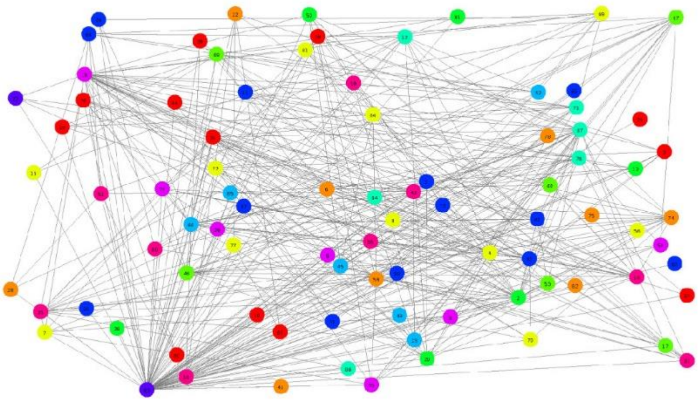
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
& ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Αλγόριθμοι και Πολυπλοκότητα

Εργασία εξαμήνου
Αλγόριθμοι Χρωματισμού Γραφημάτων



Ονοματεπώνυμο : Οικονόμου Αναστασία
Αριθμός Μητρώου : 15590 (παλαιό) 1412 (νέο)

Επιβλέπων καθηγητής : Γκόγκος Χρήστος

Άρτα 2020-2021

Περιεχόμενα

1. Εισαγωγή

Στη παρούσα εργασία στόχος μας είναι η επίλυση του προβλήματος χρωματισμού γραφημάτων με τη χρήση διαφορετικών αλγόριθμων κάθε φορά, όπως για παράδειγμα με τον First Fit, DSatur κλπ. Όμως αρχικά θα πρέπει να επισημάνουμε τη σημασία ορισμένων εννοιών, όπως για παράδειγμα τι ονομάζουμε κλάση P ,τι NP καθώς επίσης και τη σημασία της NP πληρότητας.

1.1 Κλάση P

Η γενική κλάση ερωτημάτων για το οποίο κάποιος αλγόριθμος δίνει την απάντηση σε πολυωνυμικό χρόνο ονομάζεται "κλάση P" ή απλούστερα "P". Για κάποια ερωτήματα δεν υπάρχει γνωστός τρόπος για την γρήγορη εύρεση απάντησης, αλλά αν κάποιος διαθέτει πληροφορίες που να αποδεικνύουν ποια είναι η απάντηση, είναι δυνατό να επιβεβαιώσει την απάντηση γρήγορα. Η κλάση των προβλημάτων που μπορούν να "επιβεβαιωθούν" σε πολυωνυμικό χρόνο ονομάζεται κλάση NP.[1]

1.2 Κλάση NP

Η κλάση πολυπλοκότητας NP αντιστοιχεί σε όλα τα προβλήματα για τα οποία υπάρχει πολυωνυμικός ανταιτιοκρατικός (non-deterministic) αλγόριθμος. Υπάρχει όμως ένας ισοδύναμος ορισμός που είναι εξαιρετικά πιο διαισθητικός και δείχνει ακριβώς τη σημασία αυτής της κλάσης πολυπλοκότητας. Ένα πρόβλημα P ανήκει στην κλάση N P αν έχει την ιδιότητα της πολυωνυμικής επαληθευσιμότητας. [2]

1.3 NP-πληρότητα

Σε αυτό το σημείο θα ορίσουμε την κλάση NP- C (N P-Complete) που είναι υποσύνολο της κλάσης NP. Αυτή περιλαμβάνει όλα τα NP-πλήρη προβλήματα που υπό μία έννοια είναι τα **πιο δύσκολα** της κλάσης NP. Ένα πρόβλημα Π λέγεται NP-πλήρες, όταν ισχύουν τα εξής:

