound del problema, è possibile quindi valutare la bontà della soluzione provvisoria in qualsiasi momento.

1.2.2 Algoritmo Branch and Cut

L'algoritmo *branch-and-cut* (B&C) rappresenta una versione migliorata dell'algoritmo branch-and-bound, introdotta nel 1987 da M. Padberg e G. Rinaldi [6] e ideata appositamente per la risoluzione di problemi MIP.

L'algoritmo branch-and-cut è un ibrido tra branch-and-bound, trattato in precedenza, e un algoritmo a piani di taglio puro, in cui la soluzione è ottenuta mediante raffinazioni successive dello spazio delle soluzioni mediante la progressiva aggiunta di vincoli. Queste due tecniche si rafforzano a vicenda, contribuendo al raggiungimento di prestazioni complessive superiori a quelle che otterrebbe ciascuna di esse singolarmente.

L'idea alla base di questo algoritmo è quella di "rafforzare", per ogni sotto-problema, la formulazione associata al suo rilassamento lineare mediante la generazione di piani di taglio. I vantaggi a livello risolutivo sono diversi, tra cui una maggior probabilità di ottenere soluzioni intere al rilassamento lineare o, in alternativa, di ottenere lower bound migliori e quindi più discriminanti in fase di pruning.

Nonostante l'idea alla base di questo approccio sia relativamente semplice, l'implementazione dell'algoritmo B&C è tutt'altro che scontata e richiede l'esistenza di procedure efficienti per la risoluzione del seguente problema di *separazione*: data una soluzione frazionaria x^* , trovare una disuguaglianza valida $\alpha^T x \leq \alpha_0$ (se esiste) violata da x^* , cioè tale che $\alpha^T x^* > \alpha_0$.

1.3 Vertex cover

Il problema di *vertex cover* (anche detto di *copertura dei vertici*) è un problema di ottimizzazione combinatoria che consiste nel trovare il minimo vertex cover di un grafo, ossia il più piccolo insieme di nodi del grafo tale che almeno uno dei due vertici di ogni arco sia contenuto in questo insieme. Trovare il vertex cover minimo di un generico grafo è un problema *NP-completo*, ovvero non esistono algoritmi in grado di trovarne una soluzione in un tempo polinomiale. In questo elaborato è stata considerata la formulazione indiretta e *unweighted* del problema, in cui gli archi non sono direzionati e in cui i nodi del grafo hanno tutti peso uguale e pari ad 1.

Formalmente, il vertex cover V' di un grafo $G = (V, E)$ può essere definito come un sotto-insieme dei vertici del grafo V tale che

$$uv \in E \Rightarrow u \in V' \vee v \in V'$$

Definita $\tau = |V'|$ la cardinalità del vertex cover, il vertex cover minimo può essere definito formalmente come

$$V'_{min} = \arg \min_{\tau} V'$$

Capitolo 2

Approccio sperimentale

Terminata la breve parentesi teorica introduttiva, si procede in questo capitolo alla presentazione dell'impostazione pratica che si è voluto dare alle sperimentazioni condotte. La struttura secondo cui saranno esposte le informazioni nei seguenti paragrafi ricalca la partizione logica alla base del codice sviluppato, pragmaticamente diviso in quattro moduli tra loro indipendenti:

- generazione dei grafi
- generazione delle istanze di problemi MIP
- risoluzione delle istanze di problemi MIP
- estrazione ed analisi dei dati

2.1 Generazione dei grafi

L'iniziale problema che è stato necessario affrontare nello svolgimento di questo lavoro è stata la generazione automatica e pseudo-casuale di grafi, indispensabile al fine di ottenere un numero sufficiente di istanze da cui estrarre informazioni statisticamente rilevanti. Quello della generazione di grafi pseudo-casuali è un problema noto in letteratura, in quanto in numerosi ambiti di ricerca, che spaziano dalla biologia molecolare allo studio delle reti di calcolatori, si rende necessaria la generazione casuale di queste strutture dati nello studio delle realtà di loro interesse. Di conseguenza, nel corso degli ultimi decenni sono stati molti i metodi proposti dalla comunità scientifica al fine di affrontare il problema. Tra questi ne sono stati selezionati quattro, sulla base delle proprietà e delle caratteristiche particolari di ciascuno di essi.

La gestione dei grafi a livello di codice è stata implementata utilizzando NetworkX, una libreria stabile e flessibile sviluppata appositamente per lo studio e la rappresentazione di grafi in Python [7]. NetworkX offre, oltre alla possibilità di rappresentazione, diversi algoritmi per lo studio delle proprietà e il calcolo di determinate misure su istanze di grafi fornite. La libreria comprende inoltre un insieme di metodi che implementano un gran numero di modelli standard noti in letteratura per la generazione di grafi, tra cui naturalmente quelli utilizzati in questo lavoro.

Inoltre, tutti i grafi generati in questo lavoro sono stati salvati sotto forma di *adjacency list*, in modo tale da poter essere letti ed elaborati da altri moduli del programma a posteriori. Il salvataggio è stato svolto mediante un apposito metodo della libreria, `write_adjlist(G, path)`. In una lista di adiacenze ogni nodo è rappresentato da una stringa nel formato

```
nodo [vicino] [vicino] ... [vicino]
```

2.1.1 Modello di Erdős–Rényi

Il modello di Erdős–Rényi è un modello per la generazione di grafi casuali, detti grafi Erdős–Rényi (*ER*) o binomiali, introdotto per la prima volta nel 1959 dai matematici ungheresi Paul Erdős e Alfréd Rényi [8]. Nella formulazione $G(n, p)$ del modello, un grafo di n nodi viene costruito secondo un procedimento iterativo, in cui ogni arco tra due nodi del grafo viene generato con una probabilità p fissata ed indipendentemente da quella con cui sono generati gli altri archi.

Formalmente, un grafo ER, anche detto $G(n, p)$, è un grafo (N, G) con $N = 1, 2, \dots, n$ insieme dei nodi e in cui la matrice delle adiacenze $G = (g_{ij})$ è tale per cui, per ogni coppia di nodi distinti i e j , la probabilità che $g_{ij} = 1$ è $P(g_{ij} = 1) = p$, con $p \in [0, 1]$.

I parametri di questo modello sono quindi due, il numero di nodi del grafo n e la probabilità di generazione di ogni arco p . A parità di parametri, i grafi che si possono ottenere sono in ogni caso molti, come mostrato in Figura 2.1.

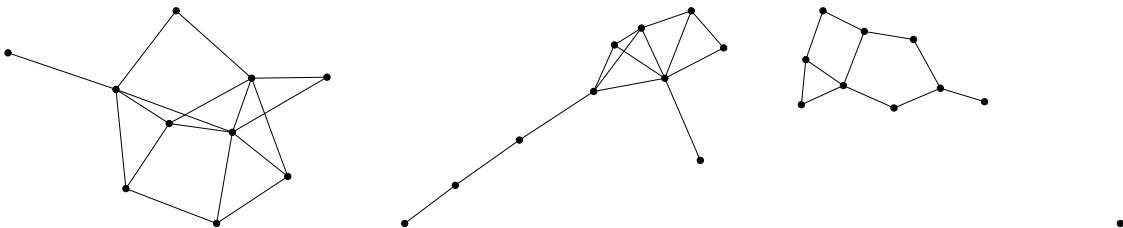


Figura 2.1: Diversi grafi $G(n, p)$ generati a partire dagli stessi parametri $n = 10$ e $p = 0.3$ con la libreria NetworkX e visualizzati graficamente con l'ausilio di un'altra libreria Python, Matplotlib [9].

Il metodo di NetworkX utilizzato nella generazione di questa tipologia di grafi è `gnp_random_graph(n, p, seed)`, che restituisce un'istanza di grafo $G(n, p)$ di dimensione n in cui gli archi hanno probabilità di essere generati pari a p , mentre il parametro `seed` regola il comportamento pseudo-casuale, in modo tale da permettere la riproducibilità degli esperimenti. Lo pseudo-codice dell'algoritmo che gestisce la generazione di questa tipologia di grafi è riportato in Algoritmo 1.

