



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

**Αλγόριθμοι Ανάλυσης Κεντρικότητας σε Κοινωνικά
Δίκτυα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΩΤΗΡΟΠΟΥΛΟΥ ΚΩΝΣΤΑΝΤΙΝΟΥ

Επιβλέπων : Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2016



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αλγόριθμοι Ανάλυσης Κεντρικότητας σε Κοινωνικά Δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΩΤΗΡΟΠΟΥΛΟΥ ΚΩΝΣΤΑΝΤΙΝΟΥ

Επιβλέπων : Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 30^η Μαρτίου 2016.

.....
Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

.....
Ευστάθιος Συκάς
Καθηγητής Ε.Μ.Π.

.....
Ιωάννα Ρουσάκη
Επίκουρη Καθηγήτρια Ε.Μ.Π.

Αθήνα, Μάρτιος 2016

(Υπογραφή)

.....

ΣΩΤΗΡΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

© 2016– All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας διπλωματικής εργασίας εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη μίας μεθοδολογίας για την ανάλυση και αποδοτική προσέγγιση/υπολογισμό μετρικών κεντρικότητας βασισμένων σε μονοπάτια, δηλαδή κεντρικότητες οι οποίες απαιτούν γνώση της πραγματικής δρομολόγησης μονοπατιών για τον υπολογισμό τους. Τέτοιες μετρικές είναι πολύ σημαντικές για την ανάλυση της δρομολόγησης και για μηχανισμούς και εφαρμογές δικτύων που βασίζονται στη δρομολόγηση.

Παρ' όλα αυτά, η δρομολόγηση που βασίζεται σε συντομότερα μονοπάτια σε αριθμό βημάτων και η εκτίμηση των μετρικών κεντρικότητας στα σύγχρονα σύνθετα δίκτυα (τα οποία αποτελούνται έως και από εκατομμύρια κόμβους και ακμές) είναι μία, μάλλον αρκετά δαπανηρή, σε σχέση με τον υπολογιστικό της χρόνο, εργασία. Για τον λόγο αυτό, χρησιμοποιήθηκε η ενσωμάτωση δικτύων στον υπερβολικό χώρο, ενώ υιοθετήθηκε η άπληστη δρομολόγηση πάνω σε υπερβολικές συντεταγμένες. Η άπληστη δρομολόγηση στον υπερβολικό χώρο έχει διπλό πλεονέκτημα: (i) Απαιτεί λιγότερη υπολογιστική προσπάθεια από την δρομολόγηση συντομότερων μονοπατιών (σε σχέση με την απόσταση-βημάτων), ενώ την ίδια στιγμή, (ii) καταλήγει σε άπληστα μονοπάτια με μήκη πολύ κοντά σε αυτά των συντομότερων μονοπατιών, για τα σύνθετα δίκτυα που μας ενδιαφέρουν, τα οποία φέρουν την ιδιότητα της ελεύθερης-κλίμακας (scale-free).

Κάνοντας χρήση αυτών των δύο στοιχείων, προτείνεται η Υπερβολική Κεντρικότητα Φόρτου Κίνησης (ΥΚΦΚ) σαν μία εναλλακτική μετρική της Κεντρικότητας Φόρτου Κίνησης (ΚΦΚ), η οποία χρησιμοποιείται για την κατάταξη των κόμβων σε σχέση με τη σημασία τους στη διαδικασία δρομολόγησης. Μέσω ανάλυσης και προσομοιώσεων, επιδείχθηκε ότι η ΥΚΦΚ απαιτεί σημαντικά λιγότερο υπολογιστικό χρόνο από την ΚΦΚ, ενώ παρ' όλο που είναι περισσότερο ενδεδειγμένη για καθεστώς άπληστης δρομολόγησης στον υπερβολικό χώρο, επιτυγχάνει μία κοντινή προσέγγιση της ΚΦΚ για δίκτυα με ιδιότητες ελεύθερης-κλίμακας όταν υιοθετείται η δρομολόγηση συντομότερων μονοπατιών.

Επιπλέον, η ΥΚΦΚ μελετήθηκε κάτω από έναν τύπο ενσωμάτωσης του γράφου του δικτύου στον υπερβολικό χώρο συντεταγμένων (που αποκαλείται «Άπληστη Ενσωμάτωση») ο οποίος εξασφαλίζει 100% επιτυχία της άπληστης δρομολόγησης. Αυτός ο τύπος ενσωμάτωσης βρίσκει πρώτα μία ενσωμάτωση ενός συνδετικού δένδρου του γράφου του δικτύου, η οποία είναι επίσης μία άπληστη ενσωμάτωση ολόκληρου του γράφου. Παρ' όλα αυτά, η επιλογή των δενδρικών ακμών του δένδρου έχει σημαντική επίδραση στις μετρικές κεντρικότητας όλων των κόμβων στο δίκτυο. Αυτή η επίδραση παρουσιάστηκε μέσω μαθηματικών σχέσεων, απτών παραδειγμάτων και στατιστικών ιδιοτήτων. Κάνοντας χρήση αυτής της επίδρασης, προτάθηκε ένα πλαίσιο σχεδιασμού και προσαρμογής της δρομολόγησης πακέτων με αξιολόγηση της σημασίας των κόμβων, χρήσιμο για λειτουργίες μεταφοράς της συμφόρησης κατά τη δρομολόγηση.

Λέξεις Κλειδιά: Κεντρικότητα ενδιαμεσικότητας, Κεντρικότητα φόρτου κίνησης, Υπερβολική γεωμετρία, Άπληστη δρομολόγηση, Ενσωμάτωση δικτύου, Συμφόρηση δρομολόγησης, Άπληστη ενσωμάτωση.

Abstract

The scope of this thesis work was to develop a methodology for the analysis and efficient approximation/computation of path-based centrality metrics, namely centralities that require knowledge of the actual routing path for their computation. Such metrics are very important for analysis of routing and routing-based network mechanisms and applications.

However, path-based routing and centrality metrics estimation in modern complex networks (that consist up to millions of nodes and edges) is a rather expensive task in terms of computational time. For this reason, we employed network embedding in hyperbolic space and greedy routing over hyperbolic coordinates. Greedy routing over hyperbolic space has a two-fold advantage: (i) it requires less computational effort than shortest path routing (in terms of hop-distances) while, at the same time, (ii) it yields greedy paths with lengths very close to the shortest ones for complex networks of interest bearing the scale-free property.

By exploiting these two facts, Hyperbolic Traffic Load Centrality (HTLC) was proposed, as a novel alternative to the Traffic Load Centrality (TLC) metric, used for ranking nodes with respect to their importance in the routing operation. Through analysis and simulation, it was demonstrated that HTLC requires significantly lower computational time than TLC, and despite being more suitable for greedy routing constraints over hyperbolic space, it nevertheless achieves a close approximation of TLC for networks with scale-free properties when assuming shortest path routing.

Furthermore, HTLC was studied under an embedding type (called “Greedy Embedding”) that achieves a 100% success in greedy routing. This embedding type finds an embedding of a spanning tree of a graph, which is also a greedy embedding for the entire graph. However, the choice of the tree-edges has a significant impact on the centrality metrics of all the nodes in the network. This impact was exhibited by mathematical equations, concrete examples and statistical properties. By evaluating this impact, a framework was proposed, useful for operations that deal with transferring routing congestion.

Keywords: Betweenness centrality, Traffic load Centrality, Hyperbolic geometry, Greedy routing, Network embedding, Routing congestion, Greedy embedding

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον καθηγητή κύριο Συμεών Παπαβασιλείου για την επίβλεψη της διπλωματικής μου εργασίας, την ευκαιρία που μου έδωσε να ασχοληθώ με ένα σύγχρονο και αρκετά ενδιαφέρον θέμα, καθώς και το συνεχές ενδιαφέρον του και τις πολύτιμες συμβουλές του.

Ένα μεγάλο ευχαριστώ οφείλω και στους μεταδιδακτορικούς ερευνητές Έλενα Στάη και Βασίλη Καρυώτη για την αμέριστη βοήθεια που μου παρείχαν, το διόλου ευκαταφρόνητο χρόνο που διέθεσαν, καθώς και την άψογη συνεργασία που είχαμε και οδήγησε στην ολοκλήρωση της παρούσας εργασίας.

Τέλος, δε θα μπορούσα παρά να μην ευχαριστήσω και την οικογένεια μου (τη μητέρα μου Στεφανία, τον πατέρα μου Παναγιώτη και την αδερφή μου Βασιλική) για την ηθική και υλική στήριξη που μου παρείχε κατά τη διάρκεια των σπουδών μου.

