μΠλ∀

«Το πρόβλημα εύρεσης του k-Πυκνότερου Υπογράφου»

Διπλωματική Εργασία

Μαρία Λιάζη

Ιούλιος 2002

Διαπανεπιστημιακό Πρόγραμμα Μεταπτυχιακών Σπουδών στη Λογική και Θεωρία Αλγορίθμων και Υπολογισμού

Περιεχόμενα

Περίληψη

Μελετάμε το πρόβλημα της εύρεσης υψηλά συνεκτικών υπογράφων σε κατευθυνόμενους και μη κατευθυνόμενους γράφους καθώς και το σχετικό πρόβλημα υπολογισμού του *k-πυκνότερου υπογράφου* ενός δοθέντος γράφου.

Παρουσιάζουμε διάφορους προσεγγιστικούς αλγόριθμους γι' αυτά τα προβλήματα βελτιστοποίησης καθώς και κάποια αποτελέσματα σχετικά με τη δυσκολία προσέγγισης του προβλήματος του *k-πυκνότερου υπογράφου.*

Λέξεις κλειδιά: *προσεγγιστικοί αλγόριθμοι, πυκνότερος υπογράφος, k-πυκνότερος υπογράφος, NP-hardness, PTAS.*

Abstract

We study the problem of finding highly connected subgraphs of undirected and directed graphs and the relative problem of computing the *densest k-subgraph* of a given graph.

We present some approximation algorithms for these optimization problems as well as some results about the NP-hardness of approximation of the *k-densest subgraph* problem.

Key words: *approximation algorithms, densest subgraph, k-densest subgraph, NP-hardness, PTAS.*

1 Εισαγωγή

Το πρόβλημα εύρεσης του πυκνότερου υπογράφου ενός δοθέντος γράφου είναι γνωστό σαν *Densest Subgraph problem* (DS) και όταν το πρόβλημα αναφέρεται σε μη κατευθυνόμενους γράφους μπορεί να λυθεί με flow techniques . Ο γρηγορότερος γνωστός αλγόριθμος για το DS δίνεται στο [7] και εκτελείται σε χρόνο $O(mn \log n^2/m)$. Εδώ παρουσιάζουμε έναν greedy 2-προσεγγιστικό αλγόριθμο για το DS για κατευθυνόμενους και μη γράφους. Ίσως κάποιος να αναρωτηθεί τη χρησιμότητα ενός τέτοιου αλγορίθμου για το DS σε μη κατευθυνόμενους γράφους τη στιγμή που υπάρχει πολυωνυμικού χρόνου αλγόριθμος για την επιλυσή του. Ο μόνος λόγος που αναφερόμαστε σε έναν τέτοιο αλγόριθμο είναι για να μας βοηθήσει να οδηγηθούμε στον αντίστοιχο αλγόριθμο που αφορά το DS για την περίπτωση των κατευθυνόμενων γράφων όπου εκεί δεν υπάρχει αντίστοιχο πολυωνυμικού χρόνου αποτέλεσμα.

Ένα σχετικό με αυτό πρόβλημα και ίσως και πιο σημαντικό είναι το *k-Densest Subgraph problem* (DkS) όπου στόχος είναι να βρούμε ένα επαγόμενο υπογράφο με k κορυφές και μέγιστο μέσο βαθμό. Το πρόβλημα αυτό είναι NP-hard. Για αυτό το πρόβλημα παρουσιάζουμε έναν προσεγγιστικό αλγόριθμο με προσεγγιστικό λόγο $O(n^{\delta})$ για κάποιο $\delta < \frac{1}{3}$. Δυστυχώς δεν ξέρουμε κάποιο συμπληρωματικό αρνητικό αποτέλεσμα που να μας δίνει ενδείξεις ότι για κάποιο $\epsilon > 0$ το να επιτύχουμε έναν προσεγγιστικό λόγο $O(n^{\epsilon})$ είναι NP-hard.

Κάποιος μπορεί να αναρωτηθεί αν μπορούμε να χρησιμοποιήσουμε κάποιες απο τις αλγοριθμικές τεχνικές που χρησιμοποιήθηκαν για την επίλυση του DS για να πετύχουμε κάποια προσέγγιση για το DkS. Κάτι τέτοιο όμως φαίνεται να έχει μεγάλες δυσκολίες. Ας θεωρήσουμε για παράδειγμα την περίπτωση των κανονικών (*regular*) γράφων. Ο πυκνότερος υπογράφος ενός κανονικού γράφου είναι ο ίδιος ο γράφος, και έτσι καμία αλγοριθμική ιδέα δεν εμπλέκεται για την επίλυση του DS σε κανονικούς γράφους. Απο την άλλη μεριά η εύρεση του πυκνότερου k-κορυφών υπογράφου παραμένει NP-hard αφού το πρόβλημα της κλίκας παραμένει NP-hard στους κανονικούς γράφους.

Επίσης θα παρουσιάσουμε κάποια αποτελέσματα σχετικά με τη δυσκολία προσέγγισης του DkS προβλήματος και κατά πόσο τα αποτελέσματα αυτά μας βοηθάνε να βγάλουμε κάποια συμπεράσματα για τη δυσκολία προσέγγισης αυτού του προβλήματος. Θα δούμε ότι ακόμα και στην περίπτωση αραιών γράφων η NP-hardness του DkS αποδεικνύεται ακόμα με αναγωγή απο το Clique.

Τέλος αναφέρουμε επιγραμματικά ένα αποτέλεσμα που αφορά το DkS αλλά με κάποιους περιορισμούς. Συγκεκριμένα εάν το $k = \Omega(n)$ και ο αριθμός των ακμών είναι $\Omega(n^2)$ τότε το DkS έχει ένα πολυωνυμικού χρόνου προσεγγιστικό σχήμα $(PTAS)$, δηλαδή για κάθε ϵ υπάρχει πολυωνυμικού χρόνου αλγόριθμος που προσεγγίζει το DkS με έναν προσεγγιστικό λόγο $1 + \epsilon$.

2 Εφαρμογές του DkS

Ένας γράφος συχνά συσχετίζεται με μια σχέση ανάμεσα σε ένα σύνολο ατόμων, δηλαδή ένα άτομο αντιστοιχεί σε μια κορυφή και μια ακμή μεταξύ δύο κορυφών υποδεικνύει κάποια 'καλή' σχέση μεταξύ αυτών των δύο ατόμων. Έτσι μια κλίκα (*clique*) είναι ένα υποσύνολο ατόμων, τέτοιο που κάθε άτομο σε αυτό το σύνολο έχει μια 'καλή' σχέση με κάθε άλλο άτομο σε αυτό το σύνολο. Ωστόσο λαμβάνοντας μια 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 συνεκτικών 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 του ορισμού της πυκνότητας που θα χρησιμοποιηθεί.

Το πρόβλημα εύρεσης υψηλά συνεκτικών υποδομών στο Web έχει λάβει αρκετή προσοχή. Ειδικότερα τέτοιες υποδομές αντιστοιχούν σε κοινότητες (*communities*) στο Web , δηλαδή συλλογές απο 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 (βλέπε [8]) αναγνωρίζει κόμβους (*hubs*) και πηγές (*authorities*) μεταξύ του συνόλου των πιθανών σελίδων σχετικών με ένα ερώτημα. Οι κόμβοι χαρακτηρίζονται απο την παρουσία μεγάλου αριθμού συνδέσμων στις πηγές και οι πηγές χαρακτηρίζονται απο την παρουσία μεγάλου αριθμού συνδέσμων απο τους κόμβους. Οι κόμβοι και οι πηγές επιδεικνύουν μια αμοιβαία ενισχύουσα σχέση (*mutually reinforcing relationship*) : ένας καλός κόμβος δείχνει σε πολλές καλές πηγές και μια καλή πηγή υποδεικνύεται απο πολλούς καλούς κόμβους.

Ο ορισμός της πυκνότητας για κατευθυνόμενους γράφους, τον οποίο αναφέρουμε παρακάτω, είναι κατάλληλος για αραιούς κατευθυνόμενους γράφους όπως είναι ο γράφος του *Web*. Αυτό παρακινείται απο την προσπάθεια διατύπωσης της ιδέας της εύρεσης συνόλων κόμβων και πηγών που έχουν μεγάλη συνεκτικότητα σε σχέση με τον υπόλοιπο γράφο.

Εκτός απο το Web γράφο κάποιος θα μπορούσε να σκεφτεί αρκετές εφαρμογές του DS και του DkS. Ενδεικτικά αναφέρουμε την εφαρμογή τους στην ασφάλεια της παραγωγής στιγιοτύπων 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 (για περισσότερες πληροφορίες βλέπε [10]).

3 Ορισμοί

3.1 Το πρόβλημα εύρεσης του πυκνότερου υπογράφου (DS)

Ορισμός 1 *Έστω $G = G(V, E)$ ένας μη κατευθυνόμενος γράφος και $S \subseteq V$. Ορίζουμε $E(S)$ να είναι οι ακμές που επάγονται απο το S ,δηλαδή*

$$E(S) = \{ij \in E : i \in S, j \in S\}.$$

Ορίζουμε την πυκνότητα $f(S)$ του υποσυνόλου S να είναι:

$$f(S) = \frac{|E(S)|}{|S|}$$

Ορίζουμε την ***πυκνότητα*** *$f(G)$* ***μη κατευθυνόμενου γράφου*** *$G = G(V, E)$ να είναι:*

$$f(G) = \max_{S \subseteq V}\{f(S)\}$$

Παρατήρηση 1 *Ο μέσος βαθμός του υπογράφου που επάγεται απο το S είναι $2f(S)$ ενώ $2f(G)$ είναι ο μέγιστος μέσος βαθμός ανάμεσα σε όλους τους επαγόμενους υπογράφους.*

Ορισμός 2 *Έστω $G = G(V, E)$ ένας κατευθυνόμενος γράφος και $S, T \subseteq V$. Ορίζουμε $E(S, T)$ να είναι το σύνολο των ακμών που πηγαίνουν απο το S στο T, δηλαδή*

$$E(S, T) = \{ij \in E : i \in S, j \in T\}.$$

Ορίζουμε την πυκνότητα $d(S, T)$ των συνόλων S, T να είναι:

$$d(S, T) = \frac{|E(S, T)|}{\sqrt{|S||T|}}$$

Ορίζουμε την ***πυκνότητα*** *$d(G)$* ***κατευθυνόμενου γράφου*** *$G = G(V, E)$ να είναι:*

$$d(G) = \max_{S, T \subseteq V}\{d(S, T)\}$$

Παρατήρηση 2 *Αν $|S| = |T|$ τότε $d(S, T)$ είναι ο μέσος βαθμός ακμών απο μια κορυφή του S στο T.*

Παρατήρηση 3 *Τα σύνολα S,T δεν είναι απαραίτητα ξένα.*

Ορισμός 3 *Το πρόβλημα εύρεσης του πυκνότερου υπογράφου* ***Densest subgraph problem*** *(DS) έχει σαν είσοδο ένα γράφο $G = G(V, E)$ (με n κορυφές). Η έξοδος είναι ένας υπογράφος του G, ο οποίος έχει μεγιστη πυκνότητα.*

3.2 Το πρόβλημα εύρεσης του k-πυκνότερου υπογράφου (DkS)

Ορισμός 4 *Η* ***πυκνότητα*** *d_G ενός γράφου $G = G(V, E)$ είναι ο μέσος βαθμός του, δηλαδή $d_G = \frac{2|E|}{|S|}$.*

Ορισμός 5 *Το πρόβλημα του k-πυκνότερου υπογράφου* ***Densest k- subgraph*** *(DkS) έχει σαν είσοδο ένα γράφο $G = G(V, E)$ (με n κορυφές) και μια παράμετρο k, $1 \leq k \leq n$. Η έξοδος είναι ένας υπογράφος του G επαγόμενος απο τις k κορυφές, ο οποίος έχει μέγιστη πυκνότητα.*

3.3 Γενικοί Ορισμοί

Ορισμός 6 *Η κλάση πολυπλοκότητας* ***NP*** *είναι η κλάση των γλώσσών που μπορούν να πιστοποιηθούν απο έναν πολυωνυμικού χρόνου αλγόριθμο. Ειδικότερα μια γλώσσα L ανήκει στην NP αν και μόνο αν υπάρχει ένας διπλής εισόδου πολυωνυμικού χρόνου αλγόριθμος A και μια σταθερά c έτσι 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.

Ορισμός 7 *Μια γλώσσα $L \subseteq \{0, 1\}^*$ είναι* ***NP-complete*** *εάν*

1. *$L \in NP$ και*
2. *$L' \leq_P L$ για κάθε $L' \in NP$, δηλαδή κάθε γλώσσα L' που ανήκει στο NP ανάγεται πολυωνυμικά στη γλώσσα L.*

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

Ορισμός 8 *Ας υποθέσουμε ότι A είναι ένα πρόβλημα βελτιστοποιήσης· 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 $OPT(x) = \min_{s \in F(x)} c(s)$ (ή $\max_{s \in F(x)} c(s)$, εάν το A είναι πρόβλημα μεγιστοποίησης). Έστω M ένας αλγόριθμος ο οποίος δοθέντος ενός στιγμιοτύπου x, επιστρέφει μια 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*

$$\frac{|c(M(x)) - OPT(x)|}{\max\{OPT(x), c(M(x))\}} \leq \epsilon.$$

Ορισμός 9 *Λέμε ότι ένας προσεγγιστικός αλγόριθμος για ένα πρόβλημα έχει έχει προσεγγιστικό λόγο (**approximation ratio**) ρ όπου $\rho \geq \frac{OPT}{A}$ για προβλήματα μεγιστοποίησης (και $\rho \geq \frac{A}{OPT}$ για προβλήματα ελαχιστοποίσης αντίστοιχα).*

Ορισμός 10 *Ένα **PTAS** (**P**olynomial **T**ime **A**pproximation **S**cheme) 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 $\frac{1}{\epsilon}$, όπως $O(n^{\frac{1}{\epsilon}})$).*

Ορισμός 11 *Μια κλίκα (**Clique**) σε ένα μη κατευθυνόμενο γράφο $G = G(V, E)$ είναι ένα υποσύνολο κορυφών $V' \subseteq V$, έτσι ώστε κάθε ζεύγος κορυφών του να συνδέεται με μια ακμή του E. Το μέγεθος της κλίκας είναι ο αριθμός των κορυφών που περιέχει. Το πρόβλημα της κλίκας είναι το πρόβλημα (βελτιστοποίσης) της εύρεσης μιας κλίκας μεγίστου μεγέθους σε ένα γράφο. Σαν πρόβλημα απόφασης απλά ρωτάμε αν υπάρχει κλίκα μεγέθους τουλάχιστον k στο γράφο, δηλαδή $CLIQUE = \{\langle G, k\rangle$:ο G έιναι ένας γράφος με κλίκα μεγέθους τουλάχιστον $k\}$. Το πρόβλημα της κλίκας είναι NP-complete.*

Ορισμός 12 *Ένα στιγμιότυπο του **3SAT** είναι μια 3CNF φόρμουλα $f = C_1, C_2, ..., C_n$ όπου κάθε clause C_i είναι της μορφής (x'_i, x'_j, x'_k). Το 3SAT είναι NP-complete.*

Ορισμός 13 *Ένα **LP** (**L**inear **P**rogramming) στην κανονική του μορφή είναι της μορφής:*

$$\min c'x$$
$$Ax \geq b$$
$$x \geq 0$$

Ορισμός 14 *Δοθέντος ενός LP στη κανονική του μορφή, που καλείται πρωτογενές (primal) , το αντίστοιχο δυϊκό του (**dual**) ορίζεται ως εξής:*

$$\max d'b$$

$$d'A \leq c'$$
$$d' \geq 0$$

4 Εύρεση πυκνότερου υπογράφου

4.1 Ακριβής αλγόριθμος για την εύρεση πυκνότερου υπογράφου σε ένα μη κατευθυνόμενο γράφο