Algoritmo 1: Generazione di un grafo di Erdős–Rényi

Input: numero di nodi n e probabilità di generazione di ogni arco p
 inizializza il grafo vuoto G ;
 aggiungi i nodi $1 \dots n$ a G ;
foreach coppia di nodi $e=(i,j)$ **do**
 $rand$ = numero casuale $\in [0, 1]$;
 if $rand \leq p$ **then**
 aggiungi l'arco e a G ;
 end
end
return G

Tuttavia, l'assunto principale su cui è basata la generazione in questo modello, ovvero che a tutti gli archi è associata una probabilità di generazione uguale ed indipendente da quella degli altri archi, potrebbe risultare poco veritiero nella generazione di grafi casuali il cui scopo è modellare certe categorie di fenomeni del mondo reale. Molte proprietà dei grafi nel mondo reale, infatti, non riescono ad essere replicate da questo particolare modello di generazione casuale.

Per questa ragione sono stati sviluppati nel corso del tempo diversi altri modelli di generazione basati su meccanismi completamente diversi, come ad esempio *rewiring* e *preferential-attachment*, di cui verranno presentati degli esempi più avanti.

2.1.2 Modello di Steger-Wormald

Nella teoria dei grafi, un grafo è detto *regolare di grado d* quando tutti i suoi vertici hanno lo stesso grado, ossia lo stesso numero di archi incidenti. Se tale grado comune è uguale a d , il grafo è detto d -regolare. Un grafo regolare casuale è un grafo selezionato da $\mathcal{G}_{n,r}$, lo spazio di probabilità di tutti i possibili grafi r -regolari di n vertici, in cui $3 \leq r < n$ e nr è pari. In questo spazio ogni grafo d -regolare di n nodi ha la stessa probabilità $\frac{1}{|\mathcal{G}_{n,r}|}$ di essere generato. Si tratta quindi di grafi random, le cui caratteristiche sono tuttavia significativamente diverse dal caso generale a causa del vincolo di regolarità.

Dal momento che la maggior parte dei grafi random (che possono essere generati ad esempio mediante il modello di Erdős–Rényi precedentemente esposto nella Sottosezione 2.1.1) non sono regolari, l'implementazione di un algoritmo imparziale per la generazione di grafi regolari casuali non è così scontata.

Un modello tipicamente utilizzato per la generazione di grafi d -regolari con d piccolo è il cosiddetto *pairing model*. Tale modello si basa sulla definizione di n insiemi, uno per ogni vertice del grafo da generare, contenenti ciascuno d elementi, tanti quanti il numero di archi che devono incidere ogni nodo. La costruzione del grafo avviene quindi selezionando iterativamente due insiemi i e j , con $i \neq j$ e per cui non esiste già un arco nel grafo che collega i due nodi corrispondenti, aggiungendo al grafo l'arco (i, j) e rimuovendo dai corrispondenti insiemi un elemento ciascuno.

Questo modello è molto intuitivo ma viene solitamente utilizzato nella generazione di istanze di grafi dove d è molto piccolo, come già accennato in precedenza. Si può dimostrare infatti che la complessità computazionale dell'algoritmo è esponenziale nella grandezza di d , dell'ordine di $nde^{(d^2-1)/4}$ [10] per valori di d fino a $n^{\frac{1}{3}}$.

In questo lavoro la generazione dei grafi regolari random è stata quindi implementata utilizzando una versione ottimizzata del modello ad accoppiamento, proposta nel 1999 dai matematici Angelika Steger e Nick Wormald [10]. Lo pseudo-codice dell'algoritmo utilizzato nel metodo proposto da questi ultimi è riportato in Algoritmo 2. Alcuni esempi di grafi regolari random generati mediante il modello di Steger-Wormald sono riportati in Figura 2.2.

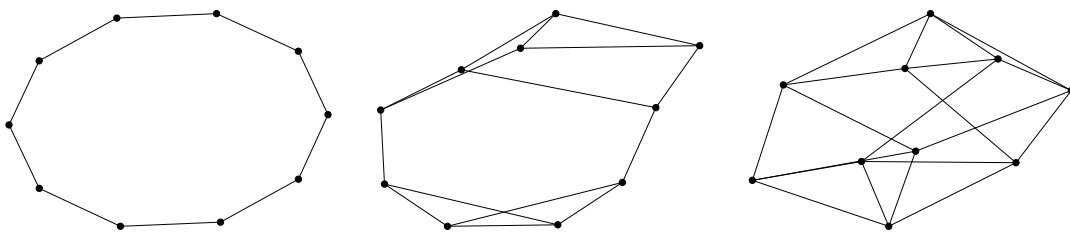


Figura 2.2: Diversi grafi regolari di dimensione 10 generati utilizzando il metodo `random_regular_graph(d, n)` della libreria NetworkX con d pari a, da sinistra verso destra, 2, 3 e 4.

Il metodo di NetworkX utilizzato nella generazione di questa tipologia di grafi è `random_regular_graph(d, n, seed)`, che restituisce un'istanza di grafo regolare di dimensione n in cui tutti i nodi hanno grado d . Il parametro `seed` anche in questo caso regola il comportamento pseudo-casuale della generazione, permettendo la replicabilità dell'esperimento.

Algoritmo 2: Generazione di un grafo regolare con il modello di Steger-Wormald

Input: numero di nodi n e numero di archi incidenti per ogni nodo d

do

- inizializza un grafo G con n nodi e nessun vertice;
- inizializza un insieme S di coppie di nodi non adiacenti e il cui grado è al massimo $d - 1$;
- do**
 - assegna ad ogni coppia (u, v) in S una probabilità di essere scelta proporzionale a $(d - d(u))(d - d(v))$;
 - scegli una coppia (u, v) da S sulla base delle probabilità calcolate;
 - aggiungi l'arco (u, v) a G ;
 - rimuovi la coppia u, v da S ;
- while** S non è vuoto;

while G non è d -regolare;

return G

2.1.3 Modello di Watts-Strogatz

Il modello di Watts-Strogatz è un modello per la generazione di grafi casuali presentato nel 1998 da Duncan J. Watts e Steven Strogatz [11], la cui caratteristica principale risiede nel possedere proprietà *small-world*.

Il concetto di small-world venne introdotto per la prima volta nel 1967 dallo psicologo statunitense Stanley Milgram nella sua serie di esperimenti mirata ad esaminare la lunghezza media dei percorsi delle reti sociali tra residenti negli Stati Uniti (popolarmente conosciuti con il nome di "*six degrees of separation*") [12]. In questi esperimenti Milgram ipotizzò l'aderenza dei contesti sociali da lui analizzati ad un modello a "*piccolo mondo*", composto da una rete di collegamenti tra persone relativamente breve.

Watts e Strogatz, partendo dalle osservazioni di Milgram, verificarono che diversi esempi di grafi che rappresentano contesti reali, tra cui il sistema nervoso del verme nematode fasmidario *Caenorhabditis elegans*, la rete elettrica degli Stati Uniti occidentali e un grafo rappresentante le collaborazioni tra diversi attori di film, mostrano alcune proprietà comuni molto importanti. Sulla base di queste osservazioni proposero quindi un modello di rete denominato *small-world network*, caratterizzato da:

- lunghezza media del cammino minimo tra due nodi molto bassa
- tendenza al raggruppamento in cluster all'interno della rete, che si traduce in regolarità e località del grafo

Nessuno dei principali metodi di generazione di grafi casuali garantiva tuttavia la ge-

nerazione di grafi con tali caratteristiche. I grafi $G(n, p)$ infatti presentano caratteristiche simili a reti small-world, ma non hanno un alto coefficiente di clusterizzazione. Al contrario, i grafi regolari presentano un alto coefficiente di clusterizzazione ma lunghezza media dei cammini minimi generalmente elevata.

Per questa ragione, Watts e Strogatz proposero un nuovo modello di generazione di grafi casuali, basato su modifiche iterative di un grafo regolare ad anello mediante un *rewiring* di archi del grafo controllato dal parametro p (che modula la regolarità o, viceversa, la casualità del grafo prodotto, come mostrato in Figura 2.3b). Queste modifiche consentono, per opportuni valori di p , la conservazione di un alto livello di clustering locale ma, allo stesso tempo, abbattano la lunghezza del cammino minimo tra nodi del grafo mediante l'aggiunta di collegamenti casuali. Un andamento di queste due proprietà al variare del valore di p è riportato in Figura 2.3a. Lo pseudo-codice di questo metodo è riportato in Algoritmo 3.

