ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Προσεγγιστικά σχήματα, κλίκες, χρώματα και πυκνότεροι υπογράφοι

Μαρία Λιάζη

ΑΘΗΝΑ
ΙΟΥΝΙΟΣ 2008

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Προσεγγιστικά σχήματα, κλίκες, χρώματα και πυκνότεροι υπογράφοι

Μαρία Λιάζη

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: **Βασίλης Ζησιμόπουλος**, Καθηγητής ΕΚΠΑ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Βασίλης Ζησιμόπουλος, Καθηγητής ΕΚΠΑ
Ηλίας Κουτσουπιάς, Καθηγητής ΕΚΠΑ
Ιωάννης Μήλης, Αναπληρωτής Καθηγητής ΟΠΑ

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Ιωάννης Εμίρης	**Ευστάθιος Ζάχος**	**Βασίλης Ζησιμόπουλος**
Καθηγητής ΕΚΠΑ	Καθηγητής ΕΜΠ	Καθηγητής ΕΚΠΑ
Παναγιώτης Κατερίνης	**Σταύρος Κολλιόπουλος**	**Ηλίας Κουτσουπιάς**
Καθηγητής ΟΠΑ	Επίκουρος Καθηγητής ΕΚΠΑ	Καθηγητής ΕΚΠΑ

Ιωάννης Μήλης

Αναπληρωτής Καθηγητής ΟΠΑ

Ημερομηνία Εξέτασης: 19 Ιουνίου 2008

ΠΕΡΙΛΗΨΗ

Σε αυτήν τη διατριβή μελετάμε το πρόβλημα του πυκνότερου k-υπογράφου ενός δοθέντος γράφου $G = (V, E)$. Παρουσιάζονται αλγόριθμοι πολυωνυμικού χρόνου καθώς και κάποια προσεγγιστικά αποτελέσματα σε ειδικές κατηγορίες γράφων.

Αναλυτικότερα, στη διατριβή παρουσιάζονται αλγόριθμοι πολυωνυμικού χρόνου για το πρόβλημα του πυκνότερου k-υπογράφου σε βεβαρημένους γράφους με μέγιστο βαθμό δύο, δέντρα με βάρη όπου η λύση δεν είναι απαραίτητα συνεκτική, καθώς και σε γράφους διαστημάτων με τομή μόνο ανά δύο διαδοχικές κλίκες.

Επιπλέον, παρουσιάζεται ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα για το πρόβλημα του πυκνότερου k-υπογράφου σε αστέρι κλικών καθώς και ένας πολυωνυμικού χρόνου αλγόριθμος σε δέντρο κλικών φραγμένου βαθμού. Στο τελευταίο μέρος της διατριβής αναλύεται ένας σταθερού λόγου προσεγγιστικός αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου στους χορδικούς γράφους.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Αλγόριθμοι

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: πυκνότερος k-υπογράφος, πολυωνυμικού χρόνου αλγόριθμοι, προσεγγιστικοί αλγόριθμοι, πολυωνυμικού χρόνου προσεγγιστικό σχήμα

ABSTRACT

In this thesis we study the problem of finding the densest k-subgraph of a given graph $G = (V, E)$. We present algorithms of polynomial time as well as approximation results on special graph classes.

Analytically, in the thesis we study polynomial time algorithms for the densest k-subgraph problem on weighted graphs of maximal degree two, on weighted trees even if the solution is disconnected, and on interval graphs with intersection only between two consecutive cliques.

Moreover, we present a polynomial time approximation scheme for the densest k-subgraph problem on stars of cliques and a polynomial time algorithm on trees of cliques of bounded degree. Finally, in the last part of the thesis we analyze a constant-factor approximation algorithm for the densest k-subgraph problem on chordal graphs.

SUBJECT AREA: Algorithms

KEYWORDS: densest k-subgraph, polynomial time algorithms, approximation algorithms, polynomial time approximation scheme

Στον μπαμπά, στη μαμά και στην Άλεξ

Ευχαριστίες

Καταρχάς, θα ήθελα να ευχαριστήσω από καρδιάς τον επιβλέποντα μου, καθηγητή Βασίλη Ζησιμόπουλο, για την καθοδήγηση, τη βοήθεια και την υποστήριξη του όλα αυτά τα χρόνια. Τον ευχαριστώ γιατί μου έδωσε πάνω από όλα την ευκαιρία να καταπιαστώ με όμορφα ζητήματα της επιστήμης αλλά και να κατανοήσω καλύτερα τόσο τα "δύσκολα" όσο και τα "βέλτιστα".

Μεγάλο ευχαριστώ χρωστάω και στον καθηγητή Ιωάννη Μήλη. Δε θα ξεχάσω ποτέ ότι η πόρτα του γραφείου του ήταν πάντα ανοικτή για εμένα προσφέροντας μου απλόχερα τη βοήθεια του.

Θα ήθελα επίσης να ευχαριστήσω θερμά τους καθηγητές Ιωάννη Εμίρη, Ευστάθιο Ζάχο, Παναγιώτη Κατερίνη, Σταύρο Κολλιόπουλο και Ηλία Κουτσουπιά για την τιμή που μου έκαναν να είναι μέλη της επταμελούς εξεταστικής επιτροπής μου.

Τις ευχαριστίες μου θα ήθελα να εκφράσω και στην Fanny Pascual για την αρμονική και γόνιμη συνεργασία μας.

Σε αυτό το σημείο θα ήθελα να αναφέρω κάποια πρόσωπα που συχνά συναντούσα στους διαδρόμους και στα γειτονικά γραφεία στο DI και μου έκαναν πάντα τη μέρα πιο ευχάριστη. Χρυσηίδα, Βασίλη, Γεράσιμε, Δημήτρη, Κυριάκο, Κατερίνα, Αντώνη, Γιώργο Τ., Γιώργο Κ., Μαρία εσάς εννοώ και σας ευχαριστώ πολύ.

Θα μου επιτρέψετε να αναφερθώ ιδιαιτέρως σε τρία αγόρια που μετά από σχεδόν τρία χρόνια στο ίδιο γραφείο έχω τη χαρά και την ευχαρίστηση να αποκαλώ φίλους μου. Ευχαριστώ πολύ τον Ορέστη, το Γιώργο και τον Ηλία γιατί, εκτός του ότι δημιούργησαν για μένα ένα πολύ ανθρώπινο, φιλικό και βιώσιμο περιβάλλον εργασίας, μου έκαναν την τιμή όλα αυτά τα χρόνια να μοιραστούν μαζί μου τα περισσότερα από 'αυτά που λένε οι άντρες μεταξύ τους'.

Ευχαριστώ πολύ το Νίκο γιατί υποστηρίζει πάντα με μεγάλη θέρμη όλες μου τις προσπάθειες και με βοηθά να αντιμετωπίζω τις καταστάσεις με περισσότερο χιούμορ από ότι 'αρμόζει στη στιγμή'.

Τελευταία αλλά όχι έσχατη, θα ήθελα να ευχαριστήσω την οικογένεια μου, τον μπαμπά μου Σωτήρη, τη μαμά μου Ειρήνη και την αδερφή μου Αλεξία γιατί πάνω από όλα είναι η κινητήριος δύναμη μου. Τους χρωστώ ευγνωμοσύνη γιατί πάντα με στηρίζουν και με βοηθούν διατηρώντας συγχρόνως μία πολύ διακριτική στάση. Ιδιαίτερο ευχαριστώ χρωστάω στη μαμά Ειρήνη που μου έμαθε να μην υποχωρώ στα "δύσκολα" αλλά να τα αντιμετωπίζω με υπομονή, επιμονή, αισιοδοξία και πολύ δουλειά.

Μαρία Λιάζη
Αθήνα, Ιούνιος 2008

Περιεχόμενα

Κατάλογος Σχημάτων

Κεφάλαιο 1

Εισαγωγή

"Θα μου πεις σε παρακαλώ, ποιο δρόμο να πάρω;"
"Αυτό εξαρτάται άμεσα από το που θες να πας" είπε ο Γάτος.
"Δεν με πολυνοιάζει που θα πάω..." είπε η Αλίκη.
"Τότε δεν έχει σημασία ποιο δρόμο θα πάρεις" είπε ο Γάτος.
"...αρκεί να πάω κάπου" πρόσθεσε η Αλίκη σαν διευκρίνιση.

LEWIS CARROLL
Alice's Adventures in Wonderland

1.1 Το πρόβλημα του πυκνότερου k-υπογράφου

Στο πρόβλημα του πυκνότερου k-υπογράφου (*Densest* k-*subgraph* - DkS) δίνεται ένας γράφος $G = (V, E)$, $|V| = n$, και ένας ακέραιος $k \leq n$, και ζητείται ένας υπογράφος του G επαγόμενος από ακριβώς k κορυφές του έτσι ώστε ο αριθμός των ακμών του υπογράφου να είναι ο μέγιστος δυνατός. Το πρόβλημα είναι *NP-hard* ως γενίκευση του προβλήματος εύρεσης της μέγιστης κλίκας (*Maximum Clique problem*). Στη βεβαρημένη εκδοχή του προβλήματος δίνεται γράφος με μη αρνητικά βάρη στις ακμές και ζητείται ο k επαγόμενος υπογράφος με το μέγιστο συνολικό βάρος στις ακμές του.

Κατά τη διάρκεια των τελευταίων χρόνων μεγάλο μέρος εργασιών [4, 5, 10, 11, 12, 20, 25, 39, 40] έχουν επικεντρωθεί στη σχεδίαση προσεγγιστικών αλγορίθμων, τόσο για το DkS πρόβλημα όσο και για τη βεβαρημένη εκδοχή του, βασισμένοι σε τεχνικές όπως οι άπληστοι αλγόριθμοι (*greedy algorithms*), οι γραμμικές χαλαρώσεις (*LP relaxations*) και ο ημιορισμένος προγραμματισμός (*semidefinite programming*). Το καλύτερο από τα γνωστά προσεγγιστικά αποτελέσματα για το DkS πρόβλημα, το οποίο είναι ανεξάρτητο της τιμής του k, είναι $O(n^{\delta})$ για κάποιο $\delta < \frac{1}{3}$ [10], ενώ ένας απλός άπληστος αλγόριθμος [4], ο οποίος αφαιρεί την ελαχίστου βαθμού κορυφή κάθε φορά, επιτυγχάνει προσεγγιστικό λόγο $O(\frac{n}{k})$ ακόμα και για τη βεβαρημένη εκδοχή του προβλήματος.

Στον αντίποδα, έχει αποδειχθεί ότι εάν $NP \not\subseteq \cap_{\epsilon>0} BPTIME(2^{n^{\epsilon}})$ τότε το DkS πρόβλημα δεν επιδέχεται πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*PTAS*) [23]. Ωστόσο δεν υπάρχει κανένα αρνητικό αποτέλεσμα που να δείχνει ότι η επίτευξη προσεγγιστικού λόγου $O(n^{\epsilon})$, για κάποιο $\epsilon > 0$, είναι *NP-hard*. Θεωρώντας ειδικές περιπτώσεις για το πρόβλημα, είναι γνωστό ότι το DkS πρόβλημα επιδέχεται πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*PTAS*) σε γράφους ελαχίστου βαθμού $\Omega(n)$ καθώς και σε πυκνούς γράφους ($\Omega(n^2)$ ακμών) όταν το k είναι $\Omega(n)$ [1]. Επιπλέον, έχουν προταθεί αλγόριθμοι με προσεγγιστικό λόγο 4 [37] και 2 [21] για τη βεβαρημένη εκδοχή του προβλήματος σε πλήρης γράφους υπό την προϋπόθεση ότι τα βάρη ικανοποιούν την τριγωνική ανισότητα.

Το DkS πρόβλημα είναι τετριμμένο σε δέντρα (*trees*) διότι κάθε υποδέντρο k κορυφών περιέχει ακριβώς $k-1$ ακμές. Η βεβαρημένη εκδοχή του DkS προβλήματος είναι πολυωνυμική σε δέντρα όταν αναζητείται συνεκτική λύση [17, 33, 35]. Είναι επίσης γνωστό ότι το DkS πρόβλημα είναι πολυωνυμικό σε γράφους μέγιστου βαθμού δύο [12] καθώς και σε *cographs*, *split* γράφους και *k-trees* [7]. Από την άλλη μεριά, το DkS πρόβλημα παραμένει *NP-hard* σε διμερείς γράφους (*bipartite graphs*) [7], ακόμα και μέγιστου βαθμού τρία [12], καθώς επίσης σε *comparability* γράφους και σε χορδικούς γράφους (*chordal graphs*) [7]. Τέλος, το DkS παραμένει *NP-hard* σε επίπεδους γράφους (*planar graphs*) όταν η λύση που ζητείται είναι συνεκτική [22].

1.2 Εφαρμογές

Ένας γράφος συχνά συσχετίζεται με μία σχέση ανάμεσα σε ένα σύνολο ατόμων, δηλαδή ένα άτομο αντιστοιχεί σε μία κορυφή και μία ακμή μεταξύ δύο κορυφών υποδεικνύει κάποια "καλή" σχέση μεταξύ αυτών των δύο ατόμων. Έτσι μία κλίκα (*clique*) είναι ένα υποσύνολο ατόμων, τέτοιο που κάθε άτομο σε αυτό το σύνολο έχει μία "καλή" σχέση με κάθε άλλο άτομο σε αυτό το σύνολο. Ωστόσο λαμβάνοντας μία κλίκα στον πραγματικό κόσμο, αυτή η συνθήκη μοιάζει να είναι αρκετά ισχυρή· έτσι είναι περισσότερο σαν ένα υποσύνολο ατόμων στο οποίο υπάρχουν σχετικά πολλά ζεύγη "καλά" συσχετισμένων ατόμων. Με τους όρους ενός γράφου, το σύνολο αυτό είναι ένας υπογράφος που περιλαμβάνει σχετικά πολλές ακμές και έναν τέτοιο υπογράφο ενδιαφερόμαστε να βρούμε.

Το πρόβλημα εύρεσης πυκνών υπογράφων ενός γράφου έχει μελετηθεί διεξοδικά. Ερευνητές έχουν διερευνήσει διαφορετικούς ορισμούς της πυκνότητας και έχουν μελετήσει προβλήματα βελτιστοποίησης που αντιστοιχούν στην εύρεση υποδομών που μεγιστοποιούν τη δοθείσα έννοια της πυκνότητας κάθε φορά. Η πολυπλοκότητα των προβλημάτων βελτιστοποίησης διαφέρει ανάλογα με τον ορισμό της πυκνότητας που θα χρησιμοποιηθεί. Ειδικότερα, εάν $G = G(V, E)$ είναι ένας μη κατευθυνόμενος γράφος τότε ορίζουμε ως πυκνότητα $f(G)$ του G, $f(G) = \max_{S \subseteq V}\{f(S)\}$, όπου $f(S) = \frac{|E(S)|}{|S|}$ η πυκνότητα του υποσυνόλου $S \subseteq V$ και $E(S)$ οι ακμές που επάγονται από το S. Σύμφωνα με αυτόν τον ορισμό της πυκνότητας ενός μη κατευθυνόμενου γράφου, το πρόβλημα εύρεσης του πυκνότερου υπογράφου ενός δοθέντος μη κατευθυνόμενου γράφου (*Densest Subgraph problem*) μπορεί να λυθεί με *flow techniques* και ο γρηγορότερος γνωστός αλγόριθμος για αυτό το πρόβλημα εκτελείται σε χρόνο $O(mn \log n^2/m)$ [14].

Το πρόβλημα εύρεσης πυκνών υποδομών στο διαδίκτυο (*Web*) έχει λάβει αρκετή προσοχή. Ειδικότερα τέτοιες υποδομές αντιστοιχούν σε κοινότητες (*communities*) στο διαδίκτυο, δηλαδή συλλογές από σελίδες (*pages*) σχετικές με το ίδιο θέμα. Περαιτέρω η παρουσία μεγάλης πυκνότητας συνδέσμων (*links*) μέσα σε ένα συγκεκριμένο σύνολο σελίδων λαμβάνεται ως ένδειξη της σπουδαιότητας των σελίδων.

Ο αλγόριθμος του *Kleinberg* [24] αναγνωρίζει κόμβους (*hubs*) και πηγές (*authorities*) μεταξύ του συνόλου των πιθανών σελίδων σχετικών με ένα ερώτημα. Οι κόμβοι χαρακτηρίζονται από την παρουσία μεγάλου αριθμού συνδέσμων στις πηγές και οι πηγές χαρακτηρίζονται από την παρουσία μεγάλου αριθμού συνδέσμων από τους κόμβους. Οι κόμβοι και οι πηγές επιδεικνύουν μία αμοιβαία ενισχύουσα σχέση (*mutually reinforcing relationship*): ένας καλός κόμβος δείχνει σε πολλές καλές πηγές και μία καλή πηγή υποδεικνύεται από πολλούς καλούς κόμβους.

Η πυκνότητα ενός κατευθυνόμενου γράφου $G = G(V, E)$, με $S, T \subseteq V$ είναι $\max_{S,T\subseteq V}\{d(S,T)\}$, όπου $d(S,T) = \frac{|E(S,T)|}{\sqrt{|S||T|}}$ και $E(S,T)$ το σύνολο των ακμών που πηγαίνουν από το σύνολο S στο σύνολο T. Ο ορισμός της πυκνότητας για κατευθυνόμενους γράφους είναι κατάλληλος για αραιούς κατευθυνόμενους γράφους όπως είναι ο γράφος του διαδικτύου. Αυτό παρακινείται από την προσπάθεια διατύπωσης της ιδέας της εύρεσης συνόλων κόμβων και πηγών που έχουν μεγάλη συνεκτικότητα σε σχέση με τον υπόλοιπο γράφο.

Εκτός από το διαδίκτυο υπάρχουν αρκετές εφαρμογές του DkS προβλήματος. Ενδεικτικά αναφέρουμε την εφαρμογή του στην ασφάλεια παραγωγής στιγμιότυπων τυχαίων δοκιμών. Όταν παράγονται τυχαία στιγμιότυπα, για να εκτιμηθεί εμπειρικά η απόδοση αλγορίθμων βελτιστοποίησης, ο αλγόριθμος θα μπορούσε να συντονιστεί έτσι ώστε να τρέχει γρήγορα, ειδικά για τη δοκιμασία των επιδόσεών του, εάν εκμεταλλευτούμε τη μέθοδο παραγωγής των δοκιμασιών [3].

1.3 Διάρθρωση Κειμένου

Στο Κεφάλαιο 2 θα παρουσιάσουμε κάποια από τα αποτελέσματα της περιοχής, που είναι σχετικά και συγκρίσιμα με τη συνεισφορά μας. Στο Κεφάλαιο 3 θα μελετήσουμε το DkS πρόβλημα σε βεβαρημένους γράφους με μέγιστο βαθμό δύο καθώς και σε δέντρα με βάρη στις ακμές. Προτείνουμε "περιγραφή του αποτελέσματος" όταν η λύση δεν αποτελείται απαραίτητα από μία συνιστώσα. Επίσης, στο Κεφάλαιο αυτό εξετάζουμε το DkS πρόβλημα σε γράφους διαστημάτων με τομή μόνο ανάμεσα σε δύο διαδοχικές κλίκες.

Στο Κεφάλαιο 4 εξετάζουμε το DkS πρόβλημα σε γράφους κλικών. Συγκεκριμένα παρουσιάζουμε ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*PTAS*) σε αστέρι κλικών και έναν πολυωνυμικού χρόνου αλγόριθμο σε δέντρο κλικών φραγμένου βαθμού.

Στο Κεφάλαιο 5, αναλύουμε ένα σταθερού λόγου προσεγγιστικό αλγόριθμο για το DkS πρόβλημα σε χορδικούς γράφους, στους οποίους το πρόβλημα παραμένει δύσκολο. Τέλος, στο Κεφάλαιο 6 συνοψίζουμε τα αποτελέσματα των προηγουμένων κεφαλαίων και θέτουμε κάποια ενδιαφέροντα ανοικτά ερωτήματα για το πρόβλημα του πυκνότερου k-υπογράφου τόσο σε γενικούς γράφους όσο και σε κάποιες ειδικές αλλά σημαντικές κατηγορίες γράφων.

Κεφάλαιο 2

Υπάρχοντα αποτελέσματα

Τα σπουδαία πράγματα δεν είναι αποτέλεσμα κάποιας παρόρμησης, αλλά μίας σειράς μικρών πραγμάτων συγκεντρωμένων μαζί.

VINCENT VAN GOGH

2.1 Ορισμοί

Καταρχάς δίνουμε κάποιους ορισμούς βασικών εννοιών που θα αναφερθούν σε ότι ακολουθεί.

Η πυκνότητα d_G ενός γράφου $G = G(V, E)$ είναι ο μέσος βαθμός του, δηλαδή, $d_G = \frac{2|E|}{|S|}$.

Η κλάση πολυπλοκότητας NP είναι η κλάση των γλωσσών που μπορούν να πιστοποιηθούν από έναν πολυωνυμικού χρόνου αλγόριθμο. Ειδικότερα μία γλώσσα L ανήκει στην κλάση NP αν και μόνο αν υπάρχει ένας διπλής εισόδου πολυωνυμικού χρόνου αλγόριθμος A και μία σταθερά c τέτοια ώστε $L = \{x \in \{0,1\}^*$, υπάρχει ένα πιστοποιητικό y με $|y| = O(|x|^c)$ τέτοιο ώστε $A(x, y) = 1\}$. Λέμε ότι ο αλγόριθμος A πιστοποιεί τη γλώσσα L σε πολυωνυμικό χρόνο.

Μία γλώσσα $L \subseteq \{0,1\}^*$ είναι *NP-complete* εάν
- $L \in NP$ και
- $L' \leq_P L$ για κάθε $L' \in NP$, δηλαδή κάθε γλώσσα L' που ανήκει στην κλάση NP

ανάγεται πολυωνυμικά στη γλώσσα L.

Εάν μία γλώσσα L ικανοποιεί μόνο τη δεύτερη ιδιότητα, αλλά όχι απαραίτητα την πρώτη ιδιότητα, λέμε ότι η L είναι *NP-hard*.

Ας υποθέσουμε ότι A είναι ένα πρόβλημα βελτιστοποίησης, αυτό σημαίνει ότι για κάθε στιγμιότυπο x έχουμε ένα σύνολο εφικτών λύσεων, έστω $F(x)$, και για κάθε τέτοια λύση $s \in F(x)$ έχουμε ένα θετικό ακέραιο, το κόστος της $c(s)$. Η βέλτιστη λύση ορίζεται ως $OPT(x) = \max_{s \in F(x)} c(s)$ (ή $\min_{s \in F(x)} c(s)$, εάν το A είναι πρόβλημα ελαχιστοποίησης). Έστω M ένας αλγόριθμος ο οποίος δοθέντος ενός στιγμιότυπου x, επιστρέφει μία εφικτή λύση $M(x) \in F(x)$. Λέμε ότι ο M είναι ένας ϵ-προσεγγιστικός αλγόριθμος (ϵ-*approximation algorithm*), όπου $1 \geq \epsilon \geq 0$, εάν για όλα τα x έχουμε $\frac{|c(M(x)) - OPT(x)|}{\max\{OPT(x), c(M(x))\}} \leq \epsilon$.

Λέμε ότι για κάθε στιγμιότυπο x ενός προβλήματος με βέλτιστη λύση OPT, ένας προσεγγιστικός αλγόριθμος A για το πρόβλημα, έχει προσεγγιστικό λόγο ρ (*approximation ratio*), όπου $\rho \geq \frac{OPT(x)}{A(x)}$ για προβλήματα μεγιστοποίησης (και $\rho \geq \frac{A(x)}{OPT(x)}$ για προβλήματα ελαχιστοποίησης αντίστοιχα).

Ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*Polynomial Time Approximation scheme - PTAS*) είναι ένας αλγόριθμος τέτοιος που για κάθε καθορισμένο $\epsilon > 0$, επιτυγχάνει προσεγγιστικό λόγο $1 + \epsilon$ σε χρόνο πολυωνυμικό σε σχέση με το μέγεθος της εισόδου (αλλά που αυξάνει εκθετικά ως προς $\frac{1}{\epsilon}$, $O(n^{\frac{1}{\epsilon}})$).

Κλίκα (*clique*) ενός γράφου $G = (V, E)$ είναι ένας πλήρης υπογράφος του G επαγόμενος από ένα υποσύνολο κορυφών του, $C \subseteq V$. Το μέγεθος $|C|$ της κλίκας είναι ο αριθμός των κορυφών της. Μεγιστική (*maximal*) είναι μία κλίκα όταν δεν περιέχεται σε μεγαλύτερη κλίκα. Η μεγαλύτερη μεγιστική κλίκα ονομάζεται μέγιστη κλίκα (*maximum clique*).

Με K_n συμβολίζουμε έναν πλήρη υπογράφο n κορυφών, με $K_{1,n}$ το γράφο αστέρι $n + 1$ κορυφών και με P_n ένα μονοπάτι n κορυφών.

Ο γράφος τομής (*intersection graph*) μίας οικογένειας, F, υποσυνόλων ενός συνόλου ορίζεται ως ο γράφος, $\mathcal{G}$, του οποίου οι κορυφές αντιστοιχούν στα υποσύνολα της F και υπάρχει ακμή μεταξύ δύο κορυφών του $\mathcal{G}$ εάν το αντίστοιχο ζευγάρι υποσυνόλων έχει κοινή τομή. Με βάση αυτούς τους ορισμούς, ο γράφος μεγιστικών κλικών (*clique graph*) ενός γράφου G ορίζεται ως ο γράφος τομής των μεγιστικών κλικών (*maximal* *cliques*) του G με βάρη στις ακμές τον αριθμό των κορυφών στην κοινή τομή. Το δέντρο των μεγιστικών κλικών (*clique tree*) του G είναι το μεγίστου βάρους δέντρο επικάλυψης (*maximum weight spanning tree*) του γράφου μεγιστικών κλικών.

Χορδικός (*chordal*) είναι ο γράφος στον οποίο κάθε κύκλος με μέγεθος μεγαλύτερο του τρία έχει τουλάχιστον μία χορδή, δηλαδή μία ακμή που ενώνει δύο μη διαδοχικές κορυφές του κύκλου.

Ένας γράφος $G = (V, E)$ ονομάζεται γράφος διαστημάτων (*interval graph*) εάν υπάρχει μία αντιστοιχία I των κορυφών του G σε σύνολα διαδοχικών ακέραιων έτσι ώστε για κάθε ζεύγος κορυφών $u, v \in V$ να ισχύει ότι $(u, v) \in E$ αν και μόνο αν $I(u) \bigcap I(v) \neq \emptyset$. Ο *proper interval* γράφος κατασκευάζεται από μία οικογένεια διαστημάτων σε μία σειρά έτσι ώστε κανένα διάστημα να μην περιέχει εξ ολοκλήρου κάποιο άλλο. Οι *proper interval* γράφοι ορίζονται και ως οι γράφοι διαστημάτων που δεν περιέχουν το $K_{1,3}$.

Ένας γράφος λέγεται κανονικός (*regular*) εάν κάθε κορυφή του έχει τον ίδιο βαθμό (*degree*), δηλαδή, τον ίδιο αριθμό γειτόνων. Ένας κανονικός γράφος με κορυφές βαθμού d λέγεται d-κανονικός γράφος (d-*regular graph*) ή κανονικός γράφος βαθμού d.

Ένας γράφος $G = (V, E)$ είναι διμερής (*bipartite*) εάν το σύνολο κορυφών του, V, μπορεί να διαμεριστεί σε δύο ανεξάρτητα υποσύνολα.

Το πρόβλημα εύρεσης της μέγιστης κλίκας (*Maximum Clique problem*) είναι το πρόβλημα εύρεσης μίας κλίκας μεγίστου μεγέθους σε ένα γράφο. Σαν πρόβλημα απόφασης (k-*clique problem*) απλά ρωτάμε αν υπάρχει κλίκα μεγέθους τουλάχιστον k στο γράφο, δηλαδή, $CLIQUE = \{\langle G, k \rangle$: ο G είναι ένας γράφος με κλίκα μεγέθους τουλάχιστον $k\}$. Το πρόβλημα εύρεσης της μέγιστης κλίκας είναι *NP-complete*.

Στο πρόβλημα του σακιδίου (*Knapsack problem*) δίνεται σαν είσοδος ένα σύνολο αντικειμένων U και μία παράμετρος B. Κάθε αντικείμενο του συνόλου, $u \in U$, έχει ένα βάρος w_u και μία αξία p_u. Ζητείται ένα υποσύνολο $U' \subseteq U$ με $\sum_{u \in U'} w_u \leq B$ έτσι ώστε να μεγιστοποιείται το $\sum_{u \in U'} p_u$.

2.2 Δυσκολία προβλήματος

Στο πρόβλημα εύρεσης του πυκνότερου k-υπογράφου (*Densest k-subgraph problem* - DkS) δίνεται ένας γράφος $G = (V, E)$, $|V| = n$, και ένας ακέραιος $k \leq n$ και ζητείται ένας υπογράφος του G επαγόμενος από ακριβώς k κορυφές του τέτοιος ώστε ο αριθμός των ακμών του να είναι ο μέγιστος δυνατός. Το πρόβλημα είναι *NP-hard* ως γενίκευση του προβλήματος εύρεσης της μέγιστης κλίκας (*Maximum Clique problem*) ενός γράφου. Εάν μπορούσαμε να λύσουμε το DkS πρόβλημα σε πολυωνυμικό χρόνο, τότε παίρνοντας k από 1 έως $|V|$ θα μπορούσαμε να βρούμε τη μέγιστη κλίκα σε πολυωνυμικό χρόνο.