Πίνακας Περιεχομένων

Πίνακας Περιεχομένων	1
Κατάλογος Σχημάτων	3
Κατάλογος Πινάκων	5
1 Εισαγωγή	6
1.1 Σύνθετα Δίκτυα	6
1.2 Αντικείμενο Διπλωματικής και Συνεισφορά	7
1.3 Οργάνωση Κειμένου	8
2 Στοιχεία της Θεωρίας Γραφημάτων	10
2.1.1 Βασικοί Ορισμοί και Συμβολισμοί	10
2.1.2 Μονοπάτια, Κύκλοι και Συνδεσιμότητα	12
2.1.3 Δένδρα	13
2.1.4 Επίπεδα Γραφήματα	14
2.1.5 Αλγεβρική Θεωρία Γραφημάτων	15
2.1.6 Τυχαίοι Γράφοι	16
3 Ανάλυση Σύνθετων και Κοινωνικών Δικτύων	18
3.1 Χαρακτηριστικά και Μετρικές των Σύνθετων Δικτύων	18
3.1.1 Κατανομή Βαθμού Κόμβων	18
3.1.2 Μέσο Μήκος Μονοπατιού	19
3.1.3 Συντελεστής Ομαδοποίησης	20
3.2 Κεντρικότητα Κόμβων	20
3.2.1 Κεντρικότητα Βαθμού	21
3.2.2 Κεντρικότητα Εγγύτητας	22

3.2.3	<i>Κεντρικότητα Ενδιαμεσικότητας.....</i>	23
3.3	Μοντέλα Αναπαράστασης και Δημιουργίας Συνθετικών Δικτύων	25
3.3.1	<i>Δίκτυα Μικρού Κόσμου</i>	26
3.3.2	<i>Δίκτυα Ελεύθερης Κλίμακας.....</i>	28
4	Ενσωμάτωση των Γράφων στην Υπερβολική Γεωμετρία	31
4.1	Ανάλυση Δεδομένων Μεγάλης Κλίμακας στον Υπερβολικό Χώρο Συντεταγμένων.....	31
4.2	Θεμελιώδεις Αρχές του Υπερβολικού Γεωμετρικού Χώρου.....	33
4.3	Ενσωμάτωση Δικτύων στον Υπερβολικό Χώρο.....	35
4.3.1	<i>Ενσωμάτωση Rigel.....</i>	36
4.3.2	<i>Άπληστη Ενσωμάτωση</i>	39
5	Υπολογισμός Μετρικών Κεντρικότητας Με Χρήση Συντεταγμένων στην Υπερβολική Γεωμετρία για Δίκτυα Μεγάλης Κλίμακας.....	45
5.1	Υπερβολική Κεντρικότητα Φόρτου για Σύνθετα Δίκτυα Μεγάλης Κλίμακας.....	45
5.1.1	<i>Ο Αλγόριθμος του Brandes για την Κεντρικότητα Ενδιαμεσικότητας.....</i>	46
5.1.2	<i>Υπερβολική Κεντρικότητα Φόρτου Κίνησης</i>	53
5.2	Επίδραση της Επιλογής Δενδρικών Ακμών Κατά την Άπληστη-Ενσωμάτωση στο Φόρτο των Κόμβων	56
6	Αποτελέσματα Υπολογισμού Μετρικών Κεντρικότητας στην Υπερβολική Γεωμετρία 60	
6.1	Υπερβολική Κεντρικότητα Φόρτου Κίνησης με Χρήση της Ενσωμάτωσης Rigel ..	61
6.2	Επίδραση Επιλογής Δενδρικών Ακμών στην Άπληστη Ενσωμάτωση.....	65
7	Επίλογος	80
7.1	Σύνοψη και Συμπεράσματα	80
7.2	Μελλοντικές Επεκτάσεις	81
	Βιβλιογραφία.....	82

Κατάλογος Σχημάτων

Σχήμα 1: Ορισμοί των υπογράφων [1].....	11
Σχήμα 2: Δένδρα	13
Σχήμα 3: Η φόρμουλα του Euler για επίπεδους γράφους.....	15
Σχήμα 4: Παραδείγματα κατανομής βαθμών κόμβων για διάφορους τύπους δικτύων (γραμμική κλίμακα) [1].....	19
Σχήμα 5: Παρουσίαση του b -μοντέλου των Watts και Strogatz. Οι παράμετροι του δικτύου είναι $n = 24, k = 4$ [1]	27
Σχήμα 6: Κατανομή του βαθμού κόμβων σε σύνθετα δίκτυα τύπου ελεύθερης κλίμακας	28
Σχήμα 7: Συν-εξέλιξη δικτύων επικοινωνιών (σύνθετων)-κοινωνικών δικτύων [10].....	31
Σχήμα 8: Η προτεινόμενη ροή ενεργειών στο χώρο της υπερβολικής γεωμετρίας για ανάλυση δεδομένων μεγάλης κλίμακας σε σύνθετα και κοινωνικά δίκτυα [10]	33
Σχήμα 9: Το μοντέλο του δίσκου του Poincaré (a) μαζί με τα συντομότερα μονοπάτια σε δύο διαστάσεις. Το (b) δείχνει τα διαγράμματα Voronoi του δίσκου του Poincaré για υπερβολικά τρίγωνα με ίσο εμβαδόν [10].....	34
Σχήμα 10: Ένα παράδειγμα ενσωμάτωσης ενός γράφου στον Ευκλείδειο Χώρο. Για παράδειγμα, το συντομότερο μονοπάτι μεταξύ των κόμβων A και B είναι 3 στο γράφημα (αριστερά) και η Ευκλείδεια απόσταση των συντεταγμένων τους είναι 3.1 (δεξιά) [12]	37
Σχήμα 11: Παράδειγμα μίας άπληστης ενσωμάτωσης ενός μη-κανονικού συνδετικού δένδρου στο μοντέλο του δίσκου Poincare [13].....	43
Σχήμα 12: Κατάλληλη τοποθέτηση του κόμβου-ρίζα για μία άπληστη ενσωμάτωση[13]	43
Σχήμα 13: Με την υπόθεση του λήμματος 4, ένας κόμβος βρίσκεται ως ενδιαμέσος πάνω στα συντομότερα μονοπάτια προς τους διαδόχους του στο δένδρο των συντομότερων μονοπατιών από την πηγή [5].....	49
Σχήμα 14: Ένα απλοποιημένο μοντέλο ενός d -κανονικού δένδρου ενός γράφου[23].....	57
Σχήμα 15: Συνθετικοί γράφοι ελεύθερης-κλίμακας: Ακρίβεια πρόβλεψης σε σύγκριση ΥΚΦΚ και ΚΦΚ	63
Σχήμα 16: Γράφοι από πραγματικά σύνολα δεδομένων: Ποσοστό επιτυχίας στην πρόβλεψη των k -υψηλότερων στην κατάταξη κόμβων	65
Σχήμα 17: ΚΦΚ των κόμβων του γράφου ελεύθερης κλίμακας	67
Σχήμα 18: ΥΚΦΚ των κόμβων με επιλογή του κόμβου 1 ως ρίζα του συνδετικού δένδρου .	67
Σχήμα 19: ΥΚΦΚ των κόμβων με επιλογή του κόμβου 35 ως ρίζα του συνδετικού δένδρου	68
Σχήμα 20: Σύγκριση μεταξύ της επιλογής των κόμβων 1 και 35 ως ρίζα του συνδετικού δένδρου	69

Σχήμα 21: ΥΚΦΚ του κόμβου 1 για διαφορετικά επίπεδά του στο δυνδευτικό δένδρο.....	69
Σχήμα 22: Τιμή της ΚΦΚ των κόμβων ενός τυχαίου γράφου.....	70
Σχήμα 23: ΥΚΦΚ των κόμβων του τυχαίου γράφου με τον κόμβο 82 ως ρίζα του δένδρου .	71
Σχήμα 24: Σύγκριση της κατανομής ΥΚΦΚ και ΚΦΚ.....	72
Σχήμα 25: Τιμή της ΚΦΚ των κόμβων του γράφου μικρού-κόσμου.....	72
Σχήμα 26: ΥΚΦΚ με τους κόμβους 37 και 83 ως ρίζες.....	73

Κατάλογος Πινάκων

3.1	ΠΕΡΙΛΗΨΗ ΤΥΠΩΝ ΚΕΝΤΡΙΚΟΤΗΤΑΣ ΣΤΗΝ ΑΝΑΛΥΣΗ ΣΥΝΘΕΤΩΝ ΔΙΚΤΥΩΝ	21
6.1	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΣΥΝΘΕΤΙΚΩΝ ΓΡΑΦΩΝ & ΣΥΓΚΡΙΣΗ ΤΟΥ ΥΠΟΛΟΓΙΣΤΙΚΟΥ ΧΡΟΝΟΥ ΜΕΤΑΞΥ ΚΦΚ & ΥΚΦΚ	62
6.2	ΠΕΡΙΓΡΑΦΗ ΠΡΑΓΜΑΤΙΚΩΝ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ & ΣΥΓΚΡΙΣΗ ΤΟΥ ΥΠΟΛΟΓΙΣΤΙΚΟΥ ΧΡΟΝΟΥ ΜΕΤΑΞΥ ΚΦΚ & ΥΚΦΚ	64
6.3	ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΕΛΕΥΘΕΡΗΣ ΚΛΙΜΑΚΑΣ ΜΕ ΤΗΝ ΥΨΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ	73
6.4	ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΕΛΕΥΘΕΡΗΣ ΚΛΙΜΑΚΑΣ ΜΕ ΤΗΝ ΧΑΜΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ	74
6.5	ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ 5 ΤΥΧΑΙΑ ΕΠΙΛΕΓΜΕΝΟΥΣ ΚΟΜΒΟΥΣ ΤΟΥ ΤΥΧΑΙΟΥ ΓΡΑΦΟΥ	75
6.6	ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΜΙΚΡΟΥ ΚΟΣΜΟΥ ΜΕ ΤΗΝ ΥΨΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ	76
6.7	ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΜΙΚΡΟΥ ΚΟΣΜΟΥ ΜΕ ΤΗ ΧΑΜΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ	77