1. $\Pi \in NP$

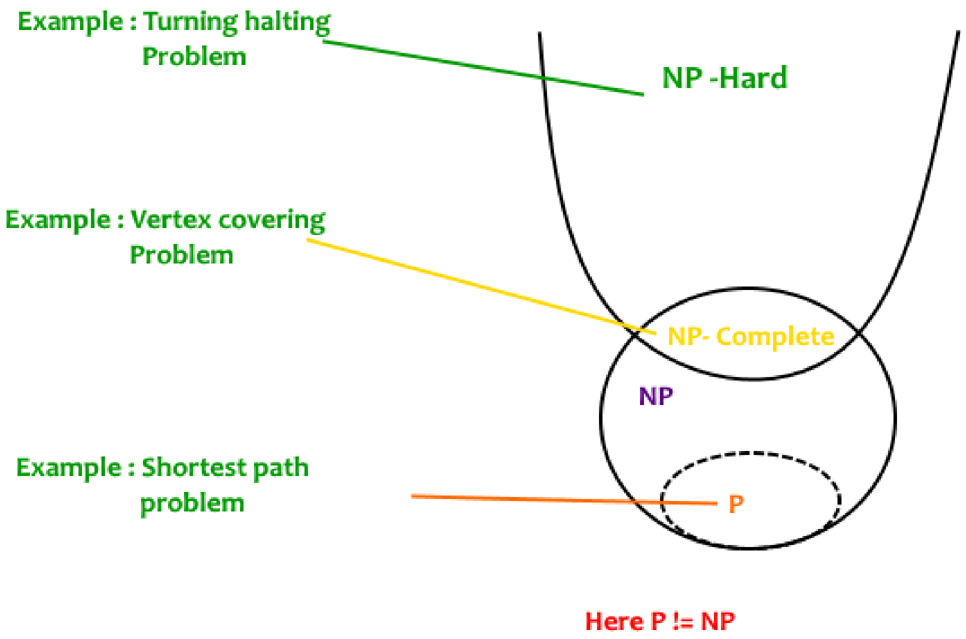
2. $\forall \Xi \in NP \Rightarrow \Xi \leq \Pi$

Για να δείξουμε, λοιπόν, ότι ένα πρόβλημα Π είναι NP-πλήρες θα πρέπει να δείξουμε ότι ανήκει στην κλάση NP κατασκευάζοντας έναν πολυωνυμικό επαληθευτή και έπειτα να δείξουμε ότι κάθε πρόβλημα στην κλάση NP ανάγεται στο Π. Το δεύτερο μέρος, όμως, είναι αρκετά πολύπλοκο και για αυτό το λόγο αν αποδείξουμε ότι ισχύει για ένα πρόβλημα, έπειτα κάνοντας αναγωγές από αυτό θα αποδεικνύουμε ότι και άλλα προβλήματα είναι NP-πλήρη. Το γνωστό θεώρημα των Cook-Levin δίνει το πρώτο τέτοιο πρόβλημα που είναι NP-πλήρες. [2]

1.4 Διαφορές μεταξύ NP-hard & NP-Complete problem

NP-hard Problems (έστω X) μπορεί να λυθεί αν και μόνο αν υπάρχει ένα NP-Complete πρόβλημα (έστω Y) μπορεί να είναι reducible στο X σε πολυωνυμικό χρόνο.	Τα προβλήματα **NP-Complete** μπορούν να επιλυθούν με ντετερμινιστικό αλγόριθμο σε πολυωνυμικό χρόνο.
Για να λύσει αυτό το πρόβλημα, πρέπει να είναι ένα πρόβλημα NP.	Για να λυθεί αυτό το πρόβλημα, πρέπει να είναι τόσο NP όσο και NP-hard πρόβλημα.
Δεν πρόκειται για πρόβλημα απόφασης.	Είναι αποκλειστικά πρόβλημα απόφασης .
Παράδειγμα: Halting problem, Vertex cover problem, Circuit-satisfiability problem κ.λπ.	Παράδειγμα: Προσδιορίστε αν ένα γράφημα έχει κύκλο Χάμιλτον, Προσδιορίστε αν ένας τύπος Boolean είναι ικανοποιητικός ή όχι, κ.λπ.