Στη βεβαρημένη εκδοχή του προβλήματος δίνεται γράφος με μη αρνητικά βάρη στις ακμές και ζητείται ο k επαγόμενος υπογράφος με το μέγιστο συνολικό βάρος στις ακμές του. Αυτή η εκδοχή του προβλήματος απαντάται στη βιβλιογραφία ως βαρύτερος k-υπογράφος (*Heaviest k-subgraph problem*).

Το αντίστοιχο πρόβλημα απόφασης του πυκνότερου k-υπογράφου (DSG) ρωτάει εάν δοθέντων ενός γράφου n κορυφών και m ακμών και δύο ακεραίων K_1 και K_2 υπάρχει υπογράφος του G ο οποίος έχει το πολύ K_1 κορυφές και τουλάχιστον K_2 ακμές. Παρατηρούμε ότι όταν $K_2 = K_1(K_1 - 1)/2$ το DSG είναι ισοδύναμο με το πρόβλημα εύρεσης της μέγιστης κλίκας και γι' αυτό το γενικό DSG είναι *NP-complete*.

Παρακάτω αναφέρουμε κάποια θεωρήματα που δείχνουν ότι το DSG παραμένει *NP-complete* ακόμα κι εάν περιορίσουμε το K_2.

Θεώρημα 1. *[2] Το DSG είναι NP-complete για το σύνολο των στιγμιότυπων (G, K_1, K_2) τέτοια ώστε $K_1 \leq \frac{n}{2}$, $K_2 \leq K_1^{(1+\epsilon)}$ και $K_2 \leq \frac{m}{4}(1 + \frac{9}{20} + o(1))$.*

Θεώρημα 2. *[2] Το DSG είναι NP-complete ακόμα και εάν $K_1 = \frac{n}{2}$ και $K_2 \leq \frac{m}{4}(1 + O(\frac{1}{\sqrt{n}}))$.*

Θεώρημα 3. *[2] Το DSG είναι NP-complete ακόμα και εάν $K_1 \leq |V|^{\varepsilon_1}$ και $K_2 \leq K_1^{1+\varepsilon_2}$ για κάθε $\varepsilon_1, \varepsilon_2 > 0$.*

Θεώρημα 4. *[2] Το DSG είναι NP-complete ακόμα κι αν ο γράφος είναι $|V|^{\varepsilon_3}-$ κανονικός (regular).*

2.3 Πολυωνυμικού χρόνου αλγόριθμοι

Το DkS πρόβλημα είναι τετριμμένο σε δέντρα (*trees*) διότι κάθε υποδέντρο k κορυφών περιέχει ακριβώς $k-1$ ακμές. Ωστόσο για να λύσουμε το DkS πρόβλημα σε ένα δάσος (*forest*) που αποτελείται από δέντρα T_i, $1 \leq i \leq m$, πρέπει να επιλύσουμε ένα πρόβλημα σακιδίου (*Knapsack problem*). Έστω n_i ο αριθμός των κορυφών και $e_i = n_i - 1$ ο αριθμός των ακμών κάθε δέντρου T_i του δάσους G.
Θεωρούμε το εξής πρόβλημα του σακιδίου:

$$\max \sum_{i=1}^{m} e_i x_i \quad \text{έτσι ώστε} \quad \sum_{i=1}^{m} n_i x_i \leq k \text{ και } x_i \in \{0, 1\}$$

Μία βέλτιστη λύση x_i^*, $1 \leq i \leq m$, αυτού του προβλήματος αποτελείται από ένα υποσύνολο δέντρων του G τέτοιο ώστε ο συνολικός αριθμός κορυφών τους να είναι το πολύ k και ο συνολικός αριθμός ακμών τους να είναι ο μέγιστος δυνατός. Ωστόσο, γενικά $k' = \sum_{i=1}^{m} n_i x_i^* < k$. Εάν συμβαίνει αυτό τότε υπάρχει κάποιο T_i τέτοιο ώστε $x_i^* = 0$ και $n_i > k - k'$, γιατί αλλιώς η βέλτιστη λύση θα μπορούσε να βελτιωθεί από την προσθήκη του T_i που είναι αντίφαση. Για το λόγο αυτό, η προσθήκη ενός υποδέντρου $k - k'$ κορυφών του T_i, και άρα των $k - k' - 1$ ακμών του, στη βέλτιστη λύση του παραπάνω προβλήματος σακιδίου οδηγεί σε μία βέλτιστη λύση για το DkS πρόβλημα. Το πρόβλημα του σακιδίου μπορεί να επιλυθεί βέλτιστα σε χρόνο $O(mk)$ και αφού $m \leq n$, όπου n το πλήθος των κορυφών του δάσους, η πολυπλοκότητα εύρεσης του DkS σε δάσος είναι $O(nk)$. Επιπλέον, αφού $k \leq n$, το πρόβλημα είναι πολυωνυμικό και λύνεται σε χρόνο το πολύ $O(n^2)$.

Αν θεωρήσουμε τώρα ότι τα δέντρα έχουν βάρη, τότε ένας $O(nk^2)$ χρόνου αλγόριθμος δυναμικού προγραμματισμού που προτείνεται ανεξάρτητα στα [35], [33] και [17] λύνει βέλτιστα το DkS πρόβλημα υπό τον περιορισμό ότι η λύση που ψάχνουμε είναι συνεκτική, δηλαδή, ένα (μοναδικό) υποδέντρο του αρχικού δέντρου. Εάν ψάχνουμε μία μη συνεκτική λύση τότε στο Κεφάλαιο 3 προτείνεται αλγόριθμος δυναμικού προγραμματισμού ο οποίος επιλύει το πρόβλημα σε κάθε περίπτωση.

Είναι επίσης γνωστό ότι το DkS πρόβλημα είναι πολυωνυμικό σε γράφους μέγιστου βαθμού δύο [12] καθώς και σε *cographs*, *split* γράφους και *k-trees* [7]. Στη συνέχεια παραθέτουμε αναλυτικότερα κάποια από τα προαναφερθέντα πολυωνυμικά αποτελέσματα.

2.3.1 Δέντρα με βάρη στις ακμές

Όπως προαναφέραμε, στα [35], [33] και [17] έχει προταθεί ένας αλγόριθμος δυναμικού προγραμματισμού, $O(nk^2)$ χρόνου, για τη βέλτιστη επίλυση του DkS προβλήματος σε δέντρα με βάρη στις ακμές υπό τον περιορισμό ότι αναζητούμε συνεκτική λύση. Συγκεκριμένα, στο άρθρο του Maffioli [33] παρουσιάζεται ένας δυναμικού προγραμματισμού αλγόριθμος για την εύρεση του ελαχίστου βάρους υποδέντρου k-κορυφών σε βεβαρημένα δέντρα. Εδώ παραθέτουμε διαφοροποιημένο αυτόν τον αλγόριθμο έτσι ώστε δοθέντος ενός δέντρου με βάρη να βρίσκει το μεγίστου βάρους υποδέντρο k-κορυφών.

Έστω $G = (V, E)$ ένα δέντρο με $|V| = n$ κορυφές και $|E| = n-1$ ακμές. Κάθε ακμή (i, j) του δέντρου έχει βάρος w_{ij}. Η τυχαία επιλογή ρίζας, r, για το δέντρο G δίνει έναν προσανατολισμό των ακμών του G από (και προς) τη ρίζα. Χρησιμοποιώντας τον προσανατολισμό από την ρίζα r ορίζουμε τα εξής:

- ρίζα του G είναι ο πρόγονος όλων των άλλων κορυφών του G,
- μία κορυφή j είναι παιδί της κορυφής i αν και μόνο αν $(i, j) \in E$,
- μία κορυφή j είναι απόγονος της κορυφής i αν και μόνο αν υπάρχει στο G ένα (προσανατολισμένο) μονοπάτι από το i στο j,
- το υποδέντρο $P(i)$ του G που περιέχει όλους τους απόγονους της κορυφής i καλείται *downtree* του G.

Τα φύλλα του G είναι *downtrees* με μηδέν ακμές, ενώ το G είναι *downtree* με ρίζα το r.

Για κάθε *downtree* P του G, το $z^P(h)$ είναι η τιμή ενός βέλτιστου υποδέντρου του P με ακριβώς h ακμές. Όταν το h είναι μεγαλύτερο από τον αριθμό των ακμών του P θέτουμε $z^P(h) = -\infty$. Επίσης, οι $z_1^P(h)$ και $z_0^P(h)$ είναι οι τιμές ενός βέλτιστου υποδέντρου του P με ακριβώς h ακμές, οι οποίες περιέχουν και δεν περιέχουν αντίστοιχα τη ρίζα του P.

Προφανώς, για κάθε h ισχύει
$z^P(h) = \max\{z_1^P(h), z_0^P(h)\}$.

Παρατηρούμε ότι για κάθε φύλλο του G ισχύει:
$z_0^P(h) = -\infty$,

$$z_1^P(h) = \begin{cases} 0, & \text{αν } h = 0 \\ -\infty, & \text{αν } h > 0. \end{cases}$$

Καταρχάς, ορίζουμε δύο λειτουργίες ADDFATHER και MERGE για τα *down-*

trees.

Κατά τη διάρκεια της $\mathrm{ADDFATHER}(a, b, f)$ αντικαθιστούμε τα δύο υποδέντρα A και B των οποίων οι ρίζες a και b είναι παιδιά της ίδιας κορυφής f, με ένα μοναδικό υποδέντρο F με ρίζα την κορυφή f. Εφόσον γνωρίζουμε τις τιμές των z_0 και z_1 για τα A και B μπορούμε να υπολογίσουμε τις τιμές των z_0 και z_1 (και άρα τη z) για το F με τις εξής σχέσεις:
$z_0^F(h) = \max\{z^A(h), z^B(h)\}$,
$z_1^F(h) = \max\{w_{fa} + z_1^A(h-1), w_{fb} + z_1^B(h-1), w_{fa} + w_{fb} + \eta\}$,
όπου $\eta := \max\{z_1^A(\alpha) + z_1^B(\beta) : \alpha + \beta = h - 2\}$.

Η $\mathrm{MERGE}(p, d)$ προσθέτει ένα *downtree* D του οποίου η ρίζα d είναι παιδί της κορυφής p στο *downtree* P με ρίζα το p, δημιουργώντας ένα μεγαλύτερο *downtree* Q με ρίζα την κορυφή p. Για την MERGE έχουμε τις εξής σχέσεις:
$z_0^Q(h) = \max\{z_0^P(h), z^D(h)\}$,
$z_1^Q(h) = \max\{z_1^P(h), w_{pd} + \vartheta\}$,
όπου $\vartheta := \max\{z_1^P(\alpha) + z_1^D(\beta) : \alpha + \beta = h - 1\}$.

Τέλος ορίζουμε μία πιο απλή λειτουργία, την ASCEND, η οποία μπορεί επίσης να θεωρηθεί και ως ειδική περίπτωση της ADDFATHER. Η $\mathrm{ASCEND}(l, c)$ προσθέτει ένα *downtree* C του οποίου η ρίζα c είναι παιδί της κορυφής l στην κορυφή l δημιουργώντας ένα μεγαλύτερο *downtree* L με ρίζα την κορυφή l. Για την $\mathrm{ASCEND}(l, c)$ θεωρούμε τις εξής σχέσεις:
$z_0^L(h) = z^C(h)$,
$z_1^L(h) = w_{lc} + z_1^C(h-1)$.

Οι ιδιότητες των *downtrees* μας επιτρέπουν να υλοποιήσουμε έναν αλγόριθμο για την εύρεση του βέλτιστου πυκνότερου k-υπογράφου όταν ο γράφος είναι δέντρο με βάρη στις ακμές και όταν η λύση που αναζητούμε είναι συνεκτική. Στον αλγόριθμο κάθε *downtree* ονομάζεται από τη ρίζα του και συνδέεται με έναν πίνακα με 3 στήλες και k σειρές. Οι τρεις στήλες αντιστοιχούν στις τιμές των $z_0(h)$, $z_1(h)$ και $z(h)$ αντίστοιχα. Κάθε σειρά αντιστοιχεί σε μία διαφορετική τιμή του h, για h από 0 έως $k-1$. Στην αρχή το σύνολο των *downtrees* περιέχει μόνο φύλλα. Με βάση τα παραπάνω περιγράφουμε έναν αλγόριθμο, πολυπλοκότητας $O(nk^2)$, για το DkS πρόβλημα σε ένα δέντρο G.

ΑΛΓΟΡΙΘΜΟΣ

αρχή

επίλεξε μία κορυφή του G σαν ρίζα r και δώσε προσανατολισμό στο G;

αρχικοποίησε το σύνολο των *downtrees* S ώστε να περιέχουν ένα *downtree* j για κάθε φύλλο j του G;

//κάθε *downtree* χαρακτηρίζεται από έναν πίνακα με 3 στήλες που αντιστοιχούν στα $z_0(h)$, $z_1(h)$ και $z(h)$ και k σειρές που αντιστοιχούν στις τιμές του $h = 0, \ldots, k-1$;//

επανέλαβε

όσο υπάρχει *downtree* c του οποίου η ρίζα είναι το μοναδικό παιδί της κορυφής l, ASCEND(l, c);

όσο υπάρχουν *downtrees* a και b των οποίων οι ρίζες έχουν τον ίδιο πατέρα f, ADDFATHER(a, b, f);

όσο υπάρχει *downtree* d του οποίου η ρίζα είναι παιδί της ρίζας του *downtree* p, MERGE(p, d)

μέχρι $|S| = 1$;

δώσε την τιμή $z^r(k-1)$

τέλος.

2.3.2 Γράφοι μέγιστου βαθμού 2

Στο άρθρο [12] αποδεικνύεται ότι το DkS πρόβλημα είναι *NP-hard* για γράφους με μέγιστο βαθμό τρία, ενώ στο ίδιο άρθρο αναφέρεται ότι υπάρχει πολυωνυμικού χρόνου αλγόριθμος δυναμικού προγραμματισμού για το DkS πρόβλημα σε γράφους με μέγιστο βαθμό δύο. Αναλύουμε περαιτέρω αυτή την αναφορά. Μία συνεκτική συνιστώσα g_i, $1 \leq i \leq m$, ενός γράφου G με μέγιστο βαθμό δύο είναι είτε ένας κύκλος (*cycle*) είτε μία αλυσίδα (*chain*). Έστω n_i ο αριθμός των κορυφών και e_i ο αριθμός των ακμών κάθε συνεκτικής συνιστώσας g_i του G (προφανώς, $e_i = n_i$, εάν g_i κύκλος, και $e_i = n_i - 1$, εάν g_i αλυσίδα).

Θεωρούμε το εξής πρόβλημα του σακιδίου (*Knapsack problem*):

$$\max \sum_{i=1}^{m} e_i x_i \quad \text{έτσι ώστε} \quad \sum_{i=1}^{m} n_i x_i \leq k \text{ και } x_i \in \{0, 1\}.$$

Μία βέλτιστη λύση x_i^*, $1 \leq i \leq m$, αυτού του προβλήματος αποτελείται από ένα υποσύνολο συνεκτικών συνιστωσών του G τέτοιο ώστε ο συνολικός αριθμός κορυφών τους να είναι το πολύ k και ο συνολικός αριθμός ακμών τους να είναι ο μέγιστος δυνατός. Ωστόσο, γενικά $k' = \sum_{i=1}^{m} n_i x_i^* < k$. Εάν συμβαίνει αυτό τότε υπάρχει κάποια g_i τέτοια ώστε $x_i^* = 0$ και $n_i > k - k'$, γιατί αλλιώς η βέλτιστη λύση θα μπορούσε να βελτιωθεί από την προσθήκη της g_i που είναι αντίφαση. Για το λόγο αυτό, η προσθήκη μίας αλυσίδας $k - k'$ κορυφών της g_i, και άρα των $k - k' - 1$ ακμών της, στη βέλτιστη λύση του παραπάνω προβλήματος σακιδίου οδηγεί σε μία βέλτιστη λύση για το DkS πρόβλημα. Το πρόβλημα του σακιδίου μπορεί να επιλυθεί βέλτιστα σε χρόνο $O(mk)$ και αφού $m \leq n$, όπου n το σύνολο κορυφών ενός γράφου G με μέγιστο βαθμό δύο, η πολυπλοκότητα εύρεσης του DkS σε γράφους με μέγιστο βαθμό δύο είναι $O(nk)$.

2.3.3 Cographs

Cographs είναι οι γράφοι οι οποίοι δεν έχουν P_4, δηλαδή, μονοπάτι μεγέθους τέσσερα, ή ισοδύναμα οι γράφοι που το συμπλήρωμα κάθε συνεκτικού επαγόμενου υπογράφου τους είναι μη συνεκτικό. Οι *cographs* έχουν τη σημαντική ιδιότητα της μοναδικής αναπαράστασης σε δέντρο. Σε αυτήν την αναπαράσταση (καλείται *cotree*) οι κορυφές του γράφου είναι τα φύλλα του δέντρου. Οι εσωτερικές κορυφές του δέντρου έχουν ετικέτα είτε (0) είτε (1). Η ρίζα r του δέντρου έχει ετικέτα (1) και σε κάθε μονοπάτι από τη ρίζα οι ετικέτες των εσωτερικών κορυφών εναλλάσσονται. Δύο κορυφές x και y του G είναι γειτονικές αν και μόνο αν το μοναδικό μονοπάτι από το x στο r πρώτα συναντά το μοναδικό μονοπάτι από το y στο r σε μία (1) κορυφή. Η μοναδική σε δέντρο αναπαράσταση οδηγεί σε πολλούς γρήγορους αλγόριθμους για τα *cographs*.

Έστω ότι ένας *cograph* G αναπαρίσταται με ένα *cotree* T. Δοθέντος ενός γράφου μπορεί να βρεθεί εάν είναι *cograph* και άρα να κατασκευαστεί το *cotree* του σε γραμμικό χρόνο. Παραθέτουμε έναν πολυωνυμικού χρόνου αλγόριθμο για το DkS πρόβλημα στους *cographs* όπως παρουσιάζεται στο άρθρο των Corneil και Perl, [7]. Ο αλγόριθμος υπολογίζει τον πυκνότερο i-υπογράφο για $i = 1, \ldots, k$. Ο πυκνότερος i-υπογράφος μπορεί εύκολα να ανακατασκευαστεί φυλάσσοντας τους κατάλληλους δείκτες καθώς εκτελείται ο αλγόριθμος.

Για κάθε κορυφή u του T υπολογίζουμε το $D(u, i)$, $1 \leq i \leq k$ που είναι ο πυκνότερος i-υπογράφος στον υπογράφο του G που αναπαρίσταται από το υποδέντρο T που έχει ρίζα το u. Εάν δεν υπάρχει τέτοιος πυκνότερος i-υπογράφος, τότε $D(u, i) = \wedge$. Προφανώς, $D(r, i)$ όπου r είναι η ρίζα του T και $1 \leq i \leq k$ είναι ο πυκνότερος i-υπογράφος του G. Ο αλγόριθμος είναι δυναμικού προγραμματισμού και υπολογίζει το D για κάθε κορυφή του T διασχίζοντας το δέντρο από τα φύλλα προς τη ρίζα.

Για κάθε φύλλο του T, θέτουμε $D(u, 0) = D(u, 1) = 0$ και $D(u, i) = \wedge$ για $1 < i \leq k$. Θεωρούμε ότι εάν ένα άθροισμα περιέχει ένα $\wedge$ ισούται με $\wedge$ και ότι στις συγκρίσεις $\wedge < 0$.

Έστω ότι η εσωτερική κορυφή v του T έχει d παιδιά $x_1, x_2, \ldots, x_d$ για τα οποία το D έχει ήδη υπολογιστεί. Λαμβάνοντας υπόψη όλες τις διαμερίσεις του i σε d μπορούμε να βρούμε το $D(v, i)$. Ωστόσο, ο αριθμός αυτών των διαμερίσεων είναι $O(i^{d-1})$, το οποίο δεν εγγυάται ότι ο υπολογισμός τους θα γίνει ακριβώς σε πολυωνυμικό χρόνο. Για να επιτύχουμε έναν πολυωνυμικού χρόνου αλγόριθμο εκτελούμε τον υπολογισμό του $D(v, i)$ με μία βήμα προς βήμα διαδικασία. Έστω $D(v, i, j)$, $1 \leq j \leq d$ ο πυκνότερος i-υπογράφος του υποδέντρου με ρίζα το v που περιέχει τα παιδιά του $x_1, x_2, \ldots, x_j$ και όλους τους απόγονους τους. Προφανώς, $D(v, i) = D(v, i, d)$.
Για $i = 1, 2, \ldots, k$ θέτουμε

$$D(v, i, 1) = D(x_1, i), \;\; D(v, 0) = 0 \;\text{ και }\; D(v, 0, m) = 0, \; 1 \leq m \leq d.$$

Για $j = 2, \ldots, d$ εάν η ετικέτα του v είναι $(v) = 0$ τότε

$$D(v, i, j) = \max_{0 \leq m \leq i} \{D(v, m, j-1) + D(x_1, i-m)\},$$

ενώ εάν η ετικέτα του v είναι $(v) = 1$ τότε

$$D(v, i, j) = \max_{0 \leq m \leq i} \{D(v, m, j-1) + D(x_j, i-m) + m(i-m)\}.$$

Η πολυπλοκότητα υπολογισμού του πυκνότερου k-υπογράφου για το d είναι $O(dk^2)$. Το άθροισμα των βαθμών όλων των κορυφών του T είναι φραγμένο από το διπλάσιο του αριθμού των φύλλων στο δέντρο που είναι $|V| = n$. Συνεπώς, ο αλγόριθμος υπολογίζει το DkS σε χρόνο $O(nk^2)$.

Στο άρθρο [7] αποδεικνύεται επίσης ότι εάν $k < n$ τότε κάθε πυκνότερος k-υπογράφος σε ένα συνεκτικό *cograph* n κορυφών είναι συνεκτικός.

2.3.4 Split γράφοι

Ένας γράφος είναι *split* γράφος εάν υπάρχει μία διαμέριση του συνόλου των κορυφών του, V, σε δύο υποσύνολα V_1 και V_2 τέτοια που ο επαγόμενος από το V_1 υπογράφος να είναι πλήρης και ο επαγόμενος από το V_2 υπογράφος να είναι κενός. Προφανώς, οι *split* γράφοι είναι χορδικοί.

Το σύνολο κορυφών, V, ενός *split* γράφου G μπορεί να διαμεριστεί σε έναν πλήρη υπογράφο C και σε ένα ανεξάρτητο σύνολο I. Εάν $k \leq c = |C|$, τότε ο πυκνότερος k-υπογράφος του G [7] είναι οποιοσδήποτε υπογράφος του G επαγόμενος από k κορυφές του C. Εάν $k > c$ τότε ο πυκνότερος k-υπογράφος του G περιέχει τον υπογράφο C και τις $k - c$ κορυφές του I με το μεγαλύτερο βαθμό.

2.3.5 h-trees

Ένα *h-tree* ορίζεται αναδρομικά ως εξής. Η κλίκα με h κορυφές, K_h, είναι ένα *h-tree*. Εάν ο G είναι ένα *h-tree* τότε και ο G' είναι *h-tree*, όπου G' είναι ο γράφος που σχηματίζεται εάν προσθέσουμε μία νέα κορυφή στον G και τη συνδέσουμε με όλες τις κορυφές μίας κλίκας K_h στον G. Παρατηρούμε ότι τα *h-trees* είναι χορδικοί γράφοι και ότι ένα *1-tree* είναι ένα δέντρο με την κλασσική του μορφή.

Ο πολυωνυμικός αλγόριθμος για το DkS στα *h-trees* είναι μία γενίκευση του αλγορίθμου του DkS προβλήματος στα δέντρα [7]. Καταρχάς δημιουργούμε το δέντρο μεγιστικών κλικών (*clique tree*) του *h-tree*. Εάν $k \leq h + 1$ τότε επιλέγουμε μία K_k, αλλιώς επιλέγουμε ένα υποδέντρο $k - h$ κορυφών στο δέντρο μεγιστικών κλικών. Ανεξάρτητα από την επιλογή του υποδέντρου είναι ξεκάθαρο ότι ο αντίστοιχος υπογράφος του *h-tree* έχει k κορυφές και ακριβώς $h(k - \frac{h+1}{2})$ ακμές και αυτό είναι το μέγιστο δυνατό.

2.4 Δυσκολία ειδικών περιπτώσεων

Στον αντίποδα, το DkS πρόβλημα παραμένει *NP-hard* για διμερείς γράφους (*bipartite graphs*) [7], ακόμα και μέγιστου βαθμού τρία [12]. Απόρροια της *NP-hardness* του DkS προβλήματος στους διμερείς γράφους είναι η *NP-hardness* στους *comparability* γράφους, αφού οι διμερείς γράφοι είναι υποκλάση των *comparability*. Επιπρόσθετα, το DkS παραμένει *NP-hard* στους χορδικούς γράφους (*chordal graphs*)

[7] σε αντίθεση με άλλα γνωστά προβλήματα συνδυαστικής βελτιστοποίησης που είναι πολυωνυμικά σε αυτή την κλάση [15]. Επιπλέον, όταν ψάχνουμε για συνεκτική λύση το πρόβλημα παραμένει *NP-hard* στους επίπεδους γράφους (*planar graphs*) [22] ενώ αξιοσημείωτο είναι ότι σε αυτή την κατηγορία γράφων το πρόβλημα εύρεσης της μέγιστης κλίκας είναι γραμμικού χρόνου [34]. Τέλος, το DkS πρόβλημα είναι *NP-hard* σε βεβαρημένους *cographs* και βεβαρημένους *split* γράφους [7]. Παραθέτουμε αναλυτικά τις αποδείξεις δυσκολίας του προβλήματος στους διμερείς και στους χορδικούς γράφους όπως αυτές παρουσιάζονται στο άρθρο των Corneil και Perl [7].

2.4.1 Διμερείς γράφοι

Ένας γράφος είναι διμερής (*bipartite graph*) εάν το σύνολο των κορυφών του, V, διαμερίζεται σε δύο ξένα ανεξάρτητα υποσύνολα S_1 και S_2, με άλλα λόγια κάθε ακμή του έχει ένα άκρο της στο υποσύνολο S_1 και ένα άκρο της στο S_2.

Comparability είναι ο γράφος που είναι *transitively orientable*, όπου *transitively orientable* σημαίνει ότι σε κάθε ακμή του ανατίθεται μία κατεύθυνση με τέτοιο τρόπο ώστε ο κατευθυνόμενος γράφος, (V, F), που θα προκύψει να ικανοποιεί τη συνθήκη: αν $ab \in F$ και $bc \in F$ τότε $ac \in F$, για κάθε $a, b, c \in V$.

Ένας γράφος είναι *triangle-free* εάν δεν περιέχει κύκλο μήκους τρία.

Θεώρημα 5. *[7] Το πρόβλημα του πυκνότερου k-υπογράφου σε διμερείς γράφους (bipartite graphs) είναι NP-complete.*

Απόδειξη. Η αναγωγή γίνεται από το πρόβλημα της k-κλίκας σε γενικούς γράφους. Δοθέντος ενός γράφου $G = (V, E)$ κατασκευάζουμε ένα διμερή γράφο $H = (W, F)$, όπου $W = V \cup E$. Έστω $k' = k + \binom{k}{2} = \binom{k+1}{2}$. Ισχυριζόμαστε ότι ο H έχει έναν πυκνότερο k'-υπογράφο με τουλάχιστον $2\binom{k}{2}$ ακμές αν και μόνο αν ο G έχει μία κλίκα k κορυφών.

Αναλυτικότερα, εάν ο G έχει μία k-κλίκα C τότε οι $\binom{k}{2}$ ακμές του C στον G αναπαριστώνται από $\binom{k}{2}$ κορυφές του W στον H και καθεμία από αυτές είναι γειτονική με δύο κορυφές στο $C \subseteq W$. Κατά συνέπεια υπάρχει ένας πυκνότερος k-υπογράφος στον H με τουλάχιστον $2\binom{k}{2}$ ακμές.

Υποθέτουμε τώρα ότι ο H περιέχει έναν πυκνότερο k'-υπογράφο L με τουλάχιστον $2\binom{k}{2}$ ακμές. Έστω $A \subseteq W$, όπου $A = L \cap V$ και $B \subseteq W$, όπου $B = L \cap E$. Εάν $|A| = a$ και $|B| = k' - a$ τότε παρατηρούμε ότι το πολύ $\binom{a}{2}$ κορυφές στο B μπορεί να έχουν βαθμό 2 και οι υπόλοιπες έχουν βαθμό το πολύ 1.

Εάν $a < k$, τότε ο αριθμός των ακμών στον L είναι το πολύ $2\binom{a}{2} + \binom{k}{2} + k - a - \binom{a}{2} = \binom{k}{2} + \binom{a}{2} + k - a < 2\binom{k}{2}$, το οποίο έρχεται σε αντίφαση με την υπόθεση ότι ο L έχει τουλάχιστον $2\binom{k}{2}$ ακμές. Εάν $a > k$, τότε $|B| < \binom{k}{2}$ και αφού όλες οι κορυφές στο B έχουν βαθμό 2, ο αριθμός των ακμών στον L είναι πάλι μικρότερος του $2\binom{k}{2}$, αντίφαση.