1 ***Εισαγωγή***

1.1 ***Σύνθετα Δίκτυα***

Η ανάπτυξη των δικτύων επικοινωνιών, παράλληλα με την ανάπτυξη πάνω σε αυτή την υποδομή και κοινωνικών δικτύων μεταξύ των χρηστών, έχει δημιουργήσει την ανάγκη ώστε αυτά τα δίκτυα να αποτελούν αντικείμενο συστηματικής μελέτης. Τα δίκτυα αυτά μπορούν να αθροιστούν κάτω από τον όρο «Σύνθετα Δίκτυα». Το αντικείμενο της επιστήμης των δικτύων που ασχολείται με την ανάλυση αυτών των σύνθετων δικτύων, τόσο σε επίπεδο ανάπτυξής τους, της δομής τους, της αλληλεπίδρασης μεταξύ των χρηστών, όσο και σε επίπεδο δρομολόγησης της πληροφορίας, έχει ονομαστεί «Ανάλυση Κοινωνικών Δικτύων». Μερικά από τα χαρακτηριστικά που απασχολούν το συγκεκριμένο επιστημονικό τομέα είναι η αξιολόγηση της επίδρασης ενός κόμβου στη ροή της πληροφορίας, είτε η δρομολόγηση της πληροφορίας προς διάφορους χρήστες στο δίκτυο. Πλέον, η ανάπτυξη αυτών των δικτύων έχει φτάσει σε τέτοιο επίπεδο που ένα δίκτυο, είτε επικοινωνιών (π.χ., αυτόνομων συστημάτων, emails, κτλ.), είτε κοινωνικό δίκτυο (π.χ., facebook, flickr κ.ο.κ.) αποτελείται από δεκάδες χιλιάδες μέχρι και εκατομμύρια χρήστες ή οντότητες. Το μέγεθος αυτών των δικτύων, επομένως, καθιστά υπολογιστικά δύσκολη- έως αδύνατη- την αξιολόγηση της σημασίας των κόμβων με παραδοσιακές μεθόδους. Επομένως, προκύπτει η ανάγκη για την όσο το δυνατόν αποδοτικότερη δρομολόγηση της πληροφορίας και αξιολόγηση της σημασίας των κόμβων στη ροή της, με μεθόδους που θα αξιοποιούν τεχνικές που δεν θα ακολουθούν πλέον την παραδοσιακή μέθοδο δρομολόγησης συντομότερων μονοπατιών, η οποία, πέρα από υπολογιστικά απαιτητική, απαιτεί και τη συνολική γνώση της δομής του δικτύου και δεν μπορεί να γίνει με κατανεμημένο τρόπο. Παραδοσιακά, η δομή αυτών των δικτύων και οι αλληλεπιδράσεις μεταξύ των χρηστών αναπαρίστανται με τη μορφή κόμβων και ακμών που συνδέουν τους κόμβους που παρουσιάζουν κάποια μορφή αλληλεπίδρασης ή σύνδεσης. Επομένως, η συνολική αναπαράσταση ενός σύνθετου δικτύου, παίρνει τη μορφή ενός γράφου (ή γραφήματος). Για το λόγο αυτό, μαζί με το γεγονός της ανάπτυξης δικτύων που δεν βασίζονται σε παραδοσιακή υποδομή, αλλά προκύπτουν λόγω της τυχαίας αλληλεπίδρασης των χρηστών (π.χ. ασύρματα αυτο-οργανούμενα δίκτυα), έχουν δημιουργηθεί τεχνικές που στοχεύουν στην ανάθεση συντεταγμένων στους κόμβους του δικτύου, ώστε η δρομολόγηση να γίνεται με τρόπο κατανεμημένο αξιοποιώντας τις συντεταγμένες κάθε κόμβου και το κατά πόσο αυτός βρίσκεται «τοπολογικά» πιο κοντά στον προορισμό της πληροφορίας.

Στην κατεύθυνση αυτή, λοιπόν, κρίνεται σημαντική η διερεύνηση αποδοτικών μεθόδων που θα εκμεταλλεύονται τις παραπάνω τεχνικές και οι οποίες θα μπορούν να οδηγούν σε αποδοτικά σχήματα δρομολόγησης της πληροφορίας και σε ταχύτερη εκτίμηση της σημασίας των κόμβων ενός δικτύου στη ροή της πληροφορίας.

1.2 Αντικείμενο Διπλωματικής και Συνεισφορά

Στα πλαίσια της διπλωματικής εργασίας, χρησιμοποιήθηκαν δύο τύποι ενσωμάτωσης, ο σκοπός των οποίων είναι η ανάθεση συντεταγμένων στον υπερβολικό χώρο στους κόμβους των σύνθετων δικτύων που εξετάστηκαν. Με τον τρόπο αυτό μπορούμε να αποφύγουμε την υπολογιστικά απαιτητική δρομολόγηση συντομότερων-μονοπατιών και να οδηγηθούμε σε σχήματα δρομολόγησης, τα οποία κάθε φορά προωθούν το πακέτο (ή διαδίδουν την πληροφορία) προς τον κόμβο εκείνο ο οποίος με βάση και τις συντεταγμένες του βρίσκεται «τοπολογικά» εγγύτερα προς τον προορισμό του πακέτου (ή της πληροφορίας). Αυτή η μέθοδος δρομολόγησης, όπου κάθε φορά αναζητούμε τον προορισμό προωθώντας την πληροφορία προς το σημείο εκείνο στο οποίο παρατηρείται τοπικά πρόοδος στην κατεύθυνση προς τον προορισμό, ονομάζεται άπληστη δρομολόγηση. Η χρήση του χώρου της υπερβολικής γεωμετρίας (έναντι άλλων γεωμετρικών χώρων) έγινε, καθώς ύστερα και από ενδελεχή μελέτη της βιβλιογραφίας, διαπιστώθηκε ότι ο χώρος αυτός παρέχει τα καλύτερα δυνατά αποτελέσματα έναντι των υπολοίπων γεωμετρικών χώρων. Δηλαδή, πολύ υψηλό ποσοστό επιτυχίας στην εύρεση του προορισμού (έως και 100% επιτυχία) και μονοπάτια που όσο το δυνατόν καλύτερα προσεγγίζουν τα συντομότερα μονοπάτια του δικτύου.

Αξιοποιώντας, λοιπόν, τη δρομολόγηση με άπληστα μονοπάτια, σκοπός της διπλωματικής εργασίας είναι να αξιολογήσει τη σημασία της επίδρασης των κόμβων ενός δικτύου στη ροή της πληροφορίας. Για το σκοπό αυτό προτείνεται ένας αλγόριθμος, ο οποίος τροποποιεί τον παραδοσιακό αλγόριθμο υπολογισμού που ποσοτικοποιεί τη σημασία ενός κόμβου στη ροή της πληροφορίας, και μας οδηγεί στον υπολογισμό της σημασίας των κόμβων υπό το καθεστώς άπληστης δρομολόγησης, σε υπολογιστικά συντομότερο χρόνο απ' ότι ο παραδοσιακός αλγόριθμος για συντομότερα μονοπάτια. Παράλληλα, εκμεταλλευόμενοι τις βιβλιογραφικές αναφορές που αποδεικνύουν ότι για συγκεκριμένους τύπους δικτύων, όταν η δρομολόγηση γίνεται με άπληστα μονοπάτια στον υπερβολικό γεωμετρικό χώρο, προκύπτουν άπληστα μονοπάτια που προσεγγίζουν τα συντομότερα μονοπάτια του δικτύου, αναμένονται αποτελέσματα που, πέρα από το ότι θα μπορούν να εξαχθούν σε μικρότερο χρόνο, θα προσεγγίζουν με ικανοποιητικό τρόπο τα αποτελέσματα, ως προς τη σημασία των κόμβων στη ροή της πληροφορίας, με τον παραδοσιακό αλγόριθμο. Για την επιβεβαίωση αυτής της υπόθεσης χρησιμοποιήθηκαν συνθετικά δίκτυα που κατασκευάστηκαν με τη χρήση συγκεκριμένων μοντέλων αναπαράστασης δικτύων, καθώς και γράφοι πραγματικών δικτύων.

Τέλος, εξετάστηκε μία μορφή ενσωμάτωσης δικτύων στον υπερβολικό χώρο, η οποία έχει αποδειχθεί ότι παρέχει 100% επιτυχία στην εύρεση του προορισμού. Η μέθοδος αυτή χρησιμοποιεί ένα δένδρο του δικτύου (δηλαδή το σύνολο των κόμβων του, όπου όμως μεταξύ δύο ζευγών κόμβων αυτή τη φορά θα υπάρχει ένα μοναδικό μονοπάτι) για να αναθέσει συντεταγμένες στον υπερβολικό χώρο στους κόμβους του δικτύου. Τα άπληστα μονοπάτια που θα ακολουθηθούν για τη δρομολόγηση της πληροφορίας στο δίκτυο αυτό, εξαρτώνται, επομένως σε μεγάλο βαθμό από την επιλογή του συγκεκριμένου δένδρου. Για τον λόγο αυτό, εξηγείται μαθηματικά η επίδραση της επιλογής του δένδρου στην άπληστη δρομολόγηση και προτείνεται ένα πλαίσιο ενεργειών το οποίο μπορεί να οδηγήσει στη μεταφορά της συμφόρησης της κίνησης στο δίκτυο.

1.3 Οργάνωση Κειμένου

Η παρούσα διπλωματική εργασία είναι οργανωμένη σε επτά κεφάλαια. Στα Κεφάλαια 2-4 παρουσιάζεται το θεωρητικό υπόβαθρο, καθώς και οι βασικές έννοιες και τεχνικές που χρησίμευσαν στη διπλωματική εργασία. Στο Κεφάλαιο 5 παρουσιάζεται το πρόβλημα του υπολογισμού μετρικών κεντρικότητας για δίκτυα ενσωματωμένα στην υπερβολική γεωμετρία και οι μέθοδοι και αλγόριθμοι που θα χρησιμοποιηθούν για τον υπολογισμό τους, ενώ στο Κεφάλαιο 6 παρουσιάζονται τα αποτελέσματα των υπολογισμών αυτών.

Ειδικότερα:

Στο Κεφάλαιο 2 παρουσιάζονται βασικά στοιχεία της Θεωρίας Γραφημάτων που είναι χρήσιμα για την κατανόηση του συγκεκριμένου πεδίου, καθώς και για την εξοικείωση με την ορολογία και το συμβολισμό που θα ακολουθηθεί στη συνέχεια της διπλωματικής εργασίας.

Στο Κεφάλαιο 3 αναλύονται βασικά χαρακτηριστικά και μετρικές των σύνθετων δικτύων, μετρικές κεντρικότητας που χρησιμεύουν για τον προσδιορισμό της σημασίας των κόμβων στην κίνηση της πληροφορίας (σχετικοί αλγόριθμοι για τον υπολογισμό τους θα αναλυθούν στο Κεφάλαιο 5), καθώς και μοντέλα αναπαράστασης που έχουν προταθεί για τη δημιουργία συνθετικών δικτύων και τα οποία χρησιμοποιήθηκαν, μαζί με γραφήματα που προκύπτουν από πραγματικά σύνολα δεδομένων, για την αξιολόγηση των αλγορίθμων και μεθόδων που προτάθηκαν.