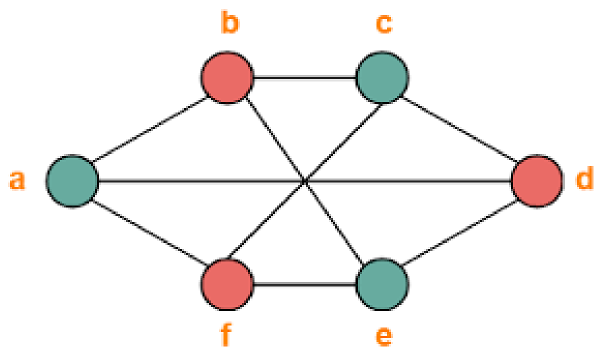
Πίνακας 1. NP-hard vs NP-Complete [3]



Εικόνα 1. Αναπαράσταση κλάσεων P, NP, NP-Complete, καθώς και NP-hard

2. Περιγραφή του προβλήματος

Το πρόβλημα του χρωματισμού γραφήματος είναι ένα NP-hard πρόβλημα συνδυαστικής βελτιστοποίησης. Αφορά την ανάθεση ενός χρώματος σε κάθε κορυφή ενός γραφήματος έτσι ώστε γειτονικές κορυφές να χρωματίζονται με διαφορετικό χρώμα όπως στην εικόνα 2, ενώ παράλληλα χρησιμοποιείται ο ελάχιστος αριθμός διαφορετικών χρωμάτων.



Εικόνα 2. Παράδειγμα graph coloring

Το πρόβλημα χρωματισμού γραφήματος τυπικά ορίζεται ως εξής. Δεδομένου ενός μη κατευθυνόμενου απλού γραφήματος $G = (V, E)$ με ένα σύνολο κορυφών V και ένα σύνολο ακμών E, ζητείται η ανάθεση σε κάθε κορυφή $v \in V$ ενός ακεραίου $c(v) \in \{1, 2, ..., k\}$ έτσι ώστε το k να ελαχιστοποιείται και να ισχύει ότι $c(v) \neq c(u)\ \forall \{v, u\} \in E$. Το πρόβλημα συναντάται σε μεγάλο αριθμό πρακτικών εφαρμογών όπως ο χρονοπρογραμματισμός εκπαιδευτικών ιδρυμάτων (educational timetabling), ο χρονοπογραμματισμός αθλητικών γεγονότων (sports scheduling), η ανάθεση συχνοτήτων (frequency assignment), η ανάθεση καταχωρητών στους μεταγλωττιστές (compiler register allocation) και άλλα. Πολλοί αλγόριθμοι χρωματισμού γραφημάτων έχουν προταθεί τα τελευταία 50 έτη. Στην παρούσα εργασία θα εξεταστούν τέσσερις αλγόριθμοι που ανήκουν στις λεγόμενες κατασκευαστικές τεχνικές (constructive techniques). Οι κατασκευαστικές τεχνικές δημιουργούν λύσεις βήμα προς βήμα, αναθέτοντας στη σειρά, σε κάθε κορυφή, ένα χρώμα, πιθανά εφαρμόζοντας οπισθοχώρηση κατά τη διαδικασία. Οι αλγόριθμοι που θα εξεταστούν είναι ο αλγόριθμος first fit, ο αλγόριθμος DSATUR, ο αλγόριθμος Recursive Largest First και ο αλγόριθμος backtracking DSATUR. Πληροφορίες για τους ανωτέρω αλγορίθμους μπορούν να βρεθούν στο άρθρο [LTMG12] καθώς και στις αναφορές του ίδιου άρθρου.