Συνεπώς $|A| = k$ και ο μόνος τρόπος για να έχουμε έναν πυκνότερο k'-υπογράφο L με τουλάχιστον $2\binom{k}{2}$ ακμές είναι καθεμία από τις $\binom{k}{2}$ κορυφές του B να συνεισφέρει 2 ακμές στον L, το οποίο υποδηλώνει ότι ο L αναπαριστά μία k-κλίκα στον G. □

Απόρροια του παραπάνω είναι ότι το (συνεκτικό) πρόβλημα του πυκνότερου k-υπογράφου είναι *NP-hard* σε διμερείς γράφους, σε *comparability* γράφους και σε *triangle-free* γράφους.

2.4.2 Χορδικοί γράφοι

Χορδικοί (*chordal graphs*) είναι οι γράφοι στους οποίους κάθε κύκλος με μέγεθος μεγαλύτερο του τρία έχει τουλάχιστον μία χορδή, δηλαδή, μία ακμή που ενώνει δύο μη διαδοχικές κορυφές του κύκλου. Η χορδικότητα είναι μία ιδιότητα που μεταφέρεται σε κάθε επαγόμενο υπογράφο του αρχικού χορδικού γράφου.
Μία κορυφή x του χορδικού γράφου G καλείται *simplicial* εάν το σύνολο των γειτονικών κορυφών της, $Adj(x)$, επάγει έναν πλήρη υπογράφο στον G, δηλαδή, το $Adj(x)$ είναι μία κλίκα (όχι απαραίτητα μεγιστική (*maximal*)). Κάθε χορδικός γράφος έχει πάντα μία *simplicial* κορυφή [8]. Επιπρόσθετα, εάν ο χορδικός γράφος δεν είναι κλίκα τότε έχει δύο μη γειτονικές *simplicial* κορυφές.
Μία επαναληπτική διαδικασία [13] για την αναγνώριση χορδικών γράφων βασίζεται στην ιδιότητα ότι η χορδικότητα μεταφέρεται και στους επαγόμενους υπογράφους και στο γεγονός ότι κάθε χορδικός γράφος έχει πάντα μία *simplicial* κορυφή. Συγκεκριμένα, ο αλγόριθμος επαναλαμβανόμενα εντοπίζει μία *simplicial* κορυφή και την αφαιρεί από το γράφο, είτε μέχρι να μη μείνει καμία κορυφή και ο γράφος να είναι χορδικός είτε σε κάποιο βήμα να μην υπάρχει καμία *simplicial* κορυφή και ο γράφος να μην είναι χορδικός.
Έστω $G = (V, E)$ ένας μη κατευθυνόμενος γράφος και έστω $\sigma = \{u_1, u_2, \ldots, u_n\}$ μία διάταξη των κορυφών του. Η σ είναι ένα *perfect elimination scheme* εάν κάθε u_i είναι μία *simplicial* κορυφή του επαγόμενου υπογράφου $G_{\{u_i,\ldots,u_n\}}$. Με άλλα λόγια,

κάθε σύνολο $X_i = \{u_j \in Adj(u_i) | j > i\}$ επάγει έναν πλήρη υπογράφο. Ένας γράφος είναι χορδικός αν και μόνο αν έχει ένα *perfect elimination scheme* [13].

Θεώρημα 6. *[7] Το πρόβλημα του πυκνότερου* k*-υπογράφου σε χορδικούς γράφους (chordal graphs) είναι NP-complete.*

Απόδειξη. Η αναγωγή γίνεται από το πρόβλημα της k-κλίκας σε γενικούς γράφους. Δοθέντος ενός γράφου $G(V, E)$, $|V| = n$ κατασκευάζουμε ένα χορδικό γράφο $H(W, F)$ ως εξής. Οι κορυφές του G σχηματίζουν έναν πλήρη γράφο στον H. Για κάθε ακμή $e = (v_i, v_j)$ στον G κατασκευάζουμε έναν πλήρη γράφο n κορυφών και συνδέουμε όλες αυτές τις νέες κορυφές και με τις δύο κορυφές v_i και v_j. Κατά συνέπεια, κάθε ακμή στον G αναπαρίσταται με μία K_{n+2} κλίκα στον H. Προφανώς ο H έχει ένα *perfect elimination scheme* και άρα είναι χορδικός.

Έστω $k' = k + \binom{k}{2} n$. Ισχυριζόμαστε ότι ο H έχει έναν πυκνότερο k'-υπογράφο με τουλάχιστον $\binom{k}{2}\binom{n+2}{2}$ ακμές αν και μόνο αν ο G έχει μία κλίκα k κορυφών (για την ακρίβεια ο H έχει έναν πυκνότερο k'-υπογράφο με ακριβώς $\binom{k}{2}\binom{n+2}{2}$).

Έστω ότι ο H περιέχει έναν πυκνότερο k'-υπογράφο L με τουλάχιστον $\binom{k}{2}\binom{n+2}{2}$ ακμές και έστω $A \subseteq W$, όπου $A = L \cap V$ και $B \subseteq W$, όπου $B = L \cap E$. Εάν $|A| = a$ και $|B| = k' - a$ τότε το πολύ $\binom{a}{2} n$ κορυφές στο B μπορεί να έχουν βαθμό 2 από το ένα σύνολο στο άλλο ενώ οι υπόλοιπες έχουν βαθμό 1 από το ένα σύνολο στο άλλο. Εάν $a < k$ τότε ο αριθμός των ακμών στο L είναι το πολύ

$$\begin{aligned}
&\binom{a}{2} + \frac{k'-a}{n}\binom{n}{2} + \binom{a}{2} 2n + k' - a - \binom{a}{2} n = \\
&\binom{a}{2} + \frac{k+\binom{k}{2}n-a}{n}\binom{n}{2} + \binom{a}{2} 2n + k + \binom{k}{2} n - a - \binom{a}{2} n = \\
&\binom{a}{2} + (k + \binom{k}{2} n - a)\frac{n-1}{2} + \binom{a}{2} n + k + \binom{k}{2} n - a = \\
&\binom{k}{2}\binom{n}{2} + \binom{k}{2} n + k\frac{n+1}{2} + \binom{a}{2}(n+1) - a\frac{n+1}{2} = \\
&\binom{k}{2}\binom{n+1}{2} + \frac{k(n+1)}{2} + \frac{a(a-2)(n+1)}{2} < \\
&\binom{k}{2}\binom{n+1}{2} + \frac{(k+k(k-2))(n+1)}{2} = \binom{k}{2}\binom{n+1}{2} + \binom{k}{2}(n+1) = \binom{k}{2}\binom{n+2}{2}
\end{aligned}$$

που έρχεται σε αντίφαση με το γεγονός ότι το L έχει τουλάχιστον $\binom{k}{2}\binom{n+2}{2}$ ακμές. Εάν $a > k$ και $d = a - k$, τότε $|B| = \binom{k}{2} n - d$. Ακόμα και αν όλες οι κορυφές του B έχουν βαθμό 2 από το ένα σύνολο στο άλλο τότε ο αριθμός των ακμών στο L είναι το πολύ

$$\begin{aligned}
&(\binom{k}{2} - 1)\binom{n}{2} + \binom{n-d}{2} + \binom{a}{2} + 2(\binom{k}{2} n - d) = \\
&\binom{k}{2}\binom{n}{2} - \binom{n}{2} + \binom{n}{2} - nd + \binom{d+1}{2} + \binom{a}{2} + 2\binom{k}{2} n - 2d = \\
&\binom{k}{2}(\binom{n}{2} + 2n) - (n+2)(a-k) + \binom{a-k+1}{2} + \binom{a}{2} = \\
&\binom{k}{2}(\binom{n}{2} + 2n) - (n+2)(a-k) + \binom{a+1}{2} - ak + \binom{k}{2} + \binom{a}{2} =
\end{aligned}$$

$\binom{k}{2}\left(\binom{n}{2}+2n+1\right)-(n+2)(a-k)+a(a-k)=$
$\binom{k}{2}\binom{n+2}{2}-(n+2-a)(a-k)<\binom{k}{2}\binom{n+2}{2}$
που έρχεται πάλι σε αντίφαση με το γεγονός ότι το L έχει τουλάχιστον $\binom{k}{2}\binom{n+2}{2}$ ακμές.

Επομένως, $a=k$ και ο μόνος τρόπος να έχουμε έναν πυκνότερο k'-υπογράφο με τουλάχιστον $\binom{k}{2}\binom{n+2}{2}$ ακμές είναι καθεμία από τις κορυφές του B να συνεισφέρει 2 ακμές από το ένα σύνολο στο άλλο στον L και το B να αποτελείται από την ένωση $\binom{k}{2}$ κλικών μεγέθους n. Αυτό όμως υποδηλώνει ότι το L αναπαριστά μία k-κλίκα στον G. □

Απότοκο του παραπάνω είναι ότι και το συνεκτικό DkS πρόβλημα είναι *NP-hard* στους χορδικούς γράφους.

2.5 Προσεγγιστικοί αλγόριθμοι

Κατά τη διάρκεια των τελευταίων χρόνων μεγάλο μέρος εργασιών έχουν επικεντρωθεί στη σχεδίαση προσεγγιστικών αλγορίθμων, τόσο για το πρόβλημα του πυκνότερου k-υπογράφου όσο και για τη βεβαρημένη εκδοχή του, βασισμένοι σε διάφορες τεχνικές όπως οι άπληστοι αλγόριθμοι (*greedy algorithms*), οι γραμμικές χαλαρώσεις (*LP relaxations*) και ο ημιορισμένος προγραμματισμός (*semidefinite programming*).

Καταρχάς, στο άρθρο [12] παρουσιάζεται προσεγγιστικός αλγόριθμος για το DkS σε γενικούς γράφους, που βασίζεται στον ημιορισμένο προγραμματισμό με λόγο $\frac{n}{k}$. Στο ίδιο άρθρο παρουσιάζεται ένα αντιπαράδειγμα το οποίο δείχνει ότι το χάσμα μεταξύ της βέλτιστης λύσης του ημιορισμένου προγραμματισμού και της πραγματικής τιμής του DkS προβλήματος είναι $\Omega(n^{\frac{1}{3}})$.

Το καλύτερο από τα γνωστά αποτελέσματα για το DkS πρόβλημα, το οποίο είναι ανεξάρτητο της τιμής του k, είναι $O(n^{\delta})$ για κάποιο $\delta<\frac{1}{3}$ [10]. Ο λόγος $O(n^{\frac{1}{3}})$ για το DkS πρόβλημα είναι βελτίωση του λόγου $O(n^{0,3885})$ που έχει προταθεί στο [25]. Το προαναφερθέν αποτέλεσμα είναι το καλύτερο γνωστό αποτέλεσμα για όλες τις τιμές του k διότι ακόμα και αν με τη βοήθεια του ημιορισμένου προγραμματισμού επιτύχουμε έναν προσεγγιστικό λόγο κοντά στο $\frac{n}{k}$ θα έχουμε μεν καλύτερο προσεγγιστικό λόγο για μεγάλες τιμές του k αλλά για μικρές τιμές του k δεν είναι γνωστό αν ένας τέτοιος αλγόριθμος αποδίδει τόσο καλά όσο ο αλγόριθμος στο [10]. Για παράδειγμα όταν το $k\simeq n^{\frac{1}{3}}$, η προσέγγιση με τη βοήθεια του ημιορισμένου προγραμματισμού δε μπορεί να διακρίνει μεταξύ γράφων που έχουν κλίκες μεγέθους k και γράφων που έχουν μόνο k-κορυφών υπογράφους με $O(k)$ ακμές, το οποίο και αποκλείει έναν προσεγγιστικό λόγο καλύτερο από $n^{\frac{1}{3}}$. Επιπλέον, ο αλγόριθμος στο άρθρο [10] μπορεί να εφαρμοστεί και σε βεβαρημένους γράφους επιφέροντας έναν επιπρόσθετο παράγοντα $O(\log n)$.

Από την άλλη μεριά, δυστυχώς δεν υπάρχει κάποιο συμπληρωματικό αρνητικό αποτέλεσμα που να δείχνει ότι η επίτευξη προσεγγιστικού λόγου $O(n^{\epsilon})$, για κάποιο $\epsilon > 0$, είναι *NP-hard*, ενώ σχετικά πρόσφατα αποδείχτηκε ότι στη γενική περίπτωση το DkS πρόβλημα δεν επιδέχεται πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*PTAS*) υπό την προϋπόθεση ότι $NP \nsubseteq \cap_{\epsilon>0} BPTIME(2^{n^{\epsilon}})$ [23]. Έχει επίσης αποδειχθεί ότι το DkS δεν επιδέχεται πολυωνυμικού χρόνου προσεγγιστικό σχήμα υποθέτοντας ότι το τυχαίο 3SAT στη μέση περίπτωση παραμένει δύσκολο [9].

Η άπληστη προσέγγιση του DkS προβλήματος, ακόμα και για βεβαρημένους γράφους, δίνει προσεγγιστικό λόγο R τέτοιο που $(\frac{1}{2}+\frac{n}{2k})^2 - O(n^{-\frac{1}{3}}) \leq R \leq (\frac{1}{2}+\frac{n}{2k})^2 + O(\frac{1}{n})$, για $\frac{n}{3} < k \leq n$ και $2(\frac{n}{k}-1) - O(\frac{1}{k}) \leq R \leq 2(\frac{n}{k}-1) + O(\frac{n}{k^2})$, για $k \leq \frac{n}{3}$ [4]. Για τη βεβαρημένη εκδοχή υπάρχει επίσης προσεγγιστικός αλγόριθμος που βασίζεται σε ημιορισμένο προγραμματισμό και έχει λόγο $\frac{n}{k}$, όταν $k \simeq \frac{n}{3}$ [39] ενώ στο [11] παρουσιάζεται ένας τυχαιοκρατικός (*randomized*) προσεγγιστικός αλγόριθμος, βασισμένος στις ιδέες που υπάρχουν στο [39], με λόγο $\frac{n}{k}$, για $k \simeq \frac{n}{2}$. Ημιορισμένη χαλάρωση (*semidefinite relaxation*) για το πρόβλημα έχουν επίσης προταθεί στα [20] και [40]. Επιπλέον, στο [5] προτείνεται προσεγγιστικός αλγόριθμος με λόγο $\max\{d, \frac{8}{9}\frac{k}{n}\}$ όπου d είναι η πυκνότητα του δοθέντος γράφου.

Σε πλήρεις βεβαρημένους γράφους που τα βάρη στις ακμές ικανοποιούν την τριγωνική ανισότητα άπληστοι αλγόριθμοι επιτυγχάνουν προσεγγιστικούς λόγους 4 [37] και 2 [21]. Επιπλέον, για την ειδική περίπτωση εύρεσης του πυκνότερου k-υπογράφου σε πυκνούς γράφους, δηλαδή, σε γράφους ελάχιστου βαθμού $\Omega(n)$ καθώς και σε πυκνούς γράφους ($\Omega(n^2)$ ακμών) όταν το k είναι $\Omega(n)$, έχει αποδειχτεί ότι υπάρχει πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*PTAS*) [1].

Αναλυτικότερα, παραθέτουμε τον προσεγγιστικό αλγόριθμο των Feige, Kortsarz και Peleg [10] που δίνει λόγο $O(n^{\delta})$ για κάποιο $\delta < \frac{1}{3}$. Επιπλέον αναφέρουμε επιγραμματικά τον άπληστο αλγόριθμο προσεγγιστικού λόγου $O(\frac{n}{k})$ των Asahiro, Iwama, Tamaki και Tokuyama [4] καθώς και το πολυωνυμικού χρόνου προσεγγιστικό σχήμα για πυκνούς γράφους των Arora, Karger και Karpinski [1].

2.5.1 $O(n^{\frac{1}{3}})$- προσεγγιστικός αλγόριθμος

Δεδομένης της δυσκολίας του DkS προβλήματος ενδιαφερόμαστε για πολυωνυμικού χρόνου προσεγγιστικούς αλγόριθμους. Με είσοδο ένα γράφο και ένα φυσικό αριθμό, (G, k), ένας τέτοιος αλγόριθμος δίνει σαν έξοδο μία λίστα k κορυφών. Έστω ότι

- $A(G, k)$ είναι η πυκνότητα του υπογράφου που επάγεται από τις k κορυφές που επιστρέφει ο αλγόριθμος A με είσοδο το γράφο (G, k).
- $d^*(G, k)$ είναι η μέγιστη πυκνότητα του υπογράφου που επάγεται από k κορυφές, δηλαδή η βέλτιστη τιμή.

Θεώρημα 7. *[10] Υπάρχει πολυωνυμικού χρόνου αλγόριθμος A που προσεγγίζει το DkS με έναν παράγοντα $2n^{\frac{1}{3}}$, δηλαδή για κάθε γράφο G και κάθε k, $1 \leq k \leq n$, $A(G, k) \geq \frac{d^*(G,k)}{2n^{\frac{1}{3}}}$.*

Ο αλγόριθμος A χρησιμοποιεί τρείς διαφορετικές διαδικασίες A_1, A_2 και A_3 για να επιλέξει έναν πυκνό υπογράφο. Επιστρέφει τον πυκνότερο των τριών υπογράφων που έχουν κατασκευαστεί από τις τρείς διαδικασίες.

Πριν παρουσιάσουμε τις τρείς διαδικασίες δίνουμε μερικούς συμβολισμούς:

- $\Delta(G)$ είναι ο μέγιστος βαθμός του γράφου G.
- d_H είναι ο μέσος βαθμός των $k/2$ κορυφών με το μεγαλύτερο βαθμό στο G. Παρατηρούμε ότι $\Delta(G) \geq d_H \geq d^*(G, k)$.
- $deg(\upsilon, S)$ είναι ο αριθμός των ακμών που συνδέουν την κορυφή υ με τις κορυφές του συνόλου S.
- $cut(A, B)$ είναι ο αριθμός των ακμών που συνδέουν τις κορυφές του συνόλου A με τις κορυφές του συνόλου B.
- Ένας περίπατος (*walk*) μήκους l είναι μία ακολουθία $l + 1$ κορυφών στην οποία οι διαδοχικές κορυφές είναι παρακείμενες μίας ακμής (δηλαδή ο περίπατος ακολουθεί l ακμές και οι κορυφές δεν είναι απαραίτητα διαφορετικές).
- $W_l(u, \upsilon)$ είναι ο αριθμός των περίπατων μήκους l που ξεκινούν από την κορυφή u και καταλήγουν στην κορυφή υ.

Χωρίς βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι ο G περιέχει τουλάχιστον $k/2$ ακμές.

ΔΙΑΔΙΚΑΣΙΑ A_1

1. Επίλεξε αυθαίρετα $k/2$ ακμές από τον G.

2. Επίστρεψε το σύνολο των κορυφών που προσπίπτουν σε αυτές τις ακμές, προσθέτοντας αυθαίρετα κορυφές εάν το μέγεθος του συνόλου είναι μικρότερο του k.

Προφανώς, $A_1(G,k) \geq 1$.

ΔΙΑΔΙΚΑΣΙΑ A_2

1. Ταξινόμησε τις κορυφές βάση του βαθμού τους.
Έστω H το σύνολο των $\frac{k}{2}$ κορυφών με το μεγαλύτερο βαθμό στον G.

2. Ταξινόμησε τις εναπομένουσες κορυφές βάση του αριθμού των γειτόνων τους στο H.
Έστω C το σύνολο των $\frac{k}{2}$ κορυφών στον $G \backslash H$ με το μεγαλύτερο αριθμό γειτόνων στο H.

3. Επίστρεψε το $H \cup C$.

Λήμμα 1. *Η διαδικασία A_2 επιστρέφει έναν υπογράφο επαγόμενο από k κορυφές, που ικανοποιεί τη σχέση $A_2(G,k) \geq \frac{kd_H}{2n}$, όπου d_H είναι ο μέσος βαθμός μίας κορυφής στο H.*

Απόδειξη. Έστω ότι m_1 είναι ο αριθμός των ακμών που έχουν και τα δύο άκρα τους μέσα στο H. Τότε $cut(H, V\backslash H) = d_H \cdot |H| - 2m_1 = d_H \frac{k}{2} - 2m_1 \geq 0$.
Λόγω του άπληστου κανόνα επιλογής του C, τουλάχιστον ένα $\frac{|C|}{|V\backslash H|} > \frac{k}{2n}$ κλάσμα των ακμών της τομής (*cut*) περιέχονται στο $H \cup C$. Επομένως ο συνολικός αριθμός ακμών στον υπογράφο που επάγεται από το $H \cup C$ είναι τουλάχιστον $(d_H \frac{k}{2} - 2m_1)\frac{k}{2n} + m_1 \geq d_H \frac{k^2}{4n}$. Συνεπώς, η πυκνότητα $A_2(G,k)$ του υπογράφου που επάγεται από τις k κορυφές που επιστρέφει ο αλγόριθμος A_2 είναι $A_2(G,k) \geq \frac{2(d_H \frac{k^2}{4n})}{k} = \frac{k\,d_H}{2n}$ □
Εφόσον $d_H \geq d^*(G,k)$ έχουμε ότι $A_2(G,k) \geq \frac{d^*(G,k)}{2n/k}$. Επομένως, η άπληστη διαδικασία A_2 προσεγγίζει το $d^*(G,k)$ με ένα λόγο το πολύ $\frac{2n}{k}$.

ΔΙΑΔΙΚΑΣΙΑ A_3

1. Υπολόγισε το $W_2(u,v)$ για όλα τα ζεύγη των κορυφών.

2. Κατασκεύασε έναν υποψήφιο γράφο H^v για κάθε κορυφή v στον G ως εξής:
   α. Ταξινόμησε τις κορυφές του G σε φθίνουσα σειρά, βάση του αριθμού των περίπατων τους μήκους 2 προς την κορυφή v, $W_2(v,w_1) \geq W_2(v,w_2) \geq \ldots$
   Έστω P_h^v το σύνολο των $\frac{k}{2}$ κορυφών που έχουν τους περισσότερους

περιπάτους μήκους 2 προς την κορυφή υ, δηλαδή το $\{w_1, w_2, ..., w_{\frac{k}{2}}\}$

β. Υπολόγισε για κάθε γείτονα x του υ τον αριθμό των ακμών που συνδέουν το x με το P_h^υ, δηλαδή το $deg(x, P_h^\upsilon)$ και κατασκεύασε ένα σύνολο B^υ που να περιέχει τους $\frac{k}{2}$ γείτονες του υ με το μεγαλύτερο $deg(x, P_h^\upsilon)$.

γ. Έστω H^υ ο υπογράφος που επάγεται από το $P_h^\upsilon \cup B^\upsilon$. Εάν ο H^υ περιέχει λιγότερες από k κορυφές συμπλήρωσε αυθαίρετα μέχρι τις k κορυφές.

3. Επίλεξε τον πυκνότερο υποψήφιο γράφο H^υ σαν έξοδο.

Έστω G^* ο βέλτιστος υπογράφος και $deg^*(\upsilon)$ ο βαθμός του υ στον G^*. Καταρχάς παρατηρούμε ότι ο αριθμός των περίπατων μήκους 2 στον βέλτιστο υπογράφο G^* είναι τουλάχιστον $k(d^*(G,k))^2$. Αυτό ισχύει γιατί κάθε $\upsilon \in G^*$ συνεισφέρει $(deg^*(\upsilon))^2$ στο άθροισμα και $\sum_{\upsilon \in G^*}(deg^*(\upsilon))^2 \geq k(d^*(G,k))^2$.
Συνεπώς υπάρχει μία κορυφή υ η οποία είναι το άκρο τουλάχιστον $(d^*(G,k))^2$ περίπατων μήκους 2 στον G^*. Λόγω της άπληστης κατασκευής του P_h^υ υπάρχουν τουλάχιστον $\frac{(d^*(G,k))^2}{2}$ περίπατοι μήκους 2 μεταξύ της υ και των κορυφών του P_h^υ. Οι κορυφές του B^υ έχουν τουλάχιστον $\frac{(d^*(G,k))^2}{2}$ ακμές που τις συνδέουν με το P_h^υ εάν $deg(\upsilon) \leq \frac{k}{2}$ και τουλάχιστον $\frac{(d^*(G,k))^2 k}{4deg(\upsilon)}$ ακμές που τις συνδέουν με το P_h^υ αλλιώς. Αφού δεν απαιτούμε το P_h^υ και το B^υ να είναι ξένα, κάθε ακμή μπορεί να έχει μετρηθεί δύο φορές. Έπεται ότι ο H^υ περιέχει τουλάχιστον $\min\{\frac{(d^*(G,k))^2}{4}, \frac{(d^*(G,k))^2 k}{8\Delta(G)}\}$ ακμές, όπου $\Delta(G)$ είναι ο μέγιστος βαθμός του γράφου. Κατά συνέπεια, $A_3(G,k) \geq \dfrac{(d^*(G,k))^2}{2\max\{k, 2\Delta(G)\}}$.

ΑΛΓΟΡΙΘΜΟΣ Α

Ο Αλγόριθμος A εφαρμόζει τις τρείς διαδικασίες που περιγράψαμε παραπάνω και αποδίδει τον πυκνότερο από τους τρείς υπογράφους που λαμβάνονται από την καθεμία από αυτές τις διαδικασίες. Οι Διαδικασίες A_1 και A_2 εφαρμόζονται στον αρχικό γράφο G ενώ η Διαδικασία A_3 εφαρμόζεται στο γράφο G_l που επάγεται από τις κορυφές του συνόλου $V \backslash H$, όπου H είναι το σύνολο των $\frac{k}{2}$ κορυφών με το μεγαλύτερο βαθμό στον G, όπως αυτό ορίστηκε στη δεύτερη διαδικασία. Άρα $\Delta(G_l) \leq d_H(G)$.
Για το ακόλουθο λήμμα υποθέτουμε ότι $k \leq \frac{2n}{3}$. Η υπόθεση αυτή μπορεί να γίνει χωρίς βλάβη της γενικότητας διότι για $k \geq \frac{2n}{3}$ η δεύτερη διαδικασία προσεγγίζει το DkS με ένα λόγο όχι χειρότερο από 3.

Λήμμα 2. *Ο γράφος G_l περιέχει έναν k-κορυφών επαγόμενο υπογράφο με μέσο βαθμό τουλάχιστον $d^*(G,k) - 2d_2$, όπου $d_2 = A_2(G,k)$.*

Απόδειξη. Έστω m ο αριθμός των ακμών του G^* που έχουν και τα δύο άκρα τους στο H και έστω l ο αριθμός των ακμών του G^* με το ένα άκρο τους στο H. Άρα ο G_l περιέχει έναν k-κορυφών επαγόμενο υπογράφο με τουλάχιστον $d^*(G,k)\frac{k}{2} - m - l$ ακμές. Για να αποδείξουμε το λήμμα χρειάζεται να δείξουμε ότι η δεύτερη διαδικασία επιστρέφει μία λύση με τουλάχιστον $\frac{m+l}{2}$ ακμές. Στην πραγματικότητα η λύση έχει τουλάχιστον τουλάχιστον $m + \frac{l}{2}$ ακμές. Αυτό συμβαίνει γιατί περιέχει τις m ακμές που είναι εσωτερικές στο $V(G^*) \cap H$ και πρέπει να υπάρχουν τουλάχιστον $\frac{l}{2}$ ακμές μεταξύ του C και του H αφού τουλάχιστον μία πιθανή επιλογή του C προσφέρει τόσες ακμές (δηλαδή παίρνοντας το C να περιέχει τις $\frac{k}{2}$ κορυφές του $V(G^*)\backslash H$ με το μεγαλύτερο αριθμό ακμών στο H). □

Επομένως, από την απόδοση που εγγυώνται οι τρείς διαδικασίες έχουμε

$$A(G,k) \geq \max\left\{1, \frac{kd_H}{2n}, \frac{(d^*(G,k) - 2d_2)^2}{2\max\{k, 2d_H\}}\right\}.$$

Για να αποδείξουμε το Θεώρημα 7 μπορούμε να υποθέσουμε ότι $d_2 \leq \frac{d^*(G,k)}{n^{\frac{1}{3}}}$, αλλιώς η έξοδος της δεύτερης διαδικασίας επιτυγχάνει το ζητούμενο προσεγγιστικό λόγο. Άρα για την τρίτη διαδικασία έχουμε ότι $d^*(G,k) - 2d_2 \simeq d^*(G,k)$ με ένα αμελητέο σφάλμα. Η απόδοση που εγγυάται ο Αλγόριθμος A είναι τουλάχιστον ο γεωμετρικός μέσος της απόδοσης που εγγυώνται οι τρείς διαδικασίες A_1, A_2 και A_3. Κατά συνέπεια, έχουμε

$$A(G,k) \geq \left(1 \cdot \frac{kd_H}{2n} \cdot \frac{(d^*(G,k))^2}{2\max\{k, 2d_H\}}\right)^{\frac{1}{3}} \geq \frac{d^*(G,k)}{2n^{\frac{1}{3}}},$$

όπου η τελευταία ανισότητα έπεται από το γεγονός ότι $k \geq d^*(G,k)$ και $d_H \geq d^*(G,k)$.