Στο Κεφάλαιο 4, παρουσιάζεται η χρησιμότητα της Υπερβολικής Γεωμετρίας στην ανάλυση δεδομένων μεγάλης κλίμακας, θεμελιώδεις αρχές της, καθώς και 2 τεχνικές για την ενσωμάτωση των κόμβων δικτύων στον υπερβολικό γεωμετρικό χώρο.

Στο Κεφάλαιο 5, παρουσιάζεται η Υπερβολική Κεντρικότητα Φόρτου Κίνησης, ως ένας αλγόριθμος που υπολογίζει την Κεντρικότητα Φόρτου Κίνησης των κόμβων ενός δικτύου σε συνθήκες άπληστης δρομολόγησης, με μικρότερο κόστος σε υπολογιστικό χρόνο. Παράλληλα,

παρουσιάζεται το μαθηματικό υπόβαθρο και προτείνεται ένα πλαίσιο ενεργειών για τη μεταφορά της συμμόρφωσης της κίνησης σε ένα δίκτυο.

Στο Κεφάλαιο 6, γίνεται η πειραματική αξιολόγηση των παραπάνω αλγορίθμων και πλαισίων ενεργειών.

Τέλος, το Κεφάλαιο 7, συνοψίζει τη διπλωματική εργασία και παρέχει ιδέες και κατευθύνσεις για την προοπτική μελλοντικής εξέλιξής της.

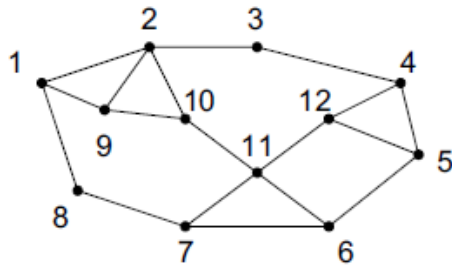
2 Στοιχεία της Θεωρίας Γραφημάτων

Στην ανάλυση των ηλεκτρονικών και κοινωνικών δικτύων (στη συνέχεια θα αναφερόμαστε σε αυτά ως «Σύνθετα Δίκτυα») μία οντότητα ενός κοινωνικού δικτύου αναπαρίσταται, κατά σύμβαση, με έναν κόμβο, ενώ κάθε αλληλεπίδραση μεταξύ δύο οντοτήτων i και j αναπαρίσταται με μία ακμή (i, j) που συνδέει τις δύο οντότητες. Επομένως, το συνολικό δίκτυο που δημιουργείται έχει τη μορφή ενός γράφου. Η Θεωρία Γραφημάτων είναι, λοιπόν, το μαθηματικό εργαλείο που μας επιτρέπει να περιγράψουμε με σαφήνεια την αλληλεπίδραση μεταξύ δύο οντοτήτων σε ένα κοινωνικό δίκτυο, αλλά και να εξετάσουμε συνολικά τη δομή και τα χαρακτηριστικά αυτού του δικτύου (όπως η σπουδαιότητα συγκεκριμένων οντοτήτων στη διάδοση της κυκλοφορίας, η ύπαρξη κοινοτήτων κοκ.). Βασικά στοιχεία της Θεωρίας Γραφημάτων που χρησιμοποιήθηκαν στην εργασία αναφέρονται στο παρόν κεφάλαιο, ώστε η ορολογία και ο συμβολισμός να γίνουν σαφή προς τον αναγνώστη. Εκτενέστερη περιγραφή και αποδείξεις θεωρημάτων μπορούν να αναζητηθούν στη σχετική βιβλιογραφία. [1] [2]

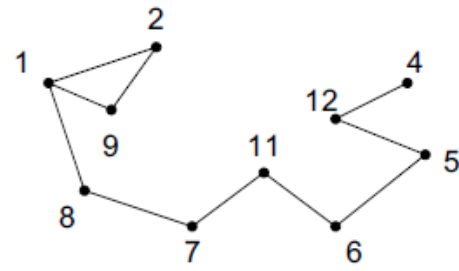
2.1.1 Βασικοί Ορισμοί και Συμβολισμοί

Ένας κατευθυνόμενος γράφος (*directed graph*) G είναι ένα διατεταγμένο ζεύγος από σύνολα (V, E) , όπου V είναι το σύνολο των κόμβων του γράφου με πληθικότητα $|V|$ και E είναι το σύνολο των ακμών πληθικότητας $|E|$. Αναπαριστούμε μία ακμή μεταξύ δύο κόμβων i, j ως (i, j) . Θα χρησιμοποιούμε τη μεταβλητή n για να υποδηλώσουμε τον αριθμό των κόμβων και m τον αριθμό των ακμών. Στην περίπτωση των μη-κατευθυνόμενων γράφων (*undirected graphs*), το ζεύγος (V, E) είναι μη διατεταγμένο.

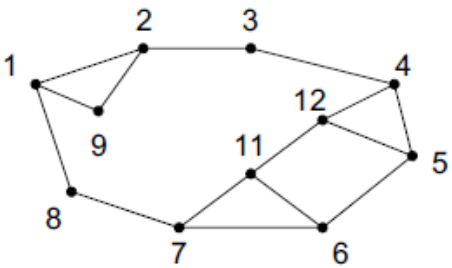
Ένας υπογράφος (*subgraph*) $G' = (V', E')$ είναι ένας υπογράφος του G , αν $V' \subseteq V$ και $E' \subseteq E$. Εάν ο G' περιέχει όλες τις ακμές του G που συνδέουν δύο κόμβους του V' , τότε ο G' καλείται επαγόμενος (*induced*) υπόγραφος. Εάν $V' = V$, τότε ο G' καλείται επικαλύπτον ή συνδετικός (*spanning*) υπογράφος του G .



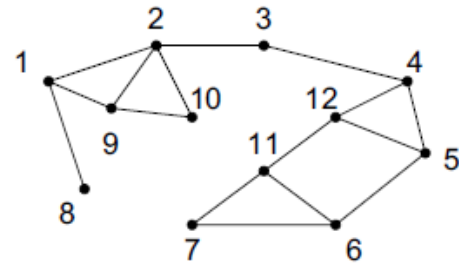
(a) αυθεντικός γράφος G



(b) υπόγραφος του G



(c) επαγόμενος υπόγραφος του G



(d) συνδετικός υπόγραφος του G

Σχήμα 1: Ορισμοί των υπογράφων [1]

Δύο κόμβοι $u, v \in V$ καλούνται *γείτονες* (*neighbors*) όταν $(u, v) \in E$. Το σύνολο των γειτόνων ενός κόμβου u συμβολίζεται $N(u)$, και ο αριθμός των γειτόνων του καλείται ο *βαθμός* (*degree*) του u και συμβολίζεται k_u . Σε έναν κατευθυνόμενο γράφο, κάθε κόμβος χαρακτηρίζεται από δύο βαθμούς: τον *προσ-τα-έσω βαθμό* (*in-degree*) k_u^{in} που μετράει όλες τις ακμές που κατευθύνονται προς τον u και τον *προσ-τα-έξω βαθμό* (*out-degree*) k_u^{out} που μετράει όλες τις ακμές που ξεκινάν από τον u .

Ένας συνηθισμένος τρόπος για την αναπαράσταση ενός γράφου είναι με τον *πίνακα γειτνίασης* (*adjacency matrix*) του, που συμβολίζεται $A = [a_{ij}]$ και έχει διάσταση $n \times n$, όπου $a_{ij} = 1$, όταν υπάρχει σύνδεση μεταξύ i και j , ενώ σε αντίθετη περίπτωση $a_{ij} = 0$. Όπως γίνεται αντιληπτό, στην περίπτωση των μη-κατευθυνόμενων γράφων ο πίνακας αυτός είναι συμμετρικός. Μία άλλη μορφή αναπαράστασης των συνδέσεων ενός γράφου είναι μέσω της *λίστας γειτνίασης* (*adjacency list*), όπου για κάθε κόμβο υπάρχει μία λίστα των συνδέσεων-γειτόνων του.

Ένας γράφος (κατευθυνόμενος ή μη-κατευθυνόμενος) είναι γράφος με *βάρη* (*weighted*), όταν μία μετρήσιμη ποσότητα (που συνήθως αναφέρεται ως *βάρος* και συμβολίζεται με w) ανατίθεται σε κάθε ακμή, $w : E \rightarrow \mathbb{R}$.

2.1.2 Μονοπάτια, Κύκλοι και Συνδεσιμότητα

Μία *διαδρομή* (walk) ενός γράφου G είναι μία διαδοχική ακολουθία από κόμβους και ακμές, π.χ., $v_0, x_1, v_1, \dots, v_{n-1}, x_n, v_n$, η οποία αρχίζει και τελειώνει με κόμβους στους οποίους κάθε ακμή προσπίπτει στους δύο κόμβους που προηγούνται και τον ακολουθούν στη συγκεκριμένη ακολουθία. Μία διαδρομή είναι *κλειστή* (closed) εάν $v_0 = v_n$, διαφορετικά είναι *ανοιχτή* (open). Εάν επιπρόσθετα όλοι οι κόμβοι (και αναγκαστικά όλες οι ακμές) είναι διαφορετικοί, τότε η διαδρομή καλείται *μονοπάτι* (path). Ένας *κύκλος* (cycle) είναι μία κλειστή διαδρομή με διαφορετικούς κόμβους και συμβολίζεται ως C_n (κύκλος σε n σημεία, $n \geq 3$).

Ένας γράφος G καλείται *συνδεδεμένος* (connected) εάν κάθε δύο κόμβοι του είναι συνδεδεμένοι με ένα μονοπάτι στο G . Σε διαφορετικά περίπτωση καλούμε το γράφο μη-συνδεδεμένο ή *συνηθέστερα αποσυνδεδεμένο* (disconnected).