Σκοπός μας είναι να βρούμε τον πυκνότερο υπογράφο ενός δοθέντος μη κατευθυνόμενου γράφου, δηλαδή να υπολογίσουμε την ποσότητα $f(G)$. Θα δείξουμε ότι το πρόβλημα υπολογισμού της $f(G)$ μπορεί να εκφρασθεί σαν ένα γραμμικό πρόγραμμα (LP) και θα παρουσιάσουμε έναν απλό greedy 2-προσεγγιστικό αλγόριθμο για αυτό το πρόβλημα. Σε αυτό το σημείο πρέπει να τονίσουμε ότι το αποτέλεσμα αυτό απο μόνο του ίσως να μην είναι τόσο ενδιαφέρον αφού όπως αναφέραμε και στην εισαγωγή υπάρχει ακριβής αλγόριθμος για τον υπολογισμό της $f(G)$ βασισμένος σε flow techniques, ωστόσο η τεχνική της απόδειξης που θα χρησιμοποιήσουμε θα μας βοηθήσει αργότερα σε πιο πολύπλοκες αποδείξεις για τον υπολογισμό της $d(G)$ σε κατευθυνόμενους γράφους.
Χρησιμοποιούμε το εξής LP:

$$\max \sum_{i,j} x_{ij} \tag{1}$$

$$\forall ij \in E \qquad x_{ij} \leq y_i \tag{2}$$

$$\forall ij \in E \qquad x_{ij} \leq y_j \tag{3}$$

$$\sum_i y_i \leq 1 \tag{4}$$

$$x_{ij}, y_i \geq 0 \tag{5}$$

Λήμμα 1 *Για κάθε $S \subseteq V$, η τιμή του LP (1)-(5) είναι τουλάχιστον $f(S)$.*

Απόδειξη: Θα δώσουμε μια εφικτή λύση για το LP με τιμή $f(S)$.Έστω $x = \frac{1}{|S|}$. Για κάθε $i \in S$, θέτουμε $\bar{y}_i = x$. Για κάθε $ij \in E(S)$, θέτουμε $\bar{x}_{ij} = x$. Όλες τις εναπομένουσες μεταβλητές τις θέτουμε ίσες με το 0.
Έτσι έχουμε, $\sum_i \bar{y}_i = |S| \cdot x = 1$. Άρα η $(\bar{x}, \bar{y})$ είναι εφικτή λύση για το LP. Η τιμή αυτής της λύσης είναι:

$$|E(S)| \cdot x = \frac{|E(S)|}{|S|} = f(S).$$

$\diamondsuit$

Λήμμα 2 *Δοθείσης μιας εφικτής λύσης για το LP (1)-(5) με τιμή υ μπορούμε να κατασκευάσουμε $S \subseteq V$ τέτοιο ώστε $f(S) \geq \upsilon$.*

Απόδειξη: Θεωρούμε μια εφικτή λύση $(\bar{x}, \bar{y})$ για το LP (1)-(5). Χωρίς να χάνεται η γενικότητα, μπορούμε να υποθέσουμε ότι για όλα τα ij, $\bar{x}_{ij} = \min(\bar{y}_i, \bar{y}_j)$. Ορίζουμε μια συλλογή συνόλων S δεικτοδοτούμενα απο μια παράμετρο $r \geq 0$. Έστω $S(r) = \{i : \bar{y}_i \geq r\}$ και $E(r) = \{ij : \bar{x}_{ij} \geq r\}$. Αφού $\bar{x}_{ij} \leq \bar{y}_i$ και $\bar{x}_{ij} \leq \bar{y}_j$, $ij \in E(r) \Rightarrow i \in S(r), j \in S(r)$. Επίσης αφού $\bar{x}_{ij} = \min(\bar{y}_i, \bar{y}_j)$, $i \in S(r), j \in S(r) \Rightarrow ij \in E(r)$. Έτσι το $E(r)$ είναι ακριβώς το σύνολο των ακμών που επάγονται απο το $S(r)$.
Έχουμε ότι $\int_0^\infty |S(r)|dr = \sum_i \bar{y}_i \leq 1$. Παρατηρούμε ότι $\int_0^\infty |E(r)|dr = \sum_{ij} \bar{x}_{ij}$, που είναι η τιμή της αντικειμενικής συνάρτησης της λύσης του LP. Έστω ότι αυτή η τιμή είναι υ.
Ισχυρισμός: Υπάρχει r τέτοιο που $\frac{|E(r)|}{|S(r)|} \geq \upsilon$.
Ας υποθέσουμε ότι δεν υπάρχει τέτοιο r. Τότε

$$\int_0^\infty |E(r)|dr < \upsilon \int_0^\infty |S(r)|dr \leq \upsilon.$$

Έτσι καταλήγουμε σε αντίφαση. Για να βρούμε ένα τέτοιο r, παρατηρούμε ότι μπορούμε να ελέγξουμε όλα τα συνδυαστικώς ξένα σύνολα $S(r)$, ελέγχοντας απλά τα σύνολα $S(r)$ που προκύπτουν θέτοντας $r = \bar{y}_i$, για κάθε $i \in V$. $\diamondsuit$

Θεώρημα 1

$$\max_{S \subseteq V} \{f(S)\} = OPT(LP) \tag{6}$$

όπου $OPT(LP)$ δηλώνει την τιμή της βέλτιστης λύσης του LP (1)-(5). Περαιτέρω ένα σύνολο S που μεγιστοποιεί την $f(S)$ μπορεί να υπολογιστεί απο τη βέλτιστη λύση του LP.

Απόδειξη: Θεωρούμε το σύνολο S που μεγιστοποιεί την $f(S)$ και έστω z η τιμή της αντικειμενικής συνάρτησης της λύσης του LP.
Απο το Λήμμα 1 έχουμε ότι για κάθε $S \subseteq V$ ισχύει ότι $f(S) \leq z$. Αφού για κάθε z ισχύει ότι $z \leq OPT(LP)$ έχουμε ότι $f(S) \leq OPT(LP)$.
Απο το Λήμμα 2 έχουμε ότι για κάθε z υπάρχει $S \subseteq V$ τέτοιο ώστε $f(S) \geq z$. Αφού ισχύει για κάθε z θα ισχύει και για την $OPT(LP)$, δηλαδή
$f(S) \geq OPT(LP)$. Άρα $f(S) = OPT(LP)$.
Η απόδειξη του λήμματος 2 δίνει μια κατασκευή του συνόλου S που μεγιστοποιεί την $f(S)$ απο τη βέλτιστη λύση του LP. $\diamondsuit$

4.2 2-προσεγγιστικός αλγόριθμος για την εύρεση πυκνότερου υπογράφου σε ένα μη κατευθυνόμενο γράφο

Σκοπός μας είναι να παράγουμε έναν υπογράφο του G με μεγάλο μέσο βαθμό. Διαισθητικά για να παράγουμε έναν τέτοιο υπογράφο πρέπει να ξεφορτωθούμε όλες τις χαμηλού βαθμού κορυφές. Αυτό μας οδηγεί σε έναν greedy αλγόριθμο.

`ΑΛΓΟΡΙΘΜΟΣ 1`

Ο αλγόριθμος διατηρεί ένα υποσύνολο S των κορυφών κάθε φορά.

1.Αρχικά $S \leftarrow V$

2.Αν το $S \neq \emptyset$ τότε

βρες το i_{min}, όπου i_{min}= κορυφή με το μικρότερο βαθμό στο S

αφαίρεσε το i_{min} απο το σύνολο S

3.Απο όλα τα σύνολα που κατασκευάστηκαν κατά τη διάρκεια εκτέλεσης του αλγορίθμου, επέστρεψε σαν έξοδο το σύνολο S που μεγιστοποιεί την $f(S)$, δηλαδή το σύνολο με το μέγιστο μέσο βαθμό.

Θα αποδείξουμε ότι ο `ΑΛΓΟΡΙΘΜΟΣ 1` δίνει μια προσέγγιση 2 για την $f(G)$. Για να αναλύσουμε τον αλγόριθμο, παράγουμε ένα άνω φράγμα για τη βέλτιστη λύση. Το άνω φράγμα έχει την εξής μορφή:

Αναθέτουμε κάθε ακμή ij είτε στην i είτε στην j κορυφή. Για κάθε κορυφή i, $d(i)$ είναι ο αριθμός των ακμών ij ή ji που ανατέθηκαν στην i. Έστω $d^{\max} = \max_i\{d(i)\}$.

Η ανάθεση των ακμών σε κάποια κορυφή γίνεται καθώς εκτελείται ο αλγόριθμος. Αρχικά όλες οι ακμές δεν έχουν ανατεθεί. Όταν η ελαχίστου βαθμού κορυφή αφαιρείται απο το S της ανατίθενται όλες οι ακμές που συνδέουν τη συγκεκριμένη κορυφή με όλες τις υπόλοιπες κορυφές του S. Διατηρούμε την αναλλοίωτη συνθήκη ότι όλες οι ακμές μεταξύ δύο κορυφών στο τρέχον σύνολο S δεν έχουν ανατεθεί ακόμα, ενώ όλες οι άλλες ακμές έχουν ανατεθεί. Στο τέλος της εκτέλεσης του αλγορίθμου όλες οι ακμές έχουν ανατεθεί σε κάποια κορυφή.

Το ακόλουθο λήμμα δείχνει ότι το $f(S)$ φράσσεται απο το $d^{\max}$.

Λήμμα 3

$$\max_{S \subseteq V}\{f(S)\} \leq d^{max}$$

Απόδειξη: Θεωρούμε το σύνολο S που μεγιστοποιεί την $f(S)$. Κάθε ακμή στο $E(S)$ πρέπει να έχει ανατεθεί σε κάποια κορυφή στο S. Έτσι έχουμε:

$$|E(S)| \leq |S| \cdot d^{\max}$$

$$f(S) = \frac{|E(S)|}{|S|} \leq d^{\max}$$

$\diamondsuit$

Λήμμα 4 *Έστω υ η μέγιστη τιμή της $f(S)$ για όλα τα σύνολα S που κατασκευάζονται κατα τη διάρκεια εκτέλεσης του greedy αλγορίθμου, τότε:*

$$d^{\max} \leq 2\upsilon$$

Απόδειξη: Θεωρούμε μια επανάληψη του αλγορίθμου. Αφού $i_{\min}$ είναι η ελαχίστου βαθμού κορυφή του S, ο βαθμός της θα είναι το πολύ $\frac{2|E(S)|}{|S|} \leq 2\upsilon$. Παρατηρούμε ότι σε μια κορυφή ανατίθενται ακμές όταν αυτή αφαιρείται απο το S. Αυτό αποδεικνύει ότι $d^{\max} \leq 2\upsilon$. $\diamondsuit$

Θεώρημα 2 *Ο greedy αλγόριθμος δίνει μια προσέγγιση 2 για την $f(G)$.*

Απόδειξη: Απο το Λήμμα 3 έχουμε: $\max_{S \subseteq V}\{f(S)\} \leq d^{max}$.
Απο το Λήμμα 4 έχουμε: $d^{\max} \leq 2\upsilon$. Άρα $\max_{S \subseteq V}\{f(S)\} \leq 2\upsilon$. $\diamondsuit$

Πολυπλοκότητα αλγορίθμου:

Εύκολα βλέπουμε πως ο αλγόριθμος μπορεί να εκτελεσθεί σε χρόνο $O(n^2)$ για ένα γράφο με n κορυφές και m ακμές. Μπορούμε να διατηρούμε τους βαθμούς των κορυφών στον υπογράφο που επάγεται απο το S. Σε κάθε επανάληψη εντοπίζεται και αφαιρείται η ελαχίστου βαθμού κορυφή και ενημερώνονται οι βαθμοί των εναπομένουσων κορυφών· και τα δύο μπορούν να γίνουν σε χρόνο $O(n)$.

Στην πραγματικότητα μπορούμε να υλοποιήσουμε τον αλγόριθμο να τρέχει σε γραμμικό χρόνο. Αφού ο βαθμός μιας κορυφής είναι ένας ακέραιος $\in \{0, n\}$, μπορούμε να διατηρούμε λίστες των κορυφών με τον ίδιο βαθμό, δηλαδή μια λίστα κορυφών βαθμού 0,1,2, κ.τ.λ. Σε κάθε επανάληψη η ελαχίστου βαθμού κορυφή αφαιρείται και οι βαθμοί των γειτονικών κορυφών ενημερώνονται, απαιτώντας να μετακινηθούν σε καινούργιες λίστες. Ο συνολικός χρόνος εργασίας γι' αυτές τις ενημερώσεις είναι $O(m)$. Παρατηρούμε ότι ο ελάχιστος βαθμός μειώνεται το πολύ κατα 1 σε κάθε επανάληψη. Εάν ο ελάχιστος βαθμός σε μια συγκεκριμένη επανάληψη ήταν d, η ελαχίστου βαθμού κορυφή για την επόμενη επανάληψη λαμβάνεται διερευνώντας τις λίστες για βαθμούς $d-1$, d, $d+1$ κ.τ.λ. Ο συνολικός χρόνος διερεύνησης στις λίστες είναι $O(n)$. Άρα ο αλγόριθμος τρέχει σε χρόνο $O(m+n)$.

Διαίσθηση πίσω απο το άνω φράγμα:

Ίσως κάποιος να αναρωτιέται για την προέλευση του άνω φράγματος της βέλτιστης λύσης που χρησιμοποιήσαμε στις προηγούμενες αποδείξεις. Το άνω φράγμα είναι στενά συνδεδεμένο με το δυϊκό του LP (1)-(5). Συγκεκριμένα το δυϊκό του LP (1)-(5) είναι το εξής:

$$\min \gamma \tag{7}$$
$$\forall ij \in E \qquad \alpha_{ij} + \beta_{ij} \geq 1 \tag{8}$$
$$\forall i \qquad \gamma \geq \sum_j \alpha_{ij} + \sum_j \beta_{ji} \tag{9}$$
$$\alpha_{ij}, \gamma \geq 0 \tag{10}$$

Το άνω φράγμα αντιστοιχεί στη λύση του δυϊκού, όπου α_{ij},β_{ij} είναι 0-1 μεταβλητές, $\alpha_{ij} = 1$ αντιστοιχεί στο ότι η ακμή ij έχει ανατεθεί στο i ενώ $\beta_{ij} = 1$ αντιστοιχεί στο ότι η ακμή ij έχει ανατεθεί στο j και άρα το γ αντιστοιχεί στο $d^{\max}$. Ουσιαστικά η παραπάνω απόδειξη κατασκευάζει μια λύση για το δυϊκό καθώς εκτελείται ο greedy αλγόριθμος. Η τιμή αυτής της δυϊκής λύσης είναι το $d^{\max}$, το άνω φράγμα σύμφωνα με το Λήμμα 3.