3. Προσεγγίσεις επίλυσης

3.1 Δεδομένα προβλήματος

Το πρόβλημα χρονοπρογραμματισμού εξετάσεων αφορά φοιτητές που έχουν πραγματοποιήσει εγγραφές σε εξετάσεις μαθημάτων. Για κάθε εξέταση διατίθεται μια λίστα από φοιτητές και κάθε φοιτητής μπορεί να είναι εγγεγραμμένος σε μια ή περισσότερες εξετάσεις. Κάθε εξέταση θα πρέπει να τοποθετηθεί σε μια περίοδο εξέτασης και η λύση του προβλήματος συνίσταται στην ανάθεση όλων των εξετάσεων στο μικρότερο δυνατό αριθμό περιόδων έτσι ώστε να μην υπάρχουν **συγκρούσεις**, δηλαδή να μην υπάρχουν φοιτητές που θα έπρεπε να συμμετάσχουν σε εξετάσεις σε περισσότερα του ενός μαθήματα στην ίδια περίοδο. Ως δεδομένα του προβλήματος θα χρησιμοποιηθούν τα δεδομένα του προβλήματος χρονοπρογραμματισμού εξετάσεων Toronto τα οποία είναι διαθέσιμα προς μεταφόρτωση στη διεύθυνση :
https://github.com/chgogos/datasets/blob/main/UETT/toronto.zip.

Τα δεδομένα Toronto αποτελούνται από 13 προβλήματα και πληροφορίες για κάθε πρόβλημα παρουσιάζονται στον Πίνακα 1.

Τα αρχεία δεδομένων (κατάληξη .stu) διαθέτουν για κάθε σπουδαστή μια γραμμή που περιέχει τους αριθμούς των μαθημάτων στα οποία είναι εγγεγραμμένος χωρισμένους μεταξύ τους με κενά. Η πρώτη γραμμή του αρχείου αντιστοιχεί στον πρώτο σπουδαστή, η δεύτερη γραμμή στο δεύτερο σπουδαστή κ.ο.κ. Για παράδειγμα το αρχείο car-f-92.stu περιέχει 18419 σειρές δεδομένων και ξεκινά με τις ακόλουθες σειρές: 0170 0156 0281 0006 0154 0156 0383 0534 0535 0536 0275 0091 0160 0164 ... που σημαίνουν ότι :

- ο φοιτητής 1 έχει εγγραφεί στο μάθημα 0170
- ο φοιτητής 2 έχει εγγραφεί στο μάθημα 0156
- ο φοιτητής 3 έχει εγγραφεί στο μάθημα 0281
- ο φοιτητής 4 έχει εγγραφεί στο μάθημα 0006
- ο φοιτητής 5 στα μαθήματα 0154 0156 κ.ο.κ

Πίνακας 1: Δεδομένα προβλημάτων

Πρόβλημα	Αρχείο Δεδομένων	Εξετάσεις	Φοιτητές	Εγγραφές
car-f-92	car-f-92.stu	543	18419	55522
car-s-91	car-s-91.stu	682	16925	56877
ear-f-83	ear-f-83.stu	190	1125	8109
hec-s-92	hec-s-92.stu	81	2823	10632
kfu-s-93	kfu-s-93.stu	461	5349	25113
lse-f-91	lse-f-91.stu	381	2726	10918
pur-s-93	pur-s-93.stu	2419	30029	120681
rye-s-93	rye-s-93.stu	486	11483	45051
sta-f-83	sta-f-83.stu	139	611	5751
tre-s-92	tre-s-92.stu	261	4360	14901
uta-s-92	uta-s-92.stu	622	21266	58979
ute-s-92	ute-s-92.stu	184	2749	11793
yor-f-83	yor-f-83.stu	181	941	6034

Θεωρώντας κάθε εξέταση ως κόμβο ενός γραφήματος και κάθε ακμή ανάμεσα σε δύο κόμβους να υποδηλώνει την ύπαρξη κοινών φοιτητών ανάμεσα στις δύο εξετάσεις που βρίσκονται στα άκρα της ακμής, το πρόβλημα μπορεί να θεωρηθεί ως πρόβλημα χρωματισμού γραφήματος όπου κάθε χρώμα είναι και μια περίοδος εξέτασης.

3.2 Στατιστικά στοιχεία προβλημάτων

Θα εμφανίζονται τα ακόλουθα στατιστικά στοιχεία για καθένα από τα 13 προβλήματα του Toronto dataset:

1. Αριθμός κορυφών.

2. Πυκνότητα.