Για να δώσει ο αλγόριθμος A έναν προσεγγιστικό λόγο τόσο κακό όσο το $\Omega(n^{\frac{1}{3}})$ πρέπει και οι τρείς διαδικασίες A_1, A_2 και A_3 να δώσουν έναν προσεγγιστικό λόγο $\Theta(n^{\frac{1}{3}})$. Αυτό γίνεται μόνο εάν $d^*(G,k) = \Theta(n^{\frac{1}{3}})$, $kd_H = \Theta(n)$ και $\max\{k, d_H\} = \Theta(n^{\frac{2}{3}})$. Εάν κάποια από τις τρείς παραπάνω συνθήκες παραβιάζεται από έναν παράγοντα n^{ϵ}, τότε ο προσεγγιστικός λόγος είναι $O(n^{\frac{1}{3}-\frac{\epsilon}{2}})$. Οι παραπάνω χειρότερης περίπτωσης συνθήκες ικανοποιούνται μόνο σε δύο περιπτώσεις:

- $d^*(G,k) = \Theta(n^{\frac{1}{3}})$, $k = \Theta(n^{\frac{1}{3}})$, $d_H = \Theta(n^{\frac{2}{3}})$,
- $d^*(G,k) = \Theta(n^{\frac{1}{3}})$, $k = \Theta(n^{\frac{2}{3}})$, $d_H = \Theta(n^{\frac{1}{3}})$.

Στο άρθρο [10] παρουσιάζονται δύο πρόσθετες διαδικασίες, καθεμία από τις οποίες δίνει έναν προσεγγιστικό λόγο καλύτερο από $O(n^{\frac{1}{3}})$ σε καθεμία από τις παραπάνω περιπτώσεις. Αυτό το αποτέλεσμα μαζί με τον αλγόριθμο A εγγυώνται έναν προσεγγιστικό λόγο $O(n^{\frac{1}{3}-\epsilon})$, για κάποιο $\epsilon > 0$, για το DkS πρόβλημα και το επόμενο θεώρημα έπεται

Θεώρημα 8. *[10] Υπάρχει πολυωνυμικού χρόνου αλγόριθμος* B *που προσεγγίζει το* DkS *με έναν παράγοντα* $n^{\frac{1}{3}-\epsilon}$, *για κάποιο* $\epsilon > 0$. *Δηλαδή, για κάθε γράφο* G *και για κάθε* $1 \leq k \leq n$,

$$B(G,k) \geq \frac{d^*(G,k)}{n^{\frac{1}{3}-\epsilon}}.$$

2.5.2 $O(\frac{n}{k})$-προσεγγιστικός αλγόριθμος

Δοθέντος ενός γράφου $G = (V, E)$ με βάρη στις ακμές περιγράφουμε έναν άπληστο αλγόριθμο για το πρόβλημα του πυκνότερου k-υπογράφου.

1. Θέσε $X = V(G)$.
2. Εάν $|X| = k$ τότε δώσε ως έξοδο το γράφο $G[X]$, αλλιώς βρες στον $G[X]$ την κορυφή με το μικρότερο βεβαρημένο βαθμό, δηλαδή, με το μικρότερο άθροισμα βαρών στις προσπίπτουσες ακμές της, στον $G[X]$ και αφαίρεσε την από το X.
3. Επανάλαβε το δεύτερο βήμα.

$G[X]$ είναι ο υπογράφος του G που επάγεται από το σύνολο κορυφών X. Στο [4] αποδεικνύονται άνω και κάτω φράγματα για τον προσεγγιστικό λόγο R του αλγόριθμου, τα οποία είναι τα καλύτερα δυνατά μέχρι και έναν επιπρόσθετο όρο $o(1)$ και όσο το $k = \omega(\sqrt{n})$. Σαν συνέπεια προκύπτει το παρακάτω θεώρημα.

Θεώρημα 9. *[4] Ο χειρότερης περίπτωσης προσεγγιστικός λόγος* R *του αλγόριθμου ικανοποιεί τις παρακάτω ανισώσεις*

$$(\frac{1}{2} + \frac{n}{2k})^2 - O(n^{-\frac{1}{3}}) \leq R \leq (\frac{1}{2} + \frac{n}{2k})^2 + O(\frac{1}{n}),$$

για $\frac{n}{3} < k \leq n$, *και*

$$2(\frac{n}{k} - 1) - O(\frac{1}{k}) \leq R \leq 2(\frac{n}{k} - 1) + O(\frac{n}{k^2}),$$

για $k \leq \frac{n}{3}$.

Παρόλο που ο αλγόριθμος εφαρμόζεται σε βεβαρημένους γράφους η ειδική περίπτωση των μη βεβαρημένων γράφων παρουσιάζει όπως είδαμε ιδιαίτερο ενδιαφέρον. Προφανώς τα άνω φράγματα ισχύουν και σε μη βεβαρημένους γράφους. Από την άλλη μεριά, εφόσον η βέλτιστη λύση έχει το πολύ $\binom{k}{2}$ ακμές, ένα τετριμμένο άνω φράγμα $O(k)$ ισχύει για κάθε αλγόριθμο που δίνει σαν έξοδο έναν υπογράφο με $\Omega(k)$ ακμές. Συνεπώς ισχύει και για τον παραπάνω αλγόριθμο, αφού η λύση του μπορεί να περιέχει το πολύ μία μεμονωμένη κορυφή (εκτός εάν ο αρχικός γράφος περιέχει περισσότερες από $n-k$ μεμονωμένες κορυφές, αλλά σε αυτή την περίπτωση ο αλγόριθμος βρίσκει τη βέλτιστη λύση). Αυτό το άνω φράγμα είναι καλύτερο από το άνω φράγμα του Θεωρήματος 9 όταν $k \leq \sqrt{n}$. Για $k \leq cn^{\frac{3}{4}}$, το κάτω φράγμα αποδεικνύεται μέσω ενός παραδείγματος που χρησιμοποιεί μη μοναδιαία βάρη. Κατά συνέπεια, για $\sqrt{n} \leq k \leq cn^{\frac{3}{4}}$ δεν γνωρίζουμε αν το άνω φράγμα του Θεωρήματος 9 είναι το καλύτερο δυνατό για μη βεβαρημένους γράφους.

2.5.3 Πολυωνυμικού χρόνου προσεγγιστικό σχήμα για πυκνούς γράφους

Η μέθοδος για την εύρεση *PTAS* για πυκνά στιγμιότυπα αρκετών *NP-hard* προβλημάτων βελτιστοποίησης βασίζεται σε δύο τεχνικές. Καταρχάς στην εξαντλητική δειγματοληψία (*exhaustive sampling*), δηλαδή, επιλέγουμε ένα μικρό τυχαίο σύνολο κορυφών, μαντεύουμε πως είναι στη βέλτιστη λύση και χρησιμοποιούμε την τοποθέτηση τους για να καθορίσουμε την τοποθέτηση όλων των άλλων και κατά δεύτερο λόγο, στη μείωση του βαθμού d των περιορισμών σε γραμμικούς περιορισμούς (προσεγγιστικά).

Αρχικά εκφράζουμε το πρόβλημα σαν ένα τετραγωνικό ακέραιο πρόγραμμα (*quadratic integer program*). Η μορφή ενός τετραγωνικού προγράμματος μοιάζει αρκετά με ένα ακέραιο γραμμικό πρόγραμμα για το οποίο υπάρχουν πολυάριθμες γνωστές προσεγγιστικές τεχνικές. Δυστυχώς οι συντελεστές στην αντικειμενική συνάρτηση δεν είναι σταθερές. Ωστόσο η εξαντλητική δειγματοληψία μας επιτρέπει να υπολογίσουμε την τιμή των συντελεστών στη βέλτιστη λύση. Φθάνουμε στην προσέγγιση σε τρία βήματα:

- Χρησιμοποιώντας την εξαντλητική δειγματοληψία υπολογίζουμε τις τιμές των συντελεστών της αντικειμενικής συνάρτησης στη βέλτιστη λύση.
- Αντικαθιστούμε κάθε συντελεστή με την τιμή που βρήκαμε παραπάνω. Αυτό

μετατρέπει το τετραγωνικό πρόγραμμα σε ένα γραμμικό ακέραιο πρόγραμμα. Η βέλτιστη λύση αυτού του γραμμικού ακέραιου προγράμματος είναι κοντά στη βέλτιστη λύση του τετραγωνικού προγράμματος.

- Χαλαρώνουμε τους περιορισμούς του ακέραιου γραμμικού προγράμματος, το λύνουμε και χρησιμοποιούμε τη μέθοδο τυχαιοκρατικής στρογγυλοποίησης (*randomized rounding*) για να μετατρέψουμε τη λύση σε ακέραια.

Τα παραπάνω αποτελέσματα γενικεύονται για τη χρήση τους σε πολυωνυμικά ακέραια προγράμματα (*polynomial integer program - PIP*). Συνεπώς, στόχος μας είναι δοθέντος ενός *PIP* για κάποια προβλήματα βελτιστοποίησης να βρούμε μία προσεγγιστικά βέλτιστη ακέραια λύση. Συνεπώς ο γενικός αλγόριθμος έχει τα εξής τρία βήματα:

1. Χρησιμοποιούμε τη μέθοδο της τυχαιοκρατικής στρογγυλοποίησης (*randomized rounding*) για να μετατρέψουμε κάθε εφικτή κλασματική λύση ενός *PIP* σε μία εφικτή ακέραια λύση.

2. Υπολογίζουμε την τιμή των πολυώνυμων χρησιμοποιώντας τη μέθοδο της εξαντλητικής δειγματοληψίας.

3. Χρησιμοποιούμε τους παραπάνω υπολογισμούς για να μετατρέψουμε τους περιορισμούς που είναι βαθμού d σε γραμμικούς περιορισμούς.

Ένας γράφος είναι δ-πυκνός εάν έχει $\frac{\delta n^2}{2}$ ακμές και είναι παντού δ-πυκνός εάν ο ελάχιστος βαθμός του είναι δn. Ένα πολυωνυμικό ακέραιο πρόγραμμα *PIP* είναι της μορφής

$$\begin{aligned} &\max p_0(x_1, ..., x_n) \\ l_i \leq\, &p_i(x) \leq u_i, \quad i = 1, ..., m \\ &x_i \in \{0, 1\}, \quad \forall i \leq n \end{aligned}$$

όπου $p_0, ..., p_m$ είναι πολυώνυμα. Όταν όλα τα p_i έχουν βαθμό το πολύ d, καλούμε το πρόγραμμα βαθμού d-*PIP*. Εφόσον τα *PIP* συμπεριλαμβάνουν τα ακέραια προγράμματα είναι ξεκάθαρο ότι το να επιλύσεις *PIPs* είναι *NP-hard*.

Ένα n-τυχαίων μεταβλητών βαθμού d πολυώνυμο έχει λειότητα (*smoothness*) c εάν η απόλυτη τιμή κάθε συντελεστή, του κάθε βαθμού i μονώνυμου είναι το πολύ cn^{d-i}. Ένα c-λείο (c-*smooth*) βαθμού d-*PIP* είναι ένα *PIP* στο οποίο η αντικειμενική συνάρτηση και οι περιορισμοί είναι c-λεία πολυώνυμα με βαθμό το πολύ d.

Λήμμα 3. *(Τυχαιοκρατική στρογγυλοποίηση) Εάν* c *και* f *θετικοί ακέραιοι και* $0 < \epsilon < 1$ *τότε το ακόλουθο αληθεύει για κάθε ακέραιο* $n \geq 0$. *Έστω* $y = (y_i)$ *διάνυσμα* n

μεταβλητών, $0 \leq y_i \leq 1$*, που ικανοποιεί έναν ορισμένο γραμμικό περιορισμό* $\alpha^T y = b$*, όπου κάθε* $|\alpha_i| \leq c$*. Κατασκευάζουμε ένα διάνυσμα* $z = (z_i)$ *τυχαία θέτοντας* $z_i = 1$ *με πιθανότητα* y_i *και* $z_i = 0$ *με πιθανότητα* $1 - y_i$*. Τότε με πιθανότητα τουλάχιστον* $1 - n^{-f}$ *έχουμε*

$$\alpha^T z \in b \pm c\sqrt{fn \ln n}$$

Λήμμα 4. *(Τυχαιοκρατική στρογγυλοποίηση για βαθμού* d *πολυώνυμα) Έστω* p *ένα* c*-λείο βαθμού* d *πολυώνυμο. Δίνοντας κλασματικές τιμές* (y_i) *τέτοιες που* $p(y_1, ..., y_n) = b$*, υποθέτουμε ότι η τυχαιοκρατική στρογγυλοποίηση εκτελείται στα* (y_i) *όπως στο παραπάνω λήμμα για να αποδώσει ένα 0,1 διάνυσμα* (z_i)*. Τότε με πιθανότητα τουλάχιστον* $1 - n^{d-f}$ *έχουμε*

$$p(z_1, ..., z_n) \in [b \pm gdn^{d-\frac{1}{2}}\sqrt{\ln n}],$$

όπου $g = 2ce\sqrt{f}$.

Θεώρημα 10. *[1] Υπάρχει ένας πιθανοτικός πολυωνυμικού χρόνου αλγόριθμος που λύνει προσεγγιστικά λεία PIPs. Συγκεκριμένα δοθέντος ενός εφικτού* c*-λείου* PIP *βαθμού* d *με* n *μεταβλητές, αντικειμενική συνάρτηση* p_0 *και* K *περιορισμούς, ο αλγόριθμος βρίσκει μία 0/1 λύση* z *που ικανοποιεί*

$$p_0(z_1, ..., z_n) \geq \mathrm{OPT} - \epsilon n^d,$$

όπου OPT *είναι η βέλτιστη λύση του* PIP*. Η λύση* z *ικανοποιεί επίσης κάθε βαθμού* d' *περιορισμό με έναν πρόσθετο παράγοντα* $\epsilon n^{d'}$ *για* $d' > 1$ *και ικανοποιεί κάθε γραμμικό περιορισμό με έναν πρόσθετο σφάλμα* $O(\epsilon\sqrt{n \ln n})$.

Η πολυπλοκότητα του αλγορίθμου είναι $O((dKn^d)^t)$, όπου $t = \frac{4c^2e^2d^2}{\epsilon^2} = O(\frac{1}{\epsilon^2})$. Ο αλγόριθμος μπορεί να γίνει ντετερμινιστικός (*derandomization*) αυξάνοντας την πολυπλοκότητα μόνο κατά έναν πολυωνυμικό παράγοντα.

Το θεώρημα γίνεται πιο ισχυρό: Το *PIP* της εισόδου δε χρειάζεται να είναι εφικτό αλλά μόνο προσεγγιστικά εφικτό (δηλαδή πρέπει να υπάρχει ένα σημείο που ικανοποιεί κάθε βαθμού d' περιορισμό με ένα πρόσθετο σφάλμα $\epsilon' n^{d'}$ για κάποιο $\epsilon' < \epsilon/2$).

Συγκεκριμένα για το DkS πρόβλημα θεωρούμε $k \geq an$. Εάν ένας γράφος είναι δ-πυκνός τότε ο γράφος που επάγεται από ένα τυχαίο υποσύνολο από k κορυφές περιέχει κατά μέσο όρο $\frac{a^2\delta n^2}{2}$ ακμές. Συνεπώς, ο πυκνότερος υπογράφος περιέχει τουλάχιστον $\frac{a^2\delta n^2}{2}$ ακμές. Μπορούμε να εκφράσουμε το DkS σαν τη βέλτιστη λύση του ακόλουθου τετραγωνικού προγράμματος

$$\max \sum_{\{i,j\}\in E} x_i x_j$$
$$x_i \in \{0,1\}$$
$$\sum_{i=1}^{n} x_i = k$$

Το παραπάνω *PIP* είναι 1-λείο. Από το Θεώρημα 10 ξέρουμε πως να βρούμε ένα προσεγγιστικά βέλτιστο 0,1 διάνυσμα x που να ικανοποιεί $\sum_{i=1}^{n} x_i \in [k \pm g\sqrt{n \ln n}]$. Μετακινούμε το πολύ $g\sqrt{n \ln n}$ κορυφές μέσα ή έξω για να πάρουμε ένα υποσύνολο μεγέθους k. Αυτό επηρεάζει τον αριθμό των ακμών που περιλαμβάνονται στον υπογράφο το πολύ κατά $gn\sqrt{n \ln n} = o(n^2)$.

Θεώρημα 11. *[1] Υπάρχει πολυωνυμικού χρόνου προσεγγιστικό σχήμα (PTAS) για πυκνά στιγμιότυπα του προβλήματος του πυκνότερου* k*-υπογράφου για* $k = \Omega(n)$.

2.6 Χρωματισμός γράφων

Το πρόβλημα του k-χρωματισμένου υπογράφου (*minimum* k*-colored subgraph problem* - MkCSP), ορίζεται ως εξής: δοθέντος ενός μη κατευθυνόμενου γράφου G, μιας συνάρτησης που αναθέτει σε κάθε ακμή ένα ή περισσότερα χρώματα από ένα σύνολο n δοθέντων χρωμάτων και ενός ακέραιου $k \leq n$, ζητείται να βρεθεί το ελάχιστο σύνολο κορυφών του G που επάγουν ακμές με τουλάχιστον k χρώματα. Το πρόβλημα αυτό αποτελεί γενίκευση δύο σπουδαίων προβλημάτων στην υπολογιστική βιολογία. Στο άρθρο [19] αναλύεται ένας απλός άπληστος αλγόριθμος ο οποίος προσεγγίζει το MkCSP με λόγο $O(\sqrt{k \ln \Delta})$.

Θεώρημα 12. *[19] Υπάρχει προσεγγιστικός αλγόριθμος με λόγο* $\sqrt{2kH(\Delta)} = O(\sqrt{k \ln \Delta})$ *για το πρόβλημα του* k*-χρωματισμένου υπογράφου, όπου* $H(\Delta) = 1 + \frac{1}{2} + \ldots + \frac{1}{\Delta}$.

Απόδειξη. Καταρχάς, περιγράφουμε τον αλγόριθμο. Έστω X το σύνολο των κορυφών που έχουν επιλεχθεί, το οποίο αρχικά είναι κενό. Όσο ο αριθμός των χρωμάτων που έχουν καλυφθεί παραμένει μικρότερος του k, επιλέγουμε μία ακμή με το μέγιστο αριθμό χρωμάτων που δεν έχουν καλυφθεί ακόμα και προσθέτουμε τις κορυφές που αποτελούν τα άκρα της στο σύνολο X (εάν δεν υπάρχουν ήδη στο X οι κορυφές αυτές). Έστω I ο αριθμός των ακμών που επιλέχθηκαν κατά τη διάρκεια αυτής της διαδικασίας. Γνωρίζουμε ότι $|X| \leq 2i$. Επιπλέον είναι ήδη γνωστό [38] ότι ο άπληστος αλγόριθμος που μόλις περιγράψαμε για το πρόβλημα της μερικής επικάλυψης συνόλων (partial set cover), δηλαδή, της εύρεσης του ελάχιστου αριθμού συνόλων που χρειάζονται για να καλυφθούν τουλάχιστον k στοιχεία, είναι ένας $H(\Delta)$-προσεγγιστικός αλγόριθμος. Αυτό σημαίνει ότι ο ελάχιστος αριθμός ακμών που χρειάζονται για να καλυφθούν τουλάχιστον k χρώματα είναι τουλάχιστον $i/H(\Delta)$. Για να επάγει τουλάχιστον $i/H(\Delta)$ ακμές η βέλτιστη λύση του MkCSP πρέπει να επιλέξει τουλάχιστον $\sqrt{2i/H(\Delta)}$ κορυφές. Ο προσεγγιστικός λόγος προκύπτει άμεσα χρησιμοποιώντας αυτό το κάτω φράγμα. □

Παραθέτουμε ένα Θεώρημα το οποίο συσχετίζει το MkCS πρόβλημα με το DkS πρόβλημα. Συγκεκριμένα, αποδεικνύεται ότι εάν υπάρχει f-προσεγγιστικός αλγόριθμος για το MkCS πρόβλημα τότε υπάρχει f^2-προσεγγιστικός αλγόριθμος για το DkS πρόβλημα. Η σχέση μεταξύ του MkCS προβλήματος και του DkS προβλήματος υποδεικνύει ότι ο προσεγγιστικός λόγος του Θεωρήματος 12 δεν μπορεί εύκολα να βελτιωθεί και ότι σημαντικές βελτιώσεις στον προσεγγιστικό λόγο του MkCSP θα απαιτούσαν νέες ιδέες.

Θεώρημα 13. *[19] Εάν υπάρχει ένας πολυωνυμικού χρόνου f-προσεγγιστικός αλγόριθμος $\mathcal{A}$ για το MkCS πρόβλημα, τότε υπάρχει ένας πολυωνυμικού χρόνου $2f^2$-προσεγγιστικός αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου .*

Απόδειξη. Δοθέντος ενός γράφου G με m ακμές, θα θέλαμε να βρούμε ένα σύνολο k κορυφών του με το μέγιστο αριθμό ακμών στον υπογράφο που επάγεται από αυτό το σύνολο κορυφών. Αναθέτουμε σε κάθε ακμή του G ένα διαφορετικό χρώμα και χρησιμοποιούμε τον αλγόριθμο $\mathcal{A}$ για να βρούμε τις προσεγγιστικές λύσεις για το MkCSP στο γράφο που προκύπτει. Ας υποθέσουμε ότι l είναι ο μέγιστος αριθμός χρωμάτων για τον οποίο ο αλγόριθμος $\mathcal{A}$ δίνει σαν έξοδο μία λύση Y με το πολύ k κορυφές. Αυτό σημαίνει ότι l χρώματα έχουν ανατεθεί στον υπογράφο που επάγεται από το Y και ότι η προσεγγιστική λύση που επιστρέφει ο $\mathcal{A}$ όταν $l+1$ χρώματα απαιτούνται για να καλυφθούν οι ακμές περιέχει τουλάχιστον $k+1$ κορυφές. Έστω ότι η βέλτιστη λύση του προβλήματος του πυκνότερου k-υπογράφου περιέχει t ακμές. Θα αποδείξουμε ότι $t \leq 2f^2 l$ και άρα το σύνολο Y είναι μία λύση για το DkS πρόβλημα η οποία προσεγγίζει τη βέλτιστη με ένα παράγοντα $\frac{1}{2f^2}$.

Λόγω της επιλογής του l και του γεγονότος ότι ο $\mathcal{A}$ είναι ένας f-προσεγγιστικός αλγόριθμος, οποιεσδήποτε $\frac{k}{f}$ κορυφές του G επάγουν το πολύ l χρώματα. Θεωρούμε ένα υποσύνολο X με k κορυφές. Ο συνολικός αριθμός χρωμάτων που επάγεται από όλα τα πιθανά υποσύνολα $\frac{k}{f}$ στοιχείων του X είναι το πολύ $\binom{k}{\frac{k}{f}} l$. Παρατηρούμε ότι κάθε ακμή μετριέται ακριβώς $\binom{k-2}{\frac{k}{f}-2}$ φορές. Συνεπώς, ο συνολικός αριθμός ακμών στο X είναι το πολύ

$$\frac{\binom{k}{\frac{k}{f}} l}{\binom{k-2}{\frac{k}{f}-2}} = \frac{k(k-1)}{\frac{k}{f}(\frac{k}{f}-1)} l \leq f^2 l \frac{k-1}{k-f} < 2f^2 l.$$

Η τελευταία ανισότητα ισχύει αφού μπορούμε να υποθέσουμε χωρίς βλάβη της γενικότητας ότι $k > 2f$, αλλιώς κάθε ακμή είναι μία $2f^2$ προσέγγιση. Εφόσον το X είναι ένα αυθαίρετο σύνολο k κορυφών, $t \leq 2f^2 l$. □

Κεφάλαιο 3

Πολυωνυμικές περιπτώσεις

Το "προφανές" είναι η πιο επικίνδυνη λέξη στα Μαθηματικά.

ERIC TEMPLE BELL

3.1 Δέντρα με βάρη

Στην ενότητα αυτή παρουσιάζουμε έναν πολυωνυμικού χρόνου αλγόριθμο δυναμικού προγραμματισμού για το πρόβλημα του πυκνότερου k-υπογράφου σε δέντρα με βάρη στις ακμές (*weighted trees*). Σε αντίθεση με τους αλγόριθμους που προτείνονται στα [35], [33] και [17], στους οποίους η βέλτιστη λύση που προκύπτει είναι συνεκτική, η βέλτιστη λύση που προκύπτει από τον αλγόριθμο που παρουσιάζεται σε αυτό το κεφάλαιο μπορεί να είναι είτε συνεκτική είτε όχι. Στην πραγματικότητα αυτός ο αλγόριθμος είναι μία γενίκευση για προβλήματα όπου απαιτείται μία μη συνεκτική λύση. Ωστόσο, και παρά αυτή τη γενίκευση, ο αλγόριθμος διατηρεί την ίδια χρονική πολυπλοκότητα.

Θεωρούμε ένα δέντρο με n κορυφές και μη αρνητικά βάρη στις ακμές του και υποθέτουμε ότι έχει ρίζα κάποια αυθαίρετη κορυφή s. Για να υπολογίσει τη βέλτιστη λύση ο αλγόριθμος αναδρομικά λύνει για όλες τις τιμές του j, $j = 1, \ldots, k$, όλα τα προβλήματα του πυκνότερου j-υπογράφου σε κάθε υποδέντρο του αρχικού δέντρου.

Έστω $C_u(j)$ η τιμή μίας βέλτιστης λύσης για το πρόβλημα του πυκνότερου j-υπογράφου στο υποδέντρο με ρίζα την κορυφή u του αρχικού δέντρου. Διαχωρίζουμε την $C_u(j)$ σε $C_u^1(j)$, που είναι η τιμή της βέλτιστης λύσης που περιλαμβάνει την ρίζα u του υποδέντρου και $C_u^0(j)$, που είναι η τιμή της βέλτιστης λύσης που δεν περιλαμβάνει τη ρίζα u. Τότε ισχύει ότι,

$$C_u(j) = \max\{C_u^0(j), C_u^1(j)\}.$$

Η βέλτιστη λύση για το πρόβλημα του πυκνότερου k-υπογράφου είναι $C_s(k) = \max\{C_s^0(k), C_s^1(k)\}$, όπου s είναι η ρίζα του αρχικού δέντρου.

Για να υπολογίσει τη βέλτιστη λύση ο αλγόριθμος υπολογίζει τις τιμές $C_u^1(j)$ και $C_u^0(j)$, $j = 1, ..., k$, για όλες τις κορυφές του δέντρου και τελικά και για τη ρίζα του. Ο αλγόριθμος ξεκινά τον υπολογισμό από τα φύλλα του δέντρου (στα οποία εφαρμόζονται οι αρχικές συνθήκες) και προχωρά προς τα επάνω κάνοντας χρήση δύο αναδρομικών συναρτήσεων.

Οι αρχικές συνθήκες για τα φύλλα του δέντρου έχουν ως εξής. Εάν η κορυφή u είναι φύλλο, τότε
$C_u^0(0) = C_u^1(1) = 0$, $C_u^0(1) = C_u^1(0) = -\infty$ και $C_u^{\epsilon}(j) = -\infty$, για $j = 2, 3, \ldots, k$.

Θεωρούμε μία κορυφή u του δέντρου και έστω $u_1, u_2, ..., u_t$ τα παιδιά της κορυφής u σε τυχαία σειρά. Έστω επίσης w_i το βάρος που αντιστοιχεί στην ακμή (u, u_i), $i = 1, 2, ..., t$ και $\epsilon \in \{0, 1\}$.

Η πρώτη αναδρομική συνάρτηση εφαρμόζεται σε δύο παιδιά της κορυφής u, έστω u_1 και u_2, και υπολογίζει τις τιμές $C_u^1(j)$ και $C_u^0(j)$, $j = 1, 2, ..., k$ ως εξής:

$$C_u^{\epsilon}(j) = \max_{\substack{k_1+k_2=j-\epsilon, \\ \epsilon_1,\epsilon_2\in\{0,1\}}} \{C_{u_1}^{\epsilon_1}(k_1) + C_{u_2}^{\epsilon_2}(k_2) + \epsilon_1 \cdot \epsilon \cdot w_1 + \epsilon_2 \cdot \epsilon \cdot w_2\}. \tag{3.1}$$

Παρατηρούμε ότι αυτή η αναδρομική συνάρτηση εφαρμόζεται και όταν το $t = 1$ θέτοντας $k_2 = 0$, $\epsilon_2 = 0$ και $C_{u_2}^{\epsilon_2}(k_2) = 0$.

Η δεύτερη αναδρομική συνάρτηση εφαρμόζεται για τα υπόλοιπα παιδιά της κορυφής u, έστω u_i, $i = 3, 4, ..., t$ και ενημερώνει τις τιμές $C_u^1(j)$ και $C_u^0(j)$, $j = 1, 2, ..., k$ ως εξής:

$$C_u^{\epsilon}(j) = \max_{\substack{k_1+k_2=j, \\ \epsilon_i\in\{0,1\}}} \{C_u^{\epsilon}(k_1) + C_{u_i}^{\epsilon_i}(k_2) + \epsilon \cdot \epsilon_i \cdot w_i\}. \tag{3.2}$$

Ο υπολογισμός κάθε τιμής $C_u^{\epsilon}(j)$ θέλει $O(k)$ χρόνο για κάθε συνδυασμό των k_1, k_2 έτσι ώστε $k_1 + k_2 = j - \epsilon$ και αφού αυτές οι τιμές υπολογίζονται για κάθε τιμή του j, $j = 1, 2, ..., k$, και για όλες τις κορυφές u του δέντρου, προκύπτει ότι η πολυπλοκότητα του παραπάνω αλγόριθμου δυναμικού προγραμματισμού είναι $O(nk^2)$. Συνακόλουθα προκύπτει το παρακάτω θεώρημα.