4.3 Ακριβής αλγόριθμος για την εύρεση πυκνότερου υπογράφου σε έναν κατευθυνόμενο γράφο

Σκοπός μας είναι να βρούμε τον πυκνότερο υπογράφο ενός δοθέντος κατευθυνόμενου γράφου, δηλαδή να υπολογίσουμε την ποσότητα $d(G)$. Υπενθυμίζουμε ότι $d(G)$ είναι η μέγιστη τιμή του $d(S,T)$ πάνω σε όλα τα υποσύνολα κορυφών S,T. Αρχικά παρουσιάζουμε ένα LP relaxation για το $d(G)$, το οποίο εξαρτάται απο την τιμή του $|S|/|T|$ για το ζεύγος S,T που μεγιστοποιεί το $d(S,T)$. Επειδή δεν ξέρουμε αυτον το λόγο εξαρχής γράφουμε ένα ξεχωριστό LP για κάθε δυνατή τιμή αυτού του λόγου· υπάρχουν $O(n^2)$ δυνατές τιμές. Για $|S|/|T| = c$, χρησιμοποιούμε το ακόλουθο LP relaxation $LP(c)$:

$$\max \sum_{i,j} x_{ij} \tag{11}$$

$$\forall ij \qquad x_{ij} \leq s_i \tag{12}$$

$$\forall ij \qquad x_{ij} \leq t_j \tag{13}$$

$$\sum_i s_i \leq \sqrt{c} \tag{14}$$

$$\sum_j t_j \leq \frac{1}{\sqrt{c}} \tag{15}$$

$$x_{ij}, s_i, t_j \geq 0 \tag{16}$$

Λήμμα 5 *Θεωρούμε $S,T \subseteq V$. Έστω $c = |S|/|T|$, τότε η βέλτιστη τιμή του $LP(c)$ είναι τουλάχιστον $d(S,T)$.*

Απόδειξη: Δίνουμε μια εφικτή λύση $(\bar{x}, \bar{s}, \bar{t})$ για το $LP(C)$ (11)-(16) με τιμή $d(S,T)$. Έστω $x = \frac{\sqrt{c}}{|S|} = \frac{1}{\sqrt{c}|T|}$. Για κάθε $i \in S$, θέτουμε $\bar{s}_i = x$. Για κάθε $j \in T$, θέτουμε $\bar{t}_j = x$. Για κάθε $ij \in E(S,T)$, θέτουμε $\bar{x}_{ij} = x$. Όλες τις εναπομένουσες μεταβλήτές τις θέτουμε ίσες με 0. Έτσι έχουμε $\sum_i \bar{s}_i = |S| \cdot x = \sqrt{c}$ και $\sum_j \bar{t}_j = |T| \cdot x = 1/\sqrt{c}$. Άρα η $(\bar{x}, \bar{s}, \bar{t})$ είναι μια εφικτή λύση για το $LP(c)$. Η τιμή αυτής της λύσης είναι:

$$|E(S,T)| \cdot x = \frac{|E(S,T)|}{\sqrt{c} \cdot |T|} = \frac{|E(S,T)|}{\sqrt{|S||T|}} = d(S,T).$$

$\diamondsuit$

Λήμμα 6 *Δοθείσης μια εφικτής λύσης για το $LP(c)$ με τιμή υ μπορούμε να κατασκευάσουμε $S,T \subseteq V$ τέτοια που $d(S,T) \geq \upsilon$.*

Απόδειξη: Θεωρούμε μια εφικτή λύση $(\bar{x}, \bar{s}, \bar{t})$ για το $LP(C)$ (11)-(16). Χωρίς να χάνεται η γενικότητα, μπορούμε να υποθέσουμε ότι για όλα τα i, j, $\bar{x}_{ij} = \min(\bar{s}_i, \bar{t}_j)$.
Ορίζουμε μια συλλογή συνόλων S, T δεικτοδοτούμενα απο μια παράμετρο $r \geq 0$. Έστω $S(r) = \{i : \bar{s}_i \geq r\}$, $T(r) = \{j : \bar{t}_j \geq r\}$ και $E(r) = \{ij : \bar{x}_{ij} \geq r\}$. Αφού $\bar{x}_{ij} \leq \bar{s}_i$ και $\bar{x}_{ij} \leq \bar{t}_j$, $ij \in E(r) \Rightarrow i \in S(r), j \in T(r)$. Επίσης αφού $\bar{x}_{ij} = \min(\bar{s}_i, \bar{t}_j)$, $i \in S(r), j \in T(r) \Rightarrow ij \in E(r)$. Έτσι το $E(r)$ είναι ακριβώς το σύνολο των ακμών που πηγαίνουν απο το $S(r)$ στο $T(r)$.
Τώρα έχουμε ότι $\int_0^\infty |S(r)|dr = \sum_i \bar{s}_i \leq \sqrt{c}$. Επίσης $\int_0^\infty |T(r)|dr = \sum_j \bar{t}_j \leq 1/\sqrt{c}$.
Απο την ανισότητα του Schwarz έχουμε,

$$\int_0^\infty \sqrt{|S(r)||T(r)}dr = \sqrt{\left(\int_0^\infty |S(r)|dr\right)\left(\int_0^\infty |T(r)|dr\right)} \leq 1.$$

Παρατηρούμε ότι $\int_0^\infty |E(r)|dr = \sum_{ij} \bar{x}_{ij}$, που είναι η τιμή της αντικειμενικής συνάρτησης της λύσης του $LP(c)$. Έστω ότι αυτή η τιμή είναι υ.
Ισχυρισμός: Υπάρχει r τέτοιο που $\frac{|E(r)|}{\sqrt{|S(r)||T)r)|}} \geq \upsilon$.
Ας υποθέσουμε ότι δεν υπάρχει τέτοιο r. Τότε

$$\int_0^\infty |E(r)|dr < \upsilon \int_0^\infty \sqrt{|S(r)||T(r)|}dr \leq \upsilon.$$

Έτσι καταλήγουμε σε αντίφαση. Για να βρούμε ένα τέτοιο r, παρατηρούμε ότι μπορούμε να ελέγξουμε όλα τα συνδυαστικώς ξένα σύνολα $S(r)$,$T(r)$ ελέγχοντας απλά τα σύνολα $S(r)$,$T(r)$ που προκύπτουν θέτοντας $r = \bar{s}_i$ και $r = \bar{t}_j$, για κάθε $i \in V, j \in V$. $\diamondsuit$

Παρατήρηση 4 *Τα ζεύγη συνόλων S, T που εγγυώνται απο την παραπάνω απόδειξη δεν ικανοποιούν τη $|S|/|T| = c$.*

Θεώρημα 3

$$\max_{S,T \subseteq V}\{d(S,T)\} = \max_{c}\{OPT(LP(c))\} \tag{17}$$

όπου $OPT(LP(c))$ δηλώνει την τιμή της βέλτιστης λύσης του $LP(c)$ (11)-(16). Περαιτέρω τα σύνολα S,T που μεγιστοποιούν την $d(S,T)$ μπορούν να υπολογιστούν απο τις βέλτιστες λύσεις του συνόλου των γραμμικών προγραμμάτων $LP(c)$.

Απόδειξη: Θεωρούμε τα σύνολα S,T που μεγιστοποιούν την $d(S,T)$.
Απο το Λήμμα 5 έχουμε ότι $\max_{S,T\subseteq V}\{d(S,T)\} \leq OPT(LP(c))$.
Άρα $\max_{S,T\subseteq V}\{d(S,T)\} \leq \max_c\{OPT(LP(c))\}$.
Θέτουμε c να είναι η τιμή που μεγιστοποιεί την $OPT(LP(c))$.
Απο το Λήμμα 6 έχουμε ότι $d(S,T) \geq \max_c\{OPT(LP(c))\}$.
Άρα $\max_{S,T\subseteq V}\{d(S,T)\} \geq d(S,T) \geq \max_c\{OPT(LP(c))\}$.
Άρα $\max_{S,T\subseteq V}\{d(S,T)\} = \max_c\{OPT(LP(c))\}$.
Η απόδειξη του λήμματος 6 δίνει μια κατασκευή των συνόλων S,T που μεγιστοποιούν την $d(S,T)$ απο την LP λύση που μεγιστοποιεί την $OPT(LP(c))$. $\diamondsuit$

Παρατήρηση 5 *Ο προταθείς αλγόριθμος περιλαμβάνει την επίλυση $O(n^2)$ LPs, ένα για κάθε πιθανή τιμή του λόγου $c = |S|/|T|$. Στην πραγματικότητα μπορούμε να μαντέψουμε αυτό το λόγο με ένα παράγοντα $(1+\epsilon)$ χρησιμοποιώντας μόνο $O(\frac{\log n}{\epsilon})$ τιμές. Δεν είναι ιδιαίτερα δύσκολο να δείξουμε ότι αυτό αποδίδει έναν $(1+\epsilon)$ παράγοντα στην προσέγγιση. Το Λήμμα 5 μπορεί να τροποποιηθεί έτσι ώστε να ενσωματώνει τον $(1+\epsilon)$ παράγοντα.*

4.4 2-προσεγγιστικός αλγόριθμος για την εύρεση πυκνότερου υπογράφου σε έναν κατευθυνόμενο γράφο

Αν εκμεταλλευτούμε την ανάλυση του greedy αλγορίθμου για την προσέγγιση του $f(G)$, τότε εξετάζοντας το δυϊκό του LP (11)-(16) θα μπορούσαμε να πάρουμε κάποιες ενδείξεις για το πως πρέπει να κατασκευαστεί και να αναλυθεί ένας greedy αλγόριθμος για το $d(G)$.
Το δυϊκό του $LP(c)$ (11)-(16) είναι το εξής:

$$\min \sqrt{c} \cdot \gamma + \frac{\delta}{\sqrt{c}} \tag{18}$$

$$\forall ij \qquad \alpha_{ij} + \beta_{ij} \geq 1 \tag{19}$$

$$\forall i \qquad \gamma \geq \sum_j \alpha_{ij} \tag{20}$$

$$\forall j \qquad \delta \geq \sum_i \alpha_{ij} \tag{21}$$

$$\alpha_{ij}, \gamma, \delta \geq 0 \tag{22}$$

Κάθε εφικτή λύση του δυϊκού είναι ένα άνω φράγμα της ακέραιας λύσης του αρχικού προβλήματος. Αυτό υποδεικνύει ένα άνω φράγμα που αντιστοιχεί στη δυϊκή λύση, όπου α_{ij}, β_{ij} είναι 0-1 μεταβλητές. Το $\alpha_{ij} = 1$ αντιστοιχεί στο ότι η ακμή ij έχει ανατεθεί στο i ενώ $\beta_{ij} = 1$ αντιστοιχεί στο ότι η ακμή ij έχει ανατεθεί στο j. Τότε το γ είναι ο μέγιστος αριθμός ακμών ij που έχουν ανατεθεί στην κορυφή i (maximum out-degree) ενώ το δ είναι ο μέγιστος αριθμός ακμών ij που έχουν ανατεθεί στην κορυφή j (maximum in-degree). Τότε η τιμή της δυϊκής λύσης $\sqrt{c} \cdot \gamma + \frac{\delta}{\sqrt{c}}$ είναι ένα άνω φράγμα επί του $d(S,T)$ για όλα τα ζεύγη συνόλων S, T τέτοια που $|S|/|T| = c$.

Greedy προσεγγιστικός αλγόριθμος

Χρησιμοποιούμε τη γνώση που κερδίσαμε εξετάζοντας το δυϊκό του $LP(c)$ για να κατασκευάσουμε και να αναλύσουμε ένα greedy προσεγγιστικό αλγόριθμο. Όπως στον ακριβή αλγόριθμο πρέπει να μαντέψουμε την τιμή του $c = |S|/|T|$. Για κάθε τέτοια τιμή του c, τρέχουμε έναν greedy αλγόριθμο. Το καλύτερο ζεύγος S, T (δηλαδή αυτό που μεγιστοποιεί το $d(S,T)$) που παράγεται απο όλους αυτούς τους greedy αλγορίθμους, είναι η έξοδος του αλγορίθμου μας.

Θα περιγράψουμε ένα greedy αλγόριθμο για μια συγκεκριμένη τιμή του c. Ο αλγόριθμος διατηρεί δύο σύνολα S και T και σε κάθε φάση αφαιρεί είτε την ελαχίστου βαθμού κορυφή απο το S είτε την ελαχίστου βαθμού κορυφή απο το T. (Ο βαθμός μιας κορυφής i στο S είναι ο αριθμός των ακμών απο το i στο T και ο βαθμός μιας κορυφής j στο T είναι ο αριθμός των ακμών απο το j στο S.)

ΑΛΓΟΡΙΘΜΟΣ 2

Ο αλγόριθμος διατηρεί δύο υποσύνολα S, T των κορυφών κάθε φορά.

1.Αρχικά $S \leftarrow V$,$T \leftarrow V$

2.Έστω ότι $i_{\min}$ είναι η κορυφή $i \in S$ που ελαχιστοποιεί το $|E(\{i\}, T)|$.
Έστω $d_S \leftarrow |E(\{i_{\min}\}, T)|$.

3.Έστω ότι $j_{\min}$ είναι η κορυφή $j \in T$ που ελαχιστοποιεί το $|E(S, \{j\})|$.
Έστω $d_T \leftarrow |E(S, \{j_{\min}\})|$.

4.Εάν $\sqrt{c} \cdot d_S \leq \frac{1}{\sqrt{c}} \cdot d_T$
τότε $S \leftarrow S - \{i_{\min}\}$
αλλιώς $T \leftarrow T - \{j_{\min}\}$.

5.Εάν αμφότερα τα S και T δεν είναι κενά, πήγαινε πίσω στο ΒΗΜΑ 2.

Απο όλα τα σύνολα S, T που παράγονται κατα τη διάρκεια εκτέλεσης του παραπάνω αλγορίθμου, το ζεύγος που μεγιστοποιεί το $d(S, T)$ επιστρέφεται σαν έξοδος του αλγορίθμου.
Για να αναλύσουμε τον αλγόριθμο, παράγουμε ένα άνω φράγμα της βέλτιστης λύσης. Το άνω φράγμα έχει την ακόλουθη μορφή, η οποία προτείνεται απο το δυϊκό του $LP(c)$:
Αναθέτουμε κάθε (κατευθυνόμενη) ακμή ij είτε στο i είτε στο j. Για μια κορυφή i, $d_{out}(i)$ είναι ο αριθμός των ακμών ij' που έχουν ανατεθεί στο i. Για μια κορυφή j, $d_{in}(j)$ είναι ο αριθμός των ακμών $i'j$ που έχουν ανατεθεί στο j. Έστω $d_{out}^{\max} = \max\{d_{out}(i)\}$ και $d_{in}^{\max} = \max\{d_{in}(j)\}$. Το ακόλουθο λήμμα δίνει το άνω φράγμα για το $d(S, T)$, για όλα τα ζεύγη S, T τέτοια ώστε $|S|/|T| = c$ και σε σχέση με τις παραμέτρους $d_{out}^{\max}$ και $d_{in}^{\max}$.

Λήμμα 7

$$\max_{|S|/|T|=c}\{d(S,T)\} \leq \sqrt{c}\cdot d_{out}^{\max} + \frac{1}{\sqrt{c}}\cdot d_{in}^{\max}$$

Απόδειξη: Η απόδειξη προκύπτει απευθείας απο το γεγονός ότι η ανάθεση των ακμών στις κορυφές αντιστοιχεί σε μια 0-1 λύση του δυϊκού του $LP(c)$. Παρατηρούμε ότι η τιμή της αντίστοιχης λύσης είναι ακριβώς $\sqrt{c}\cdot d_{out}^{\max}+\frac{1}{\sqrt{c}}\cdot d_{in}^{\max}$. Δίνουμε μια εναλλακτική συνδυαστική απόδειξη αυτού του γεγονότος:
Θεωρούμε το ζεύγος συνόλων S, T που μεγιστοποιεί το $d(S,T)$ απο όλα τα ζεύγη S, T τέτοια που $|S|/|T| = c$. Κάθε ακμή στο $E(S,T)$ πρέπει να ανατεθεί σε μια κορυφή στο S ή σε μια κορυφή στο T. Έτσι έχουμε

$$|E(S,T)| \leq |S|\cdot d_{out}^{\max} + |T|\cdot d_{in}^{\max}$$

$$\begin{aligned} d(S,T) = \frac{|E(S,T)|}{\sqrt{|S||T|}} &\leq \sqrt{\frac{|S|}{|T|}}d_{out}^{\max} + \sqrt{\frac{|T|}{|S|}}d_{in}^{\max} \\ &= \sqrt{c}\cdot d_{out}^{\max} + \frac{1}{\sqrt{c}}\cdot d_{in}^{\max} \end{aligned}$$

$\diamondsuit$

Η ανάθεση των ακμών σε ένα απο τα άκρα τους κατασκευάζεται καθώς εκτελείται ο αλγόριθμος. Παρατηρούμε ότι λαμβάνεται μια ξεχωριστή ανάθεση για κάθε διαφορετική τιμή του c. Αρχικά όλες οι ακμές δεν έχουν ανατεθεί. Όταν μια κορυφή διαγράφεται απο το S ή το T στο βήμα 4, στην κορυφή ανατίθενται όλες οι ακμές που πηγαίνουν απο την κορυφή στο άλλο σύνολο (δηλαδή εάν η $i_{\min}$ διαγραφεί της ανατίθενται όλες οι ακμές απο το $i_{\min}$ στο T και ομοίως εάν διαγραφεί η $j_{\min}$). Διατηρούμε την αναλλοίωτη συνθήκη ότι όλες οι ακμές που πηγαίνουν απο το τρέχον σύνολο S στο τρέχον σύνολο T δεν έχουν ανατεθεί σε κάποια κορυφή ακόμα· όλες οι άλλες ακμές έχουν ανατεθεί. Στο τέλος της εκτέλεσης του αλγορίθμου, όλες οι ακμές έχουν ανατεθεί σε κάποια κορυφή.