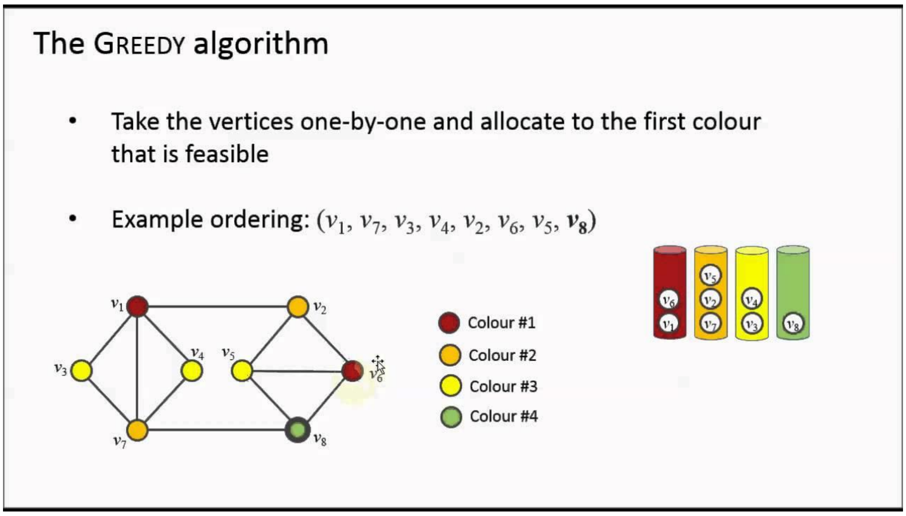
Για τον υπολογισμό της πυκνότητας θα πρέπει να κατασκευαστεί ο πίνακας συγκρούσεων. Ο πίνακας συγκρούσεων είναι ένας δισδιάστατος πίνακας c στον οποίο κάθε στοιχείο cij = 1 αν η εξέταση i βρίσκεται σε σύγκρουση με την εξέταση j ενώ ισχύει ότι cij = 0 σε άλλη περίπτωση. Η πυκνότητα συγκρούσεων υπολογίζεται διαιρώντας τον αριθμό των στοιχείων του πίνακα συγκρούσεων που έχουν την τιμή 1 με το συνολικό πλήθος των στοιχείων του πίνακα.

3. Για τους βαθμούς (degrees) των κορυφών η ελάχιστη τιμή (min), η διάμεσος τιμή (median), η μέγιστη τιμή (max), η μέση τιμή (mean) καθώς και ο συντελεστής διακύμανσης (CV=coefficient of variation) που ορίζεται ως η τυπική απόκλιση προς τη μέση τιμή.

4. Επίλυση προβλήματος.

4.1 Επίλυση προβλήματος με τη χρήση του αλγόριθμου FIRST FIT

Ο αλγόριθμος First Fit είναι ένας **άπληστος (greedy)** αλγόριθμος που λαμβάνει κάθε κορυφή και την αναθέτει στο μικρότερο αριθμό χρώματος που δεν προκαλεί σύγκρουση, δημιουργώντας έτσι νέα χρώματα όταν χρειάζεται. Οι κορυφές μπορούν αρχικά να ταξινομηθούν σε φθίνουσα σειρά βαθμού, όπως έχει προταθεί στο [WP67] και ο χρωματισμός των κορυφών να γίνει από την κορυφή με τον υψηλότερο βαθμό προς την κορυφή με το χαμηλότερο βαθμό.



Εικόνα 3. greedy algorithm

4.2Επίλυση του προβλήματος με τη χρήση του αλγορίθμου DSATUR

Ο αλγόριθμος DSATUR [Bré79] καθορίζει δυναμικά την επόμενη κορυφή που θα χρωματιστεί επιλέγοντας ανάμεσα στις κορυφές που δεν είναι χρωματισμένες εκείνη που κάθε φορά έχει το μεγαλύτερο αριθμό διαφορετικών χρωμάτων σε γειτονικές κορυφές.