Σε έναν κατευθυνόμενο γράφο μία *αδύναμα συνδεδεμένη συνιστώσα* (weakly connected component) είναι ο μεγιστικός υπογράφος του γράφου, τέτοιος ώστε για κάθε ζεύγος κόμβων u, v στον υπογράφο, υπάρχει ένα μη κατευθυνόμενο μονοπάτι από τον u στον v . *Ισχυρά συνδεδεμένη συνιστώσα* (strongly connected component) είναι ο μεγιστικός υπογράφος του γράφου, στον οποίο για κάθε ζεύγος κόμβων u, v στον υπογράφο, υπάρχει ένα κατευθυνόμενο μονοπάτι από τον u στον v και από τον v στον u . Στην περίπτωση των μη-κατευθυνόμενων γράφων η μέγιστη συνδεδεμένη συνιστώσα του καλείται *γιγαντιαία συνιστώσα* (giant component).

Η *απόσταση* (distance) $d(u, v)$ μεταξύ δύο κόμβων u, v στον G ορίζεται ως το μέγεθος του συντομότερου μονοπατιού που τους ενώνει, εάν αυτό υπάρχει. Σε αντίθετη περίπτωση, η απόσταση, όπως ορίστηκε παραπάνω, θεωρείται ως $d(u, v) = \infty$. Εάν ο αναφερόμενος γράφος G είναι συνδεδεμένος τότε η απόσταση είναι μία φυσική μετρική (proper metric), πράγμα που σημαίνει ότι για όλα τα σημεία u, v, w ισχύουν οι ακόλουθες ιδιότητες:

1. $d(u, v) \geq 0$, με $d(u, v) = 0$ αν και μόνο αν $u = v$.
2. $d(u, v) = d(v, u)$
3. $d(u, v) + d(v, w) \geq d(u, w)$ (τριγωνική ανισότητα)

Το συντομότερο μονοπάτι μεταξύ δύο κόμβων $u - v$ καλείται *γεωδαισιακό* (geodesic). Επομένως, η *διάμετρος* (diameter) $d(G)$ ενός συνδεδεμένου γράφου G είναι το μήκος του μεγαλύτερου γεωδαισιακού.

2.1.3 Δένδρα

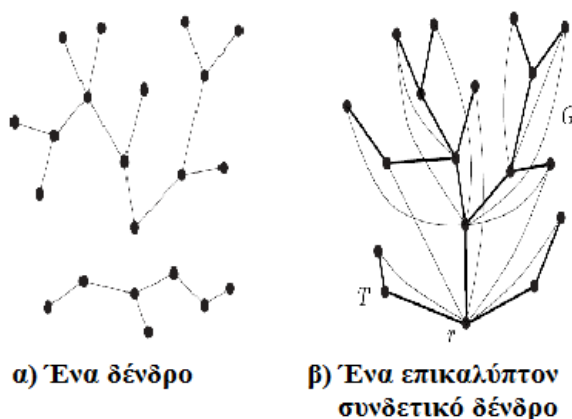
Πιθανόν ένας από τους πιο σημαντικούς ορισμούς, τόσο για θεωρητικούς λόγους, όσο και για πρακτικές εφαρμογές, είναι αυτός του δένδρου (*tree*). Υπάρχουν διάφοροι ορισμοί για την ύπαρξη ενός δένδρου γράφου και οι ακόλουθες ιδιότητες είναι όλες ισοδύναμες για έναν γράφο G :

1. Ο G είναι ένα δένδρο
2. Ο G είναι ένας συνδεδεμένος γράφος και κάθε ακμή είναι μία *γέφυρα* (*bridge*) (δηλαδή ακμή η αφαίρεση της οποίας αυξάνει τον αριθμό των συνδεδεμένων συνιστωσών ενός γράφου, ισοδύναμα δεν περιέχεται σε κανέναν κύκλο)
3. Ο G είναι ένας μεγιστικός ακυκλικός γράφος, τέτοιος ώστε ο G είναι ακυκλικός και εάν x και y είναι μη γειτονικοί κόμβοι του G , τότε ο $G + xy$ περιέχει έναν κύκλο.

Ένα δάσος (*forest*) είναι ένας γράφος του οποίου οι συνιστώσες είναι δένδρα.

Οι κόμβοι βαθμού 1 σε ένα δένδρο αποτελούν τα φύλλα (*leaves*) του δένδρου.

Ένα επικαλύπτον ή συνδετικό δένδρο (*spanning tree*) ενός γράφου G είναι ένα δένδρο που περιέχει κάθε κόμβο του G . Ένα συνδετικό δένδρο ελάχιστου βάρους (*minimum-weight spanning tree*) (για συντομία ελάχιστο συνδετικό δένδρο) σε έναν γράφο με βάρη είναι ένα συνδετικό δένδρο, το οποίο επιπροσθέτως έχει ένα σύνολο ακμών που αθροιστικά καταλήγουν στο ελάχιστο δυνατό βάρος. Δύο άπληστοι αλγόριθμοι έχουν κυρίως προταθεί για την εύρεση ελάχιστων συνδετικών δένδρων, αυτοί του Kruskal και του Prim. Στον αλγόριθμο του Kruskal (πολυπλοκότητας $O(|E|\log|V|)$) βρίσκουμε μία ακμή με το ελάχιστο δυνατό βάρος που συνδέει δύο δένδρα σε ένα δάσος, ενώ στον αλγόριθμο του Prim (πολυπλοκότητας $O(|E|\log|V|)$) γενικά, η οποία μπορεί να μειωθεί με χρήση σωρού Fibonacci σε $O(E + V\log(|V|))$ ξεκινάμε



Σχήμα 2: Δένδρα

από έναν τυχαίο κόμβο και κάθε φορά προσθέτουμε στο δένδρο μία ακμή η οποία έχει το ελάχιστο βάρος και δεν δημιουργεί κύκλο. [22]

Πολλές φορές είναι χρήσιμο να θεωρούμε έναν κόμβο του δένδρου ως ξεχωριστό: Ένας τέτοιος κόμβος τότε αποκαλείται *ρίζα* (*root*) του δένδρου. Ένα δένδρο T με μία ορισμένη ρίζα r είναι ένα *δένδρο με ρίζα* (*rooted tree*). Γράφοντας $x \leq y$ για $x \in rTy$ ορίζουμε μία μερική διάταξη στο σύνολο $V(T)$, τη *διάταξη – δένδρου* (*tree-order*) που σχετίζεται με το δένδρο T και τη ρίζα r . Η ρίζα r είναι το ελάχιστο αντικείμενο σε αυτή τη μερική διάταξη, ενώ τα φύλλα του T είναι τα μεγιστικά της στοιχεία. Οι κόμβοι σε απόσταση k από τη ρίζα r έχουν *ύψος* (*height*) $k + 1$ και αποτελούν το $(k + 1)$ – *οστό επίπεδο* (*level*) του δένδρου T .

Απόρροια του ορισμού ενός δένδρου είναι ότι κάθε συνδεδεμένος γράφος περιέχει ένα συνδετικό δένδρο.

2.1.4 Επίπεδα Γραφήματα

Μία ακόμη ενδιαφέρουσα πτυχή των γραφημάτων, εξαιρετικά χρήσιμη κυρίως στο πεδίο των εφαρμογών- όπως η τοπογραφία, η γεω-τοποθεσία και εφαρμογές οπτικοποίησης γραφημάτων- είναι η τοπολογία του γράφου σε σχέση με τη δυνατότητα να ενσωματωθεί σε συγκεκριμένους χώρους, όπως ένα επίπεδο ή μία σφαίρα.

Ένας γράφος λέγεται *ενσωματώσιμος* (*embedded*) σε μία επιφάνεια S όταν μπορεί να σχεδιασθεί πάνω στο S με τέτοιο τρόπο ώστε δύο ακμές του να μην τέμνονται.

Ένας γράφος είναι *επίπεδος* (*planar*) μόνο όταν μπορεί να ενσωματωθεί σε ένα επίπεδο.

Οι περιοχές που ορίζονται από έναν επίπεδο γράφο αναφέρονται ως *όψεις* και οι μη φραγμένες περιοχές αναφέρονται ως οι *εξωτερικές όψεις*.

Ένα από τα πιο σημαντικά αποτελέσματα που αφορούν τους επίπεδους γράφους είναι το ακόλουθο θεώρημα του Euler (για την απόδειξη του παρακάτω θεωρήματος, όπως και των υπολοίπων σε αυτή την ενότητα μπορεί κάποιος να ανατρέξει στο *Κεφάλαιο 4* του [2]):

Θεώρημα 1: Για κάθε σφαιρικό πολύεδρο με p κόμβους, q ακμές, και F όψεις, $p - q + F = 2$.

Το θεώρημα αυτό μας δείχνει ότι ο αριθμός των όψεων ενός επίπεδου γράφου είναι μία αναλλοίωτη ιδιότητα και δεν εξαρτάται από την αποτύπωση του γράφου.

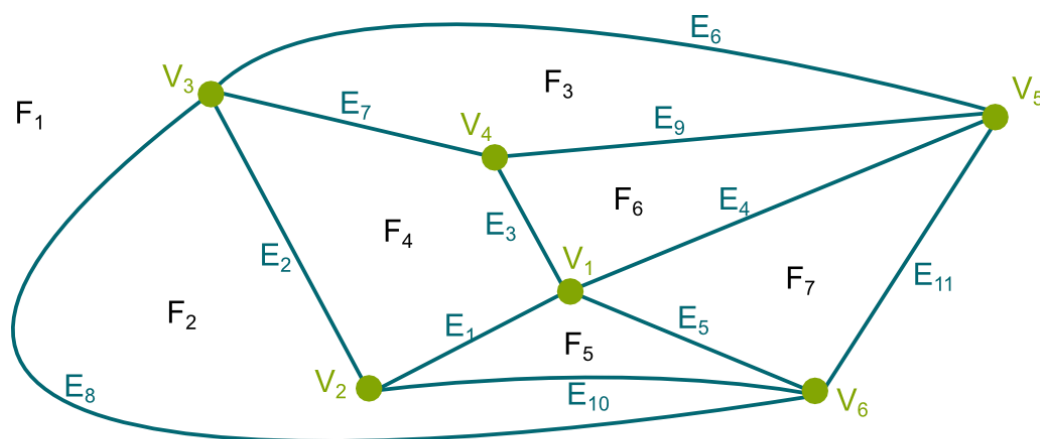
Επίσης, ένα από τα πιο σημαντικά και χρήσιμα εργαλεία για την εξακρίβωση επίπεδων δικτύων είναι το ακόλουθο θεώρημα:

Θεώρημα 2: Ένας γράφος είναι επίπεδος εάν και μόνο εάν κάθε ένα από τα μπλοκ (μεγιστικοί μη-διαχωριζόμενοι υπογράφοι) του είναι επίπεδα.