Έστω ότι τα $d_{out}^{\max}$ και $d_{in}^{\max}$ ορίζονται όπως πριν, για τη συγκεκριμένη ανάθεση που κατασκευάστηκε απο την εκτέλεση του greedy αλγορίθμου. Το ακόλουθο λήμμα συσχετίζει την τιμή της λύσης που κατασκευάζεται απο το greedy αλγόριθμο με τα $d_{out}^{\max}$ και $d_{in}^{\max}$.

Λήμμα 8 *Έστω υ η μέγιστη τιμή του $d(S,T)$ για όλα τα ζεύγη S,T που λαμβάνονται κατα τη διάρκεια εκτέλεσης του greedy αλγορίθμου για μια συγκεκριμένη τιμή του c.*
Τότε $\sqrt{c} \cdot d_{out}^{\max} \leq \upsilon$ και $\frac{1}{\sqrt{c}} \cdot d_{in}^{\max} \leq \upsilon$.

Απόδειξη: Θεωρούμε μια εκτέλεση των βημάτων 2 εως 5 σε οποιοδήποτε σημείο του αλγόριθμου. Αφού το $i_{\min}$ επιλέγεται να είναι η ελαχίστου βαθμού κορυφή στο S, ο βαθμός της είναι το πολύ $|E(S,T)|/|S|$, δηλαδή $d_S \leq |E(S,T)|/|S|$. Ομοίως $d_T \leq |E(S,T)|/|T|$. Έτσι έχουμε

$$\min(\sqrt{c} \cdot d_S, \frac{1}{\sqrt{c}} \cdot d_T) \leq \sqrt{d_S d_T} \leq \frac{|E(S,T)|}{\sqrt{|S||T|}} \leq \upsilon.$$

Εάν $\sqrt{c} \cdot d_S \leq \frac{1}{\sqrt{c}} \cdot d_T$, τότε η $i_{\min}$ διαγράφεται και τις ανατίθενται οι ακμές που πηγαίνουν απο το $i_{\min}$ στο T. Σε αυτήν την περίπτωση $\sqrt{c} \cdot d_S \leq \upsilon$. Εάν $\sqrt{c} \cdot d_S > \frac{1}{\sqrt{c}} \cdot d_T$ τότε η $j_{\min}$ διαγράφεται και τις ανατίθενται οι ακμές που πηγαίνουν απο το S στο $j_{\min}$. Σε αυτήν την περίπτωση $\frac{1}{\sqrt{c}} \cdot d_T \leq \upsilon$. Παρατηρούμε ότι σε μια συγκεκριμένη κορυφή ανατίθενται ακμές μόνο όταν αφαιρείται απο το S ή το T στο βήμα 4. Αυτό αποδεικνύει ότι $\sqrt{c} \cdot d_{out}^{\max} \leq \upsilon$ και $\frac{1}{\sqrt{c}} \cdot d_{in}^{\max} \leq \upsilon$. ♢

Άμεση συνέπεια των λημμάτων 7 και 8 είναι το παρακάτω λήμμα.

Λήμμα 9 *Έστω υ η μέγιστη τιμή του $d(S,T)$ για όλα τα ζεύγη S,T που λαμβάνονται κατα τη διάρκεια εκτέλεσης του greedy αλγορίθμου για μια συγκεκριμένη τιμή του c.*
Τότε

$$\upsilon \geq \frac{1}{2} \max_{|S||T|=c} \{d(S,T)\}.$$

Παρατήρηση 6 *Το βέλτιστο ζεύγος στο παραπάνω λήμμα δε χρειάζεται να ικανοποιεί το $|S|/|T| = c$.*

Η έξοδος του αλγορίθμου είναι το καλύτερο ζεύγος S,T που παράγεται απο το greedy αλγόριθμο απο όλες τις εκτελέσεις (για διαφορετικές τιμές του c). Έστω S^*, T^* τα σύνολα που μεγιστοποιούν το $d(S,T)$ απο όλα τα ζευγάρια S,T. Εφαρμόζοντας το παραπάνω λήμμα για τη συγκεκριμένη τιμή του $c = |S^*|/|T^*|$ παίρνουμε το ακόλουθο φράγμα στον προσεγγιστικό λόγο του αλγόριθμου.

Θεώρημα 4 *Ο greedy αλγόριθμος δίνει μια προσέγγιση 2 για το μέγιστο μέσο βαθμό $d(G)$ στο γράφο G.*

Παρατήρηση 7 *Όπως στον ακριβή αλγόριθμο που βασίζεται στο LP, αντί να τρέξουμε τον αλγόριθμο για όλες τις $\Omega(n^2)$ τιμές του c, μπορούμε να μαντέψουμε την τιμή του c στη βέλτιστη λύση με έναν $(1+\epsilon)$ παράγοντα, χρησιμοποιώντας μόνο $O(\frac{\log n}{\epsilon})$ τιμές. Δεν είναι ιδιαίτερα δύσκολο να δείξουμε ότι αυτό θα κόστιζε έναν $(1+\epsilon)$ παράγοντα στον προσεγγιστικό λόγο.*

Πολυπλοκότητα αλγορίθμου:

Όμοια με την υλοποιήση του greedy αλγορίθμου για το $f(G)$, ο greedy αλγόριθμος για το $d(G)$ για μια συγκεκριμένη τιμή του c μπορεί να υλοποιηθεί για να τρέχει σε χρόνο $O(m+n)$. Σύμφωνα με την παραπάνω παρατήρηση χρειάζεται να τρέξουμε το greedy αλγόριθμο για $O(\frac{\log n}{\epsilon})$ τιμές του c για να πάρουμε μια $(2+\epsilon)$ προσέγγιση.

5 NP-hardness του DkS

Είδαμε στα προηγούμενα ότι η εύρεση του πυκνότερου υπογράφου σε μη κατευθυνόμενους γράφους είναι ένα πολυωνυμικό πρόβλημα ενώ στους κατευθυνόμενους γράφους δώσαμε έναν προσεγγιστικό αλγόριθμο με προσεγγιστικό λόγο $2+\epsilon$. Στη συνέχεια θα εξετάσουμε το πρόβλημα του k-πυκνότερου υπογράφου.

5.1 Το DkS είναι NP-hard

Θεώρημα 5 *Το **DkS** είναι NP-hard.*

Απόδειξη: Εάν μπορούσαμε να λύσουμε το **DkS** σε πολυωνυμικό χρόνο, τότε παίρνοντας k απο 1 εως $|V|$ θα μπορούσαμε να βρούμε τη μέγιστη κλίκα σε πολυωνυμικό χρόνο. $\diamondsuit$

Το αντίστοιχο πρόβλημα απόφασης του DkS ορίζεται ως εξής:

Ορισμός 15 *Το πρόβλημα απόφασης **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.*

Παρατήρηση 8 *Όταν $K_2 = K_1(K_1-1)/2$ το **DSG** είναι ισοδύναμο με το **Clique** και γι' αυτό το γενικό **DSG** είναι NP-complete.*

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

Θεώρημα 6 *Το **DSG** είναι NP-complete για το σύνολο των στιγμιοτύπων (G, K_1, K_2) τέτοια που $K_1 \leq \frac{n}{2}$,$K_2 \leq K_1^{(1+\epsilon)}$ και $K_2 \leq \frac{m}{4}(1+\frac{9}{20}+\mathrm{o}(1))$.*

Απόδειξη: Αναγωγή απο το **3SAT**. Η απόδειξη επιτυγχάνεται δύσκολα. Για την πλήρη απόδειξη παραπέμπουμε στο [1]. $\diamondsuit$

Παρακάτω αναφέρουμε κάποια θεωρήματα. Για τις αποδείξεις των παρακάτω θεωρημάτων παραπέμπουμε στο [1].

Θεώρημα 7 *Το **DSG** είναι NP-complete ακόμα και εάν $K_1 = \frac{n}{2}$ και $K_2 \leq \frac{m}{4}(1+\mathrm{O}(\frac{1}{\sqrt{n}}))$.*

Θεώρημα 8 *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 (μικρά) $\varepsilon_1, \varepsilon_2 > 0$.*

Θεώρημα 9 *Κάτω απο τις ίδιες συνθήκες ως άνω, το **DSG** είναι NP-complete ακόμα κι αν ο γράφος είναι $|V|^{\varepsilon_3}$-κανονικός (regular).*

5.2 Το DkS όταν ο γράφος G έχει μια k-κλίκα

Όπως δείξαμε παραπάνω η NP-hardness του DkS μπορεί να αποδειχθεί με αναγωγή απο το Clique. Μπορεί αυτή η αναγωγή να χρησιμοποιηθεί απευθείας για να δείξουμε ότι το DkS είναι δύσκολο να προσεγγιστεί; Δηλαδή, είναι NP-hard να διακρίνουμε μεταξύ γράφων που έχουν μια k-κλίκα και γράφων στους οποίους ο πυκνότερος k-κορυφών υπογράφος περιέχει $(1-\epsilon)\binom{k}{2}$ ακμές για κάποιο συγκεκριμένο ϵ; Θα δείξουμε ότι αυτό το πρόβλημα δεν είναι και τόσο δύσκολο.

Εξετάζουμε το DkS, όταν γνωρίζουμε ότι ο G περιέχει μια K_k-κλίκα με $k \geq 1$ κορυφές. Θεωρείστε τον ακόλουθο αλγόριθμο:

DenseSubgraph $(G = (V, E), k)$
1.Δήλωσε με n τον αριθμό των κορυφών του G.
Εάν $n \leq k$ τότε εάν $G = K_k$ τότε επέστρεψε το G
αλλιώς επέστρεψε έναν άδειο γράφο.
2.Βρες την κορυφή v με το μικρότερο βαθμό. Έστω h ο βαθμός της.
3.Εάν $h < (1-\epsilon)n$ τότε δημιούργησε δύο γράφους: G_v και G_{-v}.
G_v είναι ένας υπογράφος του G επαγόμενος απο την v και τους γείτονες της.
G_{-v} είναι ο γράφος που απομένει αν αφαιρέσουμε το v απο τον G.
Κάλεσε τον **DenseSubgraph** (G_v, k).
Εάν ο **DenseSubgraph** επιστρέφει μη κενή έξοδο τότε σταμάτα, αλλιώς κάλεσε τον **DenseSubgraph** (G_{-v}, k).
4.Εάν $h \geq (1-\epsilon)n$ τότε επανειλημμένα αφαίρεσε απο τον G την κορυφή με τον ελάχιστο βαθμό στον εναπομένοντα κάθε φορά γράφο, μέχρι να μείνουν ακριβώς k κορυφές.
Δήλωσε με H το γράφο που επάγεται απο αυτές τις κορυφές.
Επέστρεψε το H.

Καλούμε ένα γράφο με k κορυφές και τουλάχιστον $(1-\epsilon)\binom{k}{2}$ ακμές «επιθυμητό» γράφο.

***Ισχυρισμός*:** Εάν ο G περιέχει μια k-κλίκα, τότε ο **DenseSubgraph** (G, k) επιστρέφει έναν «επιθυμητό» γράφο.
Απόδειξη: Φιξάρουμε το k. Θα αποδείξουμε έναν ισχυρότερο ισχυρισμό, δηλαδή

1. Εάν ο G περιέχει μια k-κλίκα τότε η έξοδος του αλγορίθμου είναι ένας «επιθυμητός» γράφος.

2. Εάν η έξοδος του **DenseSubgraph** (G, k) είναι ένας μη κενός γράφος, τότε αυτός είναι ένας «επιθυμητός» γράφος.

Θα αποδείξουμε τον ισχυρισμό με επαγωγή στο n. Ο ισχυρισμός είναι αληθής για $n \leq k$, αφού σε αυτήν την περίπτωση $G = K_k$ (αλλιώς η έξοδος είναι κενή). Υποθέτουμε ότι ο ισχυρισμός είναι αληθής για $n' < n$ και τον αποδεικνύουμε για n. Καταρχήν υποθέτουμε ότι

$$h \geq (1-\epsilon)n \qquad (*)$$

και ακολουθούμε το 4^o βήμα του αλγορίθμου. Χρησιμοποιούμε το ακόλουθο βοηθητικό λήμμα.

Λήμμα 10 *Εάν ο γράφος $U = (V, E)$ με $|V| = q$, έχει τουλάχιστον $(1-\epsilon)\binom{q}{2}$ ακμές, για κάποιο $1 > \epsilon > 0$ και αφαιρέσουμε απο τον U την ελαχίστου βαθμού κορυφή τότε ο εναπομένων γράφος U' έχει τουλάχιστον $(1-\epsilon)\binom{|V(U')|}{2}$ ακμές.*

Απόδειξη: Εάν αφαιρέσουμε την ελαχίστου βαθμού κορυφή, τότε θα χάσουμε το πολύ $2\frac{|E|}{q}$ ακμές, αφού ο βαθμός της θα είναι το πολύ όσος ο μέσος βαθμός. Έτσι παραμένουμε με τουλάχιστον

$$(1-\epsilon)\binom{q}{2} - \frac{2}{q}|E| \leq (1-\epsilon)\binom{q}{2} - \frac{2}{q}(1-\epsilon)\binom{q}{2} =$$

$$(1-\epsilon)\binom{q}{2}(1-\frac{2}{q}) = (1-\epsilon)\binom{q}{2}\frac{q-2}{q} =$$

$$(1-\epsilon)\binom{q-1}{2} = (1-\epsilon)\binom{|V(U')|}{2}$$

ακμές. $\diamondsuit$

Παρατηρούμε ότι εάν ισχύει η $(*)$ τότε ο γράφος G έχει τουλάχιστον $(1-\epsilon)$ $\binom{n}{2}$ ακμές και αν εφαρμόσουμε επανειλημμένα το παραπάνω λήμμα επιτυγχάνουμε τον επιθυμητό υπογράφο H. Άρα σε αυτήν την περίπτωση, όποτε ο αλγόριθμος επιστρέφει μη κενό γράφο, αυτός ο γράφος είναι «επιθυμητός».

Υποθέτουμε τώρα ότι

$$h < (1 - \epsilon)n.$$

Παρατηρούμε ότι ο αριθμός των κορυφών στους γράφους G_v και G_{-v} είναι αυστηρά μικρότερος του n. Επιπλέον ένας τουλάχιστον απο τους δύο αυτούς γράφους περιέχει μια k-κλίκα, λόγω υπόθεσης.
Έτσι η επαγωγική υπόθεση ισχύει για κάθε ένα απο αυτούς τους γράφους.
Άρα η έξοδος του **DenseSubgraph** (G, k) είναι ένας «επιθυμητός» γράφος.
Συνεπώς ο ισχυρισμός είναι αλήθής και για n. $\diamondsuit$

Ετσι έχουμε ένα ντετερμινιστικό αλγόριθμο, ο οποίος με είσοδο έναν αριθμό k και ένα γράφο G που να περιέχει μια k- κλίκα, βρίσκει έναν k-κορυφών υπογράφο με τουλάχιστον $(1-\epsilon)\binom{k}{2}$ ακμές.