5. Αποτελέσματα

Μαζική επίλυση με τον αλγόριθμο **First Fit** :

```
Name:datasets/hec-s-92.stu   |V|:81   Conflict Density:0.415485   min:9   median:32   max:62   mean:33.6543   cv:36.3264   greedy:22
Name:datasets/sta-f-83.stu   |V|:139   Conflict Density:0.143005   min:7   median:16   max:61   mean:19.8777   cv:67.3556   greedy:13
Name:datasets/yor-f-83.stu   |V|:181   Conflict Density:0.288209   min:7   median:51   max:117   mean:52.1657   cv:35.5273   greedy:26
Name:datasets/ute-s-92.stu   |V|:184   Conflict Density:0.0844754   min:2   median:13   max:58   mean:15.5435   cv:69.1351   greedy:13
Name:datasets/ear-f-83.stu   |V|:190   Conflict Density:0.265568   min:4   median:45   max:134   mean:50.4579   cv:56.1103   greedy:29
Name:datasets/tre-s-92.stu   |V|:261   Conflict Density:0.180003   min:0   median:45   max:145   mean:46.9808   cv:59.6187   greedy:29
Name:datasets/lse-f-91.stu   |V|:381   Conflict Density:0.0627579   min:0   median:16   max:173   mean:23.9108   cv:95.595   greedy:22
Name:datasets/kfu-s-93.stu   |V|:461   Conflict Density:0.0555286   min:0   median:19   max:247   mean:25.5987   cv:119.813   greedy:25
Name:datasets/rye-s-93.stu   |V|:486   Conflict Density:0.0751283   min:0   median:24   max:274   mean:36.5123   cv:111.763   greedy:28
Name:datasets/car-f-92.stu   |V|:543   Conflict Density:0.137732   min:0   median:63   max:381   mean:74.7882   cv:75.3453   greedy:44
Name:datasets/uta-s-92.stu   |V|:622   Conflict Density:0.125355   min:1   median:65   max:303   mean:77.9711   cv:73.671   greedy:43
Name:datasets/car-s-91.stu   |V|:682   Conflict Density:0.128198   min:0   median:77   max:472   mean:87.4311   cv:70.91   greedy:48
Name:datasets/pur-s-93.stu   |V|:2419   Conflict Density:0.0294831   min:0   median:47   max:857   mean:71.3196   cv:129.48   greedy:54
```

V: Vertices/ Κορυφές→ Σύνολο εξετάσεων

Conflict Density: Πυκνότητα συγκρούσεων

Για τους βαθμούς των κορυφών (degrees) (δηλαδή το κάθε μάθημα με πόσα άλλα μαθήματα συγκρούεται) :

- Min: Ελάχιστη τιμή
- Median: διάμεσος τιμή
- Max: Μέγιστη τιμή
- Mean: Μέση τιμή
- Cv: Τυπική Απόκλιση προς τη μέση τιμή.

First Fit/ greedy: Το σύνολο των χρωμάτων που χρησιμοποίησε ο αλγόριθμος για να χρωματίσει το γράφημα.

6. Συμπεράσματα

Τέλος, συμπεραίνουμε ότι όσο περισσότερα δεδομένα έχει ένα dataset τόσο περισσότερα χρώματα χρησιμοποιεί ο αλγόριθμος για να χρωματίσει το γράφημα.

ΑΝΑΦΟΡΕΣ

[1] https://el.wikipedia.org/wiki/%CE%A0%CF%81%CF%8C%CE%B2%CE%BB%CE%B7%CE%BC%CE%B1_P%3DNP

[2] https://repository.kallipos.gr/bitstream/11419/4015/1/chapter10Final.pdf

[3] https://www.geeksforgeeks.org/difference-between-np-hard-and-np-complete-problem/

[Bré79] Daniel Brélaz. New methods to color the vertices of a graph. Communications of the ACM, 22(4):251–256, 1979.

[WP67] Dominic JA Welsh and Martin B Powell. An upper bound for the chromatic number of a graph and its application to timetabling problems. The Computer Journal, 10(1):85–86, 1967.

[LTMG12] Rhyd Lewis, J Thompson, C Mumford, and J Gillard. A wide-ranging computational comparison of high-performance graph colouring algorithms. Computers & Operations Research, 39(9):1933–1950, 2012.