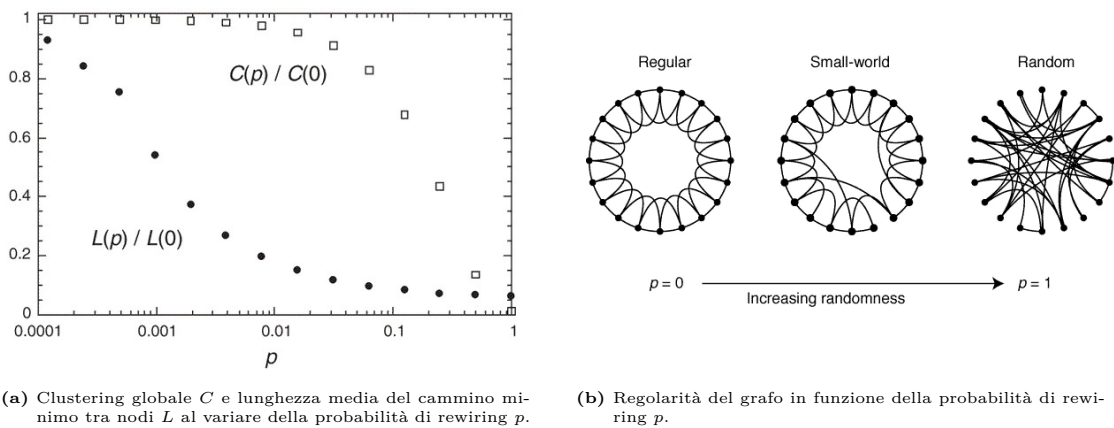


Figura 2.3: Rappresentazioni grafiche che evidenziano caratteristiche proprie del modello di generazione proposto da Watts-Strogatz [11].

Il metodo di NetworkX utilizzato nella generazione di questa tipologia di grafi è `watts_strogatz_graph(n, k, p, seed)`, che restituisce un'istanza di grafo di Watts-Strogatz di dimensione n . I nodi di questo grafo sono inizialmente organizzati in una struttura ad anello in cui ogni nodo è collegato ai suoi $k//2+1$ vicini sinistri e $k//2+1$ vicini destri ed in cui la probabilità che ciascun nodo sia "ricollegato" con un altro nodo diverso da uno dei suoi vicini attuali è p . Anche in questo caso il parametro `seed` regola il comportamento casuale della generazione, permettendo così di ottenere la riproducibilità dei risultati.

Questo modello di generazione presenta tuttavia diverse limitazioni. La principale tra queste riguarda il grado dei nodi che, a differenza della maggior parte dei grafi che rappresentano modelli reali, denota una varianza media molto bassa tra i nodi di uno

stesso grafo. Inoltre, questo modello di generazione implica l'utilizzo di grafi con un numero fisso di nodi e , di conseguenza, non può essere utilizzato nella simulazione di realtà la cui dimensione è variabile nel tempo.

Algoritmo 3: Generazione di un grafo di Watts-Strogatz

Input: numero di nodi n , numero collegamenti iniziali con i vicini k e probabilità di fare *rewiring* di ciascun arco p
 inizializza un grafo ad anello G con n nodi e nessun arco;
foreach nodo m appartenente al grafo **do**
 | aggiungi un arco tra m e i suoi $k/2$ vicini destri e $k/2$ vicini sinistri;
end
foreach arco $e=(u,v)$ tra due nodi del grafo **do**
 | $rand$ = numero casuale $\in [0, 1]$;
 if $rand > p$ **then**
 | w = nodo diverso da v scelto casualmente con probabilità uniforme;
 | $e = (u, w)$, ossia ricollega l'arco;
 end
end
return G

2.1.4 Modello di Barabási-Albert

Tutti i modelli visti fino a questo momento mancano di due aspetti fondamentali che caratterizzano invece i grafi reali. Per prima cosa, viene sempre assunta la presenza di un numero fisso di vertici nel grafo. Al contrario, molte delle reti nel mondo reale sono composte da un insieme di vertici che solitamente tende a crescere nel tempo. Alcuni esempi di grafi che hanno questo comportamento sono il World Wide Web (WWW), che cresce esponenzialmente con l'aggiunta di nuove pagine Web, e la letteratura scientifica, che cresce anch'essa nel tempo con la pubblicazione di nuovi articoli.

In secondo luogo, nessuno dei modelli trattati in precedenza tiene conto del cosiddetto *preferential attachment*, fenomeno per cui i nodi che si aggiungono alla rete sono solitamente più propensi a creare connessioni con nodi che hanno già un gran numero di connessioni con altri nodi. Ritornando agli esempi già citati del WWW e della letteratura scientifica, è più probabile che una nuova pagina Web includa link a pagine Web più conosciute e già referenziate da molte altre pagine, così come è molto probabile che un nuovo paper scientifico includa riferimenti ad altri paper influenti e già citati in molti altri lavori.

Proprio sulla base di queste osservazioni i fisici Albert-László Barabási e Réka Albert proposero nel 1999 un nuovo modello per la generazione di grafi casuali [13]. Tale metodo

è basato su di un meccanismo di *preferential attachment*, in cui i nodi vengono aggiunti uno ad uno al grafo e la probabilità di creazione di un arco tra il nuovo nodo i e uno di quelli esistenti j dipende strettamente dal grado di quest'ultimo, secondo la relazione

$$p_i = \frac{k_i}{\sum_{j \in J} k_j}$$

dove J è l'insieme dei nodi già presenti nel grafo durante l'iterazione in cui viene aggiunto il nuovo nodo e k_x corrisponde al numero di connessioni che incidono il nodo x nella stessa iterazione. Lo pseudo-codice del modello di generazione è riportato in Algoritmo 4.

Il modello di Barabási-Albert ha quindi la capacità di generare un grafo con pochi nodi, detti *hub*, che presentano un grado molto elevato se comparato a quello della maggior parte dei restanti nodi del grafo.

Un'altra caratteristica molto importante dei grafi casuali generati utilizzando questo modello è quella di essere *scale-free networks*, ovvero grafi per cui la relazione tra numero di nodi e numero di connessioni è di tipo esponenziale negativo, e quindi invariante ai cambiamenti di scala. Un confronto tra la distribuzione del grado dei nodi in un grafo $G(n, p)$ e in un grafo *scale-free* è riportato di seguito in Figura 2.4.

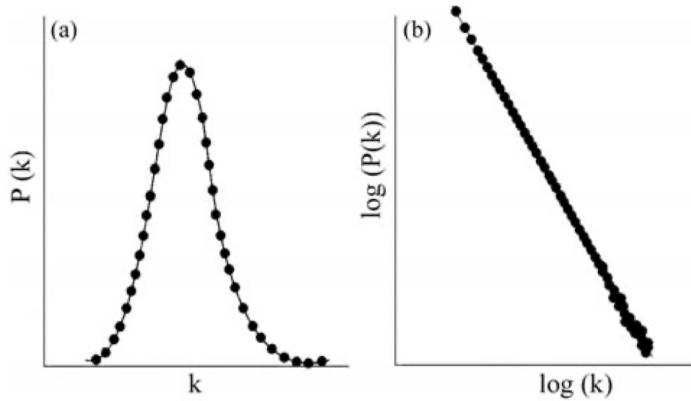


Figura 2.4: Confronto tra la distribuzione del grado dei nodi in un grafo $G(n, p)$ (a sinistra) e in un grafo *scale-free* (a destra) [14].

A livello pratico, questo significa che confrontando il numero di due tipi di nodi, ad esempio quelli di grado 5 e 15, il rapporto tra i due risulta essere pari a

$$e^{-a(Nb - Na)}$$

dove Na e Nb sono rispettivamente il numero di nodi del denominatore e del numeratore mentre a è un parametro del tipo di rete considerato.

Il metodo di NetworkX utilizzato nella generazione di questa tipologia di grafi è `barabasi_albert_graph(n, m, seed)`, che restituisce un'istanza di grafo Barabási-Albert di dimensione n in cui ogni nuovo nodo viene collegato nel procedimento iterativo a m nodi esistenti. Anche in questo caso il parametro `seed` regola il comportamento casuale della generazione, agendo in particolare sulla scelta dell'insieme di archi da creare ad ogni iterazione, permettendo così la replicabilità dell'esperimento.

Algoritmo 4: Generazione di un grafo di Barabási-Albert

Input: numero di nodi n e numero di archi m da creare tra ogni nuovo nodo ed i nodi esistenti
inizializza un grafo G con m nodi;
 $target$ = lista inizialmente contenente gli m nodi iniziali;
 r = lista di nodi inizialmente vuota;
while ci sono ancora nodi da aggiungere **do**
 aggiungi un nuovo nodo al grafo;
 genera un arco tra il nuovo nodo e ogni nodo in $target$;
 aggiungi il nuovo nodo m volte a r ;
 aggiungi tutti i nodi contenuti in $target$ a r ;
 $target$ = sottoinsieme casuale di r di cardinalità m ;
end
return G

Sebbene rappresenti molto fedelmente per certi aspetti grafi reali, grazie soprattutto alla proprietà dell'invarianza di scala, il modello di generazione proposto da Barabási ed Albert non riesce a replicare altrettanto bene altre proprietà di queste reti, come ad esempio l'alto indice di clustering globale (proprietà che, al contrario, viene ricreata perfettamente dal modello di Watts-Strogatz). Per questo motivo sia il modello di Watts-Strogatz che quello di Barabási-Albert devono essere considerati come modelli solo parzialmente realistici.