Θεώρημα 14. *[30] Υπάρχει ένας $O(nk^2)$ χρόνου αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε δέντρα με βάρη.*

Για το πρόβλημα του πυκνότερου k-υπογράφου σε δάση με βάρη (*weighted forests*) καταρχάς συνδέουμε τα δέντρα του δάσους σε ένα δέντρο, προσθέτοντας ακμές μηδενικού βάρους μεταξύ των ριζών τους. Έπειτα, εφαρμόζουμε τον παραπάνω αλγόριθμο δυναμικού προγραμματισμού σε αυτό το δέντρο. Τελικά, εάν μία ακμή μηδενικού βάρους από αυτές που προσθέσαμε περιλαμβάνεται στη βέλτιστη λύση, απλά την αφαιρούμε αφού η τιμή του συνολικού βάρους δεν αλλάζει.

Πόρισμα 1. *Υπάρχει ένας $O(nk^2)$ χρόνου αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε ένα δάσος με βάρη.*

Για τον υπολογισμό του πυκνότερου k-υπογράφου σε αλυσίδες με βάρη αρκεί η πρώτη αναδρομική συνάρτηση του αλγόριθμου δυναμικού προγραμματισμού. Έστω $u_1, u_2, ..., u_n$ οι κορυφές της αλυσίδας. Έστω $\epsilon \in \{0,1\}$ και w_i το βάρος της ακμής (u_i, u_{i+1}), $i = 1, 2, ..., n-1$. Τότε για κάθε $i = 1, 2, ..., n-1$ και $j = 1, 2, ..., k$ η σχέση 3.1 γίνεται

$$C^{\epsilon}_{u_{i+1}}(j) = \max_{\epsilon_i \in \{0,1\}} \{C^{\epsilon_i}_{u_i}(j-\epsilon) + \epsilon_i \cdot \epsilon \cdot w_i\}. \tag{3.3}$$

Πόρισμα 2. *Υπάρχει ένας $O(nk)$ χρόνου αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε αλυσίδες με βάρη.*

3.2 Βεβαρημένοι γράφοι μέγιστου βαθμού 2

Σε αυτή την ενότητα παρουσιάζουμε έναν πολυωνυμικού χρόνου αλγόριθμο ο οποίος βρίσκει μία βέλτιστη λύση (είτε συνεκτική είτε όχι) για το πρόβλημα του πυκνότερου k-υπογράφου σε βεβαρημένους γράφους μέγιστου βαθμού 2 . Συμβολίζουμε με g_i, $1 \leq i \leq m$, μία συνεκτική συνιστώσα (έναν κύκλο (*cycle*) ή μία αλυσίδα (*chain*)) ενός βεβαρημένου γράφου G με μέγιστο βαθμό δύο, με n_i τον αριθμό των κορυφών (όπου, $\sum_{i=1}^{m} n_i = n$) και με e_i τον αριθμό των ακμών κάθε συνιστώσας g_i (όπου $e_i = n_i$, εάν g_i είναι κύκλος και $e_i = n_i - 1$, εάν g_i είναι αλυσίδα).

Για το πρόβλημα του πυκνότερου k-υπογράφου αναφέρεται στο [12] ότι υπάρχει αλγόριθμος δυναμικού προγραμματισμού πολυωνυμικού χρόνου σε γράφους με μέγιστο βαθμό δύο. Στην πραγματικότητα, σε αυτή την περίπτωση το DkS πρόβλημα μπορεί να γραφτεί σαν το ακόλουθο πρόβλημα σακιδίου (*Knapsack problem*):

$$\max \sum_{i=1}^{m} e_i x_i, \text{ έτσι ώστε } \sum_{i=1}^{m} n_i x_i \leq k \text{ και } x_i \in \{0,1\}.$$

Αυτό το πρόβλημα σακιδίου μπορεί να λυθεί βέλτιστα σε χρόνο $O(nk)$.

Για το πυκνότερου k-υπογράφου πρόβλημα σε βεβαρημένους γράφους μέγιστου βαθμού 2 ο αλγόριθμος στην πρώτη φάση υπολογίζει βέλτιστα, για κάθε συνιστώσα g_i του G, όλα τα j κορυφών πυκνότερου υπογράφου προβλήματα για κάθε τιμή j, $1 \leq j \leq \min\{n_i, k\}$. Εάν η συνιστώσα g_i είναι αλυσίδα αυτό το βήμα επιτυγχάνεται χρησιμοποιώντας το Πόρισμα 2. Εάν η συνιστώσα g_i είναι κύκλος παραθέτουμε το επόμενο λήμμα.

Λήμμα 5. *Υπάρχει ένας $O(nk)$ χρόνου αλγόριθμος για το πρόβλημα του k-πυκνότερου υπογράφου σε κύκλο n κορυφών με βάρη.*

Απόδειξη. Έστω OPT η βέλτιστη λύση για τον κύκλο $C: u_1, u_2, ..., u_n, u_1$. Έστω $e = (u_n, u_1)$ μία ακμή του κύκλου με βάρος $w(e)$. Αφαιρούμε την e από τον κύκλο και εφαρμόζουμε το δυναμικό προγραμματισμό του Πορίσματος 2 στην εναπομένουσα αλυσίδα $P: u_1, u_2, ..., u_n$. Έστω OPT_1 η λύση που προκύπτει.

Εάν και οι δύο κορυφές u_1 και u_n ανήκουν στην OPT_1 τότε $OPT = OPT_1 + w(e)$. Αλλιώς εφαρμόζουμε ακόμα μία φορά το δυναμικό προγραμματισμό του Πορίσματος 2 στην ίδια αλυσίδα αλλά τροφοδοτώντας τον αλγόριθμο με τις κατάλληλες αρχικές συνθήκες έτσι ώστε να περιλαμβάνει και τις δύο κορυφές u_1 και u_n και άρα και το μεταξύ τους βάρος $w(e)$. Έστω OPT_2 η βέλτιστη λύση που περιλαμβάνει την ακμή e.

Για να υπολογίσουμε την OPT_2 χρησιμοποιούμε τη σχέση 3.3 θεωρώντας τις ακόλουθες αρχικές συνθήκες για την κορυφή u_1:
$C^0_{u_1}(j) = -\infty$ για $j \geq 0$,
$C^1_{u_1}(2) = w(e)$, και
$C^1_{u_1}(j) = -\infty$ για $j = 0, 1$ και $j \geq 3$.
Οι παραπάνω οριακές συνθήκες αναγκάζουν τον αλγόριθμο να συμπεριλάβει την ακμή e σε κάθε περίπτωση.

Εάν u_n είναι η τελευταία κορυφή της αλυσίδας τότε

$$OPT_2 = C^1_{u_n}(k) = \max_{\epsilon_i \in \{0,1\}} \{C^{\epsilon_i}_{u_{n-1}}(k) + \epsilon_i \cdot w_{u_{n-1}}\}$$

Παρατηρούμε ότι η κορυφή u_n έχει ήδη υπολογιστεί στις αρχικές συνθήκες.

Η βέλτιστη λύση, OPT, για το πρόβλημα του πυκνότερου k-υπογράφου σε βεβαρημένο κύκλο είναι το μέγιστο μεταξύ των λύσεων OPT_1 και OPT_2. Από το Πόρισμα 2 προκύπτει ότι η OPT υπολογίζεται σε χρόνο $O(nk)$. □

Χρησιμοποιώντας το Πόρισμα 2 και το Λήμμα 5 για κάθε συνεκτική συνιστώσα g_i, $1 \leq i \leq m$, του G έχουμε βέλτιστες λύσεις σε κάθε j-κορυφών πυκνότερου υπογράφου πρόβλημα για $1 \leq j \leq \min\{n_i, k\}$. Συμβολίζουμε με C_{ij}, $1 \leq i \leq m$, $1 \leq j \leq k$, την τιμή κάθε τέτοιας βέλτιστης λύσης, δηλαδή, την τιμή της βέλτιστης λύσης για το j- κορυφών πυκνότερου υπογράφου πρόβλημα της συνεκτικής συνιστώσας g_i. Για τις συνεκτικές συνιστώσες όπου $k > n_i$ θέτουμε $C_{ij} = 0$ για $j > n_i$. Οι βέλτιστες λύσεις για την ίδια συνεκτική συνιστώσα, αλλά για διαφορετικές τιμές του j, μπορεί να συμπίπτουν.

Για κάθε συνεκτική συνιστώσα g_i του G χρειαζόμαστε $O(kn_i)$ χρόνο για να υπολογίσουμε τις βέλτιστες λύσεις για όλα τα j- κορυφών πυκνότερου υπογράφου προβλήματα για $1 \leq j \leq \min\{n_i, k\}$, αφού όλες αυτές οι λύσεις υπολογίζονται από τον αλγόριθμο δυναμικού προγραμματισμού που επιλύει το πρόβλημα για $j = \min\{n_i, k\}$. Άρα, για όλες τις συνεκτικές συνιστώσες, η πρώτη φάση του αλγόριθμου παίρνει χρόνο $O(k \sum_{i=1}^{m} n_i)$ και άρα $O(kn)$, εφόσον $\sum_{i=1}^{m} n_i = n$.

Στη δεύτερη φάση του αλγόριθμου θέλουμε να επιλέξουμε το πολύ μία βέλτιστη λύση (για κάποια τιμή του j) από κάθε συνεκτική συνιστώσα έτσι ώστε ο συνολικός αριθμός κορυφών τους να είναι k και ο συνολικός αριθμός ακμών τους να είναι ο μέγιστος δυνατός. Χρησιμοποιώντας μία μεταβλητή y_{ij} που παίρνει τιμή 1 εάν μία βέλτιστη λύση με τιμή C_{ij} έχει επιλεγεί και 0 αλλιώς, το πρόβλημα γράφεται ως εξής:

$$\max \sum_{i=1}^{m} \sum_{j=1}^{k} C_{ij}\, y_{ij} \tag{3.4}$$

$$\sum_{j=1}^{k} y_{ij} \leq 1, \qquad 1 \leq i \leq m \tag{3.5}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{k} j\, y_{ij} \leq k \tag{3.6}$$

$$y_{ij} \in \{0,1\}, \quad 1 \leq i \leq m, \quad 1 \leq j \leq k \tag{3.7}$$

Ουσιαστικά αυτό το πρόβλημα είναι μία γενίκευση του προβλήματος του σακιδίου (*Knapsack problem*), στο οποίο μας δίνεται μία διαμέριση των αντικειμένων σε ομάδες και έχουμε τον περιορισμό να επιλέξουμε το πολύ ένα αντικείμενο από κάθε ομάδα. Είναι γνωστό ότι αυτό το πρόβλημα επιλύεται σε χρόνο $O(mk^2)$ από έναν αλγόριθμο δυναμικού προγραμματισμού που προτείνεται στο άρθρο [26]. Εφόσον το m είναι $O(n)$ η δεύτερη φάση του αλγόριθμου χρειάζεται χρόνο $O(nk^2)$.

Συνδυάζοντας τις δύο φάσεις του αλγόριθμου προκύπτει το παρακάτω θεώρημα.

Θεώρημα 15. *[30] Υπάρχει ένας $O(nk^2)$ χρόνου αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε βεβαρημένους γράφους μέγιστου βαθμού δύο.*

Τα αποτελέσματα του Πορίσματος 2 και του Λήμματος 5 εξάγονται και σαν έμμεσα συμπεράσματα από ένα αποτέλεσμα για το πρόβλημα του τετραγωνικού 0-1 σακιδίου (*quadratic 0-1 knapsack problem*) σε σειριακούς-παράλληλους ακμών γράφους (*edge series-parallel graphs*) [36].

3.2.1 Το τετραγωνικό 0-1 σακίδιο σε σειριακούς - παράλληλους ακμών γράφους

Το πρόβλημα του τετραγωνικού 0-1 σακιδίου (*quadratic 0-1 knapsack problem*) αποτελεί γενίκευση του κλασσικού προβλήματος του σακιδίου (*knapsack problem*). Στο πρόβλημα του τετραγωνικού 0-1 σακιδίου ζητείται να μεγιστοποιηθεί η συνάρτηση $F(x) = \sum_{i \in V} c_i x_i + \sum_{(i,j) \in E} c_{ij} x_i x_j$, υπό τον περιορισμό ότι $\sum_{i \in V} a_i x_i \leq b$ και $x \in \{0,1\}^n$, όπου $V = \{1, \ldots, n\}$ είναι το σύνολο των αντικειμένων και a_i, c_i είναι το βάρος και η αξία του αντικειμένου i αντίστοιχα. Κάθε αντικείμενο i έχει μία διαφορετικής αξίας, c_{ij}, συσχέτιση με κάθε αντικείμενο j και E είναι το σύνολο των συσχετίσεων.

Ο κλασσικός τρόπος ορισμού των σειριακών-παράλληλων ακμών γράφων (*edge series-parallel graphs* - ESP) είναι μέσω των 2-ακροδεκτών (*2-terminal*) γράφων. Ένας 2-ακροδεκτών γράφος $G = (V, E)$ είναι ένας γράφος με δύο ειδικές κορυφές που καλούνται αριστερός και δεξιός ακροδέκτης. Για δύο 2-ακροδεκτών γράφους $G_i = (V_i, E_i)$, $i = 1, 2$, ορίζουμε τις εξής πράξεις:

- Η σειριακή (*series*) σύνθεση $G_s = G_1 * G_2$ των G_1 και G_2 προκύπτει από τη σύνδεση του δεξιού ακροδέκτη του G_1 με τον αριστερό ακροδέκτη του G_2. Ο γράφος G_s που προκύπτει θεωρείται ως 2-ακροδεκτών γράφος με αριστερό ακροδέκτη τον αριστερό ακροδέκτη του G_1 και δεξιό ακροδέκτη το δεξιό ακροδέκτη του G_2.
- Η παράλληλη (*parallel*) σύνθεση $G_p = G_1 || G_2$ των G_1 και G_2 προκύπτει από τη σύνδεση των δύο δεξιών ακροδεκτών μεταξύ τους και των δύο αριστερών ακροδεκτών μεταξύ τους. Οι κορυφές ακροδέκτες του G_p είναι αυτοί οι διασυνδεδεμένοι ακροδέκτες.

Η κλάση των ESP γράφων ορίζεται με τους εξής τρείς κανόνες: (1) Ο γράφος που αποτελείται από δύο ακροδέκτες (*terminals*) συνδεδεμένους με μία ακμή είναι ESP. (2) Εάν G_1 και G_2 είναι ESP, τότε ο $G_1 * G_2$ και ο $G_1 || G_2$ είναι ESP. (3) Κανένας άλλος γράφος εκτός αυτών που ορίστηκαν στα (1) και (2) δεν είναι ESP.

Στο άρθρο [36] προτείνονται δύο δυναμικού προγραμματισμού αλγόριθμοι για την ειδική περίπτωση του τετραγωνικού 0-1 σακιδίου σε γράφους ESP. Ο ένας αλγόριθμος είναι ψευδοπολυωνυμικός στα βάρη των αντικειμένων ενώ ο άλλος στις αξίες. Συνεπώς αποδεικνύονται τα εξής θεωρήματα.

Θεώρημα 16. *[36] Η ειδική περίπτωση του τετραγωνικού 0-1 σακιδίου (quadratic 0-1 knapsack problem) σε γράφους ESP μπορεί να λυθεί σε ψευδοπολυωνυμικό χρόνο* $O(nb^2)$.

Θεώρημα 17. *[36] Η ειδική περίπτωση του τετραγωνικού 0-1 σακιδίου (quadratic 0-1 knapsack problem) σε γράφους ESP μπορεί να λυθεί σε ψευδοπολυωνυμικό χρόνο* $O(nU^2)$, *όπου* $U = \sum_{i \in V} |c_i| + \sum_{(i,j) \in E} |c_{ij}|$.

Στην περίπτωση των αλυσίδων (*chains*) και των κύκλων (*cycles*) που είναι υποκλάσεις των ESP η χρονική πολυπλοκότητα του αλγορίθμου του Θεωρήματος 16 γίνεται $O(nb)$. Το τετραγωνικό 0-1 σακίδιο (*quadratic 0-1 knapsack problem*) μετατρέπεται στο DkS πρόβλημα σε βεβαρημένους γράφους για $b = k$, $a_i = 1$, $c_i = 0$ και $c_{ij} =$

το βάρος που αντιστοιχεί στην ακμή (i,j). Συνεπώς ο δυναμικού προγραμματισμού αλγόριθμος που προτείνεται στο Θεώρημα 16 για το DkS σε βεβαρημένες αλυσίδες και βεβαρημένους κύκλους έχει χρονική πολυπλοκότητα $O(nk)$ και συμπίπτει με τα αποτελέσματα του Πορίσματος 2 και του Λήμματος 5.

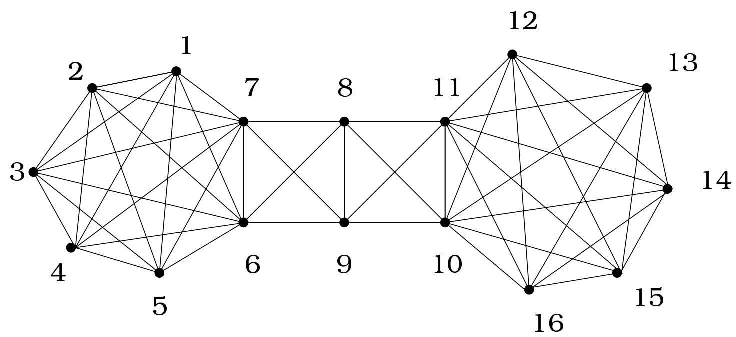
3.3 Υποκλάση γράφων διαστημάτων

Στην ενότητα αυτή παρουσιάζουμε ένα δυναμικού προγραμματισμού αλγόριθμο ο οποίος βρίσκει μία βέλτιστη λύση για το πρόβλημα του πυκνότερου k-υπογράφου σε μία υποκλάση των γράφων διαστημάτων (*interval graphs*). Υπενθυμίζουμε ότι το DkS πρόβλημα παραμένει δύσκολο (*NP-hard*) στους χορδικούς γράφους (*chordal graphs*) [7] και ότι η πολυπλοκότητα του στους γράφους διαστημάτων παραμένει ανοικτό πρόβλημα.

Ένας γράφος $G = (V, E)$ ονομάζεται γράφος διαστημάτων (*interval graph*) εάν υπάρχει μία αντιστοιχία I των κορυφών του G σε σύνολα διαδοχικών ακέραιων έτσι ώστε για κάθε ζεύγος κορυφών $u, v \in V$, $(u, v) \in E \iff I(u) \bigcap I(v) \neq \emptyset$. Υπάρχει αλγόριθμος γραμμικού χρόνου για την αναγνώριση γράφων διαστημάτων [6]. Οι γράφοι διαστημάτων έχουν πολλές και σπουδαίες εφαρμογές [18].

Ο γράφος μεγιστικών κλικών (*clique graph*) ενός γράφου G είναι ένας βεβαρημένος γράφος του οποίου οι κορυφές αντιστοιχούν στις μεγιστικές κλίκες (*maximal cliques*) του G. Υπάρχει ακμή μεταξύ δύο κορυφών του γράφου μεγιστικών κλικών αν και μόνο αν οι αντίστοιχες μεγιστικές κλίκες (*maximal cliques*) τέμνονται. Το βάρος μίας τέτοιας ακμής ισούται με τον αριθμό των κορυφών στην τομή. Το δέντρο μεγιστικών κλικών (*clique tree*) του G είναι το μέγιστου βάρους δέντρο επικάλυψης (*maximum weight spanning tree*) του γράφου μεγιστικών κλικών. Είναι γνωστό ότι όλες οι μεγιστικές κλίκες ενός γράφου διαστημάτων μπορούν να βρεθούν σε γραμμικό χρόνο [16]. Είναι επίσης γνωστό ότι το δέντρο μεγιστικών κλικών ενός γράφου διαστημάτων είναι ένα απλό μονοπάτι (*path*).

Σε αυτή την ενότητα εστιάζουμε σε μία υποκλάση των γράφων διαστημάτων για τους οποίους ο γράφος μεγιστικών κλικών είναι ένα απλό μονοπάτι. Επί παραδείγματι, στο Σχήμα 3.1 δίνεται ένας τέτοιος γράφος $G = (V, E)$ με $|V| = 16$ στον οποίο ζητάμε τον πυκνότερο υπογράφο με $k = 15$ κορυφές. Ο αλγόριθμος θα πρέπει να δώσει σαν βέλτιστη λύση για $k = 15$ την τιμή 46 η οποία προκύπτει από τον αρχικό γράφο χωρίς μία από τις κορυφές 8 ή 9. Προφανώς, σε αυτή την κλάση υπάρχουν τομές μόνο μεταξύ δύο διαδοχικών μεγιστικών κλικών στο μονοπάτι. Στη γενική κλάση των γράφων διαστημάτων υπάρχουν κοινές κορυφές μεταξύ περισσοτέρων από δύο διαδοχικών μεγιστικών κλικών.



Σχήμα 3.1: Γράφος $G = (V, E)$, $|V| = 16$, του οποίου ο γράφος μεγιστικών κλικών είναι ένα απλό μονοπάτι.

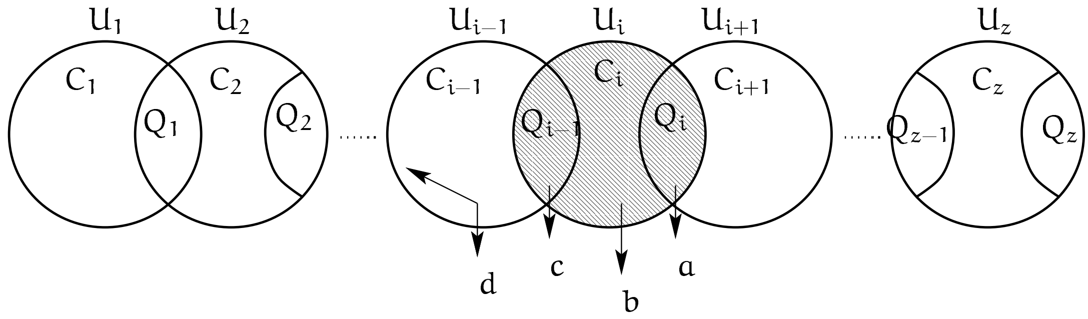
Έστω $G(V, E)$, $|V| = n$, ένας τέτοιος γράφος διαστημάτων και έστω $U_1, \ldots, U_z$ οι μεγιστικές κλίκες του (*maximal cliques*). Διαχωρίζουμε κάθε κλίκα U_i, $i = 2, \ldots, z$ σε τρείς κλίκες C_i, Q_i και Q_{i-1} ως εξής (Σχήμα 3.2):

- Q_i είναι οι κορυφές που ανήκουν στην τομή της U_i με την U_{i+1}, δηλαδή, $Q_i = U_i \cap U_{i+1}$ για $i = 1, \ldots, z-1$. Με Q_z συμβολίζουμε μία κορυφή στη U_z αλλά όχι στη U_{z-1}.
- C_i είναι οι κορυφές της U_i που δεν ανήκουν σε καμία άλλη τομή, δηλαδή, $C_1 = U_1 - Q_1$, $C_i = U_i - (Q_{i-1} \cup Q_i)$ για $i = 2, \ldots, z$.

Ο αλγόριθμος διασχίζει το γράφο μεγιστικών κλικών (*clique graph*) μέσω των κλικών U_i, $i = 1, \ldots, z$, και σε κάθε βήμα υπολογίζει μία βέλτιστη λύση για όλα τα DjS προβλήματα, $j = 1, \ldots, k$, στον υπογράφο $G_i = \bigcup_{m=1}^{i} U_m$. Λόγω της δομής του γράφου είναι προφανές ότι μία τέτοια βέλτιστη λύση στον υπογράφο G_i θα έχει την ίδια τιμή για κάθε κορυφή που ανήκει στην κλίκα Q_i. Για το λόγο αυτό θεωρούμε οποιαδήποτε κορυφή u_i στην κλίκα Q_i, $i = 1, \ldots, z$ και ορίζουμε ως $f_{u_i}(j)$ την τιμή της βέλτιστης λύσης του DjS προβλήματος στον υπογράφο G_i.

Ως $f_{u_i}(j, a)$ ορίζουμε την τιμή της βέλτιστης τιμής του DjS προβλήματος στον υπογράφο G_i η οποία περιλαμβάνει ακριβώς a κορυφές της κλίκας Q_i. Είναι φανερό ότι

$$f_{u_i}(j) = \max_{0 \leq a \leq |Q_i|} \{f_{u_i}(j, a)\}.$$



Σχήμα 3.2: Γράφος διαστημάτων με τομές μόνο μεταξύ δύο μεγιστικών διαδοχικών κλικών.

Εάν $a = 0$, τότε καμία κορυφή της κλίκας Q_i (και άρα και της κλίκας C_i) δεν ανήκει στη βέλτιστη λύση του αντίστοιχου DjS προβλήματος στον υπογράφο G_i. Άρα, η τιμή αυτής της λύσης είναι η ίδια με την τιμή της βέλτιστης λύσης του DjS προβλήματος στον υπογράφο G_{i-1}, δηλαδή,

$$f_{u_i}(j, a) = f_{u_{i-1}}(j).$$

Εάν $a > 0$, τότε μία βέλτιστη λύση για το DjS πρόβλημα που περιλαμβάνει a κορυφές της κλίκας Q_i μπορεί επίσης να περιλαμβάνει $b \geq 0$ κορυφές της κλίκας C_i, $c \geq 1$ κορυφές* της κλίκας Q_{i-1}, και $d \geq 0$ κορυφές από τον υπογράφο $\bigcup_{m=1}^{i-1} U_m - Q_{i-1}$. Τότε,

$$f_{u_i}(j, a) = \begin{cases} \dfrac{j(j-1)}{2}, & \text{αν } j \leq |Q_{i-1}| + |C_i| + a \\ \max\limits_{a+b+c+d=j}\{f_{u_{i-1}}(c+d, c) + \dbinom{a+b}{2} + c(a+b)\}, & \text{αλλιώς} \end{cases}$$

Για $i = 1$ ισχύουν οι ακόλουθες οριακές συνθήκες για $u_1 \in Q_1$ και $1 \leq j \leq k$:

Εάν $a = 0$, τότε $f_{u_1}(j, a) = 0$.

Εάν $1 \leq a \leq |Q_1|$, τότε

$$f_{u_1}(j, a) = \begin{cases} \dfrac{j(j-1)}{2}, & \text{αν } j \leq |C_1| + a \\ -\infty, & \text{αλλιώς} \end{cases}$$

*ή $c \geq 0$ για μη συνεκτική λύση.

Ο αλγόριθμος τερματίζει υπολογίζοντας την τιμή της $f_{u_z}(k)$, δηλαδή, την τιμή της βέλτιστης λύση για το DkS πρόβλημα στον αρχικό γράφο G. Εξ' ορισμού η Q_z περιλαμβάνει μία και μοναδική κορυφή, συνεπώς $f_{u_z}(k) = \max_{a=0,1}\{f_{u_z}(k, a)\}$.

Για να υπολογιστεί η τιμή κάθε συνάρτησης υπολογισμού $f_{u_i}(j)$ χρειάζεται το πολύ $O(k^2)$ χρόνος. Ο χρόνος προκύπτει λόγω του ότι η $f_{u_i}(j)$ υπολογίζεται για όλους τους δυνατούς συνδυασμούς των a, b, c και d έτσι ώστε $a + b + c + d = j$ και με δεδομένο ότι ο συνδυασμός των $c + d$ προκύπτει άμεσα από τους υπολογισμούς στα προηγούμενα βήματα. Ο αλγόριθμος υπολογίζει την τιμή $f_{u_i}(j)$, για κάθε $j = 1, 2, ..., k$ και κάθε u_i, $i = 1, 2, ..., z$. Εφόσον στη χειρότερη περίπτωση το z είναι $O(n)$, το επόμενο θεώρημα έπεται.

Θεώρημα 18. *[29, 30] Υπάρχει ένας $O(nk^3)$ χρόνου αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε γράφους διαστημάτων που έχουν ένα απλό μονοπάτι ως γράφο μεγιστικών κλικών.*

Η πολυπλοκότητα του DkS προβλήματος στη γενική κλάση των γράφων διαστημάτων παραμένει ένα ενδιαφέρον ανοικτό πρόβλημα. Παρατηρούμε ότι ένας γράφος διαστημάτων στην υποκλάση που εδώ θεωρούμε δεν περιέχει σαν υπογράφο το $K_{1,3}$ (γιατί αλλιώς ο γράφος μεγιστικών κλικών του δεν είναι απλό μονοπάτι). Οι γράφοι διαστημάτων που δεν περιέχουν το $K_{1,3}$ είναι γνωστοί ως *proper interval* γράφοι. Αυτή η κλάση ορίζεται και ως η υποκλάση των γράφων διαστημάτων οι οποίοι μπορούν να αναπαρασταθούν σαν ένα σύνολο από διαστήματα ακεραίων όπου κανένα δεν περιέχει εξ ολοκλήρου κάποιο άλλο [18]. Ωστόσο η κλάση που εδώ θεωρούμε δεν περιέχει όλους τους γράφους διαστημάτων που δεν περιέχουν το $K_{1,3}$ και άρα δεν συμπίπτει με την κλάση των *proper interval* γράφων. Η πολυπλοκότητα του DkS προβλήματος στους *proper interval* γράφους παραμένει επίσης ένα ενδιαφέρον ανοικτό πρόβλημα.