Μερικά ακόμη χρήσιμα θεωρήματα για διάφορες εφαρμογές είναι τα ακόλουθα:

Θεώρημα 3: Κάθε 2-συνδεδεμένος επίπεδος γράφος μπορεί να ενσωματωθεί σε ένα επίπεδο έτσι ώστε κάθε συγκεκριμένη όψη του να είναι εξωτερική.

Θεώρημα 4: Κάθε 3-συνδεδεμένος επίπεδος γράφος είναι μοναδικά ενσωματώσιμος σε μία σφαίρα.



$$F - E + V = 7 - 11 + 6 = 2$$

Σχήμα 3: Η φόρμουλα του Euler για επίπεδους γράφους

2.1.5 Αλγεβρική Θεωρία Γραφημάτων

Ένας γράφος μπορεί να οριστεί πλήρως από τους κόμβους του και τις ακμές που προσπίπτουν σε αυτούς και αυτή η αναπαράσταση, όπως αναφέρθηκε προηγουμένως, μπορεί να πάρει τη μορφή πινάκων.

Για μη-κατευθυνόμενους γράφους ο πίνακας αυτός είναι συμμετρικός και επομένως έχει μία ορθοκανονική βάση $v_0, v_1, \dots, v_{n-1}$ με ιδιοτιμές $\mu_0, \mu_1, \dots, \mu_{n-1}$ τέτοια ώστε για όλα τα i , $Av_i = \mu_i v_i$, για έναν $n \times n$ πίνακα γειτνίασης A .

Ο αριθμός των μονοπατιών μήκους k βημάτων μεταξύ δύο κόμβων i και j σε έναν γράφο με πίνακα γειτνίασης A είναι το (i, j) στοιχείο του πίνακα A^k . Επιπλέον, ο πίνακας $S_k = A + A^2 + \dots + A^k$ μπορεί να χρησιμοποιηθεί για να υπολογίσουμε το μήκος του συντομότερου μονοπατιού μεταξύ κάθε ζεύγους κόμβων του δικτύου. Ο S_k παρέχει τον αριθμό των μονοπατιών από τον i στο j σε k βήματα ή λιγότερα. Επομένως, για αυξανόμενο k , η πρώτη τιμή του k για την οποία ο S_k είναι μη-μηδενικός είναι το μήκος του συντομότερου μονοπατιού μεταξύ i και j . Επιπρόσθετα, ο πίνακας S_k μπορεί να χρησιμοποιηθεί για να ελέγξουμε κατά

πόσο ο γράφος είναι συνδεδεμένος. Εάν ο γράφος έχει n κόμβους, τότε η μεγαλύτερη πιθανή απόσταση μεταξύ δύο κόμβων θα είναι n βήματα. Οπότε, υπολογίζοντας τον S_n , εάν υπάρχουν μηδενικά στοιχεία σε αυτόν θα είναι αδύνατον για τους αντίστοιχους κόμβους να είναι συνδεδεμένοι σε n ή λιγότερα βήματα, έτσι το συγκεκριμένο ζεύγος δε θα είναι δυνατόν να είναι συνδεδεμένο και επομένως από τον ορισμό ο γράφος θα είναι μη-συνδεδεμένος.

Η *συνδυαστική Λαπλασιανή* (Combinatorial Laplacian) του G είναι απλώς $L = D - A$, όπου $D = [D_{ij}]$ είναι ο διαγώνιος πίνακας στον οποίο D_{ii} είναι ο βαθμός $d(v_i)$ του κόμβου v_i . Η δεύτερη μικρότερη ιδιοτιμή του L είναι ιδιαίτερα σημαντική. Όσο μεγαλύτερη είναι η ιδιοτιμή αυτή, τόσο πιο δύσκολο είναι να χωρίσουμε τον G σε συνιστώσες και τόσο πιο πολύ ο G επεκτείνεται.

2.1.6 Τυχαίοι Γράφοι

Οι *τυχαίοι γράφοι* (random graphs) θεωρούνται ένα από τα πιο σύγχρονα πεδία της Θεωρίας Γραφημάτων. Οφείλουν την ύπαρξη τους σε μία σειρά δημοσιεύσεων από τους Erdős και Rényi ανάμεσα στο 1940 και 1950.

Σε σύγκριση με την παραδοσιακή Θεωρία Γραφημάτων οι τυχαίοι γράφοι εξετάζουν ακριβώς το ίδιο πρόβλημα από μία άλλη όμως οπτική, κατάλληλη για τη μελέτη της δυναμικής συμπεριφοράς των δικτύων. Οι τυχαίοι γράφοι επίσης προσφέρουν ένα εναλλακτικό μοντέλο δικτύου για περιπτώσεις όπου επικρατεί η στοχαστική συμπεριφορά.

Σε γενικές γραμμές συναντώνται τρία βασικά και στενά συνδεδεμένα μοντέλα τυχαίων γράφων στη βιβλιογραφία. Ο πιθανοτικός χώρος σε κάθε περίπτωση αποτελείται από γράφους ενός σταθερού συνόλου από n διακριτούς κόμβους $V = [n] = \{1, 2, \dots, n\}$. Στη συνέχεια θα χρησιμοποιήσουμε τον συμβολισμό M, N για να αναπαραστήσουμε τον αριθμό των κόμβων.

Για $0 \leq M \leq N$, τότε ο χώρος $G(n, M)$ αποτελείται από όλους τους $\binom{N}{M}$ υπογράφους του K_n με M ακμές. Έτσι, το $G(n, M)$ μοντέλο περιγράφει τους γράφους που θα μπορούσαν να εξαχθούν από τον υποχώρο του K_n , ο οποίος περιέχει μόνο γράφους με M ακμές. Η πιθανότητα ότι ο G_M είναι ακριβώς ένας σταθερός γράφος H στο $[n]$ με M ακμές είναι:

$$P_M(G_M = H) = \binom{N}{M}^{-1} \quad 2.1$$

Ο χώρος $G(n, p)$ ($G(n, \mathbb{P}(\text{edge}) = p)$) ορίζεται με πιθανότητα $0 \leq p \leq 1$. Ένα τυχαίο στοιχείο αυτού του χώρου αντιστοιχίζεται στην ανεξάρτητη επιλογή ακμών με

πιθανότητα p , για όλες τις πιθανές υπάρχουσες ακμές σε ένα γράφο n κόμβων. Η πιθανότητα ενός συγκεκριμένου γράφου H στο $[n]$ με M ακμές είναι:

$$p^m(1-p)^{N-m} \quad 2.2$$

όπου κάθε μία από τις m ακμές του H πρέπει να επιλεγθεί και καμία από τις $N - m$ δεν επιτρέπεται να επιλεγεί.

Τέλος, ο χώρος $G(n, 1/2)$ αποκτάται επιλέγοντας τυχαία έναν από τους 2^N γράφους στο $[n]$, δηλαδή είναι ο χώρος των Τυχαίων Γράφων τάξης n , όπου όλοι οι γράφοι θεωρούνται ισοπίθανοι.

3 *Ανάλυση Σύνθετων και Κοινωνικών Δικτύων*

3.1 *Χαρακτηριστικά και Μετρικές των Σύνθετων Δικτύων*

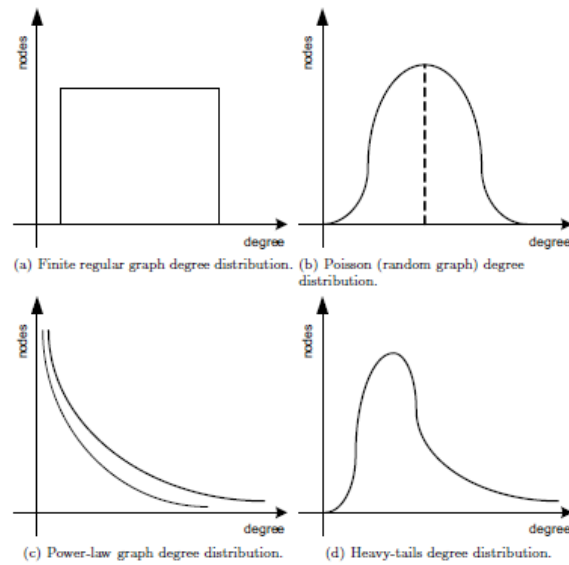
Στην ενότητα αυτή θα περιγραφούν βασικά χαρακτηριστικά και μετρικές που χρησιμεύουν στην Ανάλυση Σύνθετων Δικτύων. Τα χαρακτηριστικά και οι μετρικές αυτές παρέχουν μία ποσοτική περιγραφή της δομής ενός δικτύου, τόσο κατανεμημένα σε επίπεδο κάθε κόμβου, όσο και ολιστικά. Ιδιαίτερα ενδιαφέρον είναι το γεγονός ότι στην επιστήμη της Ανάλυσης Σύνθετων Δικτύων έχουν παρουσιαστεί συγκεκριμένοι τύποι δικτύων (οι οποίοι θα αναλυθούν στην Ενότητα 3.3.), όπου όλοι οι γράφοι που βρίσκονται στην ίδια κατηγορία παρουσιάζουν κοινά χαρακτηριστικά ως προς τις παραπάνω μετρικές.