2.2 Generazione delle istanze di problemi MIP

Terminata la generazione dei grafi, costituenti il *dataset* grezzo alla base di questo lavoro, è stato necessario ricavare a partire da ciascuno di essi il corrispondente problema di programmazione lineare intera del vertex cover minimo. Per fare questo è stato inizialmente necessario ricavare una modellazione algebrica del problema stesso, definendo variabili, vincoli e funzione obiettivo con cui rappresentarlo formalmente. Solo in un secondo momento è stato quindi possibile procedere con l'implementazione del modello nel software e la costruzione delle singole istanze, mediante l'ausilio di un'apposita libreria Python.

2.2.1 Modellazione algebrica

Un *modello algebrico* di un problema di programmazione lineare intera è un modello che descrive le caratteristiche della soluzione ottima al problema mediante relazioni matematiche [15]. Un modello algebrico è composto dai seguenti elementi:

- *insiemi*, che raggruppano gli elementi del sistema permettendo l'indicizzazione delle grandezze trattate nel problema
- *parametri*, ossia quantità costanti e definite dalla formulazione del problema stesso, che solitamente sono delle proprietà degli insiemi precedentemente definiti
- *variabili*, ossia le grandezze incognite del sistema, su cui l'algoritmo può agire (nei limiti dei vincoli imposti dal loro dominio) per ottimizzare il valore della soluzione
- *vincoli*, ossia relazioni matematiche che permettono di verificare l'ammissibilità delle soluzioni e, di conseguenza, definire quali assegnazioni alle variabili sono accettabili nel contesto del problema e quali invece non lo sono
- *funzione obiettivo*, che permette la traduzione di una soluzione del problema in un valore numerico, rendendone possibile di conseguenza l'ottimizzazione

Un modello algebrico di un problema di programmazione lineare permette quindi la definizione in forma dichiarativa delle caratteristiche della soluzione cercata, piuttosto che definire una strategia con cui eseguire la ricerca della soluzione stessa. Un modello di programmazione matematica potrebbe essere quindi paragonato ad un linguaggio dichiarativo, che descrive il risultato che si vuole ottenere, piuttosto che ad un linguaggio procedurale, che indica invece la sequenza di passi da percorrere per arrivare a quel risultato.

Durante la fase di modellazione formale del problema viene persa parte della potenza descrittiva caratteristica del linguaggio naturale, a vantaggio però della possibilità di utilizzare algoritmi particolarmente efficienti nella risoluzione dello stesso, che possono essere applicati solo quando il problema è definito in questa forma.

La modellazione algebrica non è stata, nel caso del problema analizzato in questo lavoro, particolarmente onerosa.

Per prima cosa sono stati identificati gli insiemi del problema, ovvero

V : insieme dei vertici del grafo

E : insieme degli archi del grafo

Successivamente, è stato necessario identificare i parametri del problema. Avendo studiato la formulazione *unweighted* del problema di vertex cover, in cui tutti i vertici hanno un peso associato $c_v = 1$, non è stato necessario definire alcun parametro. Infine sono state definite le variabili del problema, indicizzate dall'insieme dei vertici V

$$x_v = \begin{cases} 1 & \text{se il vertice } v \in V \text{ è compreso nel vertex cover} \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso il vincolo imposto dal problema è uno solo, ossia che almeno uno dei due vertici collegati da un arco sia presente nel vertex cover, per ognuno degli archi del grafo. Questo vincolo può essere rappresentato in forma dichiarativa mediante le variabili e gli insiemi precedentemente definiti come

$$x_u + x_v \geq 1 \quad \forall (u, v) \in E, u \in V, v \in V$$

La definizione della funzione obiettivo $f(x)$, che dovrà essere minimizzata dal risolutore, viene a questo punto naturale

$$f(x) = \sum_{v \in V} x_v$$

Il modello algebrico del problema di programmazione lineare del minimo vertex cover può quindi essere complessivamente definito come

$$\begin{aligned} & \min (\sum_{v \in V} x_v) \\ & x_u + x_v \geq 1 \quad \forall (u, v) \in E, u \in V, v \in V \\ & x_v \in 0, 1 \quad \forall v \in V \end{aligned} \tag{2.1}$$

2.2.2 Modellazione nel software

Al fine di tradurre le istanze di grafo generate con NetworkX in istanze di problemi MIP nella forma (2.1) che potessero essere risolte da CPLEX è stata utilizzata un'altra libreria Python, Pyomo [16][17]. Si tratta anche in questo caso di una libreria open-source, stabile e largamente utilizzata per formulare, risolvere ed analizzare modelli per l'ottimizzazione.

Una caratteristica molto importante di questa libreria, che è stata sfruttata in questo lavoro, è la separazione che viene mantenuta tra modello e dati. In particolare, Pyomo permette la definizione di un modello astratto, chiamato `AbstractModel`, che è definito secondo una struttura molto simile al modello algebrico e in cui non è richiesto che sia specificato nessun dato del problema. La definizione di questi ultimi è stata necessaria solo in un momento successivo, quando per ciascuno dei grafi generati si è creato un'istanza reale del modello stesso, chiamata `ConcreteModel`.

Le istanze generate in questo modo sono state salvate in file LP, formato utilizzato da CPLEX per il salvataggio di istanze di problemi di programmazione lineare in forma algebrica, che ha permesso di separare le fasi di generazione e risoluzione delle istanze stesse.

2.3 Risoluzione delle istanze di problemi MIP

La risoluzione delle istanze di problemi di vertex cover precedentemente create è stata svolta utilizzando il software IBM ILOG CPLEX Optimization Studio [18]. Si tratta di una suite sviluppata dall'azienda francese ILOG, di proprietà di IBM, che fornisce strumenti per la modellazione e la risoluzione di problemi di ottimizzazione e che rappresenta di fatto lo stato dell'arte nell'ambito della programmazione lineare. CPLEX prende il nome dal metodo del simplesso (*simplex method*) implementato in C, sebbene al giorno d'oggi la suite sia arrivata a comprendere diversi algoritmi aggiuntivi utilizzati nel campo della programmazione matematica e diverse altre interfacce verso altri ambienti e linguaggi diversi dal C.

Proprio una di queste interfacce, che permette di far dialogare software Python con il risolutore C, è stata utilizzata al fine di automatizzare il processo di risoluzione delle numerose istanze di problema. Si è infatti sviluppato un apposito script che permette di leggere i vari file `.lp` contenenti le istanze MIP e risolverli sequenzialmente, implementando in questo modo un processo di risoluzione di tipo *batch*.

Gli output prodotti da CPLEX nella risoluzione di ciascuna istanza sono stati memorizzati su disco in file separati (uno per ciascuna istanza di problema), in modo tale da permettere in un secondo momento l'estrazione dei dati ritenuti interessanti mediante l'utilizzo di un parser sviluppato ad-hoc.

Oltre agli output prodotti dal risolutore, per ogni istanza del problema sono stati memorizzati nel medesimo file anche due informazioni aggiuntive non comprese nell'output standard del risolutore, ovvero il gap tra lower bound del problema e soluzione trovata, utile nel caso in cui il problema vada in *time limit* per valutare l'ottimalità della soluzione trovata, e il valore della funzione obiettivo terminata la risoluzione, corrispondente alla cardinalità dell'insieme vertex cover trovato come soluzione.

2.4 Estrazione ed analisi dei risultati

L'estrazione dei dati dai file di testo contenenti gli output del risolutore è stata anch'essa implementata mediante un semplice script dedicato, sviluppato in linguaggio Python. Lo script permette, in maniera sequenziale, di leggere tutti i file contenenti gli output del risolutore in una specifica cartella e produrre, a partire da questi ultimi, un file CSV (*Comma-Separated Values*, formato di file basato su file di testo molto utilizzato nell'importazione ed esportazione di una tabella di dati) per ciascuna delle classi di grafo analizzate.