Κεφάλαιο 4

Γράφοι κλικών

> Παρόλο που μπορεί να φαίνεται παράδοξο, όλες οι θετικές επιστήμες κυριαρχούνται από την ιδέα της προσέγγισης.
>
> BERTRAND RUSSELL

4.1 Ορισμοί

Όπως αναφέραμε και στο προηγούμενο κεφάλαιο ένα ενδιαφέρον ανοικτό πρόβλημα είναι η πολυπλοκότητα του προβλήματος του πυκνότερου k-υπογράφου στους γράφους διαστημάτων καθώς και ο λόγος προσέγγισης του στους χορδικούς γράφους για τους οποίους το πρόβλημα παραμένει δύσκολο (*NP-hard*). Διερευνώντας το πρόβλημα ως προς αυτή την κατεύθυνση σε αυτό το κεφάλαιο εστιάζουμε σε χορδικούς γράφους και γράφους διαστημάτων οι οποίοι έχουν εξειδικευμένους γράφους μεγιστικών κλικών (*clique graphs*) με σκοπό να εντοπίσουμε τη διαχωριστική γραμμή μεταξύ δύσκολων και πολυωνυμικά επιλύσιμων ή προσεγγίσιμων περιπτώσεων του DkS προβλήματος.

Μία κλίκα (*clique*) ενός μη κατευθυνόμενου γράφου, $G = (V, E)$, είναι ένα υποσύνολο των κορυφών του G που επάγουν έναν πλήρη υπογράφο στο G. Μεγιστική (*maximal*) είναι μία κλίκα όταν δεν περιέχεται σε μεγαλύτερη κλίκα. Ο γράφος τομής (*intersection graph*) μίας οικογένειας, F, υποσυνόλων ενός συνόλου ορίζεται ως ο γράφος, $\mathcal{G}$, του οποίου οι κορυφές αντιστοιχούν στα υποσύνολα της F, και υπάρχει ακμή μεταξύ δύο κορυφών του $\mathcal{G}$ εάν το αντίστοιχο ζευγάρι υποσυνόλων έχει κοινή τομή. Με βάση αυτούς τους ορισμούς, ο γράφος μεγιστικών κλικών (*clique graph*) ενός γράφου G ορίζεται ως ο γράφος τομής των μεγιστικών κλικών (*maximal cliques*) του G. Είναι γνωστό ότι όλες οι μεγιστικές κλίκες, και άρα και ο γράφος μεγιστικών κλικών, ενός χορδικού γράφου μπορεί να βρεθεί σε πολυωνυμικό χρόνο [15]. Πολλές φορές είναι πιο βολικό να μελετάμε το DkS πρόβλημα στον γράφο μεγιστικών κλικών ενός χορδικού γράφου αντί στον ίδιο το γράφο.

Στις επόμενες ενότητες παρουσιάζουμε ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*PTAS*) σε γράφους που έχουν σαν γράφο μεγιστικών κλικών ένα αστέρι κλικών (*star of cliques*) και έναν $O(nk^{m+1})$ χρόνου αλγόριθμο δυναμικού προγραμματισμού σε γράφους που έχουν σαν γράφο μεγιστικών κλικών ένα δέντρο κλικών (*tree of cliques*) μέγιστου βαθμού m. Παρατηρούμε ότι γενικά τα αστέρια κλικών και τα δέντρα κλικών δεν είναι γράφοι με ελάχιστο βαθμό $\Omega(n)$ ούτε πυκνοί γράφοι ($\Omega(n^2)$ ακμών) για τους οποίους ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα είναι ήδη γνωστό [1].

4.2 Πολυωνυμικού χρόνου προσεγγιστικό σχήμα σε αστέρι κλικών

Σε αυτή την ενότητα μελετάμε γράφους που έχουν σαν γράφο μεγιστικών κλικών ένα αστέρι κλικών. Έστω $C_0, C_1, \ldots, C_{m-1}$ οι μεγιστικές κλίκες ενός τέτοιου γράφου έτσι ώστε η κλίκα C_0 να έχει τομή με κάθε άλλη κλίκα και καμία άλλη τομή να μην υπάρχει (σε ότι ακολουθεί συμβολίζουμε με C_i τόσο την κλίκα C_i όσο και το σύνολο κορυφών της). Εφόσον ένα τέτοιος γράφος είναι ο γράφος μεγιστικών κλικών ενός γράφου G, δεν υπάρχει ακμή του G μεταξύ κορυφών που ανήκουν σε διαφορετικές κλίκες.

Συμβολίζουμε με C_0 την κεντρική κλίκα και με C_i, $1 \leq i \leq m-1$, τις εξωτερικές κλίκες. Για κάθε εξωτερική κλίκα C_i συμβολίζουμε με a_i τον αριθμό των κορυφών της στην τομή της με τη C_0, δηλαδή, $a_i = |C_i \cap C_0|$ και με b_i τον αριθμό των κορυφών της εκτός C_0, δηλαδή, $b_i = |C_i| - a_i > 0$. Με C_0' συμβολίζουμε την κλίκα που περιέχει τις κορυφές της C_0 που δεν ανήκουν σε καμία άλλη κλίκα, δηλαδή, $C_0' = C_0 \setminus \bigcup_{i=1}^{m-1} C_i$.

Με S συμβολίζουμε μία λύση για το πρόβλημα του πυκνότερου k-υπογράφου, δηλαδή, ένα υποσύνολο με $|S| = k$ κορυφές, και με $E(S)$ συμβολίζουμε τον αριθμό των ακμών σε ένα υπογράφο επαγόμενο από το σύνολο κορυφών S. Με S^* συμβολίζουμε μία βέλτιστη λύση για το DkS πρόβλημα. Με $n > k$ συμβολίζουμε το συνολικό αριθμό κορυφών σε όλες τις κλίκες.

Λέμε ότι μία κλίκα C_i, $0 \leq i \leq m-1$, είναι πλήρως σε μία λύση S εάν όλες οι κορυφές της ανήκουν στην S. Από την άλλη μεριά, λέμε ότι οι κλίκες C_0 και C'_0 είναι μερικώς σε μία λύση S εάν ένα μη κενό υποσύνολο των κορυφών τους, αλλά όχι όλες, ανήκει στην S. Επίσης λέμε ότι μία εξωτερική κλίκα C_i, $1 \leq i \leq m-1$, είναι μερικώς στην S εάν ένα μη κενό υποσύνολο των $C_i \setminus C_0$ κορυφών της, αλλά όχι όλες, ανήκουν στην S. Διαχωρίζουμε τον ορισμό της μερικώς συμπερίληψης σε μία λύση S για μία εξωτερική κλίκα C_i διότι εάν μόνο κάποιες από τις $C_i \cap C_0$ κορυφές της ανήκουν στην S, μπορούν να θεωρηθούν και ως κορυφές της κλίκας C_0. Γενικά λέμε ότι μία κλίκα συμμετέχει σε μία λύση S εάν είναι είτε πλήρως είτε μερικώς στη λύση S.

Θεωρώντας μία βέλτιστη λύση S^* παρατηρούμε ότι εάν μία εξωτερική κλίκα C_i είναι μερικώς στη λύση S^*, τότε όλες οι $|C_i \cap C_0| = a_i$ κορυφές της ανήκουν στην S^*. Διαφορετικά αντικαθιστώντας μία κορυφή $y \in C_i \setminus C_0$, $y \in S^*$ με μία κορυφή $x \in C_i \cap C_0$, $x \notin S^*$ προκύπτει μία μεγαλύτερη λύση, που είναι άτοπο.

Σε ότι ακολουθεί υποθέτουμε ότι:

(i) $k > |C_i|$, $i = 0, 1, \ldots, m-1$. Αλλιώς η S^* είναι οποιοδήποτε υποσύνολο k κορυφών κάποιας κλίκας για την οποία ισχύει $|C_i| \geq k$.

(ii) $m > 2$. Για $m = 1$ ισχύει το (i). Για $m = 2$, εάν $k > |C_0| \geq |C_1|$, τότε η S^* αποτελείται από τις κορυφές της κλίκας C_0 και οποιοδήποτε υποσύνολο $k - |C_0|$ κορυφών της $C_1 \setminus C_0$.

Χρησιμοποιώντας τους παραπάνω ορισμούς δίνουμε στις προτάσεις που ακολουθούν μερικές χρήσιμες δομικές ιδιότητες μίας βέλτιστης λύσης S^*.

Πρόταση 1. *Το πολύ μία από τις κλίκες $C'_0, C_1, \ldots, C_{m-1}$ είναι μερικώς σε μία βέλτιστη λύση.*

Απόδειξη. Καταρχάς αποδεικνύουμε ότι το πολύ μία από τις εξωτερικές κλίκες είναι μερικώς στη βέλτιστη λύση S^*. Υποθέτουμε ότι δύο εξωτερικές κλίκες C_i, C_j, $1 \leq i \neq j \leq m-1$, είναι μερικώς στην S^* και υποθέτουμε, χωρίς βλάβη της γενικότητας, ότι $|S^* \cap C_i| \geq |S^* \cap C_j|$.

Έστω $x \notin S^*$ μία κορυφή στην κλίκα $C_i \setminus C_0$ και $y \in S^*$ μία κορυφή στην κλίκα $C_j \setminus C_0$. Θεωρούμε τη λύση S η οποία προκύπτει με αντικατάσταση της κορυφής y

με την κορυφή x. Τότε, επειδή $E(S) = E(S^*) - (|S^* \cap C_j| - 1) + |S^* \cap C_i| \geq E(S^*) + 1$, οδηγούμαστε σε αντίφαση αφού η S^* είναι βέλτιστη.

Για να ολοκληρώσουμε την απόδειξη αρκεί να αποδείξουμε ότι δεν είναι δυνατό η κλίκα C_0' και μία εξωτερική κλίκα C_j να είναι συγχρόνως μερικώς στην S^*. Αυτό όμως προκύπτει χρησιμοποιώντας το προηγούμενο επιχείρημα, αλλά θεωρώντας την C_0 αντί για την C_i και ότι η κορυφή $x \notin S^*$ είναι κορυφή της C_0'. □

Πρόταση 2.

(i) Εάν η C_0 είναι η μεγαλύτερου μεγέθους κλίκα, δηλαδή, $|C_0| > |C_i|$, $1 \leq i \leq m-1$, τότε η C_0 ανήκει πλήρως σε κάθε βέλτιστη λύση.

(ii) Εάν η C_0 είναι μερικώς σε μία βέλτιστη λύση S^*, τότε $|C_0| \leq |C_i|$ για κάθε κλίκα C_i που συμμετέχει στην S^*.

Απόδειξη.

(i) Υποθέτουμε ότι η S^* δεν περιέχει $q > 0$ κορυφές της C_0 και θεωρούμε μία λύση S που προκύπτει από την S^* αντικαθιστώντας q κορυφές εξωτερικών κλικών που δεν ανήκουν στην C_0 με τις q κορυφές της C_0 που δεν ανήκουν στην S^*. Συμβολίζουμε με E^- και E^+ τον αριθμό των ακμών που αφαιρέθηκαν και προστέθηκαν αντίστοιχα στο $E(S^*)$ λόγω της παραπάνω αντικατάστασης. Τότε, $E(S) = E(S^*) - E^- + E^+$. Το E^- ισούται με τον αριθμό των ακμών που οι q κορυφές των εξωτερικών κλικών συνεισφέρουν στο $E(S^*)$. Αυτός ο αριθμός, ακόμα και αν όλες οι q κορυφές ανήκουν στην ίδια εξωτερική κλίκα, είναι αυστηρά μικρότερος από $\binom{q}{2} + (|C_0| - q)q$. Από την άλλη πλευρά, το E^+ ισούται με τον αριθμό των ακμών που οι q κορυφές της C_0, που δεν ανήκουν στην S^*, θα συνεισφέρουν στο $E(S)$. Αυτός ο αριθμός ισούται με $\binom{q}{2} + (|C_0| - q)q$. Άρα λόγω των παραπάνω, $E(S) > E(S^*)$, το οποίο είναι αντίφαση αφού η S^* είναι βέλτιστη.

(ii) Υποθέτουμε ότι υπάρχει μία εξωτερική κλίκα C_i στην S^* τέτοια ώστε $|C_0| > |C_i|$ και ότι η S^* δεν περιέχει $q > 0$ κορυφές της κλίκας C_0. Παρατηρούμε ότι $|C_0| - q > a_i$, αφού εάν $|C_0| - q \leq a_i$ τότε καμία άλλη εξωτερική κλίκα δεν συμμετέχει στη S^*, δηλαδή, η S^* είναι μέρος μίας μοναδικής κλίκας (είτε της C_i είτε της C_0).

Θεωρούμε τη λύση S που προκύπτει από την S^* αντικαθιστώντας τις κορυφές της $S^* \cap C_i$ που δεν ανήκουν στη C_0 με κορυφές της C_0 που δεν ανήκουν στην S^*. Έστω $b_i' = |S^* \cap (C_i \setminus C_0)|$, $0 < b_i' \leq b_i$. Χρησιμοποιώντας τα E^- και E^+ όπως στο (i) έχουμε ότι $E(S) = E(S^*) - E^- + E^+$. Διαχωρίζουμε δύο περιπτώσεις με βάση τις τιμές των q και b'_i.

Εάν $q \geq b'_i$ τότε το E^- ισούται με τον αριθμό των ακμών που οι b'_i κορυφές της εξωτερικής κλίκας C_i συνεισφέρουν στο $E(S^*)$ ενώ το E^+ ισούται με τον αριθμό των ακμών που οι b'_i κορυφές της C_0 που δεν ανήκουν στην S^* θα συνεισφέρουν στο $E(S)$. Τότε, $E(S) = E(S^*) - E^- + E^+ = E(S^*) - (\binom{b'_i}{2} + b'_i a_i) + (\binom{b'_i}{2} + b'_i(|C_0| - q)) = E(S^*) + b'_i((|C_0| - q) - a_i) > E(S^*)$, το οποίο είναι αντίφαση αφού η S^* είναι βέλτιστη.

Εάν $q < b'_i$ τότε το E^- ισούται με τον αριθμό των ακμών που οι q κορυφές της εξωτερικής κλίκας C_i συνεισφέρουν στο $E(S^*)$ ενώ το E^+ ισούται με τον αριθμό των ακμών που οι q κορυφές της C_0 που δεν ανήκουν στην S^* θα συνεισφέρουν στο $E(S)$. Τότε, $E(S) = E(S^*) - E^- + E^+ = E(S^*) - (\binom{q}{2} + q(a_i + b'_i - q)) + (\binom{q}{2} + q(|C_0| - q)) = E(S^*) + q(|C_0| - (a_i + b'_i)) \geq E(S^*) + q(|C_0| - |C_i|) > E(S^*)$, το οποίο είναι αντίφαση αφού η S^* είναι βέλτιστη. □

Παρά τις χρήσιμες δομικές ιδιότητες της βέλτιστης λύσης στις Προτάσεις 1 και 2 πολλά άπληστα (*greedy*) κριτήρια, που βασίζονται είτε στο μέγεθος των κλικών είτε στο μέγεθος των τομών είτε και στα δύο, αποτυγχάνουν να δώσουν μία βέλτιστη λύση. Παρακάτω δίνουμε έναν πολυωνυμικού χρόνου αλγόριθμο δυναμικού προγραμματισμού για την περίπτωση που η κεντρική κλίκα είναι πλήρως στη βέλτιστη λύση και ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα για τη γενική περίπτωση.

Η κλίκα C_0 ανήκει πλήρως στη βέλτιστη λύση.

Λήμμα 6. *Εάν η κλίκα C_0 είναι πλήρως στη βέλτιστη λύση, τότε υπάρχει ένας $O(nk^2)$ χρόνου αλγόριθμος δυναμικού προγραμματισμού για το πρόβλημα του πυκνότερου k-υπογράφου σε ένα αστέρι κλικών.*

Απόδειξη. Εφόσον η κλίκα C_0 είναι πλήρως στη βέλτιστη λύση έχουμε μόνο να επιλέξουμε $k' = k - |C_0|$ κορυφές από τις εξωτερικές κλίκες. Εάν επιλέξουμε q κορυφές από μία εξωτερική κλίκα C_j, τότε αυτές συνεισφέρουν $q \cdot a_j + \binom{q}{2}$ ακμές στη λύση.

Έστω $f(i, j)$ ο μέγιστος αριθμός ακμών σε μία λύση που έχει επιλέξει i κορυφές από τις j πρώτες εξωτερικές κλίκες (υπενθυμίζουμε ότι υπάρχουν $m - 1$ εξωτερικές κλίκες). Επομένως για $i = 0, 1, \ldots, k'$ και $j = 2, 3, \ldots, m - 1$

$$f(i, j) = \max_{0 \leq q \leq \min\{i, b_j\}} \{f(i - q, j - 1) + q \cdot a_j + \binom{q}{2}\}.$$

Για $j = 1$ οι παρακάτω οριακές συνθήκες ισχύουν για $0 \leq i \leq \min\{k', b_1\}$

$$f(i,1) = \begin{cases} \binom{i}{2} + i \cdot a_1, & \text{αν } \; i \leq \min\{k', b_1\} \\ -\infty, & \text{αλλιώς.} \end{cases}$$

Η πολυπλοκότητα αυτού του αλγόριθμου δυναμικού προγραμματισμού είναι $O(nk^2)$. Ο υπολογισμός της τιμής μίας $f(i,j)$ θέλει $O(k)$ χρόνο λόγω των πιθανών τιμών του q ($0 \leq q \leq \min\{i, b_j\} \leq k' < k$) και κάθε $f(i,j)$ υπολογίζεται για κάθε $i \leq k' < k$ και κάθε $j \leq m - 1 < n$. Η βέλτιστη λύση για το DkS πρόβλημα είναι $f(k', m-1) + \binom{|C_0|}{2}$. □

Παρατηρούμε ότι εάν η κλίκα C_0 είναι η μεγαλύτερη κλίκα τότε, από την Πρόταση 2(i), η C_0 ανήκει πλήρως σε κάθε βέλτιστη λύση και εφαρμόζεται ο παραπάνω αλγόριθμος δυναμικού προγραμματισμού.

Γενική περίπτωση.
Στη γενική περίπτωση, η κλίκα C_0 είναι μερικώς στη βέλτιστη λύση και από την Πρόταση 2(i) γνωρίζουμε ότι υπάρχουν εξωτερικές κλίκες μεγαλύτερες από τη C_0. Έστω c ο αριθμός αυτών των κλικών με μέγεθος τουλάχιστον $|C_0|$. Επιπλέον, από την Πρόταση 2(ii), γνωρίζουμε ότι οι κλίκες που συμμετέχουν στη βέλτιστη λύση είναι κάποιες από αυτές τις c κλίκες. Η επόμενη πρόταση δίνει ένα (μη αυστηρό) άνω φράγμα για τον αριθμό c.

Πρόταση 3. *Εάν η κλίκα C_0 είναι μερικώς στη βέλτιστη λύση, τότε ο αριθμός των εξωτερικών κλικών μεγέθους τουλάχιστον $|C_0|$ είναι το πολύ $\sqrt{n}$.*

Απόδειξη. Ο αριθμός, c, των εξωτερικών κλικών είναι μικρότερος ή ίσος από το $|C_0|$, αφού $C_i \cap C_0 \neq \emptyset$ και $C_j \cap C_i = \emptyset$, $1 \leq i \neq j \leq m-1$. Επομένως, εάν $|C_0| \leq \sqrt{n}$, τότε $c \leq \sqrt{n}$. Αλλιώς εάν $|C_0| > \sqrt{n}$ τότε, ο συνολικός αριθμός κορυφών σε αυτές τις c κλίκες είναι τουλάχιστον $c \times \sqrt{n}$ και το πολύ n. Κατά συνέπεια, $c \leq \sqrt{n}$. □

Για να αποδείξουμε ότι υπάρχει πολυωνυμικού χρόνου προσεγγιστικό σχήμα επιχειρηματολογούμε περαιτέρω στον αριθμό των εξωτερικών κλικών μεγέθους τουλάχιστον $|C_0|$. Ορίζουμε ως $r = \lfloor \frac{k}{|C_0|} \rfloor$. Ο αριθμός των εξωτερικών κλικών μεγέθους τουλάχιστον $|C_0|$ που μπορεί να εμπλέκονται σε μία βέλτιστη λύση είναι το πολύ r.

Έστω δ ένας προκαθορισμένος αριθμός τον οποίο θα ορίσουμε αργότερα. Συγκρίνοντας τον αριθμό r με τον αριθμό δ διαχωρίζουμε δύο περιπτώσεις.

Περίπτωση 1: $r < \delta$

Εάν ο αριθμός r είναι "μικρός", τότε προχωράμε με έναν εξαντλητικό τρόπο. Εξετάζουμε όλα τα πιθανά σύνολα r κλικών από τις c κλίκες μεγέθους τουλάχιστον $|C_0|$, δηλαδή, $\binom{c}{r}$ σύνολα κλικών. Μία χρήσιμη τεχνική λεπτομέρεια είναι ότι η κλίκα C_0' πρέπει να θεωρείται σαν μία από τις c κλίκες. Αυτό μπορεί να επιτευχθεί θεωρώντας την κλίκα C_0' σαν μία εξωτερική κλίκα με μηδέν κορυφές εκτός της κλίκας C_0.

Από την Πρόταση 3 προκύπτει ότι ο αριθμός όλων των $\binom{c}{r}$ συνόλων κλικών είναι $O(n^{\frac{r}{2}})$. Για καθένα από αυτά τα σύνολα r κλικών υπολογίζουμε τις k κορυφές που μεγιστοποιούν τον αριθμό των ακμών ως εξής:

Έστω R ένα σύνολο r κλικών. Από την Πρόταση 1 το πολύ μία από τις κλίκες του R είναι μερικώς στην S^*. Θεωρούμε όλα τα $2^r - 1$ υποσύνολα του R. Έστω R_i ένα από αυτά τα υποσύνολα και έστω C_i^j η j-οστή, $1 \leq j \leq |R_i|$, κλίκα του συνόλου R_i. Εάν $\sum_{j=1}^{|R_i|} |C_i^j| < k$, απορρίπτουμε το σύνολο R_i. Αλλιώς, έστω $k(j) = \sum_{t=1, t\neq j}^{|R_i|} |C_i^t|$, για κάθε $j = 1, 2, \ldots, |R_i|$. Εάν $k(j) > k$ τότε απορρίπτουμε αυτό το j. Αλλιώς (εάν $k(j) \leq k$) έχουμε μία k-κορυφών λύση παίρνοντας $k - k(j)$ κορυφές από την κλίκα C_i^j, ξεκινώντας από τις κορυφές που ανήκουν στην τομή της με την κλίκα C_0.

Θεωρούμε τώρα όλες τις λύσεις που κατασκευάστηκαν για κάθε $j = 1, 2, \ldots, |R_i|$, και κάθε $R_i \subseteq R$. Από κατασκευής, αυτές οι λύσεις είναι όλες οι πιθανές k-κορυφών λύσεις για το σύνολο R των κλικών, υπό τον περιορισμό ότι το πολύ μία από αυτές είναι μερικώς στη λύση. Συνεπώς, για να βρούμε τη βέλτιστη λύση απλά πρέπει να διαλέξουμε αυτή με το μέγιστο αριθμό ακμών.

Για ένα σύνολο R, r κλικών, υπάρχουν $2^r - 1$ υποσύνολα R_i, και για κάθε υποσύνολο υπάρχουν r το πολύ πιθανές λύσεις. Επομένως, ο αριθμός των λύσεων που πρέπει να συγκρίνουμε είναι $O(r\,2^r)$.

Υπενθυμίζοντας ότι πρέπει να εξετάσουμε $O(n^{\frac{r}{2}})$ σύνολα r κλικών το επόμενο λήμμα έπεται.

Λήμμα 7. *Για την περίπτωση που το $r < \delta$, όπου δ μία προκαθορισμένη μεταβλητή, μία βέλτιστη λύση για το πρόβλημα του πυκνότερου k-υπογράφου σε ένα αστέρι κλικών μπορεί να βρεδεί σε χρόνο $O(r\,2^r\,n^{\frac{r}{2}})$.*

Περίπτωση 2: $r \geq \delta$

Εάν ο αριθμός r είναι "μεγάλος", τότε προχωράμε με έναν άπληστο τρόπο. Θεωρούμε τη λύση, S, που προκύπτει από τον επόμενο απλό αλγόριθμο:
Έστω $C_1 \geq C_2 \geq \ldots \geq C_{m-1}$ και t ο μεγαλύτερος ακέραιος αριθμός τέτοιος ώστε $k \geq \sum_{i=1}^{t} |C_i| = k'$.
Ο αλγόριθμος επιστρέφει όλες τις κορυφές των κλικών $C_1 \geq C_2 \geq \ldots \geq C_t$ και $k-k'$ κορυφές της κλίκας C_{t+1}.

Η πρόταση που ακολουθεί για την περίπτωση ανεξάρτητων κλικών είναι χρήσιμη για να φράξουμε τη λύση που επιστρέφει ο άπληστος αλγόριθμος.

Πρόταση 4. *Έστω R_1 και R_2 δύο σύνολα ανεξάρτητων κλικών με όλες τις κλίκες στο R_1 μεγέθους τουλάχιστον L και όλες τις κλίκες στο R_2 μεγέθους ακριβώς L. Για κάθε ζεύγος συνόλων k κορυφών S_1 και S_2 από τα R_1 και R_2 αντίστοιχα, τέτοιο που και στα δύο σύνολα το πολύ μία κλίκα ανήκει μερικώς σε αυτά, ισχύει ότι $E(S_1) \geq E(S_2)$.*

Απόδειξη. Μετατρέπουμε το S_1 σε ένα ισοδύναμο του συνόλου S_2 ως εξής. Καταρχάς, αφαιρούμε από κάθε κλίκα του S_1 κάποιες κορυφές έτσι ώστε κάθε κλίκα στο S_1 να έχει τώρα μέγεθος ακριβώς L. Έστω k' ο αριθμός των κορυφών που αφαιρέθηκαν. Αντικαθιστούμε τις k' κορυφές με $\lceil \frac{k'}{L} \rceil$ κλίκες, όλες εκτός μίας, μεγέθους ακριβώς L. Όλες οι κορυφές που αφαιρέθηκαν έχουν τώρα βαθμό το πολύ $L-1$ ενώ στο S_1 είχαν βαθμό τουλάχιστον $L-1$. Κατά συνέπεια, $E(S_1) \geq E(S_2)$. □

Θεωρούμε τώρα τη λύση S που προέκυψε από τον αλγόριθμο μας. Από την Πρόταση 2(ii), η βέλτιστη λύση S^* περιέχει εξωτερικές κλίκες μεγέθους τουλάχιστον $|C_0|$. Εφόσον ο αλγόριθμος μας βρίσκει μία λύση S επιλέγοντας k κορυφές από τις μεγαλύτερες κλίκες, προκύπτει ότι όλες οι κλίκες της S είναι μεγέθους τουλάχιστον $|C_0|$.

Επιπλέον, αφού $r = \lfloor \frac{k}{|C_0|} \rfloor$, χρειαζόμαστε τουλάχιστον r κλίκες μεγέθους $|C_0|$ για να επιτύχουμε το k. Επομένως, η επιλογή k κορυφών από ένα σύνολο ανεξάρτητων κλικών μεγέθους $|C_0|$ οδηγεί σε τουλάχιστον $rE(C_0)$ κορυφές. Άρα, από την Πρόταση 4, προκύπτει ότι $E(S) \geq rE(C_0)$.

Προφανώς, μία βέλτιστη λύση, S^*, μπορεί να περιέχει κλίκες μικρότερου μεγέθους από αυτές που επέλεξε ο δικός μας αλγόριθμος. Αυτές οι μικρές κλίκες επιλέχθηκαν από την S^* λόγω των ακμών μεταξύ των τομών τους με την κλίκα C_0. Αφού αυτές οι ακμές ανήκουν και στην κλίκα C_0, η βέλτιστη λύση δεν μπορεί να είναι μεγαλύτερη από το $E(S)$ συν τις ακμές της C_0, δηλαδή,

$$E(S^*) \leq E(S) + E(C_0) \leq E(S) + \frac{E(S)}{r} \leq E(S) + \frac{E(S)}{\delta} = E(S)\frac{\delta+1}{\delta}$$

και το επόμενο λήμμα έπεται.