3.1.1 *Κατανομή Βαθμού Κόμβων*

Όπως έχει αναφερθεί προηγουμένως, ο βαθμός ενός κόμβου αναφέρεται στον αριθμό των γειτόνων του. Η κατανομή του βαθμού των κόμβων $P(k)$, επομένως, είναι η πιθανότητα ένας τυχαία επιλεγμένος κόμβος να έχει ακριβώς k γείτονες. Στην ανάλυση των σύνθετων κοινωνικών δικτύων, η κατανομή του βαθμού των κόμβων χρησιμοποιείται ευρέως για την ταξινόμηση των δικτύων.

Στους τυχαίους γράφους, η πλειοψηφία των κόμβων έχει περίπου τον ίδιο βαθμό, κοντά στη μέση τιμή των βαθμών των κόμβων, $\langle k \rangle$. Η κατανομή του βαθμού των κόμβων για έναν τυχαίο γράφο αντιστοιχίζεται με μία κατανομή Poisson με κορυφή στο $P(\langle k \rangle)$ [3].

Στην πλειοψηφία, όμως, των σύνθετων δικτύων που συναντούμε, η κατανομή του βαθμού των κόμβων αποκλίνει σημαντικά από μία κατανομή Poisson. Συγκεκριμένα, για έναν μεγάλο αριθμό δικτύων, όπως ο Παγκόσμιος Ιστός, το Διαδίκτυο κ.α. [3] η κατανομή του βαθμού των κόμβων ακολουθεί την κατανομή νόμου-δύναμης (power-law) (με μία “ουρά” προς τα δεξιά) $P(k) \sim k^{-\gamma}$, δηλαδή ένας πολύ μικρός αριθμός των κόμβων του δικτύου έχει μεγάλο αριθμό



Σχήμα 4: Παραδείγματα κατανομής βαθμών κόμβων για διάφορους τύπους δικτύων (γραμμική κλίμακα) [1]

γειτόνων, ενώ η συντριπτική πλειοψηφία των κόμβων έχει μικρό βαθμό. Τέτοια δίκτυα αποκαλούνται δίκτυα ελεύθερης κλίμακας (scale free) .

3.1.2 Μέσο Μήκος Μονοπατιού

Σε αντίθεση με τον βαθμό κόμβων που είναι μία κατανεμημένη μετρική που χαρακτηρίζει έναν συγκεκριμένο κόμβο, το μέσο μήκος μονοπατιού (*average path length*) είναι μία μετρική που χαρακτηρίζει συνολικά ένα δίκτυο. Ορίζεται ως ο μέσος όρος του μήκους των συντομότερων μονοπατιών μεταξύ όλων των ζεύγων κόμβων σε ένα δίκτυο. Υποδεικνύει, έτσι, την αναμενόμενη απόσταση που θα συναντήσει κάποιος μεταξύ ενός τυχαία επιλεγμένου ζεύγους κόμβων σε ένα δίκτυο.

Μία θεμελιώδης και σημαντική πτυχή για τον υπολογισμό του συντομότερου μήκους μονοπατιού είναι ο ορισμός του μέτρου της απόστασης που εμπλέκεται για τον υπολογισμό των συντομότερων μονοπατιών. Η πιο συνηθισμένη εκδοχή που υιοθετείται για τον ορισμό της “απόστασης” είναι ο αριθμός των βημάτων (*hops*) μεταξύ δύο κόμβων, που δηλώνει τον ελάχιστο αριθμό ζεύξεων (ακμών, αν αναπαραστήσουμε ένα δίκτυο ως γράφημα) που χωρίζουν δύο κόμβους. Σε ορισμένες περιπτώσεις (όπως για παράδειγμα σε γράφους με βάρη) το μέσο μήκος μονοπατιού που συνδέει δύο κόμβους ορίζεται ως το μονοπάτι που οι ζεύξεις (ακμές) του συνεισφέρουν αθροιστικά στο ελάχιστο βάρος, επομένως μπορεί να διαφέρει από το συντομότερο μονοπάτι μεταξύ των ίδιων κόμβων, όταν αυτό ορίζεται στην περίπτωση του μικρότερου αριθμού βημάτων μεταξύ των κόμβων αυτών.

3.1.3 Συντελεστής Ομαδοποίησης

Μία συνηθισμένη ιδιότητα των κοινωνικών δικτύων είναι η δημιουργία κλικών, που αναπαριστούν κύκλους φίλων ή γνωστών στους οποίους κάθε μέλος γνωρίζει κάθε άλλο μέλος. Αυτή η έμφυτη τάση ομαδοποίησης ποσοτικοποιείται από το *Συντελεστή Ομαδοποίησης (Clustering Coefficient)* [4], μία ιδέα που έχει τις ρίζες της στον τομέα της Κοινωνιολογίας. Για έναν επιλεγμένο κόμβο i στο δίκτυο ο οποίος έχει k_i ακμές που τον συνδέουν σε k_i άλλους κόμβους, εάν οι κοντινότεροι γείτονες του αρχικού κόμβου αποτελούσαν μέρος μίας κλίκας, τότε θα υπήρχαν $k_i(k_i - 1)/2$ ακμές μεταξύ τους. Ο λόγος μεταξύ του αριθμού E_i των ακμών που πραγματικά υπάρχουν μεταξύ αυτών των k_i κόμβων και του συνολικού αριθμού $k_i(k_i - 1)/2$ μας δίνει την τιμή του συντελεστή ομαδοποίησης του κόμβου i ,

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad 3.1$$

Ο συντελεστής ομαδοποίησης ολόκληρου του δικτύου είναι ο μέσος όρος όλων των μεμονωμένων C_i .

Σε έναν Τυχαίο Γράφο, από τη στιγμή που οι ακμές κατανέμονται τυχαία, ο συντελεστής ομαδοποίησης είναι $C = p$.

3.2 Κεντρικότητα Κόμβων

Μία σημαντική πτυχή της θεωρίας γραφημάτων και της ανάλυσης κοινωνικών δικτύων είναι η αναγνώριση της σημασίας και της επίδρασης που έχει ένας κόμβος στο δίκτυο. Για το λόγο αυτό έχει υιοθετηθεί η έννοια της “κεντρικότητας”, η οποία είναι μία συνάρτηση που αποδίδει σε κάθε κόμβο μία πραγματική τιμή με βάση τη θέση του στο συνολικό δίκτυο. Η αναγνώριση της σημασίας ενός κόμβου στο συνολικό δίκτυο και η κατάταξη αυτών είναι ένα ζήτημα, ως ένα βαθμό, υποκειμενικό. Για το λόγο αυτό έχουν προταθεί διάφοροι τύποι κεντρικότητας που εξετάζουν κάθε φορά από μία διαφορετική οπτική το βαθμό της κεντρικότητας και της κατάταξης των κόμβων ως προς τη σημασία τους με βάση τον συγκεκριμένο τύπο κεντρικότητας που επιλέγεται. Ο παρακάτω πίνακας συνοψίζει κάποιους από τους πιο διαδεδομένους τύπους κεντρικότητας, οι οποίοι θα αναλυθούν στη συνέχεια:

ΠΙΝΑΚΑΣ 3.1

ΠΕΡΙΛΗΨΗ ΤΥΠΩΝ ΚΕΝΤΡΙΚΟΤΗΤΑΣ ΣΤΗΝ ΑΝΑΛΥΣΗ ΣΥΝΘΕΤΩΝ ΔΙΚΤΥΩΝ

Τύπος	Μετρική	Υπολογισμός
Βαθμού	Βαθμός Κόμβου	Κατανεμημένος
Εγγύτητας	Απόσταση ζεύγους κόμβων	Κεντρικοποιημένος
Ενδιαμεσικότητας	Συντομότερα μονοπάτια	Κεντρικοποιημένος / προσεγγίσεις
Ιδιοδιανυσμάτων	Ιδιοτιμές του πίνακα γειτνίασης	Κατανεμημένος

3.2.1 Κεντρικότητα Βαθμού

Ο πιο απλός τύπος κεντρικότητας που έχει οριστεί είναι αυτός της *κεντρικότητας βαθμού* (*degree centrality*). Διαισθητικά, μπορούμε να θεωρήσουμε ότι όσο πιο πολλές συνδέσεις με άλλους κόμβους έχει ένας κόμβος, τόσο πιο “δημοφιλής” και άρα σημαντικός είναι στη ροή της πληροφορίας μέσα σε ένα συγκεκριμένο δίκτυο. Η πιο απλή τοπολογία στην οποία μπορούμε να αναφερθούμε για να δείξουμε τη σημασία ενός τέτοιου κόμβου είναι η τοπολογία αστέρα, όπου ο πιο “κεντρικός” κόμβος είναι πράγματι αυτός με το μεγαλύτερο βαθμό, αφού η αφαίρεσή του θα οδηγούσε σε πλήρη αποσύνδεση του δικτύου που ακολουθεί τη συγκεκριμένη δομή.

Η πιο απλή συνάρτηση με την οποία θα μπορούσαμε να συνδέσουμε το βαθμό ενός κόμβου με την “κεντρικότητα” του στο δίκτυο είναι μία απλή γραμμική συνάρτηση. Υποθέτοντας ότι ο $A = [a_{ij}]$ είναι ο πίνακας γειτνίασης μίας τοπολογίας δικτύου, τότε η *κεντρικότητα βαθμού* και η *σχετική κεντρικότητα βαθμού* (*relative degree centrality*) ενός κόμβου k ορίζονται ως εξής:

$$C_D(k) = \sum_{i=1}^n a_{ik} \quad 3.2$$

$$C'_D(k) = \frac{\sum_{i=1}^n a_{ik}}{n-1} \quad 3.3$$

Το μέγεθος του $C_D(k)$ είναι κατά βάση μία συνάρτηση που εξαρτάται από το μέγεθος του δικτύου για το οποίο υπολογίζεται. Σε ορισμένες περιπτώσεις, όμως, απαιτείται ο υπολογισμός ενός μεγέθους το οποίο θα είναι ανεξάρτητο από το μέγεθος του δικτύου. Για το λόγο αυτό

ορίζεται και η σχετική κεντρικότητα $C_D'(k)$ η οποία κανονικοποιεί την τιμή ενός κόμβου με βάση το συνολικό αριθμό κόμβων στο δίκτυο (δηλαδή το μέγεθος του δικτύου).