I file CSV sono stati utilizzati per salvare tutti i dati utilizzati poi nella successiva fase di elaborazione in un formato facilmente interpretabile da moltissimi software di elaborazione dati. Un esempio della struttura di uno dei quattro file CSV ottenuti come output è riportato in Figura 2.5.

	name	n	p	seed	time	ticks	sol_nodes	gap	time_lim	edges	cnnect_cmp	avg_clust	avg_deg	std_deg
0	gnp_0000	100	0.1	1	0.44	99.51	69.0	0.0	False	508	1	0.113342	10.16	3.273286
1	gnp_0001	100	0.1	2	0.45	100.66	68.0	0.0	False	484	1	0.097705	9.68	2.989582
2	gnp_0002	100	0.1	3	0.54	138.64	69.0	0.0	False	487	1	0.083568	9.74	2.851736

Figura 2.5: Esempio di struttura di un file CSV contenente i risultati della risoluzione di tutti i problemi di vertex cover di una certa tipologia di grafi, in questo caso corrispondente ai grafi $G(n, p)$.

Come si può notare dalla figura precedente, per ognuna delle istanze sono stati salvati diversi dati sulla risoluzione e sul grafo associato al problema di ottimizzazione.

L'analisi dei dati estratti è stata infine svolta con il supporto di un Jupyter Notebook [19], applicazione web open-source che permette di coniugare in un unico documento testo ed esecuzione di codice.

Capitolo 3

Risultati sperimentali

Conclusa la presentazione dell'approccio adottato nelle sperimentazioni, si procede in questo capitolo con l'esposizione dei risultati ottenuti.

La macchina utilizzata nella risoluzione delle diverse istanze in questi esperimenti presenta le seguenti specifiche tecniche:

- Intel(R) Xeon(R) CPU E5-2623 v3 @ 3.00GHz quad-core
- 16GB RAM
- Python 3.6.8
- IBM ILOG CPLEX Optimization Studio versione 12.9

3.1 Risultati per grafi $G(n, p)$

La prima categoria di grafi di cui sono stati analizzati i risultati sono i grafi $G(n, p)$, generati utilizzando il modello di Erdős–Rényi (esposto nella Sottosezione 2.1.1).

Per prima cosa è stata studiata la correlazione tra la probabilità di generazione di ogni arco e il tempo impiegato da CPLEX nella soluzione dell'istanza del problema di vertex cover, per ciascuno dei parametri di dimensione utilizzati. Per fare questo è stato disegnato un grafico, riportato di seguito in Figura 3.1.

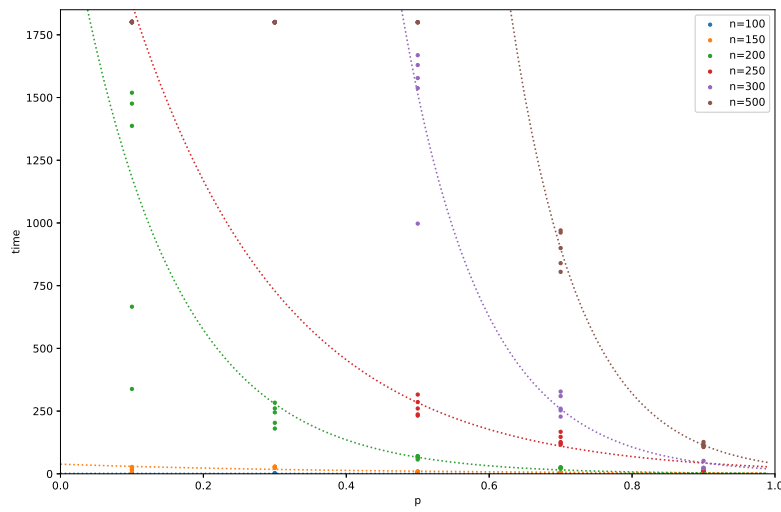


Figura 3.1: Correlazione tra tempo di risoluzione *time* (ordinata) e probabilità di generazione di ciascun arco p (ascissa).

Oltre al semplice plot dei dati, in questo grafico è stata estrapolata una curva da ciascuno degli insiemi di misurazioni corrispondenti ad ogni dimensione (applicando la regressione lineare), in modo tale da riuscire ad approssimare l'andamento. La regressione è naturalmente stata fatta esclusivamente per le misurazioni di *time* inferiori al time limit. Il grafico risultante evidenzia una chiara relazione di tipo esponenziale inverso tra p e *time*, per qualsiasi valore del parametro grandezza utilizzato.

Anche la dimensione dell'istanza di grafo sembra giocare un ruolo relativamente importante nella determinazione della complessità del problema di vertex cover associato, essendo apparentemente responsabile per la velocità con cui l'esponenziale satura. Per approfondire meglio il ruolo della dimensione nella determinazione della complessità di risoluzione si è rappresentato graficamente l'andamento del tempo di risoluzione al variare della dimensione del grafo e a parità di valori del parametro p . I risultati ottenuti sono riportati in Figura 3.2.

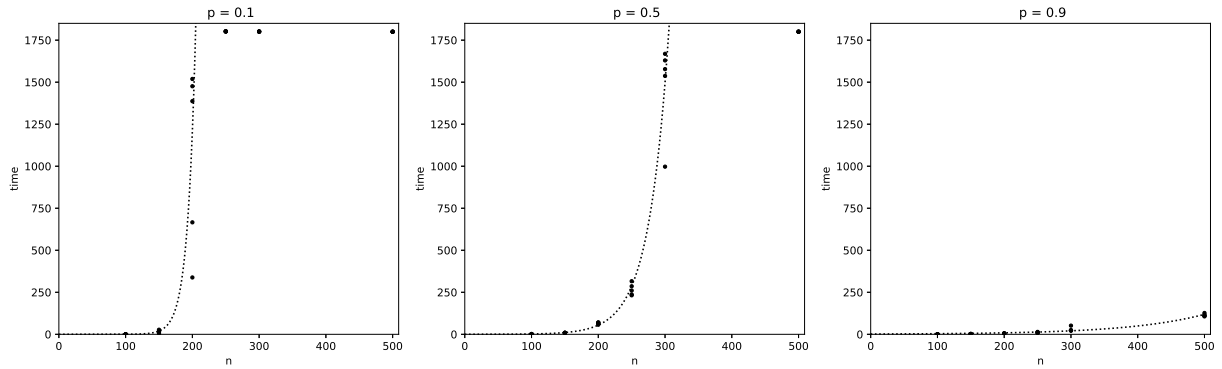


Figura 3.2: Correlazione tra tempo di risoluzione *time* (ordinata) e dimensione del grafo n (ascissa) per diversi valori di probabilità fissata $p = 0.1$, $p = 0.5$ e $p = 0.9$.

Come si può evincere dal grafo precedente, istanze di grafi più grandi diventano computazionalmente intrattabili molto prima rispetto a istanze dalla dimensione più contenuta, a parità di p . Tuttavia, come dimostra il grafico in Figura 3.2 ottenuto per $p = 0.9$, il contributo dato dal numero di nodi alla complessità di risoluzione sembra essere secondario rispetto al ruolo giocato dalla probabilità di generazione di ogni arco p .

In Figura 3.3 viene infine riportato un grafico tridimensionale riassuntivo, che rappresenta in un'unica immagine il variare della complessità del problema rispetto ad entrambi i parametri utilizzati nella generazione del grafo.

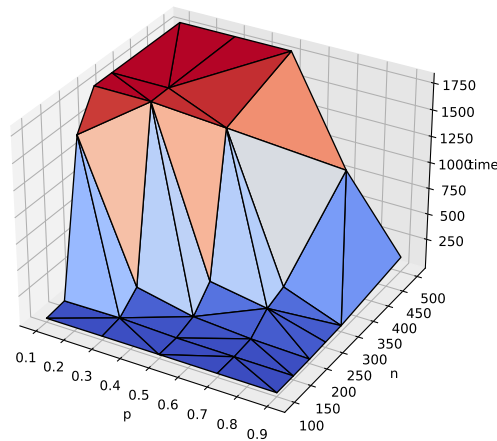


Figura 3.3: Rappresentazione tridimensionale dell'andamento della complessità del problema di vertex cover per grafi $G(n, p)$ al variare dei parametri p ed n . Colori più caldi indicano una complessità maggiore ed il colore rosso scuro è generalmente sinonimo di raggiungimento del time limit imposto al risolutore

3.2 Risultati per grafi regolari

La seconda categoria di grafi per cui è stato studiato l'andamento della complessità computazionale è stata quella dei grafi regolari, generati secondo il modello di Steger-Wormald già descritto nella Sottosezione 2.1.2.