Πολυπλοκότητα αλγορίθμου:
Το 4^o βήμα του αλγόριθμου παίρνει πολυωνυμικό χρόνο. Στο 3^o βήμα διαιρούμε το πρόβλημα σε δύο υποπροβλήματα, ένα μεγέθους $(n-1)$ και ένα μεγέθους το πολύ $(1-\epsilon)n+1$. Έτσι, αν $T(n)$ είναι ο χρόνος του αλγόριθμου για γράφους με n κορυφές έχουμε:

$$T(n) \leq n^c + T(n-1) + T((1-\epsilon)n+1)$$

όπου ο όρος n^c περιλαμβάνει όλους τους πολυωνυμικούς υπολογισμούς διαφορετικών σταδίων του αλγόριθμου.
Άρα ο χρόνος του αλγόριθμου είναι το πολύ $n^{O((1+\log\frac{n}{k})/\epsilon)}$

Θεώρημα 10 *Δοθέντων k,$\epsilon > 0$ και ενός γράφου G με μια k-κλίκα, υπάρχει ένας $n^{O((1+\log\frac{n}{k})/\epsilon)}$ χρόνου αλγόριθμος που επιστρέφει έναν υπογράφο επαγόμενο απο k κορυφές και με τουλάχιστον $(1-\epsilon)\binom{k}{2}$ ακμές.*

Άρα δεν είναι NP-hard το να διακρίνουμε μεταξύ γράφων που περιέχουν μια k-κλίκα και γράφων στους οποίους ο πυκνότερος k-κορυφών υπογράφος έχει τουλάχιστον $(1-\epsilon)\binom{k}{2}$ ακμές, εκτός εάν $NP \subseteq$TIME $(n^{O(logn)})$.

5.3 Το DkS σε αραιούς γράφους

Θα δείξουμε ότι το DkS παραμένει NP-hard σε γράφους με μέγιστο βαθμό 3. Για γράφους με μέγιστο βαθμό 2 υπάρχει πολυωνυμικού χρόνου αλγόριθμος που βασίζεται στο δυναμικό προγραμματισμό.

Θεωρούμε το πρόβλημα DkS, όπου ο γράφος που δίνεται σαν είσοδος του προβλήματος έχει μέγιστο βαθμό 3. Καλούμε αυτό το πρόβλημα $3DkS$. Θα δείξουμε ότι το αντίστοιχο πρόβλημα απόφασης είναι NP-complete και κατά συνέπεια το $3DkS$ θα είναι NP-hard .

Είναι γνωστό ότι το πρόβλημα της κλίκας είναι NP-complete . Θα ανάγουμε το Clique πρόβλημα στο $3DkS$. Ας υποθέσουμε ότι έχουμε (G, k) -ένα στιγμιότυπο του Clique με $G = (V, E), |V| = n, V = (v_1, v_2, ..., v_n)$. Δημιουργούμε ένα στιγμιότυπο του $3DkS$ ως εξής:
Για κάθε $1 \leq i \leq n$ δημιουργούμε ένα C^i, δηλαδή έναν κύκλο χωρίς χορδές με n κορυφές $(v_1^i, v_2^i, ..., v_n^i)$. Θα καλούμε τις κορυφές του «κυκλικές» κορυφές. Για κάθε ακμή του G, $e = \{v_i, v_j\}$ με $i < j$, δημιουργούμε ένα απλό μονοπάτι P^{ij} μεγέθους $kn + 3$ με $P^{ij} = (v_j^i, u_1^{ij}, .., u_{kn+1}^{ij}, v_i^j)$. Καλούμε τις κορυφές του κορυφές «μονοπατιού». Έστω ότι η ένωση όλων των κορυφών που έχουν οριστεί ως «κυκλικές» και «μονοπατιών» αποτελούν το γράφο H. Θα πάρουμε την τριάδα $(H, t := kn + \binom{k}{2}(kn+1), m := kn + \binom{k}{2}(kn+2))$ σαν ένα στιγμιότυπο του $3DkS$, στο οποίο αναζητούμε έναν υπογράφο με t κορυφές και m ακμές.Η κατασκευή παίρνει πολυωνυμικό χρόνο και κάθε κορυφή του H έχει το πολύ 3 γείτονες.

Θα δείξουμε ότι στον G υπάρχει κλίκα k κορυφών αν και μόνο αν το H έχει έναν υπογράφο t κορυφών με τουλάχιστον m ακμές.

$\hookrightarrow$ Υποθέτουμε ότι ο G έχει μια κλίκα k κορυφών. Χωρίς να χάνεται η γενικότητα, έστω ότι αυτές είναι οι $v_1, v_2, ..., v_k$. Τότε ο υπογράφος του H που επάγεται απο την ένωση των

$$C^1, C^2, ..., C^k, \bigcup_{1 \leq i < j \leq k} P^{ij}$$

έχει t κορυφές και m ακμές.

$\hookleftarrow$ Υποθέτουμε ότι ο H έχει ένα επαγόμενο υπογράφο με $kn + \binom{k}{2}(kn+1)$ κορυφές και τουλάχιστον $kn + \binom{k}{2}(kn+2)$ ακμές. Έστω ότι $S = (V(S), E(S))$ δηλώνει έναν τέτοιον υπογράφο. Σε ότι ακολουθεί παρακάτω στην απόδειξη όλος ο συλλογισμός σχετικά με τους βαθμούς των κορυφών θα είναι σχετικός με τον S.

Θα χρησιμοποιήσουμε το ακόλουθο βοηθητικό γεγονός:
Ισχυρισμός: Δεν υπάρχουν δύο διαφορετικές κορυφές $\upsilon_1, \upsilon_2 \in V(S)$ βαθμού 3 και μια κορυφή $u \in V(S)$ βαθμού 1, τέτοιες που $\upsilon_1 \rightsquigarrow u$, $\upsilon_2 \rightsquigarrow u$. Εδώ το $\alpha \rightsquigarrow b$ σημαίνει ότι υπάρχει ένα απλό μονοπάτι $\alpha \equiv x_1, x_2, ..., x_k \equiv b \in V(S)$ τέτοιο που ο βαθμός κάθε κορυφής κατα μήκος του μονοπατιού (εκτός των α και b) είναι 2.

Καλούμε μια κορυφή υ βαθμού 3 στο S «καλή» εάν δεν υπάρχει καμία κορυφή u βαθμού 1 στο S τέτοια που $\upsilon \rightsquigarrow u$.

Πρόταση 1 *Έστω $U \subseteq V(S)$ το σύνολο των «καλών» κορυφών βαθμού 3 στο S. Τότε $|U| \geq 2\binom{k}{2}$.*

Απόδειξη: Έστω ότι $m_1 = |U| < 2\binom{k}{2}$. Υπολογίζουμε το D, δηλαδή το άθροισμα των βαθμών των κορυφών του S. Εξ ορισμού , $D = 3n_3 + 2n_2 + n_1$, όπου n_i είναι ο αριθμός των κορυφών βαθμού i. $n_3 = m_1 + m_2$, όπου m_2 είναι ο αριθμός των κορυφών βαθμού 3 που δεν είναι «καλές». Λόγω του παραπάνω ισχυρισμού $m_2 \leq n_1$, έτσι έχουμε:

$$D = 3m_1 + 3m_2 + 2n_2 + n_1 = 3m_1 + 2m_2 + 2n_2 + m_2 + n_1 \leq$$

$$3m_1 + 2m_2 + 2n_2 + 2n_1 \leq 2t + m_1 <$$

$$2\binom{k}{2} + 2(kn + \binom{k}{2}(kn+1)) = 2(kn + \binom{k}{2}(kn+2)) = 2m.$$

Άρα ο αριθμός των ακμών του S είναι $\frac{D}{2} < m$, κάτι που αντιτίθεται στην επιλογή του S. Συνεπώς υποθέτοντας ότι $m_1 = |U| < 2\binom{k}{2}$ καταλήξαμε σε αντίφαση, άρα η πρόταση ισχύει. $\diamondsuit$

Λόγω της κατασκευής του H, μόνο οι κορυφές που ανήκουν στο $\bigcup_{i=1}^{n} C^i$ μπορεί να έχουν βαθμό 3. Επομένως κάθε κορυφή του U βρίσκεται σε κάποιον απο αυτούς τους κύκλους.

Έστω $\upsilon \in U$. Εκ κατασκευής υπάρχουν i, j τέτοια που $\upsilon \in C^i, \upsilon \in P^{ij}$. Έστω u_1 είναι ο γείτονας του υ σε αυτό το μονοπάτι (υπάρχει μόνο μια τέτοια κορυφή). $Deg(\upsilon) = 3$ επομένως $u_1 \in S$. Λόγω της επιλογής του υ και της δομής του S, $deg(u_1) = 2$. Έτσι ο δεύτερος γείτονας του u_1, ο u_2 ανήκει κι αυτός στο S. Προχωρώντας με αυτόν τον τρόπο ανακαλύπτουμε ότι όλα τα μονοπάτια P^{ij} ανήκουν στο S. Παρατηρούμε ότι το S περιέχει το πολύ $\binom{k}{2}$ τέτοια μονοπάτια, διότι το μήκος κάθε μονοπατιού (χωρίς τις «κυκλικές» κορυφές του) είναι $kn + 1$.

Απο την άλλη μεριά, όπως αναφέρθηκε παραπάνω, το S πρέπει να περιέχει τουλάχιστον ένα τέτοιο μονοπάτι για κάθε κορυφή απο το U και κάθε μονοπάτι περιέχει το πολύ δύο κορυφές απο το U. Άρα το S περιέχει ακριβώς $\binom{k}{2}$ τέτοια μονοπάτια και μεταξύ «κυκλικών» κορυφών μόνο κορυφές του U μπορούν να είναι κορυφές «μονοπατιού».

Τώρα επιστρέφουμε στις γειτονικές κορυφές του v στον κύκλο C^i. Το γεγονός ότι το $v \in U$ υπονοεί ότι και οι δύο ανήκουν στο S και ότι καθεμία έχει βαθμό τουλάχιστον 2. Για καθεμία απο αυτές υπάρχουν δύο πιθανότητες:

- Να είναι κορυφή «μονοπατιού». Τότε, όπως αναφέρθηκε παραπάνω, πρέπει να ανήκει στο U και άρα να έχει βαθμό 3. Έτσι οι «κυκλικές» γειτονικές κορυφές της ανήκουν στο S.

- Να μην είναι κορυφή «μονοπατιού». Τότε ο βαθμός της είναι ακριβώς 2 και έτσι οι «κυκλικές» γειτονικές κορυφές της ανήκουν επίσης στο S.

Προχωρώντας κατα μήκος του C^i θα πάρουμε τελικά ότι όλες οι κορυφές του ανήκουν στο S. Κάθε τέτοιος κύκλος C^i αντιστοιχεί στην κορυφή v_i του G και κάθε P^{ij} αντιστοιχεί στην ακμή $\{v_i, v_j\}$ του G. Όπως προκύπτει απο τα παραπάνω, κατά μήκος κάθε τέτοιας «ακμής» το S περιέχει και τις δύο «κορυφές» της «ακμής». Έχουμε $\binom{k}{2}$ «ακμές». Άρα το S πρέπει να περιέχει τουλάχιστον k διαφορετικές «κορυφές». Αλλά κάθε τέτοια «κορυφή» σημαίνει n κορυφές του S και ένα αριθμήσιμο επιχείρημα μας επιτρέπει να συμπεράνουμε ότι το S αποτελείται ακριβώς απο k τέτοιες «κορυφές»-κύκλους και ότι μεταξύ κάθε δύο «κορυφών» υπάρχει μια «ακμή». Συνεπώς το G έχει μια k-κορυφών κλίκα. $\diamondsuit$

Θεώρημα 11 *Το DkS παραμένει NP-hard για διμερείς (bipartite) γράφους με μέγιστο βαθμό 3.*

Απόδειξη: Διασπώντας κάθε ακμή του H με μια πρόσθετη κορυφή και αλλάζοντας ασήμαντα τα t και m παίρνουμε αμέσως το παραπάνω θεώρημα. $\diamondsuit$

Στην ενότητα 5.2 δείξαμε ότι για να αποδείξουμε δυσκολία στα προσεγγιστικά αποτελέσματα είναι πιο ελπιδοφόρο να εξετάσουμε αραιά στιγμιότυπα του DkS. Δυστυχώς παρά το χαμηλό βαθμό αποδείξαμε, με αναγωγή απο το Clique, ότι το πρόβλημα παραμένει NP-hard. Επομένως δεν περιμένουμε να πάρουμε αποτελέσματα για δυσκολία στην προσέγγιση του DkS στους ειδικούς αραιούς γράφους που κατασκευάσαμε για την απόδειξη του παραπάνω θεωρήματος.

Το παραπάνω θεώρημα έχει το ακόλουθο πόρισμα.

Πόρισμα 1 *Για κάθε* $1 > \epsilon > 0$, *το* DkS *παραμένει NP-complete όταν* $m = k + k^{\epsilon}$.

Απόδειξη: Υποθέτουμε ότι ο

$$z = \frac{\binom{k}{2}^{\frac{1}{\epsilon}} - kn}{\binom{k}{2}}$$

είναι φυσικός αριθμός (αλλιώς θεωρούμε τον $\ulcorner z \urcorner$). Έστω ότι το μήκος των P^{ij}, δηλαδή των μονοπατιών στην απόδειξη ορθότητας της αναγωγής (παράγραφος 5.2) είναι $z+3$ αντί $kn+3$. Τότε $m = k + k^{\epsilon}$. $\diamondsuit$

Αυτό το αποτέλεσμα βελτιώνει το αποτέλεσμα που αναφέραμε στην παράγραφο 5.1, δηλαδή ότι το DkS είναι NP-hard όταν το $m = k^{1+\epsilon}$.

6 $O(n^{1/3})$-προσεγγιστικός αλγόριθμος για το DkS

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

- $A(G, k)$ είναι η πυκνότητα του υπογράφου που επάγεται απο 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 απο k κορυφές, δηλαδή η βέλτιστη τιμή.

Θεώρημα 12 *Υπάρχει πολυωνυμικού χρόνου αλγόριθμος A που προσεγγίζει το DkS με έναν παράγοντα $2n^{1/3}$, δηλαδή για κάθε γράφο G και κάθε k, $1 \le k \le n$,*

$$\boldsymbol{A(G,k) \geq \frac{d^*(G,k)}{2n^{1/3}}}.$$

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 απο τις τρεις 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 $\Delta(G) \geq d_H \geq d^*(G, k)$
- $deg(v, S)$ είναι ο αριθμός των ακμών που συνδέουν την κορυφή v 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 μια ακολουθία $l + 1$ κορυφών στην οποία οι διαδοχικές κορυφές είναι παρακείμενες μιας ακμής, (δηλαδή ο περίπατος ακολουθεί l ακμές και οι κορυφές δεν είναι απαραίτητα διαφορετικές).

- $W_l(u,v)$ είναι ο αριθμός των περιπάτων μήκους l που ξεκινούν απο την κορυφή u και καταλήγουν στην κορυφή v.

6.1 Διαδικασία A_1

Χωρίς να χάνεται η γενικότητα μπορούμε να υποθέσουμε ότι ο G περιέχει τουλάχιστον $k/2$ ακμές.