Λήμμα 8. *Για την περίπτωση που το $r \geq \delta$, όπου $\delta = \frac{1-\epsilon}{\epsilon}$, $0 < \epsilon < 1$, υπάρχει ένας $(1-\epsilon)$-προσεγγιστικός αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε ένα αστέρι κλικών.*

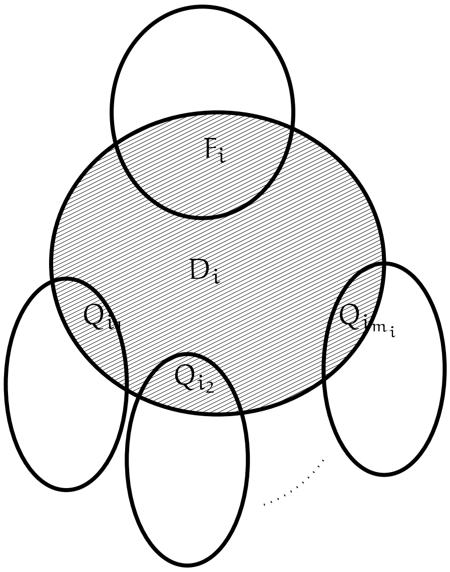
Η πολυπλοκότητα του άπληστου προσεγγιστικού αλγόριθμου του Λήμματος 8 είναι $O(n \log n)$. Η πολυπλοκότητα του εξαντλητικού βέλτιστου αλγόριθμου του Λήμματος 7 είναι εκθετική στο $r \leq \delta = \frac{1-\epsilon}{\epsilon}$, δηλαδή, εκθετική στο $\frac{1}{\epsilon}$. Συνεπώς προκύπτει ότι

Θεώρημα 19. *[27, 28] Υπάρχει πολυωνυμικού χρόνου προσεγγιστικό σχήμα για το πρόβλημα του πυκνότερου k-υπογράφου σε αστέρια κλικών.*

4.3 Πολυωνυμικός αλγόριθμος σε δέντρο κλικών φραγμένου βαθμού

Σε αυτή την ενότητα παρουσιάζουμε ένα δυναμικού προγραμματισμού αλγόριθμο ο οποίος βρίσκει μία βέλτιστη λύση για το πρόβλημα του πυκνότερου k-υπογράφου σε γράφους που έχουν ως γράφο μεγιστικών κλικών δέντρο. Έστω $C_1, C_2, ..., C_t$ οι κλίκες ενός τέτοιου δέντρου και m ο μέγιστος βαθμός του. Θεωρούμε ότι $|C_i| < k$, $i = 1, \ldots, t$, αλλιώς το πρόβλημα είναι τετριμμένο.

Θεωρούμε ότι το δέντρο έχει ως ρίζα μία κλίκα φύλλο, έστω την κλίκα C_t. Με αυτόν τον τρόπο η κλίκα ρίζα C_t έχει τουλάχιστον μία κορυφή εκτός των τομών της με τις κλίκες παιδιά της. Έστω C_i μία εσωτερική κλίκα με $m_i \geq 1$ παιδιά, $C_{i_1}, \ldots, C_{i_{m_i}}$. Συμβολίζουμε με Q_h την τομή της C_i με τη h-οστή κλίκα παιδί της, C_h, για $h = i_1, \ldots, i_{m_i}$, δηλαδή, $Q_h = C_i \cap C_h$. Συμβολίζουμε με F_i την τομή της κλίκας C_i με την κλίκα πατέρα της, C_f, στο δέντρο, δηλαδή, $F_i = C_i \cap C_f$ και με D_i τις κορυφές της κλίκας C_i που δεν ανήκουν σε καμία τομή, δηλαδή,



Σχήμα 4.1: Εσωτερική κλίκα C_i με m_i παιδιά.

$D_i = C_i - F_i - \bigcup_{h=i_1}^{i_{m_i}} Q_h$. Θεωρούμε, χωρίς βλάβη της γενικότητας, ότι η F_t, για την κλίκα ρίζα C_t, αποτελείται από μία και μοναδική κορυφή, δηλαδή, $|F_t| = 1$.

Ο αλγόριθμος διασχίζει το δέντρο των κλικών ξεκινώντας από τις κλίκες φύλλα. Σε κάθε βήμα υπολογίζει μία βέλτιστη λύση για όλα τα j-κορυφών πυκνότερου υπογράφου (DjS) προβλήματα, για $j = 1, \ldots, k$, στο υποδέντρο που έχει ρίζα την κλίκα C_i.

Συμβολίζουμε με $f_i(j)$ την τιμή της βέλτιστης λύσης του DjS προβλήματος στο υποδέντρο που έχει ως ρίζα την κλίκα C_i. Με $f_i(j, a)$ συμβολίζουμε την τιμή μίας βέλτιστης λύσης του DjS προβλήματος σε ένα υποδέντρο που έχει ρίζα την κλίκα C_i και περιλαμβάνει ακριβώς a κορυφές από την κλίκα F_i. Είναι προφανές ότι,

$$f_i(j) = \max_{0 \le a \le |F_i|} \{f_i(j, a)\}.$$

Για να υπολογίσουμε την τιμή της $f_i(j, a)$ για μία εσωτερική κλίκα C_i θεωρούμε τις κλίκες παιδιά της $C_{i_1}, \ldots, C_{i_{m_i}}$, $m_i \geq 1$. Έστω $f_h(j_h, a_h)$ η τιμή της βέλτιστης λύσης του j_h-κορυφών πυκνότερου υπογράφου, για $j_h = 1, 2, ..., k$, στο υποδέντρο που έχει ως ρίζα την κλίκα C_h, $h = i_1, \ldots, i_{m_i}$, χρησιμοποιώντας a_h κορυφές της F_h. Παρατηρούμε ότι για την τομή της κλίκας C_i με το παιδί της C_h ισχύει ότι $Q_h = C_i \cap C_h = F_h$. Υπολογίζουμε την τιμή της $f_i(j, a)$, ως εξής:

Εάν $a = 0$, τότε καμία κορυφή της $F_i \cup D_i$ δεν ανήκει στη βέλτιστη λύση του αντίστοιχου DjS προβλήματος στο υποδέντρο που έχει ως ρίζα την κλίκα C_i. Συνεπώς, η τιμή αυτής της λύσης είναι η ίδια με την τιμή της βέλτιστης λύσης του DjS προβλήματος στον υπογράφο που είναι η ένωση των υποδέντρων που έχουν ως ρίζες τα παιδιά της κλίκας C_i συν τις ακμές μεταξύ των κορυφών στα Q_h, δηλαδή,

$$f_i(j,a) = \max_{\sum_{h=i_1}^{i_{m_i}} j_h = j} \{\sum_{h=i_1}^{i_{m_i}} f_h(j_h, a_h) + \sum_{\substack{i,j=i_1 \\ i\neq j}}^{i_{m_i}} \frac{a_i \cdot a_j}{2}\}.$$

Εάν $a > 0$, τότε η βέλτιστη λύση του DjS προβλήματος που περιλαμβάνει a κορυφές της F_i, περιλαμβάνει επίσης $b \geq 0$ κορυφές της κλίκας D_i, και $a_h \geq 1$ κορυφές[1] από κάθε Q_h, $h = i_1, \ldots, i_{m_i}$.

$$f_i(j,a) = \begin{cases} \binom{j}{2}, & \text{αν } j \leq \sum_{h=i_1}^{i_{m_i}} |Q_h| + |D_i| + a \\ \max_{a+b+\sum_{h=i_1}^{i_{m_i}} j_h = j} \{\sum_{h=i_1}^{i_{m_i}} f_h(j_h, a_h) + \binom{a+b}{2} + (a+b)\sum_{h=i_1}^{i_{m_i}} a_h + \\ \qquad \sum_{\substack{i,j=i_1 \\ i\neq j}}^{i_{m_i}} \frac{a_i \cdot a_j}{2}\}, & \text{αλλιώς.} \end{cases}$$

Για όλες τις κλίκες C_i που είναι φύλλα στο δέντρο οι παρακάτω αρχικές συνθήκες ισχύουν για $1 \leq j \leq k$:

Εάν $a = 0$, τότε $f_i(j,a) = 0$.

Εάν $1 \leq a \leq |F_i|$, τότε

$$f_i(j,a) = \begin{cases} \binom{j}{2}, & \text{αν } j \leq |D_i| + a \\ -\infty, & \text{αλλιώς.} \end{cases}$$

Ο αλγόριθμος τερματίζει υπολογίζοντας την τιμή της $f_t(k)$ για την κλίκα ρίζα C_t. Υπενθυμίζουμε ότι θεωρούμε $|F_t| = 1$ και επομένως η βέλτιστη λύση για το k-κορυφών πυκνότερου υπογράφου πρόβλημα είναι $f_t(k) = \max_{a=0,1}\{f_t(k,a)\}$.

Ο υπολογισμός κάθε τιμής $f_i(j)$ για μία κλίκα C_i με m_i παιδιά χρειάζεται $O(k^{m_i+1})$ χρόνο, λόγω των συνδυασμών των a, b και $\sum_{h=i_1}^{i_{m_i}} j_h$, έτσι ώστε $a + b +$

[1] ή $a_h \geq 0$ για μη συνεκτική λύση.

$\sum_{h=i_1}^{i_{m_i}} j_h = j$. Ο αλγόριθμος υπολογίζει την τιμή $f_i(j)$ για κάθε $j = 1, 2, \ldots, k$ και κάθε $i = 1, 2, \ldots, t$. Στη χειρότερη περίπτωση το t είναι $O(n)$ και $\max_i\{m_i\} = m - 1$ και το επόμενο θεώρημα έπεται:

Θεώρημα 20. *[27, 28] Υπάρχει ένας $O(nk^{m+1})$ χρόνου αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε δέντρο κλικών μέγιστου βαθμού m.*

Το επόμενο πόρισμα προκύπτει άμεσα από το Θεώρημα 20.

Πόρισμα 3. *Υπάρχει ένας $O(nk^3)$ χρόνου βέλτιστος αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε ένα μονοπάτι κλικών.*

Κεφάλαιο 5

Χορδικοί γράφοι

Αιώνες τώρα ρωτούν οι μάγοι μα οι αστέρες
αποκρίνονται κατά προσέγγιση.

Οδυσσεας Ελυτης

5.1 Ορισμοί - Ιδιότητες χορδικών γράφων - Συμβολισμός

Κλίκα (*clique*) ενός γράφου, $G = (V, E)$, είναι ένας πλήρης υπογράφος του G επαγόμενος από ένα υποσύνολο των κορυφών, $C \subseteq V$. Το μέγεθος $|C|$ της κλίκας είναι ο αριθμός των κορυφών της. Μεγιστική (*maximal*) είναι μία κλίκα όταν δεν περιέχεται σε μεγαλύτερη κλίκα. Η μεγαλύτερη μεγιστική κλίκα ονομάζεται μέγιστη κλίκα (*maximum clique*). Μία κορυφή του γράφου καλείται *simplicial* όταν οι γειτονικές κορυφές της επάγουν έναν πλήρη υπογράφο.

Χορδικός (*chordal*) είναι ο γράφος στον οποίο κάθε κύκλος με μέγεθος μεγαλύτερο του τρία έχει τουλάχιστον μία χορδή, δηλαδή μία ακμή που ενώνει δύο μη διαδοχικές κορυφές του κύκλου. Σε ότι ακολουθεί, $G = (V, E)$ είναι ένας χορδικός γράφος. Για έναν χορδικό γράφο, $G = (V, E)$, ισχύουν τα παρακάτω:

(i) Ο G έχει τουλάχιστον $m \leq |V|$ μεγιστικές κλίκες, $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$, τις οποίες μπορούμε να βρούμε σε πολυωνυμικό χρόνο [15].

(ii) Ο G έχει μία *simplicial* κορυφή. Στην πραγματικότητα, εάν ο G δεν είναι κλίκα, τότε έχει δύο μη γειτονικές *simplicial* κορυφές [8].

Με G_A συμβολίζουμε τον υπογράφο του G που επάγεται από ένα υποσύνολο $A \subseteq V$ κορυφών του (*vertex-induced*) και με G^F συμβολίζουμε τον υπογράφο του G που επάγεται από ένα υποσύνολο $F \subseteq E$ ακμών του (*edge-induced*). Άμεση συνέπεια του ορισμού της κλάσης των χορδικών γράφων είναι το γεγονός ότι η χορδικότητα είναι μία ιδιότητα που κληρονομείται από κάθε υπογράφο G_A επαγόμενο από τις κορυφές του G, αλλά όχι από κάθε υπογράφο G^F επαγόμενο από τις ακμές του G. Είναι επίσης προφανές ότι για κάθε μεγιστική κλίκα C_i ενός υπογράφου επαγόμενου είτε από τις κορυφές είτε από τις ακμές του G, υπάρχει τουλάχιστον μία μεγιστική κλίκα C_j του G τέτοια ώστε $C_i \subseteq C_j$.

Με $E(A)$ συμβολίζουμε το σύνολο των ακμών ενός υπογράφου G_A του G, ενώ με $E(A, B)$ συμβολίζουμε το σύνολο των ακμών μεταξύ δύο ξένων υποσυνόλων $A, B \subseteq V$ κορυφών του G, δηλαδή το σύνολο των ακμών με το ένα άκρο τους στο A και το άλλο στο B.

Με S συμβολίζουμε μία λύση για το πρόβλημα του πυκνότερου k-υπογράφου, δηλαδή ένα υποσύνολο $S \subseteq V$ κορυφών του G τέτοιο ώστε $|S| = k$, ενώ με S^* συμβολίζουμε τη βέλτιστη λύση, δηλαδή τη λύση S για την οποία το $|E(S)|$ μεγιστοποιείται.

Τέλος, υποθέτουμε ότι το $k > |C_i|$, $1 \leq i \leq m$, γιατί αλλιώς το S^* αποτελείται από οποιοδήποτε υποσύνολο k κορυφών κάποιας κλίκας για την οποία ισχύει $|C_i| \geq k$.

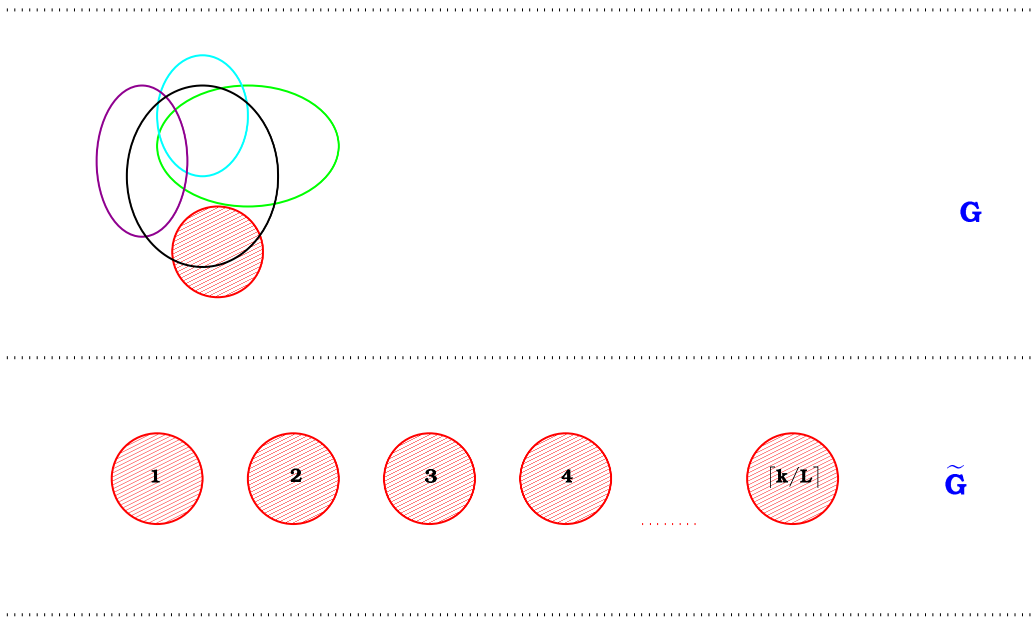
5.2 Αλγόριθμος - Ανάλυση

Αφού όλες οι μεγιστικές κλίκες ενός χορδικού γράφου $G = (V, E)$ μπορούν να βρεθούν σε πολυωνυμικό χρόνο είναι φυσικό επόμενο να μελετάμε το πρόβλημα του πυκνότερου k-υπογράφου σε αυτές τις μεγιστικές κλίκες αντί στον ίδιο το γράφο G. Σε αυτή την ενότητα αναλύουμε τον παρακάτω άπληστο (*greedy*) αλγόριθμο για την εύρεση προσεγγιστικής λύσης του προβλήματος του πυκνότερου k-υπογράφου σε ένα χορδικό γράφο G.

1. Έστω $C_1, C_2, \ldots C_m$ οι μεγιστικές κλίκες του G, διατεταγμένες σε μη αύξουσα σειρά των μεγεθών τους.
2. Βρες τον μεγαλύτερο ακέραιο t τέτοιον ώστε $k > |\bigcup_{i=1}^{t-1} C_i| = k'$.
3. Επίστρεψε τη λύση S που αποτελείται από όλες τις κορυφές των $C_1, C_2, \ldots, C_{t-1}$ και $k - k' > 0$ κορυφές της κλίκας C_t.

Λόγω του ότι το μέγεθος της μεγιστικής κλίκας C_t παίζει σπουδαίο ρόλο στην ανάλυση μας θα την συμβολίσουμε με $L = |C_t|$.

Καταρχάς αποδεικνύουμε ένα κάτω φράγμα για τον αριθμό των ακμών $|E(S)|$ της λύσης S που επιστρέφει ο άπληστος αλγόριθμος για τον χορδικό γράφο G. Αυτό το φράγμα αποδεικνύεται συσχετίζοντας τη λύση S με τη λύση που ο άπληστος αλγόριθμος επιστρέφει σε ένα γράφο που αποτελείται από ανεξάρτητες κλίκες μεγέθους L (Σχήμα 5.1). Για ένα χορδικό γράφο G και για την παράμετρο L, όπως αυτή ορίστηκε, θεωρούμε το χορδικό γράφο $\widetilde{G}$ που αποτελείται από τουλάχιστον $\lceil \frac{k}{L} \rceil$ ανεξάρτητες κλίκες μεγέθους L.



Σχήμα 5.1: Γράφοι G και $\widetilde{G}$.

Λήμμα 9. *Έστω S και $\widetilde{S}$ οι λύσεις που ο άπληστος αλγόριθμος επιστρέφει για το πρόβλημα του πυκνότερου k-υπογράφου στους γράφους G και $\widetilde{G}$, αντίστοιχα. Ισχύει ότι* $|E(S)| \geq |E(\widetilde{S})| = \dfrac{k(L-1) - b(L-b)}{2}$, *όπου* $b = k \bmod L$.

Απόδειξη. Για να αποδείξουμε το φράγμα του λήμματος θεωρούμε τις λύσεις S' και $\widetilde{S}'$ που ο άπληστος αλγόριθμος επιστρέφει για το πρόβλημα του πυκνότερου k'-υπογράφου στους γράφους G και $\widetilde{G}$ αντίστοιχα. Θυμίζουμε ότι $k' = |\bigcup_{i=1}^{t-1} C_i|$.

Καταρχάς θεωρούμε τις λύσεις S και S'. Η λύση S αποτελείται από τις $t-1$ μεγαλύτερες μεγιστικές κλίκες του G, k' κορυφών, και ένα σύνολο $A \subseteq C_t$, $a = k - k'$

κορυφών, τέτοιο ώστε $\bigcup_{i=1}^{t-1} C_i \cap A = \emptyset$. Προφανώς, $0 < a \leq |C_t| = L$. Επιπλέον, $S \setminus S' = A$ και συνεπώς,

$$|E(S)| \geq |E(S')| + \frac{a(a-1)}{2}.$$

Έπειτα θεωρούμε τις λύσεις S' και $\widetilde{S}'$. Η λύση S' αποτελείται από τις $t-1$ μεγαλύτερες μεγιστικές κλίκες του G, k' κορυφών. Αφού όλες οι μεγιστικές κλίκες της S' είναι μεγέθους τουλάχιστον L, προκύπτει ότι όλες οι κορυφές της S' έχουν βαθμό τουλάχιστον $L-1$. Η λύση $\widetilde{S}'$ αποτελείται από $q' = \lfloor k'/L \rfloor$ ανεξάρτητες κλίκες μεγέθους L και μία ακόμη ανεξάρτητη κλίκα, B', μεγέθους $b' = k' \bmod L$, δηλαδή, $|\widetilde{S}'| = q'L + b'$. Καθεμία από τις b' κορυφές στην κλίκα B' της $\widetilde{S}'$ έχει βαθμό $b'-1$. Η λύση S' περιέχει τουλάχιστον $b'(L-1-(b'-1)) = b'(L-b')$ παραπάνω ακμές από τη λύση $\widetilde{S}'$, δηλαδή, $|E(S')| \geq |E(\widetilde{S}')| + b'(L-b')$. Άρα,

$$|E(S)| \geq |E(\widetilde{S}')| + b'(L-b') + \frac{a(a-1)}{2}.$$

Τελικά θεωρούμε τις λύσεις $\widetilde{S}'$ και $\widetilde{S}$. Η λύση $\widetilde{S}$ αποτελείται από $q = \lfloor k/L \rfloor$ ανεξάρτητες κλίκες μεγέθους L και μία ακόμα ανεξάρτητη κλίκα, B, μεγέθους $b = k \bmod L$, δηλαδή, $|\widetilde{S}| = qL + b$. Επιπλέον, $|\widetilde{S}| - |\widetilde{S}'| = a$.

Εάν $a \leq b$, τότε η λύση $\widetilde{S}'$ προκύπτει από τη λύση $\widetilde{S}$, αφαιρώντας a κορυφές από την κλίκα B της $\widetilde{S}$. Σε αυτή την περίπτωση $|\widetilde{S}'| = qL + b'$, όπου $b' = b - a$, και $|E(\widetilde{S})| = |E(\widetilde{S}')| + \frac{a(a-1)}{2} + b'a$. Άρα,

$$|E(S)| \geq |E(\widetilde{S})| - b'a + b'(L-b') = |E(\widetilde{S})| + b'(L-b),$$

και αφού $L > b$, η ανισότητα του λήμματος προκύπτει.

Εάν $a > b$, τότε η λύση $\widetilde{S}'$ προκύπτει από τη λύση $\widetilde{S}$, αφαιρώντας όλες τις b κορυφές της κλίκας B και $x = a - b$ κορυφές της q-οστής κλίκας μεγέθους L της $\widetilde{S}$. Σε αυτή την περίπτωση $|\widetilde{S}'| = (q-1)L + b'$, όπου $a - b = L - b'$, και $|E(\widetilde{S})| = |E(\widetilde{S}')| + \frac{b(b-1)}{2} + \frac{x(x-1)}{2} + xb' = |E(\widetilde{S}')| + \frac{a(a-1)}{2} - xb + xb' = |E(\widetilde{S}')| + \frac{a(a-1)}{2} + (a-b)(b'-b)$. Άρα,

$$|E(S)| \geq |E(\widetilde{S})| - (a-b)(b'-b) + b'(L-b') = |E(\widetilde{S})| + (L-b')b,$$

και αφού $L > b'$ η ανισότητα του λήμματος προκύπτει.

Η λύση $\widetilde{S}$ αποτελείται από $\frac{k-b}{L}$ ολόκληρες κλίκες και $b < L$ κορυφές από την κλίκα B. Ο αριθμός των ακμών στη λύση $\widetilde{S}$ είναι

$$|E(\widetilde{S})| = \frac{k-b}{L} \cdot \frac{L(L-1)}{2} + \frac{b(b-1)}{2} = \frac{k(L-1)-b(L-b)}{2}. \qquad \square$$

Το επόμενο λήμμα, που είναι ανεξάρτητου ενδιαφέροντος, δίνει ένα άνω φράγμα στον αριθμό των ακμών ενός χορδικού γράφου συναρτήσει του μεγέθους της μέγιστης κλίκας του.

Λήμμα 10. *Έστω* $c \geq 2$ *το μέγεθος μίας μέγιστης κλίκας ενός χορδικού γράφου* $G = (V, E)$. *Ισχύει ότι* $|E| \leq (c-1)(|V| - \frac{c}{2})$ *και αυτό το φράγμα είναι το καλύτερο δυνατό.*

Απόδειξη. Αποδεικνύουμε το λήμμα με (διπλή) επαγωγή στα c και $|V|$. Για $c = 2$ και για κάθε $|V|$, είναι προφανές ότι ο G είναι είτε δέντρο είτε δάσος (ο κύκλος τριών κορυφών είναι κλίκα με $c = 3$ και ο κύκλος τεσσάρων ή περισσοτέρων κορυφών δεν επιτρέπεται αφού ο G είναι χορδικός). Συνακόλουθα, $|E| \leq |V| - 1$ και το λήμμα ισχύει.

Υποθέτουμε ότι το λήμμα ισχύει για $c = q$ και αποδεικνύουμε ότι ισχύει και για $c = q + 1$, με επαγωγή στο $|V| > c$.

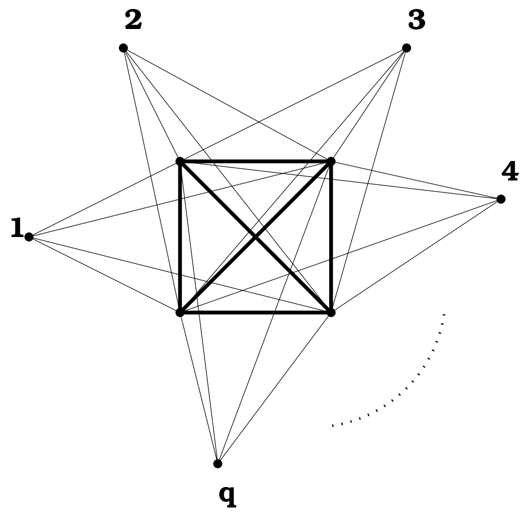
Για $c = q + 1$ και $|V| = q + 2$, ο G αποτελείται από μία κλίκα C μεγέθους $q + 1$ και από μία κορυφή u εκτός της κλίκας C. Η κορυφή u έχει το πολύ q γειτονικές κορυφές στην κλίκα C, γιατί αλλιώς ο G είναι κλίκα μεγέθους $c = q + 2$. Συνακόλουθα, $|E(V)| \leq \frac{q(q+1)}{2} + q = q(q+2-\frac{q+1}{2}) = (c-1)(|V| - \frac{c}{2})$, και το λήμμα ισχύει.

Υποθέτουμε ότι το λήμμα ισχύει για $c = q + 1$ και $|V| = r > q + 2$ και αποδεικνύουμε ότι ισχύει για $c = q + 1$ και $|V| = r + 1$. Έστω u η κορυφή ελαχίστου βαθμού στο χορδικό γράφο $G = (V, E)$, $|V| = r + 1$, με μέγεθος μέγιστης κλίκας $c = q + 1$. Θεωρούμε το γράφο $G' = (V \setminus \{u\}, E')$, ο οποίος είναι επίσης χορδικός. Ο G' αποτελείται από r κορυφές και η μέγιστη κλίκα του είναι μεγέθους το πολύ $c = q + 1$. Από την επαγωγική υπόθεση ισχύει ότι $|E'| \leq q(r - \frac{q+1}{2})$.

Επιπλέον, ο G σαν χορδικός γράφος έχει τουλάχιστον μία *simplicial* κορυφή, δηλαδή, μία κορυφή s τέτοια που οι γειτονικές της κορυφές είναι κλίκα, έστω C_s. Αφού η μεγιστική κλίκα του γράφου G είναι μεγέθους $c = q + 1$ ο βαθμός της *simplicial* κορυφής s είναι το πολύ q. Άρα, $d_u \leq d_s \leq q$, αφού η u είναι η ελαχίστου βαθμού κορυφή στο G. Γι' αυτό το λόγο, προκύπτει ότι $|E| = |E'| + d_u \leq q(r - \frac{q+1}{2}) + q =$

$q(r+1-\frac{q+1}{2})$ και το λήμμα ισχύει.

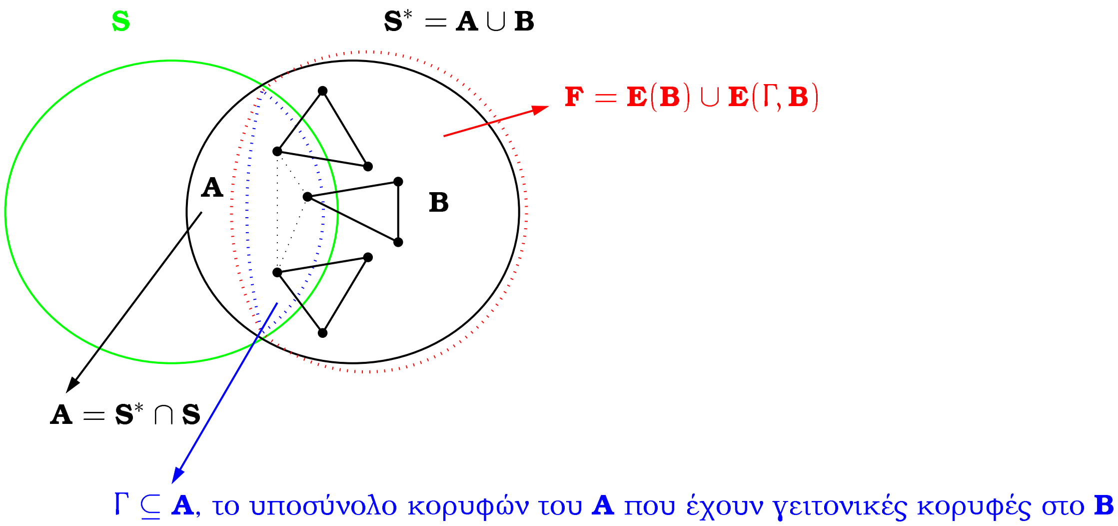
Για να αποδείξουμε ότι αυτό το φράγμα είναι το καλύτερο δυνατό θεωρούμε το χορδικό γράφο $G = (V, E)$ που αποτελείται από μία κλίκα, C, μεγέθους $c-1$ και $|V|-c+1$ ανεξάρτητες κορυφές καθεμία από τις οποίες είναι γειτονική με όλες τις κορυφές της κλίκας C. Παρατηρούμε ότι μία μέγιστη κλίκα στον G αποτελείται από την κλίκα C και μία από τις ανεξάρτητες κορυφές, και είναι μεγέθους c (Σχήμα 5.2). Για αυτόν το γράφο G ισχύει ότι $|E| = \frac{(c-1)(c-2)}{2}+(|V|-c+1)(c-1) = (c-1)(|V|-\frac{c}{2})$. Παρατηρούμε ότι εάν $c = |V|$, τότε ο G είναι πλήρης γράφος. $\square$



Σχήμα 5.2: Παράδειγμα γράφου με μέγιστη κλίκα 5 κορυφών.