In questo caso, una prima analisi è stata fatta sull'andamento della complessità computazionale del problema di vertex cover associato al variare del grado di ciascun nodo d , per ognuna delle dimensioni di grafo specificate. Il grafico risultante da questa analisi è riportato in Figura 3.4. A differenza dei grafi $G(n, p)$, in cui a parità di parametri di generazione n e p il tempo di risoluzione dimostrava oscillazioni quasi trascurabili per i diversi semi utilizzati nella gestione del comportamento casuale, nel caso dei grafi regolari le oscillazioni tra istanze generate a partire da semi diversi sono risultate essere molto marcate. Per questa ragione è stato considerato per ognuna delle coppie (n, d) il valore medio del tempo di risoluzione.

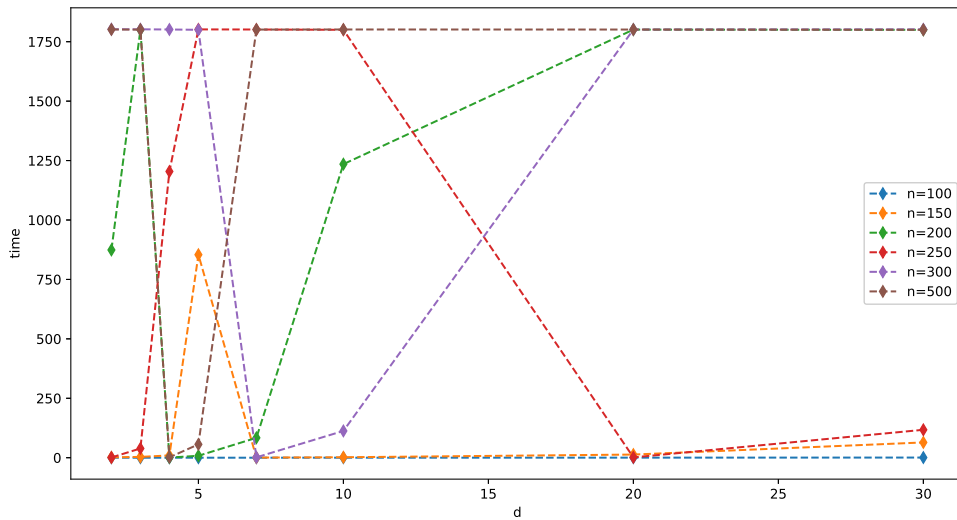


Figura 3.4: Andamento della complessità del problema di vertex cover per grafi regolari al variare del grado di ciascun nodo d , per diversi valori di dimensione n .

In questo caso risulta molto difficile, se non impossibile, riuscire a trovare un pattern ricorrente nei comportamenti delle diverse istanze. Una debole ipotesi sulla correlazione tra dimensione del grafo e complessità del problema potrebbe essere avanzata per questa particolare tipologia di grafi, sebbene esistano due valori di dimensione ($n = 200$ e $n = 250$) che si comportano esattamente al contrario di quanto previsto. Questa anomalia potrebbe essere tuttavia dovuta al fatto che i due valori di dimensione si trovano in una zona di confine, in cui il comportamento della complessità del problema è instabile e viene fortemente influenzato dalla forma delle specifiche istanze.

Una proprietà comune alle istanze di tutte le dimensioni (escluse quelle di dimensione più piccola generate per $n = 100$) sembra poi essere un'inversione di tendenza della complessità del problema (sia in meglio che in peggio) in una finestra di valori del parametro d compresa tra 3 e 10. Questo fenomeno fa sì che istanze di grande dimensione che per tutti gli altri valori di d risultano perennemente irrisolvibili diventino molto facili e, al contrario, istanze di piccola dimensione che risultano essere sempre risolte in tempi molto piccoli vadano invece in time limit. Una rappresentazione tridimensionale dell'andamento della complessità al variare dei due parametri in contemporanea, che permette di apprezzare meglio questo fenomeno, è riportata in Figura 3.5.

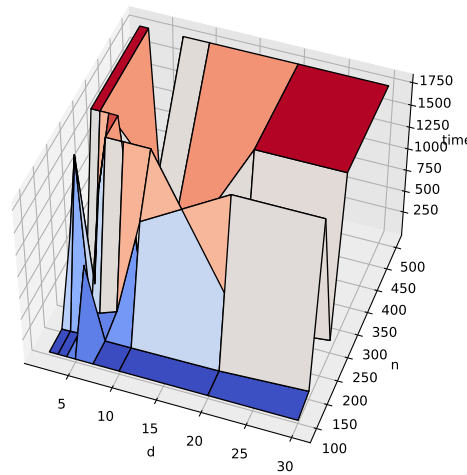


Figura 3.5: Rappresentazione tridimensionale dell'andamento della complessità del problema di vertex cover per grafi regolari al variare dei parametri d ed n . Anche in questo caso colori più caldi indicano una complessità maggiore, ed il colore rosso scuro è generalmente sinonimo di raggiungimento del time limit imposto al risolutore

Anche in questo caso, dare una spiegazione a questo fenomeno risulta molto complicato e sarebbe necessaria un'analisi più approfondita sulla forma delle specifiche istanze dei grafi e sullo sulle loro proprietà.

3.3 Risultati per grafi di Watts-Strogatz

La terza categoria di grafi per cui è stata analizzato l'andamento della complessità di risoluzione è stata quella dei grafi di Watts-Strogatz, generati secondo l'omonimo modello precedentemente trattato nella Sottosezione 2.1.3.

Essendo in questo caso 3 i parametri variabili del modello, è stato necessario fissare in tutte le analisi uno di essi, per permettere di visualizzare i dati in uno spazio a 2 o 3 dimensioni. La scelta del parametro da fissare è ricaduta in maniera piuttosto naturale su p , ossia la probabilità con cui viene eseguito il *rewiring* di ciascuno degli archi della rete ad anello iniziale. Questo parametro regola, come già visto in precedenza, la casualità (o, al contrario, la regolarità) del grafo prodotto e, fissandone il valore arbitrariamente, si è potuto svolgere un'indagine considerando ogni volta insiemi di grafi con un equivalente grado di regolarità.

Il valore del parametro p si è dimostrato influenzare notevolmente quella che è la complessità del problema associato, come si può facilmente intuire dai grafici riportati in Figura 3.6. I grafi generati per valori di p più piccoli, e quindi più simili al grafo regolare ad anello iniziale, in generale non vanno mai in time limit, cosa che invece accade sempre per valori di $p \geq 0.1$.

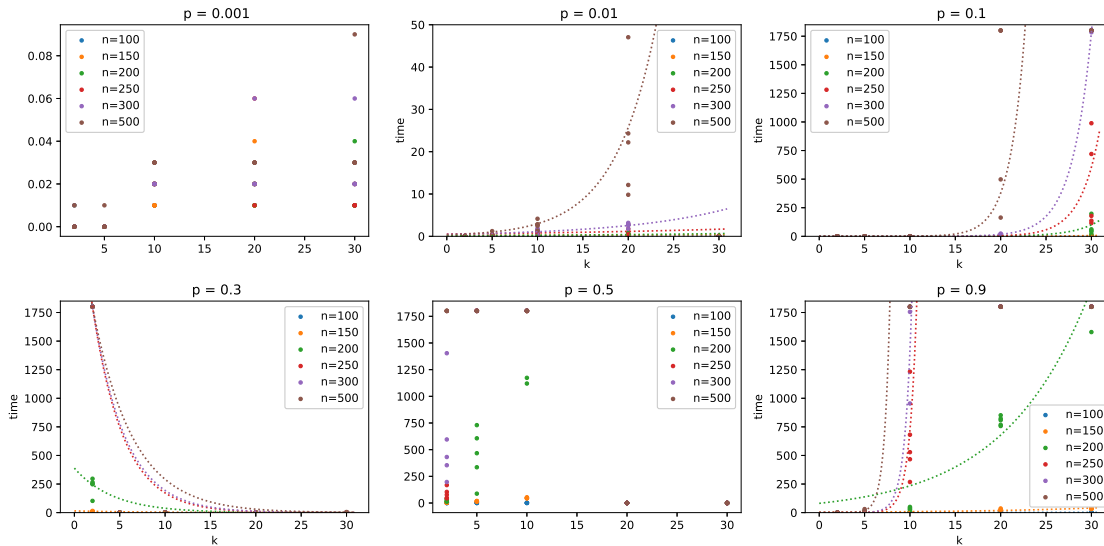


Figura 3.6: Andamento della complessità del problema di vertex cover per grafi di Watts-Strogatz al variare del numero di collegamenti iniziali k per diversi valori di dimensione n , per valori della probabilità di *rewiring* p fissati.