ΔΙΑΔΙΚΑΣΙΑ 1
1.Επέλεξε αυθαίρετα $k/2$ ακμές απο τον G.
2.Επέστρεψε το σύνολο των κορυφών που προσπίπτουν σε αυτές τις ακμές, προσθέτοντας αυθαίρετα κορυφές εάν το μέγεθος του συνόλου είναι μικρότερο του k.

Ολοφάνερα,

$$\boldsymbol{A_1(G,k) \geq 1.}$$

6.2 Διαδικασία A_2

ΔΙΑΔΙΚΑΣΙΑ 2
1.Ταξινόμησε τις κορυφές βάση του βαθμού τους.
Έστω H το σύνολο των $\frac{k}{2}$ κορυφών με το μεγαλύτερο βαθμό στο G.
2.Ταξινόμησε τις εναπομένουσες κορυφές βάση του αριθμού των γειτόνων τους στο H.
Έστω C το σύνολο των $\frac{k}{2}$ κορυφών στο $G \backslash H$ με το μεγαλύτερο αριθμό γειτόνων στο H.
3.Επέστρεψε το $H \bigcup C$.

Λήμμα 11 *Η διαδικασία 2 επιστρέφει έναν υπογράφο επαγόμενο απο k κορυφές, που ικανοποιεί*

$$\boldsymbol{A_2(G,k) \geq \frac{kd_H}{2n},}$$

όπου d_H ο μέσος βαθμός μιας κορυφής στο H.

Απόδειξη: Έστω ότι m_1 δηλώνει τον αριθμό των ακμών που έχουν και τα δύο άκρα τους μέσα στο H. Τότε

$$cut(H, V\setminus H) = d_H \cdot |H| - 2m_1 = d_H \frac{k}{2} - 2m_1 \geq 0.$$

Λόγω του greedy κανόνα επιλογής του C, τουλάχιστον ένα $\frac{|C|}{|V\setminus H|} > \frac{k}{2n}$ κλάσμα των ακμών του cut περιέχονται στο $H \bigcup C$ (βλέπε παρατήρηση 9). Έτσι ο συνολικός αριθμός ακμών στον υπογράφο που επάγεται απο το $H \bigcup C$ είναι τουλάχιστον

$$(d_H \frac{k}{2} - 2m_1)\frac{k}{2n} + m_1 \geq d_H \frac{k^2}{4n}$$

$\diamondsuit$

Αφού $d_H \geq d^*(G,k)$ έχουμε ότι $A_2(G,k) \geq \frac{d^*(G,k)}{2n/k}$. Άρα η greedy διαδικασία A_2 προσεγγίζει το $d^*(G,k)$ με ένα λόγο το πολύ $2n/k$.

Παρατήρηση 9 *Θα δείξουμε ότι όντως ένα κλάσμα* $\frac{|C|}{|V\setminus H|}$ *των ακμών του* cut *περιέχονται στο* $H \bigcup C$. *Έστω* C' *το σύνολο των κορυφών που ανήκουν στο* cut *αλλά όχι στο* C *και έστω* C'' *το σύνολο των κορυφών εντός του* $|V\setminus H|$ *και εκτός του* cut, *τότε* $|V\setminus H| = |C \bigcup C' \bigcup C''|$. *Θα πρέπει να ισχύει*

$$\sum_{j=1}^{|C|} \delta_j \geq \frac{|C|}{|V\setminus H|} cut \Rightarrow \sum_{j=1}^{|C|} \delta_j \geq \frac{|C|}{|V\setminus H|}(\sum_{j=1}^{|C|} \delta_j + \sum_{r=1}^{|C'|} \delta_r) \Rightarrow$$

$$\sum_{j=1}^{|C|} \delta_j/|C| \geq (\sum_{j=1}^{|C|} \delta_j + \sum_{r=1}^{|C'|} \delta_r)/|C \bigcup C'| \geq (\sum_{j=1}^{|C|} \delta_j + \sum_{r=1}^{|C'|} \delta_r)/|C \bigcup C' \bigcup C''| \Rightarrow$$

$$\sum_{j=1}^{|C|} \delta_j/|C| \geq (\sum_{j=1}^{|C|} \delta_j + \sum_{r=1}^{|C'|} \delta_r)/(|C| + |C'|) \Rightarrow |C'| \cdot \sum_{j=1}^{|C|} \delta_j \geq |C| \cdot \sum_{r=1}^{|C'|} \delta_r.$$

Λόγω ταξινόμησης $\delta_1 \geq ... \geq \delta_C \geq ... \geq \delta_{C'}$.
Άρα

$$\sum_{j=1}^{|C|} \delta_j \leq |C| \cdot \delta_1$$

και

$$\sum_{r=1}^{|C'|} \delta_r \geq |C'| \cdot \delta_{C'}.$$

Συνεπώς αφού $\delta_1 \geq \delta_{C'}$ *και*

$$|C'| \cdot |C| \cdot \delta_1 \geq |C'| \cdot \sum_{j=1}^{|C|} \delta_j \geq |C| \cdot \sum_{r=1}^{|C'|} \delta_r \geq |C| \cdot |C'| \cdot \delta_{C'}$$

η ανισότητα ισχύει.

6.3 Διαδικασία A_3

ΔΙΑΔΙΚΑΣΙΑ 3

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

2. Κατασκεύασε έναν υποψήφιο γράφο H^{υ} για κάθε κορυφή υ στο G ως εξής:

- ***α.*** Ταξινόμησε τις κορυφές του G σε φθίνουσα σειρά, βάση του αριθμού των περιπάτων τους μήκους 2 προς την κορυφή υ, δηλαδή $W_2(\upsilon, w_1) \geq W_2(\upsilon, w_2) \geq ....$ Έστω P_h^{υ} το σύνολο των $k/2$ 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 απο το $P_h^{\upsilon} \bigcup B^{\upsilon}$. Εάν ο H^{υ} περιέχει λιγότερες απο k κορυφές συμπλήρωσε αυθαίρετα μέχρι τις k κορυφές.

3. Επέλεξε τον πυκνότερο υποψήφιο γράφο H^{υ} σαν έξοδο.

Αναλύουμε τον προσεγγιστικό λόγο της διαδικασίας. Έστω G^* ο βέλτιστος υπογράφος και $deg^*(\upsilon)$ ο βαθμός του υ στον G^*. Καταρχήν παρατηρούμε ότι ο αριθμός των περιπάτων 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 μια κορυφή υ η οποία είναι το άκρο τουλάχιστον $(d^*(G,k))^2$ περιπάτων μήκους 2 στον G^*. Λόγω της greedy κατασκευής του P_h^{υ} υπάρχουν τουλάχιστον $(d^*(G,k))^2/2$ περίπατοι μήκους 2 μεταξύ της υ και των κορυφών του P_h^{υ}. Οι κορυφές του B^{υ} έχουν τουλάχιστον$(d^*(G,k))^2/2$ ακμές που τις συνδέουν με το P_h^{υ} εάν $deg(\upsilon) \leq k/2$ και τουλάχιστον $(d^*(G,k))^2/4deg(\upsilon)$

ακμές που τις συνδέουν με το P_h^v αλλιώς. Αφού δεν απαιτούμε το P_h^v και το B^v να είναι ξένα, κάθε ακμή μπορεί να έχει μετρηθεί δύο φορές. Έπεται ότι ο H^v περιέχει τουλάχιστον $\min[(d^*(G,k))^2/4, (d^*(G,k))^2/8\Delta(G)]$ ακμές, όπου $\Delta(G)$ είναι ο μέγιστος βαθμός του γράφου. Άρα

$$\boldsymbol{A_3(G,k) \geq \frac{(d^*(G,k))^2}{2\max[k, 2\Delta(G)]}}.$$

6.4 Αλγόριθμος A

Ο Αλγόριθμος A εφαρμόζει τις τρείς διαδικασίες που περιγράψαμε παραπάνω και αποδίδει τον πυκνότερο απο τους τρείς υπογράφους που λαμβάνονται απο την καθεμία απο αυτές τις διαδικασίες. Οι Διαδικασίες 1 και 2 εφαρμόζονται στον αρχικό γράφο G ενώ η Διαδικασία 3 εφαρμόζεται στο γράφο G_l που επάγεται απο τις κορυφές του συνόλου $V\backslash H$, όπου H είναι το σύνολο των $k/2$ κορυφών με το μεγαλύτερο βαθμό στον G, όπως ορίστηκε στη δεύτερη διαδικασία. Άρα $\Delta(G_l) \leq d_H(G)$.
Για το ακόλουθο λήμμα υποθέτουμε ότι $k \leq 2n/3$. Η υπόθεση αυτή μπορεί να γίνει χωρίς να χάνεται η γενικότητα, διότι για $k \geq 2n/3$ η δεύτερη διαδικασία προσεγγίζει το DkS με ένα λόγο όχι χειρότερο απο 3.

Λήμμα 12 *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 $d^*(G,k)k/2 - m - l$. Για να αποδείξουμε το λήμμα χρειάζεται να δείξουμε ότι η δεύτερη διαδικασία επιστρέφει μια λύση με τουλάχιστον $(m+l)/2$ ακμές. Στην πραγματικότητα η λύση έχει τουλάχιστον τουλάχιστον $m + l/2$ ακμές. Αυτό συμβαίνει γιατί περιέχει τις m ακμές που είναι εσωτερικές στο $V(G^*) \bigcap H$ και πρέπει να υπάρχουν τουλάχιστον $l/2$ ακμές μεταξύ του C και του H αφού τουλάχιστον μια πιθανή επιλογή του C προσφέρει τόσες ακμές (δηλαδή παίρνοντας το C να περιέχει τις $k/2$ κορυφές του $V(G^*)\backslash H$ με το μεγαλύτερο αριθμό ακμών στο H). $\diamondsuit$

Άρα απο την απόδοση που εγγυώνται οι τρείς διαδικασίες έχουμε

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

Για να αποδείξουμε το Θεώρημα 12 μπορούμε να υποθέσουμε ότι $d_2 \le \frac{d^*(G,k)}{n^{1/3}}$ (αλλιώς η έξοδος της δεύτερης διαδικασίας επιτυγχάνει το ποθητό 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 1, 2 και 3. Άρα έχουμε

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

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

7 Βελτιώνοντας πάνω απο το $O(n^{1/3})$

Ο προσεγγιστικός λόγος για τον αλγόριθμο A ήταν άνω φραγμένος σαν ο γεωμετρικός μέσος τριών προσεγγιστικών λόγων. Για να δώσει ο αλγόριθμος A έναν προσεγγιστικό λόγο τόσο κακό όσο το $\Omega(n^{1/3})$ πρέπει και οι τρείς διαδικασίες 1, 2 και 3 να δώσουν έναν προσεγγιστικό λόγο $\Theta(n^{1/3})$. Αυτό γίνεται μόνο εάν $d^*(G,k) = \Theta(n^{1/3})$, $kd_H = \Theta(n)$ και $\max[k, d_H] = \Theta(n^{2/3})$. Εάν κάποια απο τις τρεις παραπάνω συνθήκες παραβιάζεται απο έναν παράγοντα n^ϵ, τότε ο προσεγγιστικός λόγος είναι $O(n^{1/3-\epsilon/2})$. Οι παραπάνω χειρότερης περίπτωσης συνθήκες ικανοποιούνται μόνο στις δύο περιπτώσεις:

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

Παρουσιάζουμε δύο πρόσθετες διαδικασίες, η καθεμία απο τις οποίες δίνει έναν προσεγγιστικό λόγο καλύτερο απο $O(n^{1/3})$ σε κάποια απο τις παραπάνω περιπτώσεις. Αυτό μαζί με τον αλγόριθμο A εγγυάται έναν προσεγγιστικό λόγο $O(n^{1/3-\epsilon})$, για κάποιο $\epsilon > 0$, για το DkS πρόβλημα.

Θεώρημα 13 *Υπάρχει πολυωνυμικού χρόνου αλγόριθμος B που προσεγγίζει το DkS με έναν παράγοντα $n^{1/3-\epsilon}$, για κάποιο $\epsilon > 0$. Δηλαδή για κάθε γράφο G και για κάθε $1 \leq k \leq n$,*

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

Λήμμα 13 *Έστω G ένας γράφος με n κορυφές και μέσο βαθμό d. Υπάρχουν δύο κορυφές $v_i, v_j \in V$ τέτοιες που*

$$W_l(v_i, v_j) \geq \frac{d^l}{n}.$$

Θα αναλύσουμε παρακάτω καθεμία απο τις παραπάνω δύο περιπτώσεις. Και στις δύο περιπτώσεις υποθέτουμε ότι το ακόλουθο βήμα έχει εκτελεστεί:
Αφαιρούμε το H, δηλαδή το σύνολο των $k/2$ μεγαλύτερου βαθμού κορυφών, και παραμένουμε με το γράφο G_l.

Θα χρησιμοποιήσουμε το γεγονός ότι $\Delta(G_l) \leq d_H$. Επιπλέον υποθέτουμε ότι το $d^*(G,k)$ παραμένει σχεδόν αμετάβλητο μετά το παραπάνω βήμα. Αυτή η υπόθεση μπορεί να γίνει χωρίς να χάνεται η γενικότητα επειδή αποτυγχάνει να διατηρηθεί μόνο εάν η δεύτερη διαδικασία επιτύχει έναν προσεγγιστικό λόγο καλύτερο απο $n^{1/3-\epsilon}$. Έστω G_l^* ο k-κορυφών επαγόμενος υπογράφος με τη μεγαλύτερη πυκνότητα στον G_l.

7.1 Διαδικασία A_4

Παρουσιάζουμε μια διαδικασία για την πρώτη απο τις παραπάνω περιπτώσεις, δηλαδή όταν $d^*(G,k) = \Theta(n^{1/3})$, $k = \Theta(n^{1/3})$, $d_H = \Theta(n^{2/3})$. Η ανάλυση της βασίζεται στο παρακάτω λήμμα.

Λήμμα 14 *Υπάρχουν δύο κορυφές (όχι απαραίτητα ξένες) $v_i, v_j \in V$ τέτοιες που ο υπογράφος του G_l^* που επάγεται απο το $N(v_i) \bigcup N(v_j)$ έχει τουλάχιστον $(d^*(G,k))^3/2k$ ακμές.*

Απόδειξη: Θεωρούμε το λήμμα 13 με $l = 3$ και το εφαρμόζουμε για το γράφο G_l^*. Έστω v_i, v_j οι δύο κορυφές με $W_3[v_i, v_j] \geq (d^*(G,k))^3/k$. Θεωρούμε το πολυσύνολο των μεσαίων ακμών όλων των μήκους 3 περιπάτων μεταξύ των v_i και v_j. Μια ακμή μπορεί να εμφανίζεται σε αυτό το πολυσύνολο το πολύ δύο φορές (δηλαδή μια σαν (v_k, v_l) και μια σαν (v_l, v_k), εάν και το v_k και το v_l είναι στο $N(v_i) \bigcap N(v_j)$). $\diamondsuit$

ΔΙΑΔΙΚΑΣΙΑ 4

1. Για όλα τα ζεύγη κορυφών $v_i, v_j \in G_l$ εφάρμοσε τον αλγόριθμο $A(N(v_i) \bigcup N(v_j), k)$.

2. Επέστρεψε τον πυκνότερο απο τους υπογράφους που επιστρέφονται απο οποιαδήποτε απο τις $O(n^2)$ εφαρμογές του αλγόριθμου A.