Ας συσχετίσουμε τώρα τη λύση S του άπληστου αλγόριθμου με τη βέλτιστη λύση S^* για το πρόβλημα του πυκνότερου k-υπογράφου σε ένα χορδικό γράφο G. Έστω $S^* = A \cup B$, όπου $A = S^* \cap S$ είναι το υποσύνολο των κορυφών της S^* που ανήκουν και στη λύση S και $B = S^* \setminus A$ είναι το υποσύνολο των κορυφών της S^* που δεν ανήκουν στη λύση S. Έστω επίσης $\Gamma \subseteq A$ το υποσύνολο κορυφών του A που έχουν γειτονικές κορυφές στο B και $F = E(B) \cup E(\Gamma, B)$ (Σχήμα 5.3). Προφανώς, $\Gamma \cap B = \emptyset$ και $|E(S^*)| = |E(A)| + |E(B)| + |E(\Gamma, B)| = |E(A)| + |F|$.

Για να φράξουμε τον αριθμό των ακμών στη βέλτιστη λύση S^* θεωρούμε τον από ακμές επαγόμενο υπογράφο $G^F = (\Gamma \cup B, F)$ καθώς και τον από κορυφές επαγόμενο



Σχήμα 5.3: Λύσεις S και S^*.

υπογράφο $G_{B\cup\Gamma}$ του G. Παρατηρούμε ότι ο $G_{B\cup\Gamma}$, ως από κορυφές επαγόμενος υπογράφος του G, είναι χορδικός γράφος, ενώ ο G^F, ως από ακμές επαγόμενος υπογράφος του G, είναι γενικά ένας μη χορδικός γράφος. Η επόμενη πρόταση δίνει χρήσιμες δομικές ιδιότητες για τους υπογράφους $G_{B\cup\Gamma}$ και G^F του G.

Πρόταση 5.
(i) 'Ολες οι μεγιστικές κλίκες του γράφου $G^F = (\Gamma \cup B, F)$ *είναι μεγέθους το πολύ* L.
(ii) Ο βαθμός μίας simplicial κορυφής του γράφου $G_{B\cup\Gamma}$ *είναι το πολύ* $L - 1$.

Απόδειξη. Η λύση S του άπληστου αλγόριθμου περιέχει κορυφές από μεγιστικές κλίκες του γράφου G μεγέθους τουλάχιστον L και οι κορυφές στο B δεν ανήκουν στην S. Συνεπώς, οι κορυφές στο B ανήκουν σε μεγιστικές κλίκες του G μεγέθους το πολύ L.

(i) Καταρχάς θεωρούμε το σύνολο B των κορυφών του υπογράφου G^F. Εφόσον ανήκουν σε μεγιστικές κλίκες του G μεγέθους το πολύ L, δημιουργούν και στον G^F μεγιστικές κλίκες μεγέθους το πολύ L.
Έπειτα θεωρούμε το σύνολο Γ των κορυφών του υπογράφου G^F. Από τον ορισμό του υπογράφου G^F προκύπτει ότι οι κορυφές στο Γ: α) είναι ανεξάρτητες στον G^F και β) κάθε μία από αυτές έχει τουλάχιστον μία γειτονική κορυφή στο B. Υποθέτουμε ότι μία κορυφή του Γ ανήκει σε μία μεγιστική κλίκα, K, του G^F μεγέθους μεγαλύτερου του L. Καθώς οι κορυφές του Γ είναι ανεξάρτητες στον G^F, προκύπτει ότι τουλάχιστον

μία κορυφή του B ανήκει επίσης σε μία τέτοια κλίκα K του G^F, άρα και σε μία μεγιστική κλίκα του G μεγέθους μεγαλύτερου του L, το οποίο έρχεται σε αντίφαση με τον ορισμό του συνόλου B. Συνεπώς, οι κορυφές του Γ ανήκουν σε μεγιστική κλίκα του G^F μεγέθους το πολύ L.
Άρα, όλες οι μεγιστικές κλίκες του γράφου G^F είναι μεγέθους το πολύ L.

(ii) Ο γράφος $G_{B\cup\Gamma}$, ως από κορυφές επαγόμενος υπογράφος του G, είναι χορδικός και έχει τουλάχιστον μία *simplicial* κορυφή, έστω $s \in B\cup\Gamma$. Ο βαθμός της s είναι ίσος με $|C_s|$, όπου C_s είναι η κλίκα που επάγεται από τις γειτονικές κορυφές του s. Υποθέτουμε ότι ο βαθμός της s είναι τουλάχιστον L. Η κορυφή s μαζί με την κλίκα C_s δημιουργούν την κλίκα $C_s \cup \{s\}$ μεγέθους τουλάχιστον $L+1$. Αφού κάθε κορυφή του Γ έχει τουλάχιστον μία γειτονική κορυφή στο B, προκύπτει ότι τουλάχιστον μία κορυφή της κλίκας $C_s \cup \{s\}$ (είτε η ίδια η s είτε μία από τις γειτονικές κορυφές της) ανήκει στο B, το οποίο έρχεται σε αντίφαση με τον ορισμό του συνόλου B. $\square$

Χρησιμοποιώντας την Πρόταση 5 μπορούμε τώρα να δείξουμε ότι το φράγμα του Λήμματος 10 ισχύει και για το γράφο G^F.

Λήμμα 11. *Για τον από ακμές επαγόμενο υπογράφο* $G^F = (\Gamma \cup B, F)$ *του* G *ισχύει ότι* $|F| \leq (L-1)(|\Gamma \cup B| - \dfrac{L}{2})$.

Απόδειξη. Δουλεύουμε όμοια με το Λήμμα 10 κάνοντας (διπλή) επαγωγή στο $|\Gamma|$ και στο $|B|$. Παρατηρούμε ότι $|\Gamma \cup B| = |\Gamma| + |B|$.

Για $|\Gamma| = 1$ και για κάθε $|B|$, ο G^F είναι ένας χορδικός γράφος αφού συμπίπτει με τον από κορυφές επαγόμενο γράφο $G_{\Gamma\cup B}$ του G. Έτσι, από την Πρόταση 5(i) και το Λήμμα 10, προκύπτει ότι $|F| \leq (L-1)(|\Gamma \cup B| - \frac{L}{2})$ και το λήμμα ισχύει.

Υποθέτουμε ότι το λήμμα ισχύει για $|\Gamma| = r$ και για κάθε $|B|$ και αποδεικνύουμε ότι ισχύει για $|\Gamma| = r+1$, με επαγωγή στο $|B|$.

Για $|\Gamma| = r+1$ και $|B| = 1$, ο G^F είναι ένας γράφος αστέρι (***star graph***) ο οποίος αποτελείται από μία κορυφή στο B και $r+1$ κορυφές στο Γ. Η μέγιστη κλίκα στον G^F είναι μεγέθους 2, και το λήμμα ισχύει, αφού $|F| \leq r+1 = (L-1)(|\Gamma|+|B|-\frac{L}{2})$.

Υποθέτουμε ότι το λήμμα ισχύει για $|\Gamma| = r+1$ και $|B| = m$ και αποδεικνύουμε ότι ισχύει για $|\Gamma| = r+1$ και $|B| = m+1$. Θεωρούμε τις ακόλουθες κορυφές στους γράφους G^F και $G_{B\cup\Gamma}$:

- Έστω u, μία κορυφή ελαχίστου βαθμού στον G^F, με βαθμό $d^F(u)$.
- Έστω w, μία κορυφή ελαχίστου βαθμού στον $G_{B\cup\Gamma}$, με βαθμό $d_{B\cup\Gamma}(w)$. Αφού ο G^F είναι ένας από ακμές επαγόμενος υπογράφος του $G_{B\cup\Gamma}$ ισχύει ότι $d^F(u) \leq d_{B\cup\Gamma}(w)$.
- Έστω s, μία ***simplicial*** κορυφή του $G_{B\cup\Gamma}$, με βαθμό $d_{B\cup\Gamma}(s)$. Αφού η w είναι μία κορυφή ελαχίστου βαθμού στον $G_{B\cup\Gamma}$ ισχύει ότι $d_{B\cup\Gamma}(w) \leq d_{B\cup\Gamma}(s)$.

Χρησιμοποιώντας την Πρόταση 5(ii) προκύπτει ότι $d^F(u) \leq d_{B\cup\Gamma}(w) \leq d_{B\cup\Gamma}(s) \leq L-1$.

Αφαιρώντας την κορυφή u από τον G^F προκύπτει ο γράφος $G^{F'} = ((\Gamma' \cup B'), F')$ είτε με $|\Gamma'| = r+1$ και $|B'| = m$ (εάν $u \in B$) είτε με $|\Gamma'| = r$ και $|B'| = m+1$ (εάν $u \in \Gamma$). Και στις δύο περιπτώσεις, από τις επαγωγικές υποθέσεις, ισχύει ότι $|F'| \leq (L-1)(r+1+m-\frac{L}{2})$. Συνεπώς, $|F| = |F'| + d^F(u) \leq (L-1)(r+1+m-\frac{L}{2}) + (L-1) = (L-1)(r+1+m+1-\frac{L}{2})$ και το λήμμα ισχύει. □

Εφαρμόζοντας το Λήμμα 11 στον υπογράφο G^F, (με $|\Gamma \cup B| \leq |S^*| = k$) προκύπτει ότι

$$|F| = |E(B)| + |E(\Gamma, B)| \leq (L-1)(k - \frac{L}{2}).$$

Για το σύνολο ακμών $E(A)$ ισχύει ότι $|E(A)| \leq |E(S)|$, αφού $A \subseteq S$. Επίσης γνωρίζουμε, από το Λήμμα 9, ότι

$$|E(S)| \geq \frac{k(L-1) - b(L-b)}{2}.$$

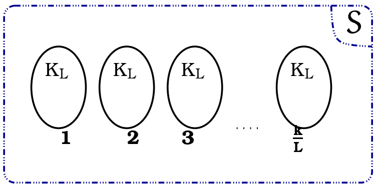
Κατά συνέπεια,

$$\frac{|E(S^*)|}{|E(S)|} = \frac{|E(A)| + |F|}{|E(S)|} \leq 1 + \frac{|F|}{|E(S)|} \leq 1 + \frac{(L-1)(2k-L)}{k(L-1) - b(L-b)}.$$

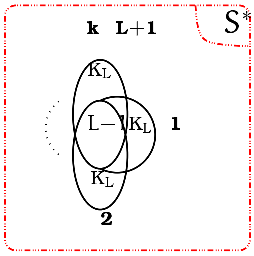
Υπενθυμίζοντας ότι $b = k \bmod L \leq L-1$ και διαχωρίζοντας μεταξύ δύο περιπτώσεων για το b ($b \leq L/2$ και $b > L/2$) είναι εύκολο να δείξουμε ότι $\frac{(L-1)(2k-L)}{k(L-1) - b(L-b)} \leq 2$. Έτσι προκύπτει το επόμενο θεώρημα.

Θεώρημα 21. *[31, 32] Υπάρχει ένας 3-προσεγγιστικός αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε χορδικούς γράφους.*

Σε αυτό το κεφάλαιο παρουσιάστηκε ένας 3-προσεγγιστικός άπληστος αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε χορδικούς γράφους. Αναφορικά με το κατά πόσο μπορεί να βελτιωθεί αυτός ο προσεγγιστικός λόγος, καταφέραμε να κατασκευάσουμε αντιπαραδείγματα για τα οποία ο άπληστος αυτός



$$E(S) = \frac{k}{L}\frac{L(L-1)}{2} = \frac{k(L-1)}{2}$$



$$E(S^*) = \frac{(L-1)(L-2)}{2} + (k-L+1)(L-1) = (k-\frac{L}{2})(L-1)$$

Σχήμα 5.4: Αντιπαράδειγμα

αλγόριθμος δίνει μία λύση με τουλάχιστον τις μισές ακμές της βέλτιστης λύσης. Συγκεκριμένα, στο σχήμα 5.4 φαίνεται ένας γράφος που αποτελείται από $\frac{k}{L}$ ανεξάρτητες κλίκες μεγέθους L και $k-L+1$ κλίκες μεγέθους L με $L-1$ κοινές κορυφές. Ο άπληστος αλγόριθμος στη χειρότερη περίπτωση θα επιλέξει τις $\frac{k}{L}$ ανεξάρτητες κλίκες ενώ η βέλτιστη λύση αποτελείται από τις $k-L+1$ κλίκες μεγέθους L με τις $L-1$ κοινές κορυφές. Ο αριθμός των ακμών της λύσης, $E(S)$, του άπληστου αλγόριθμου είναι $k(L-1)/2$ ενώ ο αριθμός των ακμών της βέλτιστης λύσης, $E(S^*)$, είναι $(k-L/2)(L-1)$. Συνεπώς, ο λόγος της βέλτιστης λύσης προς τη λύση του άπληστου αλγόριθμου είναι το πολύ δύο.

Κεφάλαιο 6

Κατακλείδα

Μπορεί να μην πήγα εκεί που σκόπευα να πάω, αλλά νομίζω ότι κατέληξα εκεί που σκόπευα να είμαι.

DOUGLAS ADAMS

6.1 Σύνοψη

Σε αυτήν τη διδακτορική διατριβή μελετήσαμε το πρόβλημα του πυκνότερου k-υπογράφου ενός δοθέντος γράφου $G = (V, E)$. Στο πρόβλημα του πυκνότερου k-υπογράφου (*Densest* k-*subgraph* - DkS) δίνεται ένας γράφος $G = (V, E)$, $|V| = n$, και ένας ακέραιος $k \leq n$, και ζητείται ένας υπογράφος του G επαγόμενος από ακριβώς k κορυφές του έτσι ώστε ο αριθμός των ακμών του υπογράφου να είναι ο μέγιστος δυνατός. Το πρόβλημα είναι *NP-hard* ως γενίκευση του προβλήματος εύρεσης της μέγιστης κλίκας (*Maximum Clique problem*). Στη βεβαρημένη εκδοχή του προβλήματος δίνεται γράφος με μη αρνητικά βάρη στις ακμές και ζητείται ο k επαγόμενος υπογράφος με το μέγιστο συνολικό βάρος στις ακμές του.

Καταρχάς παρουσιάσαμε κάποια από τα υπάρχοντα αποτελέσματα της περιοχής, που είναι σχετικά και συγκρίσιμα με τη συνεισφορά μας. Έπειτα παρουσιάσαμε αλγόριθμους πολυωνυμικού χρόνου για το DkS πρόβλημα σε βεβαρημένους γράφους με μέγιστο βαθμό δύο, σε δέντρα με βάρη στις ακμές, όπου η λύση δεν είναι απαραίτητα συνεκτική καθώς και σε γράφους διαστημάτων (*interval graphs*) όπου ο γράφος μεγιστικών κλικών (*clique graph*) είναι μονοπάτι. Ο αλγόριθμος σε δέντρα με βάρη αποτελεί γενίκευση των αλγορίθμων στα [35],[33] και [17] όπου η λύση είναι συνεκτική. Συγκεκριμένα, αναπτύξαμε έναν $O(nk^2)$ χρόνου αλγόριθμο δυναμικού προγραμματισμού σε δέντρα με βάρη, ο οποίος δίνει ως λύση είτε δέντρο είτε δάσος. Επιπλέον, αποδείξαμε ότι υπάρχει ένας $O(nk^2)$ χρόνου αλγόριθμος για το πρόβλημα του πυκνότερου k-υπογράφου σε βεβαρημένους γράφους μέγιστου βαθμού δύο και τέλος αναπτύξαμε έναν $O(nk^3)$ χρόνου αλγόριθμο δυναμικού προγραμματισμού σε γράφους διαστημάτων που έχουν ένα απλό μονοπάτι ως γράφο μεγιστικών κλικών (*clique graph*).

Παρουσιάσαμε επίσης ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα (*PTAS*) για το DkS πρόβλημα σε αστέρι κλικών (πολυπλοκότητας $O(\frac{1}{\epsilon}n^{\frac{1}{\epsilon}})$) καθώς και έναν πολυωνυμικού χρόνου αλγόριθμο σε δέντρο κλικών φραγμένου βαθμού. Ειδικότερα, ο αλγόριθμος που αναπτύχθηκε για το DkS πρόβλημα σε δέντρο κλικών είναι $O(nk^{m+1})$ χρόνου, όπου n είναι ο συνολικός αριθμός κορυφών σε όλες τις κλίκες και m ο μέγιστος βαθμός του δέντρου κλικών. Κατά συνέπεια, όταν το δέντρο κλικών εκφυλίζεται σε μονοπάτι κλικών ο αλγόριθμος υπολογίζει τον πυκνότερο k-υπογράφο σε χρόνο $O(nk^3)$.

Τελικά, αναλύσαμε ένα σταθερού λόγου προσεγγιστικό αλγόριθμο για το DkS πρόβλημα σε χορδικούς γράφους (*chordal graphs*), ο οποίος είναι ο πρώτος σταθερού λόγου προσεγγιστικός αλγόριθμος σε ειδικές κατηγορίες γράφων στις οποίες το DkS παραμένει *NP-hard*. Ο αλγόριθμος που αναλύεται είναι ένας άπληστος (*greedy*) αλγόριθμος, ο οποίος δίνει ως λύση του προβλήματος τις μεγιστικές κλίκες (*maximal cliques*) του γράφου μεγιστικών κλικών (*clique graph*) του αρχικού χορδικού γράφου κατά μη αύξουσα σειρά μεγέθους, και επιτυγχάνει προσεγγιστικό λόγο τρία. Στο συγκεκριμένο κεφάλαιο περιέχεται και ένα λήμμα ανεξάρτητου ενδιαφέροντος, στο οποίο φράσσεται ο αριθμός των ακμών ενός χορδικού γράφου συναρτήσει του μεγέθους μίας μέγιστης κλίκας του γράφου. Πρέπει να σημειωθεί ότι σε όλα τα παραδείγματα που καταφέραμε να κατασκευάσουμε ο αλγόριθμος δίνει μία λύση με τουλάχιστον το μισό αριθμό ακμών της βέλτιστης λύσης. Συνεπώς μία διαφορετική προσέγγιση της ανάλυσης του άπληστου αλγόριθμου θα μπορούσε να οδηγήσει σε προσεγγιστικό λόγο δύο.

6.2 Μελλοντική εργασία

Παρά το εκτενές ενδιαφέρον για το πρόβλημα του πυκνότερου k-υπογράφου και το μεγάλο αριθμό εργασιών που έχουν εκπονηθεί, ενδιαφέροντα ερωτήματα ακόμη παραμένουν ανοικτά. Καταρχάς παρόλο που έχουν αναπτυχθεί, μελετηθεί και αναλυθεί διάφοροι προσεγγιστικοί αλγόριθμοι για το DkS πρόβλημα, οι οποίοι βασίζονται σε διαφορετικές τεχνικές προσέγγισης, ο καλύτερος προσεγγιστικός λόγος παραμένει $O(n^{\delta})$ για κάποιο $\delta < \frac{1}{3}$ [10]. Κανένας από τους αλγόριθμους που μετέπειτα αναπτύχθηκαν δεν οδηγεί σε βελτίωση του συγκεκριμένου αποτελέσματος. Ωστόσο στη διεθνή βιβλιογραφία έχει διατυπωθεί η εικασία ότι κάτι τέτοιο είναι αδύνατο. Κατά συνέπεια το πιο ενδιαφέρον ανοικτό ερώτημα για το DkS πρόβλημα παραμένει η βελτίωση του λόγου προσέγγισης του για γενικούς γράφους ή η απόδειξη ότι κάτι τέτοιο είναι αδύνατο.

Μία ειδική κατηγορία γράφων, η μελέτη της οποίας θα μπορούσε να δώσει κάποιες ενδείξεις για τη μελέτη του προβλήματος σε γενικούς γράφους, είναι οι γράφοι οι οποίοι είναι μεν διμερείς (*bipartite*) αλλά που τα ανεξάρτητα υποσύνολα κόμβων αντικαθίστανται από κλίκες. Το DkS πρόβλημα παραμένει *NP-hard* σε αυτή την κατηγορία γράφων με αναγωγή από το DkS πρόβλημα σε διμερείς γράφους.

Το ερώτημα της προσεγγισιμότητας προκύπτει και για ενδιαφέρουσες ειδικές κατηγορίες γράφων, στις οποίες έχει αποδειχθεί ότι το DkS πρόβλημα παραμένει *NP-hard*. Τέτοιες κατηγορίες γράφων είναι οι διμερείς γράφοι (*bipartite graphs*), οι *comparability* γράφοι, οι κανονικοί γράφοι (*regular graphs*), οι φραγμένου βαθμού γράφοι (*bounded degree graphs*) ή ακόμα και οι φραγμένου βαθμού διμερείς γράφοι. Επιπλέον, ενδιαφέρον παρουσιάζει και η βελτίωση του σταθερού προσεγγιστικού λόγου που παρουσιάστηκε στο Κεφάλαιο 5 στους χορδικούς γράφους (*chordal graphs*) καθώς και ο λόγος προσέγγισης του DkS , όταν η λύση που ζητάμε είναι συνεκτική, στους επίπεδους γράφους (*planar graphs*).

Επιπροσθέτως υπάρχουν κατηγορίες γράφων για τις οποίες η πολυπλοκότητα του προβλήματος παραμένει ανοικτή. Τέτοιες κατηγορίες είναι οι γράφοι διαστημάτων (*interval graphs*) ή ακόμα και οι *proper interval* γράφοι, οι οποίοι είναι γράφοι διαστημάτων στους οποίους κανένα διάστημα δεν εμπεριέχεται πλήρως σε κάποιο άλλο. Επίσης, οι *permutation* γράφοι καθώς και οι διμερείς χορδικοί γράφοι (*chordal bipartite graphs*), δηλαδή, οι διμερείς γράφοι στους οποίους κάθε κύκλος με μέγεθος τουλάχιστον έξι έχει μία χορδή, είναι ενδιαφέρουσες κλάσεις στις οποίες η πολυπλοκότητα του DkS προβλήματος παραμένει ανοικτό πρόβλημα. Τέλος να αναφέρουμε ότι ανοικτό ερώτημα παραμένει και η πολυπλοκότητα του DkS, όταν ζητάμε μη συνεκτική λύση, σε επίπεδους γράφους (*planar graphs*).

Βιβλιογραφία

[1] S. Arora, D. Karger, and M. Karpinski. Polynomial time approximation schemes for dense instances of NP-hard problems. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, pages 284–293, 1995.

[2] Y. Asahiro and K. Iwama. Finding dense subgraphs. In *Proceedings of the 6th International Symposium on Algorithms and Computation*, pages 102–111, 1995.

[3] Y. Asahiro, K. Iwama, and E. Miyano. Random generation of test instances with controlled attributes. In *Series in Discrete Mathematics and Theoretical Computer Science*, volume 26, pages 377–393, 1996.

[4] Y. Asahiro, K. Iwama, H. Tamaki, and T. Tokuyama. Greedily finding a dense subgraph. *Journal of Algorithms*, 34(2):203–221, 2000.

[5] A. Billionnet and F. Roupin. A deterministic algorithm for the densest k-subgraph problem using linear programming. Technical Report No486, CEDRIC, CNAM-IIE, Paris, 2004.

[6] K. S. Booth and G. S. Leuker. Testing for the consecutive ones property, interval graphs, and graph planarity using pq-tree algorithms. *Journal of Computational Systems Science*, 13(3):335–379, 1976.

[7] D. G. Corneil and Y. Perl. Clustering and domination in perfect graphs. *Discrete Applied Mathematics*, 9:27–39, 1984.

[8] G. A. Dirac. On rigid circuit graphs. *Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg*, 25:71–76, 1961.

[9] U. Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 534–543, 2002.

[10] U. Feige, G. Kortsarz, and D. Peleg. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001.

[11] U. Feige and M. Langberg. Approximation algorithms for maximization problems arising in graph partitioning. *Journal of Algorithms*, 41(2):174–211, 2001.

[12] U. Feige and M. Seltser. On the densest k-subgraph problem. Technical Report CS97-16, Weizmann Institute, 1997.

[13] D. R. Fulkerson and O. A. Gross. Incidence matrices and interval graphs. *Pacific Journal of Mathematics*, 15:835–855, 1965.

[14] G. Gallo, M.D. Grigoriadis, and R.E. Tarjan. A fast parametric maximum flow algorithm and applications. *SIAM Journal on Computing*, 18(1):30–55, 1989.

[15] F. Gavril. Algorithms for minimum coloring, maximum clique, minimum covering by cliques and maximum independent set of chordal graph. *SIAM Journal on Computing*, 1(2):180–187, 1972.

[16] P. C. Gilmore and A. J. Hoffman. A characterization of comparability graphs and of interval graphs. *Canadian Journal of Mathematics*, 16:539–548, 1964.

[17] O. Goldschmidt and D. Hochbaum. k-edge subgraph problems. *Discrete Applied Mathematics and Combinatorial Operations Research and Computer Science*, 74(2):159–169, 1997.

[18] M. C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, New York, 1980.

[19] M.T. Hajiaghayi, K. Jain, L.C. Lau, I.I. Mandoiu, A. Russell, and V.V. Vazirani. Minimum multicolored subgraph problem in multiplex PCR primer set selection and population haplotyping. In *Proceedings of the International Workshop on Bioinformatics Research and Applications*, pages 758–766, 2006.

[20] Q. Han, Y. Ye, and J. Zhang. An improved rounding method and semidefinite programming relaxation for graph partition. *Mathematical Programming*, 92(3):509–535, 2002.

[21] R. Hassin, S. Rubinstein, and A. Tamir. Approximation algorithms for maximum dispersion. *Operations Research Letters*, 21(3):133–137, 1997.

[22] J. M. Keil and T. B. Brecht. The complexity of clustering in planar graphs. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 9:155–159, 1991.

[23] S. Khot. Ruling out PTAS for graph min-bisection, densest subgraph and bipartite clique. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 136–145, 2004.

[24] J. Kleinberg. Authoritative sources in a hyperlinked enviroment. *Journal of the ACM*, 46:604–632, 1999.

[25] G. Kortsarz and D. Peleg. On choosing a dense subgraph. In *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 692–701, 1993.

[26] N. Laoutaris, V. Zissimopoulos, and I. Stavrakakis. Joint object placement and node dimensioning for internet content distribution. *Information Processing Letters*, 89(6):273–279, 2004.

[27] M. Liazi, I. Milis, F. Pascual, and V. Zissimopoulos. The densest k-subgraph problem on clique graphs. *Journal of Combinatorial Optimization*, 14(4):465–474, 2007.

[28] M. Liazi, I. Milis, F. Pascual, and V. Zissimopoulos. The densest k-subgraph problem on clique graphs. In *International Combinatorics, Geometry and Computer Science Conference*, Marseille, 2007.

[29] M. Liazi, I. Milis, and V. Zissimopoulos. Polynomial variants of the densest/heaviest k-subgraph problem. In *APPOL II, Final Workshop on Approximation and on line Algorithms*, Athens, 2004.

[30] M. Liazi, I. Milis, and V. Zissimopoulos. Polynomial variants of the densest/heaviest k-subgraph problem. In *20th British Combinatorial Conference*, Durham, 2005.

[31] M. Liazi, I. Milis, and V. Zissimopoulos. The densest k-subgraph problem on chordal graphs. In *2nd Athens Colloquium on Algorithms and Complexity*, Athens, 2007.

[32] M. Liazi, I. Milis, and V. Zissimopoulos. A constant approximation algorithm for the densest k-subgraph problem on chordal graphs. *Information Processing Letters*, 108(1):29–32, 2008.

[33] F. Maffioli. Finding a best subtree of a tree. Technical Report 91.041, Politechnico di Milano, Dipartimento di Elektronica, 1991.

[34] C. H. Papadimitriou and M. Yannakakis. The clique problem for planar graphs. *Information Processing Letters*, 13(4/5):131–133, 1981.

[35] Y. Perl and Y. Shiloach. Efficient optimization of monotonic functions on trees. *SIAM Journal on Algebraic and Discrete Methods*, 4(4):512–516, 1983.

[36] D. J. Jr. Rader and G. J. Woeginger. The quadratic 0-1 knapsack problem with series-parallel support. *Operation Research Letters*, 30(3):159–166, 2002.

[37] S. S. Ravi, D. J. Rosenkrantz, and G. K. Tayi. Heuristic and special case algorithms for dispersion problems. *Operations Research*, 42(2):299–310, 1994.

[38] P. Slavík. Improved performance of the greedy algorithm for partial cover. *Information Processing Letters*, 64(5):251–254, 1997.

[39] A. Srivastav and K. Wolf. Finding dense subgraphs with semidefinite programming. In *Proceedings of the International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 181–191, 1998.

[40] Y. Ye and J. Zhang. .519 approximation of dense-n/2-subgraph. Working Paper, Department of Management Sciences, Henry B. Tippie College of Business, The University of Iowa, 1999.