Nei grafi riportati in figura precedente si è cercato di estrapolare, nei limiti del possibile, una curva per ognuno degli andamenti legati ad uno specifico valore di dimensione. Un fenomeno interessante che si può notare è che per i valori di p più piccoli (tranne

che per $p = 0.001$, probabilmente troppo piccolo per riuscire ad apprezzare in maniera sensibile differenze di tempo di esecuzione tra le varie istanze) la correlazione tra k e tempo di risoluzione è di tipo esponenziale crescente, ovvero il tempo di risoluzione cresce esponenzialmente all'aumentare del valore di k .

Per un breve intervallo di p compreso tra 0.3 e 0.5 poi la tendenza si inverte e diventa di tipo esponenziale decrescente, ovvero al crescere di k il tempo necessario alla risoluzione diminuisce. In particolare, il caso $p = 0.5$ presenta un comportamento non ben definito, in cui gli andamenti non sono univoci e a cui non è stato possibile applicare il metodo della regressione. L'andamento è infine ritornato ad essere di tipo esponenziale crescente per $p = 0.9$.

Gli stessi andamenti sono stati rappresentati, in maniera più compatta, mediante l'ausilio di grafici tridimensionali, che hanno permesso la visualizzazione della complessità di risoluzione del problema al variare contemporaneo dei due parametri n e k , per una serie di valori p fissati. I risultati di queste elaborazioni grafiche sono riportati in Figura 3.7.

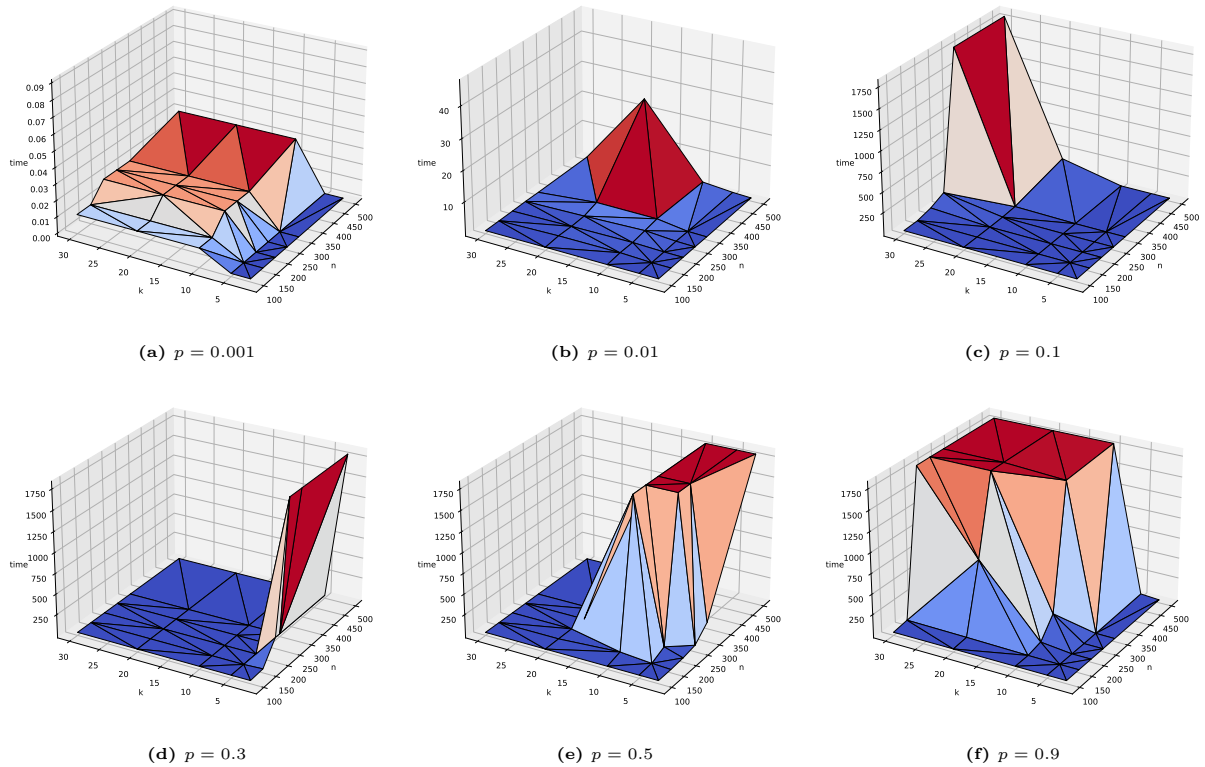


Figura 3.7: Andamento del tempo di risoluzione impiegato da CPLEX al variare dei due parametri n e k , per valori di p fissati.

3.4 Risultati per grafi di Barabási-Albert

L'ultima categoria di grafi di cui sono stati studiati i risultati sperimentali sono stati quelli generati mediante il modello di Barabási-Albert, già trattato nella Sottosezione 2.1.4.

Per prima cosa è stata analizzata la correlazione tra dimensione del grafo e complessità nella risoluzione del problema, estrapolando una curva per ognuno degli insiemi di misurazioni legate ad uno specifico valore del parametro d . L'andamento ottenuto si è dimostrato essere marcatamente esponenziale (crescente). È stato possibile anche in questo caso applicare il metodo della regressione lineare al fine di ottenere delle curve in grado di approssimare sufficientemente bene l'andamento del tempo impiegato dal risolutore ad ottenere una soluzione. Il grafico che è risultato da questo studio dei risultati è riportato in Figura 3.8.

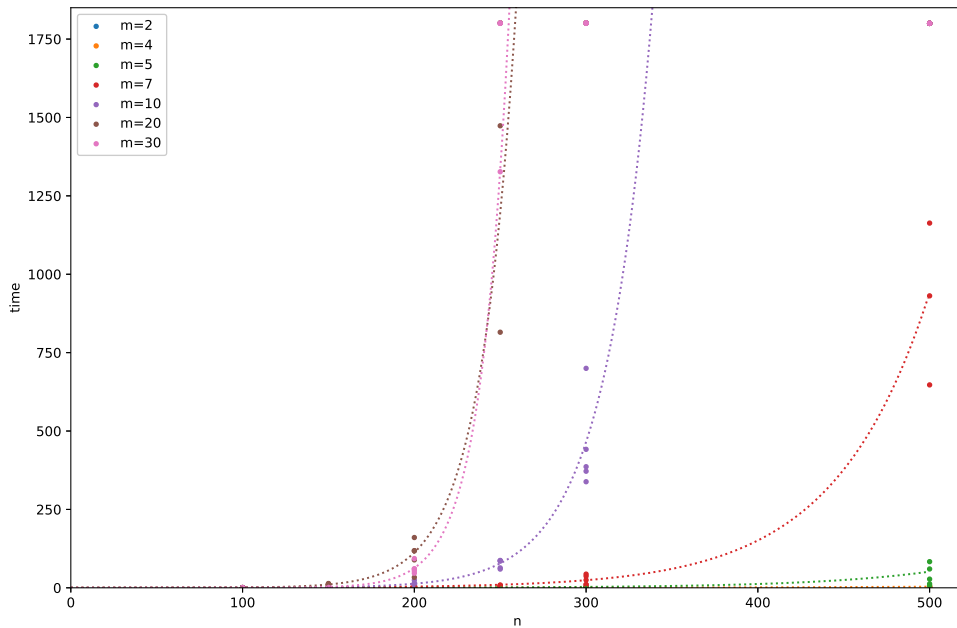


Figura 3.8: Andamento del tempo impiegato da CPLEX per risolvere l'istanza di problema (ordinata) in funzione della grandezza del grafo (ascissa), per diversi valori del parametro m , corrispondente al numero di archi creati tra ogni nuovo nodo ed i nodi esistenti nel processo di generazione.

In seconda istanza è stato analizzato il contributo del parametro m , mantenendo questa volta costante il numero dei nodi del grafo. Come si può evincere dai grafici riportati in Figura 3.9, il parametro m sembra avere un impatto marginale sulla complessità del problema, o almeno non paragonabile a quello determinato dalla dimensione del problema stesso. Valori anche grandi di m infatti non causano mai da soli la divergenza delle tempo di risoluzione, ma al limite la favoriscono nel caso di grafi dalla dimensione marcata.