Λήμμα 15 *Η απόδοση που εγγυάται η διαδικασία 4 ικανοποιεί την $A_4(G_l,k) \geq A(G',k)$, όπου G' είναι ένας γράφος, με το πολύ $n' = 2d_H$, κορυφές που περιέχει έναν k-κορυφών υπογράφο με μέσο βαθμό τουλάχιστον $d' = (d^*(G,k))^3/k^2$.*

Απόδειξη: Έστω v_i και v_j οι δύο κορυφές του G_l^* στις οποίες εφαρμόζεται το λήμμα 14. Τότε το $N(v_i) \bigcup N(v_j)$ περιέχει έναν k-κορυφών επαγόμενο υπογράφο με τουλάχιστον $(d^*(G,k)^3/2k$ ακμές, υπονοώντας ότι ο μέσος βαθμός είναι τουλάχιστον $(d^*(G,k)^3/k^2$. Επιπλέον $|N(v_i) \bigcup N(v_j)| \leq 2d_H$. $\diamondsuit$

Για την πρώτη περίπτωση παίρνουμε $A_4(G_l,k) \geq A(G',k)$ με $n' = O(n^{2/3})$ και $d' = \Theta(n^{1/3}) = \Theta(d^*(G,k))$. Ο Αλγόριθμος A επιτυγχάνει προσεγγιστικό λόγο $O((n')^{1/3})$, ο οποίος είναι σίγουρα καλύτερος απο $O(n^{1/3})$.

7.2 Διαδικασία A_5

Αναλύουμε τώρα τη δεύτερη περίπτωση με παραμέτρους $d^*(G,k) = \Theta(n^{1/3})$, $k = \Theta(n^{2/3})$, $d_H = \Theta(n^{1/3})$ και άρα $\Delta(G_l) = O(n^{1/3})$. Σε αυτή την ενότητα οι παραπάνω παράμετροι είναι καθορισμένες. Παρουσιάζουμε ένα σκιαγράφημα της διαδικασίας που χρησιμοποιείται στη συγκεκριμένη περίπτωση. Αργότερα θα συμπληρώσουμε τις λεπτομέρειες που λείπουν (το πως εκτελείται το πρώτο βήμα). Σε ότι ακολουθεί η $\epsilon > 0$ είναι μια μικρή καθολική σταθερά.

ΔΙΑΔΙΚΑΣΙΑ 5

1. Επέλεξε έναν υπογράφο που επάγεται απο $O(n^{2/3})$ κορυφές, με μέσο βαθμό $\Omega(n^\epsilon)$. Αφαίρεσε το απο το G_l για να λάβεις ένα νέο γράφο.

2. Επανέλαβε το ακόλουθο βήμα επιλογής υποσυνόλων κορυφών και αφαίρεσης τους απο το γράφο της εισόδου μέχρι μια απο τις ακόλουθες συνθήκες παύσης συμβούν:

- ***α.*** Ένα σύνολο $n^{2/3}$ κορυφών έχει επιλεγεί.
- ***β.*** Ο εναπομένων γράφος δεν περιέχει πλέον έναν $\Theta(n^{2/3})$-κορυφών επαγόμενο υπογράφο με μέσο βαθμό $\Omega(n^{1/3})$.

3. Επέστρεψε τον υπογράφο που επάγεται απο την ένωση των κορυφών που επιλέχθηκαν απο την εφαρμογή του πρώτου βήματος. Εάν συμβεί η συνθήκη παύσης 2β, συμπλήρωσε μέχρι τις $n^{2/3}$ κορυφές με ένα greedy κανόνα όμοιο με αυτόν που χρησιμοποιήθηκε στη δεύτερη διαδικασία για την επιλογή του C.

Για τις παραμέτρους της δεύτερης περίπτωσης, η διαδικασία 5 εγγυάται έναν προσεγγιστικό λόγο $O(n^{1/3}/n^\epsilon)$. Αυτό είναι ξεκάθαρο εάν η πρώτη συνθήκη παύσης συμβεί, διότι τότε ο μέσος βαθμός του υπογράφου που βρίσκει είναι $\Omega(n^\epsilon)$. Ισχύει το ίδιο και στην περίπτωση που συμβεί η δεύτερη συνθήκη παύσης αφού έχουν επιλεχθεί $n^{2/3}/2$ κορυφές. Η μόνη μη τετριμμένη περίπτωση είναι όταν η δεύτερη συνθήκη παύσης συμβεί πριν επιλεχθούν οι $n^{2/3}/2$ κορυφές, αλλά τότε ο προσεγγιστικός λόγος μπορεί να αποδειχθεί ότι είναι σταθερά. Ο λόγος είναι ότι σε αυτή την περίπτωση, όλες εκτός απο ένα μικρό κλάσμα των ακμών του G_l^* έχουν τουλάχιστον ένα άκρο τους στις επιλεγμένες κορυφές. Όσο υπάρχουν $\Omega(n)$ ακμές του G_l^* που δεν έχουν και τα δύο άκρα τους στις επιλεγμένες κορυφές, πρέπει να υπάρχει κάποια κορυφή του G_l^* η οποία δεν έχει επιλεχθεί ακόμη και η οποία έχει $\Omega(n^{1/3})$ γείτονες στις επιλεγμένες κορυφές. Ο greedy κανόνας για την επιλογή του C εξασφαλίζει ότι μια κορυφή βαθμού

$\Omega(n^{1/3})$ θα επιλεχθεί. Ο μέσος βαθμός του τελικού υπογράφου είναι $\Omega(n^{1/3})$.

Θα αναφερθούμε τώρα στο πρώτο βήμα της διαδικασίας 5 που το αφήσαμε ανεξήγητο. Η ανάλυση του βασίζεται στους περιπάτους μήκους 5. Για τις παραμέτρους της δεύτερης περίπτωσης εφαρμόζουμε το λήμμα 13 στον G_l^* και λαμβάνουμε:

Ισχυρισμός: Υπάρχουν δύο κορυφές u, v στον G_l^* με

$$W_5(u,v) \geq \frac{(d^*(G,k))^5}{k} = \Omega(n).$$

Έστω u και v οι δύο κορυφές με $\Omega(n)$ περιπάτους μήκους 5 απο το u στο v. Έστω N_1, (N_2, N_3, N_4 αντίστοιχα) το σύνολο των κορυφών που είναι πρώτες (δεύτερες, τρίτες, τέταρτες αντίστοιχα) κατα μήκος αυτών των περιπάτων. Έστω F ο υπογράφος που επάγεται απο την ένωση αυτων των συνόλων.

Παρατηρούμε ότι μια κορυφή w μπορεί να εμφανίζεται σε πολλά απο αυτά τα σύνολα, δηλαδή η w μπορεί να είναι γείτονας της v αλλά μπορεί επίσης να βρίσκεται σε ένα μονοπάτι μήκους 2 απο τη v. Το γεγονός αυτό μπορεί να γίνει η αιτία κάποιες ακμές να μετρηθούν πολλές φορές και να επηρεαστεί η σταθερά στην αναλυσή μας. Αυτό το αποτέλεσμα τακτοποιείται απο τις O, Θ, Ω συναρτήσεις που χρησιμοποιούμε.

Απο την υπόθεση ότι $d_H = O(n^{1/3})$ το βήμα 1 της διαδικασίας 5 εφαρμόζεται σε γράφους με μέγιστο βαθμό $\Delta = O(n^{1/3})$. Έπεται ότι $|N_1|, |N_4| = O(n^{1/3})$ και $|N_2|, |N_3| = O(n^{2/3})$.

$1^η$ ΥΠΟΠΕΡΙΠΤΩΣΗ

Κάνουμε κάποιες υποθέσεις όσον αφορά τη δομή του F. Κάθε υπόθεση αιτιολογείται απο το γεγονός ότι μπορεί είτε να επιβληθεί στο F είτε να βρεθεί ένας υπογράφος με μέσο βαθμό $\Omega(n^{\epsilon})$.

Υπόθεση 1: $cut(N_2, N_3) < n^{2/3+\epsilon}$.

Αιτιολόγηση: Αλλιώς παίρνουμε το $N_2 \bigcup N_3$.

Υπόθεση 2: Για κάθε $w \in N_2$, $W_3(w,v) \leq n^{1/3+\epsilon}$ και για κάθε $w \in N_3$, $W_3(w,v) \leq n^{1/3+\epsilon}$.

Αιτιολόγηση: Θεωρούμε την περίπτωση που το $w \in N_2$ και $W_3(w,v) > n^{1/3+\epsilon}$. Παρατηρούμε ότι όλοι οι περίπατοι μήκους 3 μεταξύ του w και του v πρέπει να περνάν δια μέσου του N_3 και του N_4. Θεωρούμε το γράφο που επάγεται απο τους γείτονες του w στο N_3 και το σύνολο N_4. Αφού το w έχει $O(n^{1/3})$ γείτονες στο N_3, ο γράφος αυτός περιέχει $O(n^{1/3})$ κορυφές και

$\Omega(n^{1/3+\epsilon})$ ακμές. Άρα το πρώτο βήμα ολοκληρώθηκε.

Υπόθεση 3: Κάθε ακμή μεταξύ του N_2 και του N_3 βρίσκεται σε τουλάχιστον $\Omega(n^{1/3-2\epsilon})$ περιπάτους απο το v στο u.
Αιτιολόγηση: Αφαιρούμε κάθε ακμή μεταξύ του N_2 και του N_3 που βρίσκεται σε λιγότερους απο $n^{1/3-2\epsilon}$ περιπάτους μήκους 5 απο το v στο u. Αφού ο αριθμός των ακμών μεταξύ του N_2 και του N_3 είναι μικρότερος απο $n^{2/3+\epsilon}$ (απο την πρώτη υπόθεση) «σκοτώνουμε» το πολύ $O(n^{1-\epsilon})$ περιπάτους διατηρώντας ότι $W_5(u,v) = \Omega(n)$.

2^{η} ΥΠΟΠΕΡΙΠΤΩΣΗ

Έστω $e = (w, z)$ μια αυθαίρετη ακμή μεταξύ του $w \in N_2$ και του $z \in N_3$. Απο την υπόθεση 3 η ακμή e βρίσκεται σε $p = \Omega(n^{1/3-2\epsilon})$ περιπάτους απο το u στο v. Ολοφάνερα

$$p = deg(w, N_1) \cdot deg(z, N_4).$$

Έτσι είτε $deg(w, N_1) \geq n^{1/6-\epsilon}$ είτε $deg(z, N_4) \geq n^{1/6-\epsilon}$. Εάν $deg(z, N_4) \geq n^{1/6-\epsilon}$, καλούμε τη z «καλή» κορυφή της e. Αλλιώς καλούμε τη w «καλή» κορυφή.

Στο σημείο αυτό εκκινούμε την ακόλουθη διαδικασία. Η διαδικασία επιλέγει δύο υποσύνολα $S_2 \subseteq N_2$ και $S_3 \subseteq N_3$ «καλών» κορυφών, δηλαδή κορυφών με μεγάλους βαθμούς. Επανέλαβε τα ακόλουθα 3 βήματα:

1. Επέλεξε μια ακμή e μεταξύ του N_2 και του N_3. Έστω w είναι η «καλή» της κορυφή.

2. Εάν $w \in N_2$ πρόσθεσε το w στο S_2, αλλιώς πρόσθεσε το w στο S_3.

3. Αφαίρεσε απο το F όλες τις ακμές μεταξύ του N_2 και του N_3 που πρόσκεινται στη w.

Παρατηρούμε ότι στο βήμα 3 απορρίπτουμε τους περιπάτους μήκους 5 απο το v στο u στο F που διέρχονται απο το w. Υποθέτουμε ότι $w \in N_2$. Απο την υπόθεση 2, $W_3(w, v) \leq n^{1/3+\epsilon}$. Ο αριθμός των περιπάτων μεταξύ του u και του w (οι οποίοι είναι ίσοι με $deg(w, N_1)$) είναι άνω φραγμένος απο το $n^{1/3}$. Έτσι ο αριθμός των περιπάτων μεταξύ του v και u που διέρχονται απο το w είναι φραγμένος απο $O(n^{1/3}) \cdot n^{1/3+\epsilon} = n^{2/3+\epsilon}$.

Αφού έχουμε $\Omega(n)$ περιπάτους μεταξύ του v και του u και αφού κάθε επανάληψη αφαιρεί μόνο $n^{2/3+\epsilon}$ απο αυτούς, ο αριθμός των επαναλήψεων μπορεί να επιλεχθεί να είναι $\Theta(n/n^{2/3+\epsilon}) = \Theta(n^{1/3-\epsilon})$. Έτσι ο συνολικός αριθμός «καλών» κορυφών που βρίσκονται απο τον αλγόριθμο είναι $\Theta(n^{1/3-\epsilon})$.

Υποθέτουμε ότι $|S_2| \geq |S_3|$. Θεωρούμε τον υπογράφο που επάγεται απο το $S_2 \bigcup N_1$. Περιέχει $O(n^{1/3})$ κορυφές, απο τις οποίες $\Theta(n^{1/3-\epsilon})$ κορυφές έχουν βαθμό τουλάχιστον $deg(w, N_1) \geq n^{1/6-2\epsilon}$. Έτσι ο μέσος βαθμός είναι $\Omega(n^{1/6-2\epsilon}) \geq n^{\epsilon}$ για $\epsilon \leq 1/18$. Έτσι λαμβάνουμε το εξής:

Λήμμα 16 *Για τις παραμέτρους* $k = \Theta(n^{2/3})$, $d^*(G,k) = \Theta(n^{1/3})$ *και* $d_H = \Theta(n^{1/3})$ *η διαδικασία 5 επιτυγχάνει προσεγιστικό λόγο* $O(n^{5/18})$.

ΑΛΓΟΡΙΘΜΟΣ Β

Ο Αλγόριθμος Β εφαρμόζει τον Αλγόριθμο A και τις διαδικασίες 4 και 5. Για να δούμε ότι επιτυγχάνει προσεγγιστικό λόγο $O(n^{1/3-\epsilon})$ για κάποιο $\epsilon > 0$, παρατηρούμε ότι η ανάλυση των διαδικασιών 4 και 5 μπορεί να αντιστέκεται σε μικρές αλλαγές στις παραμέτρους της εισόδου. Για παράδειγμα, εάν $d_H \leq n^{1-6\epsilon}$ και $d^*(G,k) \geq k/n^{\epsilon/3}$ (που υπονοεί ότι $d' \geq k/n^{\epsilon}$) τότε η διαδικασία 4 έχει προσεγγιστικό λόγο $O(n^{1/3-\epsilon})$.

Παρατήρηση 10 *Στην περίπτωση που* $d^*(G,k) = \Omega(k)$ *είναι πιθανόν να εναλλασσόμαστε μεταξύ των διαδικασιών 2 και 4 και να λάβουμε έναν αλγόριθμο ο οποίος για κάθε δοθέν* ϵ *βρίσκει μια προσέγγιση* $O(n^{\epsilon})$ *για το* DkS *με πολυπλοκότητα* $n^{O(1/\epsilon)}$. *Σε κάθε επανάληψη ο αλγόριθμος πρώτα εφαρμόζει τη διαδικασία 2 και σταματά εάν παράγει έναν υπογράφο με μέσο βαθμό* k/n^{ϵ}. *Εάν η διαδικασία 2 αποτύχει να παράγει έναν τέτοιο υπογράφο τότε αφαιρούνται* $k/2$ *κορυφές με το μεγαλύτερο βαθμό και εφαρμόζεται η διαδικασία 4. Αυτό έχει σαν αποτέλεσμα* $O(n^2)$ *νέα* DkS *προβλήματα που πρέπει να λυθούν, αλλά σε καθένα απο αυτά ο αριθμός των κορυφών έχει ελαττωθεί κατά έναν παράγοντα* $\Omega(n^{\epsilon})$. *Τουλάχιστον ένα απο αυτά τα μικρότερα προβλήματα πρέπει να περιέχει έναν επαγόμενο απο κάποιες κορυφές υπογράφο με πυκνότητα προσεγγιστικά* $\Theta(d^*(G,k))$. *Η διαδικασία μπορεί να επαναληφθεί σε καθένα απο τα μικρότερα προβλήματα. Μετά απο* $O(1/\epsilon)$ *επαναλήψεις, απομένουμε με* DkS *προβλήματα σε γράφους με* k *ή λιγότερες κορυφές και παίρνουμε τον πυκνότερο απο αυτούς τους γράφους.*
Μια απλούστερη εκδοχή του παραπάνω επιχειρήματος είδαμε στην ενότητα 5.2 που δείξαμε ότι εάν ο γράφος G *περιέχει μια κλίκα* k*-κορυφών, τότε για κάθε* $\epsilon > 0$ *μπορεί να βρεθεί ένας* k*-υπογράφος με μέσο βαθμό* $(1-\epsilon)(k-1)$, *σε χρόνο* $n^{O(\frac{1}{\epsilon}\log\frac{n}{k})}$.

Παρατήρηση 11 *Στην εκδοχή του* DkS *προβλήματος όπου οι ακμές έχουν θετικά βάρη ο στόχος είναι να βρούμε έναν* k*-κορυφών επαγόμενο υπογράφο* *με μέγιστο συνολικό βάρος ακμών. Το πρόβλημα αυτό μπορεί να αναχθεί στο χωρίς βάρη DkS πρόβλημα με μια απώλεια το πολύ* $O(\log n)$ *στον προσεγγιστικό λόγο. Περιγράφουμε επιγραμματικά πως γίνεται:*

1. *Κλιμακώνουμε τα βάρη των ακμών έτσι ώστε το μέγιστο βάρος πλευράς να είναι* n^2.

2. *Στρογγυλεύουμε προς τα πάνω κάθε βάρος ακμής στην πλησιέστερη δύναμη του δύο.*

3. *Λύνουμε* $2\log n$ *DkS προβλήματα, ένα για κάθε βάρος ακμής, (με όλες τις άλλες ακμές απομακρυσμένες).*

4. *Διαλέγουμε την καλύτερη απο τις* $O(\log n)$ *λύσεις.*

8 Ένα PTAS για το DkS

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

Καταρχήν εκφράζουμε το πρόβλημα σαν ένα τετραγωνικό ακέραιο πρόγραμμα. 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:

1. Χρησιμοποιώντας την εξαντλητική δειγματοληψία υπολογίζουμε τις τιμές των συντελεστών της αντικειμενικής συνάρτησης στη βέλτιστη λύση.

2. 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.

3. Χαλαρώνουμε τους περιορισμούς του ακέραιου γραμμικού προγράμματος, το λύνουμε και χρησιμοποιούμε τη μέθοδο randomized rounding για να μετατρέψουμε τη λύση σε ακέραια.

Τέλος τα παραπάνω αποτελέσματα γενικεύονται για τη χρήση τους σε πολυωνυμικά ακέραια προγράμματα (PIP). Άρα στόχος μας είναι δοθέντος ενός PIP για κάποια προβλήματα βελτιστοποίησης να βρούμε μια προσεγγιστικά βέλτιστη ακέραια λύση. Συνεπώς ο γενικός αλγόριθμος έχει τα εξής τρία βήματα:

1. Χρησιμοποιούμε τη μέθοδο randomized rounding για να μετατρέψουμε κάθε εφικτή κλασματική λύση ενός PIP σε μια εφικτή ακέραια λύση.

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

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

8.1 Γενικά

Ορισμός 16 *Ένας γράφος είναι δ-πυκνός εάν έχει $\delta n^2/2$ ακμές και είναι παντού δ-πυκνός εάν ο ελάχιστος βαθμός του είναι δn.*

Θεώρημα 14 *Υπάρχουν PTASs για πυκνά στιγμιότυπα των ακόλουθων προβλημάτων: $MAX-CUT$, $MAX-DICUT$, $MAX-k-SAT$ για κάθε σταθερά k, DkS για $k=\Omega(n)$, $MAX-HYPERCUT(d)$ για σταθερά d και κάθε $MAX-SNP$ πρόβλημα.*

Ορισμός 17 *Ένα πολυωνυμικό ακέραιο πρόγραμμα (PIP) είναι της μορφής*

$$\max p_0(x_1,...,x_n) \tag{23}$$

$$l_i \le p_i(x) \le u_i \quad (i=1,...,m) \tag{24}$$

$$x_i \in \{0,1\} \quad \forall i \le n \tag{25}$$

όπου $p_0,...,p_m$ είναι πολυώνυμα.
Όταν όλα τα p_i έχουν βαθμό το πολύ d, καλούμε το πρόγραμμα βαθμού d PIP.

Παρατήρηση 12 *Εφόσον τα PIP συμπεριλαμβάνουν τα ακέραια προγράμματα είναι ξεκάθαρο ότι το να επιλύεις $PIPs$ είναι NP-hard.*

Ορισμός 18 *Ένα n-τυχαίων μεταβλητών, βαθμού d πολυώνυμο έχει λειότητα (smoothness) c εάν η απόλυτη τιμή κάθε συντελεστή, του κάθε βαθμού i μονωνύμου είναι το πολύ $c \cdot n^{d-i}$.*

Ορισμός 19 *Ένα c-λείο (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 17 ***(Randomized Rounding)*** *Εάν c και f θετικοί ακέραιοι και $0<\epsilon<1$ τότε το ακόλουθο αληθεύει για κάθε ακέραιο $n \ge 0$. Έστω $y=(y_i)$ διάνυσμα n μεταβλητών, $0 \le y_i \le 1$, που ικανοποιεί έναν ορισμένο γραμμικό περιορισμό $\alpha^T y = b$, όπου κάθε $|\alpha_i| \le c$. 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 18 ***(Randomized Rounding για βαθμού*** d ***πολυώνυμα)*** *Έστω* p *ένα* c*-λείο βαθμού* d *πολυώνυμο. Δίνοντας κλασματικές τιμές* (y_i) *τέτοιες που* $p(y_1, ..., y_n) = b$*, υπόθετουμε ότι η randomized rounding εκτελείται στα* (y_i) *όπως στο παραπάνω λήμμα για να αποδώσει ένα 0,1 διλανυσμα* (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 15 *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 μια 0/1 λύση* z *που ικανοποιεί*

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

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* $t = 4c^2e^2d^2/\epsilon^2 = O(1/\epsilon^2)$*.*
Ο αλγόριθμος μπορεί να γίνει ντετερμινιστικός (derandomization) αυξάνοντας την πολυπλοκότητα μόνο κατά έναν πολυωνυμικό παράγοντα.

Παρατήρηση 13 *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* $\epsilon' < \epsilon/2$*).*

8.2 *DkS*

Έστω $k \geq an$. Εάν ένας γράφος είναι δ-πυκνός, τότε ο γράφος που επάγεται απο ένα τυχαίο υποσύνολο απο k κορυφές περιέχει $a^2\delta n^2/2$ πλευρές κάτα μέσο όρο. Έτσι ο πυκνότερος υπογράφος περιέχει τουλάχιστον $a^2\delta n^2/2$ πλευρές.

Μπορούμε να εκφράσουμε το DkS σαν τη βέλτιστη λύση του ακόλουθου τετραγωνικού (quadratic) προγράμματος

$$\max \sum_{\{i,j\}\in E} x_i x_j \tag{26}$$

$$x_i \in \{0,1\} \tag{27}$$

$$\sum_{i=1}^{n} x_i = k \tag{28}$$

Το παραπάνω PIP είναι 1-λείο. Απο το παραπάνω θεώρημα 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 $gn\sqrt{n \ln n} = o(n^2)$.

9 Συμπεράσματα

Καταρχήν είδαμε το DS πρόβλημα που αφορά την επιλογή ενός υποσυνόλου κορυφών (αυθαίρετου μεγέθους) τέτοιων που ο επαγόμενος απο αυτές υπογράφος να έχει μέγιστο μέσο βαθμό. Στο 4^o κεφάλαιο είδαμε κάποιους αλγόριθμους που αφορούν αυτό το πρόβλημα για κατευθυνόμενους και μη γράφους. Όλοι οι αλγόριθμοι μπορούν να γενικευτούν για την περίπτωση που ο γράφος έχει βάρη στις πλευρές. Στη με βάρη εκδοχή του προβλήματος δε μεταφέρεται η γραμμική εκτέλεση του αλγορίθμου αφού αυτή εξαρτάται απο το γεγονός ότι οι βαθμοί των κορυφών είναι ακέραιοι φραγμένοι απο το n. Ωστόσο οι αλγόριθμοι μπορούν να υλοποιηθούν χρησιμοποιώντας Fibonacci στοίβες για να καθορίσουμε την ελαχίστου βαθμού κορυφή σε κάθε επανάληψη του αλγορίθμου. Στην περίπτωση αυτή τόσο ο greedy αλγόριθμος για το $f(G)$ όσο και ο greedy αλγόριθμος για το $d(G)$ (για μια μοναδική τιμή του c) τρέχουν σε χρόνο $O(m + n \log n)$.

Έπειτα είδαμε το DkS πρόβλημα που αφορά την εύρεση του k-κορυφών πυκνότερου υπογράφου. Δείξαμε ότι το πρόβλημα παραμένει NP-hard όταν περιοριζόμαστε σε γράφους με μέγιστο βαθμό 3. Αυτό το αποτέλεσμα είναι το καλύτερο δύνατο αφού το πρόβλημα επίλυεται σε πολυωνυμικό χρόνο σε γράφους με μέγιστο βαθμό 2. Επίσης δείξαμε ότι εάν ένας γράφος έχει μια κλίκα k κορυφών τότε για κάθε $1 > \epsilon > 0$ είναι εφικτό να λάβουμε έναν k κορυφών υπογράφο με τουλάχιστον $(1-\epsilon)\binom{k}{2}$ ακμές με έναν $n^{O((1+\log \frac{n}{k})/\epsilon)}$ χρόνου αλγόριθμο. Έτσι εκτός αν το $NP \subseteq TIME(n^{O(\log n)})$ δεν είναι NP-hard να διακρίνουμε μεταξύ γράφων που περιέχουν μια k-κλίκα και γράφων στους οποίους ο k-πυκνότερος υπογράφος έχει τουλάχιστον $(1-\epsilon)\binom{k}{2}$ ακμές.

Ο αλγόριθμος που αφορά το DkS πρόβλημα και που παρουσιάστηκε στο 6^o κεφάλαιο μπορεί να θεωρηθεί το καλύτερο μέχρι στιγμής αποτέλεσμα ανεξαρτήτου k. Διότι ακόμα και αν με τη βοήθεια του ημιορισμένου προγραμματισμού επιτύχουμε έναν προσεγγιστικό λόγο κοντά στο 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 του $6^{ου}$ κεφαλαίου. Για παράδειγμα όταν το $k \simeq n^{1/3}$ 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 (βλέπε [4]). Ειδικά αυτό αποκλείει έναν προσεγγιστικό λόγο καλύτερο απο $n^{1/3}$. Τέλος ο αλγόριθμος του $6^{ου}$ κεφαλαίου μπορεί να επεκταθεί για να χειριστούμε τη με βάρη εκδοχή του DkS προβλήματος επιφέροντας έναν πρόσθετο $O(\log n)$ παράγοντα.

10 Ανοιχτά Ερωτήματα

- Στο 4^o κεφάλαιο είδαμε στον ορισμό της πυκνότητας για κατευθυνόμενους γράφους πως τα σύνολα S και T δεν ήταν απαραίτητο να είναι ξένα. Ποιά είναι η πολυπλοκότητα ενός αλγορίθμου που υπολογίζει μια ελαφρώς τροποποιημένη πυκνότητα $d'(G)$ όπου μεγιστοποιούμε το $d(S,T)$ πάνω σε ξένα μεταξύ τους σύνολα S και T;
  Παρατητούμε ότι κάθε α-προσεγγιστικός αλγόριθμος για το $d(G)$ μπορεί να χρησιμοποιηθεί για να λάβουμε έναν $O(n)$-προσεγγιστικό αλγόριθμο για το $d'(G)$.

- Αναφέραμε ότι το $f(G)$ υπολογίζεται ακριβώς με αλγόριθμο βασισμένο σε $flow$ τεχνικές. Πως θα μπορούσε αυτό να χρησιμοποιηθεί για να λάβουμε έναν αλγόριθμο υπολογισμού του $d(G)$ βασισμένο σε $flow$ τεχνικές;

- Μπορεί να βελτιωθεί ο προσεγγιστικός λόγος $O(n^\delta)$ για κάποιο $\delta < 1/3$ που προκύπτει απο τον αλγόριθμο που είδαμε στο 6^o κεφάλαιο;

- Στο 6^o κεφάλαιο παρουσιάσαμε έναν προσεγγιστικό αλγόριθμο για το DkS με προσεγγιστικό λόγο $O(n^\delta)$ για κάποιο $\delta < 1/3$. Υπάρχει κάποιο $\epsilon > 0$ τέτοιο που το να επιτύχουμε έναν προσεγγιστικό λόγο $O(n^\epsilon)$ για το DkS να είναι NP-hard;

- Είναι NP-hard το να επιτύχουμε έναν προσεγγιστικό αλγόριθμο για το DkS με προσεγγιστικό λόγο $1+\epsilon$ για κάποιο $\epsilon > 0$;

- Μπορεί να χρησιμοποιηθεί η τεχνική του $8^{ου}$ κεφαλαίου για την εύρεση ενός $PTAS$ σε μη πυκνά στιγμιότυπα του DkS προβλήματος;

Αναφορές

[1] *Finding Dense Subgraphs*, Y. Asahiro - K. Iwama, Proc.6th International Symposium on Algorithms and Computation (ISAAC), LNCS, 102-111 (1995)

[2] *Greedily Finding a Dense Subgraph*, Y. Asahiro - K. Iwama - H. Tamaki - T. Tokuyama, Journal of Algorithms, 34(2):203-221 (2000)

[3] *The Dense k-Subgraph Problem*, U. Feige - G. Kortsarz - D. Peleg, Algorithmica, 29(3):410-421 (2001)

[4] *On the Densest k-Subgraph Problem*, U. Feige - M. Seltser, Weizmann Institute Technical Report CS 97-16 (1997)

[5] *Greedy Approximation Algorithms for Finding Dense Components in a Graph*, Moses Charikar, Proceedings of APPROX (2000)

[6] *Polynomial Time Approximation Schemes for Dense Instances of NP-hard problems*, S. Arora - D. Karger - M. Kaprinski, Proceedings of the 27th Annual ACM Symposium on Theory of Computing, pp.284-293 (1995)

[7] *A Fast Parametric Maximum Flow Algorithm and Applications*, G. Gallo - M. Grigoriadis - R. Tarzan, SIAM Journal on Computation, 18:30-55 (1989)

[8] *Authoritative Sources in a Hyperlinked Enviroment*, J. Kleinberg, Journal of the ACM, 46:604-632 (1999)

[9] *Inferring Web Communities from Link Topology*, D. Gibson - J. Kleinberg - P. Raghavan, Proc. HYPERTEXT, 225-234 (1998)

[10] *Random Generation of Test Instances with Controlled Attributes*, Y. Asahiro - K. Iwama - E. Miyano, DIMACS Series in Discrete Math. and Theor. Comput. Sci., 26:377-393 (1996)

[11] *Introduction to Algorithms*, T. Cormen - C. Leiserson - R. Rivest, The Massachusetts Institute of Technology (1990)

[12] *Computational Complexity*, C. Papadimitriou, Addison-Wesley Publishing Company,Inc (1994)

[13] *The Web as a graph: measurments, models and methods*, J. Kleinberg - R. Kumar - P. Raghavan - S. Rajagopalan - A. Tomkins, Proc. 5th Annual International Conference on Computing and Combinatorics (COCOON), 1-17 (1999)

[14] *Trawlling the Web for emerging cyber-communities*, R. Kumar - P. Raghavan - S. Rajagopalan - A. Tomkins, Proc. 8th WWW Conference, Computer Networks, 31(11-16):1481-1493 (1999)