È inoltre interessante notare come, per il valore $m = 30$, il tempo di risoluzione sia mediamente inferiore (per le istanze di problema che non vanno in time limit) rispetto al caso $m = 20$, in netto contrasto con l'andamento strettamente crescente evidenziato fino a quel momento. Ciò potrebbe essere dovuto ad una particolare configurazione di forma assunta dai grafi per tale valore di m o potrebbe trattarsi invece di una vera e propria inversione di tendenza. In entrambi i casi, uno studio per un insieme molto più esteso di valori del parametro m si renderebbe necessario per verificare l'ipotesi.

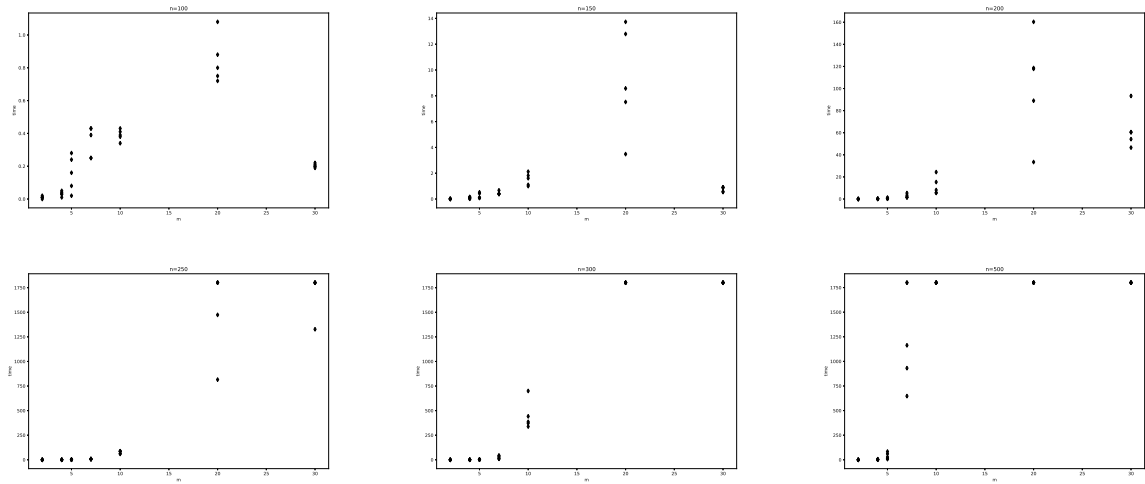


Figura 3.9: Andamento del tempo di risoluzione in funzione del parametro m , che rappresenta il numero di nodi esistenti a cui è collegato un nuovo nodo nel processo iterativo di costruzione del grafo, per valori di n fissati.

Infine, anche per i risultati ottenuti dall'analisi di grafi di Barabási-Albert è stato elaborato un grafico tridimensionale per visualizzare l'andamento della complessità del problema di vertex cover associato al variare di entrambi i parametri di generazione m e n . Tale grafico è riportato in Figura 3.10.

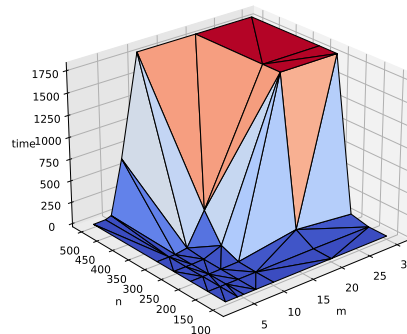


Figura 3.10: Rappresentazione tridimensionale dell'andamento della complessità del problema di vertex cover per grafi di Barabási-Albert al variare dei parametri m ed n . Anche in questo caso colori più caldi indicano una complessità maggiore, ed il colore rosso scuro è generalmente sinonimo di raggiungimento del time limit imposto al risolutore.

Capitolo 4

Conclusioni

In questo lavoro è stato affrontato lo studio di un particolare problema di ottimizzazione, il problema del vertex cover minimo. In particolare, ne è stato sperimentalmente analizzato l'andamento della complessità di risoluzione al variare dei parametri utilizzati nella generazione, per ognuno dei modelli di generazione di grafi casuali considerati.

Per alcuni modelli è stato possibile individuare degli andamenti della complessità nella risoluzione del problema chiari e coerenti, come ad esempio nel caso dei grafi di Barabási-Albert e di Erdős-Rényi. Per altre tipologie di modelli di generazione di grafi casuali l'individuazione di un andamento chiaro è stato possibile solo in parte (grafi di Watts-Strogatz), mentre per altri ancora non è stata possibile in alcuna misura (grafi regolari).

Gli esperimenti presentati in questo lavoro sono stati tuttavia svolti su di un insieme di combinazioni di parametri relativamente ristretto. Un interessante sviluppo delle sperimentazioni potrebbe consistere nell'estensione dell'insieme di combinazioni di parametri utilizzati, per verificare la validità delle considerazioni proposte in questo lavoro ed eventualmente per produrne di nuove, e l'estensione del time limit imposto al risolutore CPLEX per la soluzione dei problemi, al fine di aumentare la discriminazione nell'insieme di istanze che al momento vanno in time limit.

Bibliografia

- [1] Guido Van Rossum e Fred L Drake Jr. *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.
- [2] <https://github.com/giacomocamposampiero/bachelor-thesis>.
- [3] Domenico Salvagnin. *Cenni di Programmazione Lineare Intera*. Università degli Studi di Padova, 2020.
- [4] Domenico Salvagnin. *Introduzione all'ottimizzazione discreta*. Università degli Studi di Padova, 2018.
- [5] A. H. Land e A. G. Doig. «An Automatic Method of Solving Discrete Programming Problems». In: *Econometrica* 28.3 (1960), pp. 497–520. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/1910129>.
- [6] M. Padberg e G. Rinaldi. «Optimization of a 532-city symmetric traveling salesman problem by branch and cut». In: *Operations Research Letters* 6.1 (1987), pp. 1–7. ISSN: 0167-6377. DOI: [https://doi.org/10.1016/0167-6377\(87\)90002-2](https://doi.org/10.1016/0167-6377(87)90002-2). URL: <https://www.sciencedirect.com/science/article/pii/0167637787900022>.
- [7] Aric A. Hagberg, Daniel A. Schult e Pieter J. Swart. «Exploring Network Structure, Dynamics, and Function using NetworkX». In: *Proceedings of the 7th Python in Science Conference*. A cura di Gaël Varoquaux, Travis Vaught e Jarrod Millman. Pasadena, CA USA, 2008, pp. 11–15.
- [8] P. Erdős e A. Rényi. «On Random Graphs I». In: *Publicationes Mathematicae Debrecen* 6 (1959), p. 290.
- [9] J. D. Hunter. «Matplotlib: A 2D graphics environment». In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.
- [10] A. STEGER e N. C. WORMALD. «Generating Random Regular Graphs Quickly». In: *Combinatorics, Probability and Computing* 8.4 (1999), pp. 377–396. DOI: 10.1017/S0963548399003867.

- [11] Duncan J. Watts e Steven H. Strogatz. «Collective dynamics of ‘small-world’ networks». In: *Nature* 393.6684 (1998), pp. 440–442. DOI: 10.1038/30918.
- [12] Stanley Milgram. «The Small-World Problem». In: *Psychology Today* 1.1 (1967), pp. 61–67.
- [13] Albert-László Barabási e Réka Albert. «Emergence of Scaling in Random Networks». In: *Science* 286.5439 (ott. 1999), pp. 509–512. ISSN: 1095-9203. DOI: 10.1126/science.286.5439.509. URL: <http://dx.doi.org/10.1126/science.286.5439.509>.
- [14] L. D. Costa, F. Rodrigues e A. Cristino. «Complex networks: The key to systems biology». In: *Genetics and Molecular Biology* 31 (2008), pp. 591–601.
- [15] Laura Brentegani Luigi De Giovanni. *Modelli di Programmazione Lineare*. Università degli Studi di Padova.
- [16] Michael L. Bynum et al. *Pyomo–optimization modeling in python*. Third. Vol. 67. Springer Science & Business Media, 2021.
- [17] William E Hart, Jean-Paul Watson e David L Woodruff. «Pyomo: modeling and solving mathematical programs in Python». In: *Mathematical Programming Computation* 3.3 (2011), pp. 219–260.
- [18] IBM, ILOG CPLEX Optimization Studio 12.9.0. <https://www.ibm.com/docs/en/icos/12.9.0>.
- [19] Thomas Kluyver et al. «Jupyter Notebooks - a publishing format for reproducible computational workflows». In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. A cura di Fernando Loizides e Birgit Schmidt. Netherlands: IOS Press, 2016, pp. 87–90. URL: <https://eprints.soton.ac.uk/403913/>.