Η κεντρικότητα βαθμού, παρ' όλο που παρέχει έναν εύκολο και βολικό τρόπο για να ορίσουμε έναν τύπο που θα παρέχει μία ολιστική εικόνα κεντρικότητας για ένα γράφημα, δεν μπορεί να καταγράψει τη δυναμική της ροής πληροφορίας σε ένα δίκτυο. Για παράδειγμα, ένας κόμβος που έχει μία υψηλή τιμή βαθμού κόμβου μπορεί, για διάφορους λόγους, να μην είναι κομβικός στη ροή της πληροφορίας. Έτσι, προκύπτει η ανάγκη για ορισμό διαφορετικών τύπων κεντρικότητας που μπορεί να είναι πιο απαιτητικοί σε υπολογιστικό χρόνο, αλλά παρέχουν μία καλύτερη κατάταξη της σημασίας των κόμβων με βάση τον έλεγχο τους πάνω στη ροή της πληροφορίας.

3.2.2 Κεντρικότητα Εγγύτητας

Η κεντρικότητα εγγύτητας (*Closeness Centrality*) υπολογίζει τη σημασία των κόμβων σε ένα δίκτυο με βάση το κατά πόσο αυτοί είναι χωρικά οι πιο κεντρικοί κόμβοι στο δίκτυο αυτό. Συγκεκριμένα, αναθέτει υψηλές τιμές κεντρικότητας στους κόμβους που είναι σχετικά σε μία κεντρική θέση στην τοπολογία του δικτύου λαμβάνοντας υπόψιν την απόσταση των κόμβων από τους υπόλοιπους κόμβους του δικτύου με βάση μία συγκεκριμένη μετρική απόστασης.

Ο ορισμός της κεντρικότητας εγγύτητας βασίζεται στην απόσταση κάθε κόμβου από όλους τους υπόλοιπους, θεωρώντας ως πιο κεντρικό κόμβο αυτόν που αθροιστικά έχει τη μικρότερη απόσταση από τους υπόλοιπους κόμβους του δικτύου. Επομένως, περιγράφει ποσοτικά το πόσο “κοντά” είναι ένας κόμβος από όλους τους υπόλοιπους. Ως εθίσται, ως απόσταση θεωρούμε το μήκος του συντομότερου μονοπατιού από έναν κόμβο προς έναν άλλο για τη δεδομένη τοπολογία δικτύου που εξετάζουμε.

Επομένως, συμβολίζοντας ως $d(i, k)$ την απόσταση (ή το μήκος) του συντομότερου μονοπατιού μεταξύ δύο κόμβων i και k , η κεντρικότητα εγγύτητας και η *σχετική κεντρικότητα εγγύτητας* (*relative closeness centrality*) ορίζονται αντιστοίχως ως:

$$C_p(k) = \left(\sum_{i=1}^n d(i, k) \right)^{-1} \quad 3.4$$

$$C'_p(k) = \left(\frac{\sum_{i=1}^n d(i, k)}{n-1} \right)^{-1} = \frac{n-1}{\sum_{i=1}^n d(i, k)} \quad 3.5$$

Από τον ορισμό μπορούμε να αντιληφθούμε ότι αυτή η μετρική κεντρικότητας ορίζει ως πιο κεντρικό κόμβο ενός δικτύου, αυτόν από τον οποίο μπορούμε να φτάσουμε στους υπόλοιπους κόμβους με το ελάχιστο αθροιστικά κόστος.

Στην περίπτωση της σχετικής κεντρικότητας εγγύτητας γίνεται απλώς κανονικοποίηση ως προς τον αριθμό των κόμβων του δικτύου. Έτσι ένας κεντρικός κόμβος λαμβάνει τιμή κοντά στη μονάδα (με την τιμή 1 να δίνεται σε έναν κόμβο v στην ιδανική περίπτωση που συνδέεται με όλους τους υπόλοιπους κόμβους του δικτύου) με την τιμή αυτή να μειώνεται όσο αυξάνεται η μέση απόσταση ενός κόμβου από τους υπόλοιπους κόμβους.

Τόσο η κεντρικότητα βαθμού που αναφέρθηκε προηγουμένως, όσο και η κεντρικότητα εγγύτητας έχουν ως βασικό τους μειονέκτημα ότι ορίζονται αποκλειστικά για συνδεδεμένα δίκτυα. Εξ' ορισμού σε μη-συνδεδεμένα δίκτυα το συντομότερο μονοπάτι μεταξύ δύο κόμβων (όταν αυτό στην πραγματικότητα δεν υφίσταται) λαμβάνει τιμή στο άπειρο, οδηγώντας έτσι σε αδυναμία υπολογισμού της συγκεκριμένης μετρικής κεντρικότητας. Επομένως, ακόμη κι αν η κεντρικότητα εγγύτητας μπορεί να καλύψει σε ένα βαθμό τη διαδικασία ροής και διάχυσης της πληροφορίας, ανακύπτουν υπολογιστικά προβλήματα για απομονωμένους κόμβους του δικτύου ή για δίκτυα με δυναμική-μεταβλητή δομή (όπως τα δίκτυα με ανοχή στην καθυστέρηση στα δίκτυα επικοινωνιών). Για το λόγο αυτό έχει προταθεί η κεντρικότητα ενδιαμεσικότητας (betweenness centrality) η οποία δεν παρουσιάζει τα παραπάνω προβλήματα και μπορεί να υπολογιστεί τόσο για συνδεδεμένα δίκτυα όσο και για μη-συνδεδεμένα δίκτυα.

3.2.3 Κεντρικότητα Ενδιαμεσικότητας

Ίσως η πιο σημαντική μετρική κεντρικότητας είναι η *κεντρικότητα ενδιαμεσικότητας* (betweenness centrality). Για τον υπολογισμό της εμπλέκονται και πάλι, όπως προηγουμένως, τα συντομότερα μονοπάτια που υπάρχουν σε ένα δίκτυο. Συγκεκριμένα, με βάση τον ορισμό της μετρικής κεντρικότητας ενδιαμεσικότητας, η κεντρικότητα ενός κόμβου ισούται με τον αριθμό των συντομότερων μονοπατιών (από όλους τους κόμβους του δικτύου προς όλους τους κόμβους του δικτύου) που διέρχονται από το συγκεκριμένο κόμβο. Με βάση αυτό τον υπολογισμό, λοιπόν, ένας κόμβος ο οποίος βρίσκεται με μεγάλη συχνότητα πάνω στα συντομότερα μονοπάτια που προκύπτουν ανάμεσα σε δύο κόμβους (για κάθε ζεύγος κόμβων στο δίκτυο) είναι και σημαντικός στην πραγματική ροή της πληροφορίας πάνω σε ένα δίκτυο, εάν θεωρήσουμε ότι η ροή της πληροφορίας (ή η δρομολόγηση της) ακολουθεί το μονοπάτι

ελάχιστου μήκους ανάμεσα σε όλα τα πιθανά μονοπάτια ανάμεσα σε δύο κόμβους (κάτι που συμβαίνει συχνά στα δίκτυα επικοινωνιών).

Στην περίπτωση, όμως, που υπάρχουν περισσότερα του ενός γεωδαισιακά (συντομότερα μονοπάτια) που συνδέουν ένα ζεύγος σημείων, δηλαδή εναλλακτικά μονοπάτια που έχουν το ίδιο-ελάχιστο- μήκος, ο ορισμός αυτός οφείλει να τροποποιηθεί. Τότε ένας κόμβος που βρίσκεται ανάμεσα σε κάποια, αλλά όχι όλα, τα γεωδαισιακά που συνδέουν δύο κόμβους έχει πιο περιορισμένο έλεγχο στη ροή της πληροφορίας. Σε μία τέτοια περίπτωση εισάγεται ο ορισμός της μερικής κεντρικότητας ενδιαμεσικότητας που είναι και αυτή που κατά βάση συναντούμε στη βιβλιογραφία και χρησιμοποιείται για τον υπολογισμό της κεντρικότητας των κόμβων.

Δεδομένου ενός κόμβου k και ενός μη διατεταγμένου ζεύγους κόμβων $\{i, j\}$, όπου $i \neq j \neq k$, η μερική κεντρικότητα ενδιαμεσικότητας $b_{ij}(k)$ ενός κόμβου k σε σχέση με το ζεύγος (i, j) μπορεί να οριστεί με τον ακόλουθο τρόπο: I) Εάν δεν υπάρχει κάποιο μονοπάτι με το οποίο μπορούν οι κόμβοι i και j να προσεγγίσουν ο ένας τον άλλο, τότε $b_{ij}(k) = 0$. II) Εάν οι κόμβοι i και j συνδέονται με ένα ή περισσότερα μονοπάτια, τότε η πιθανότητα να χρησιμοποιήσουμε κάποιο από τα πιθανά εναλλακτικά γεωδαισιακά μονοπάτια που συνδέουν τους i και j είναι $\frac{1}{g_{ij}}$, όπου g_{ij} είναι ο αριθμός των γεωδαισιακών που συνδέουν τους i και j . Η δυνατότητα του κόμβου k στον έλεγχο της ροής της πληροφορίας μεταξύ των i και j μπορεί να οριστεί ως η πιθανότητα ο κόμβος k να βρίσκεται “πάνω” σε ένα τυχαία επιλεγμένο γεωδαισιακό μονοπάτι που συνδέει τους i και j . Εάν $g_{ij}(k)$ είναι ο αριθμός των γεωδαισιακών που συνδέουν τους κόμβους i και j που περιέχουν τον k , τότε η προαναφερόμενη πιθανότητα είναι:

$$b_{ij}(k) = \frac{g_{ij}(k)}{g_{ij}} \quad 3.6$$

με την παραπάνω σχέση να λαμβάνει την τιμή 1 όταν ο κόμβος k βρίσκεται “πάνω” στο μοναδικό γεωδαισιακό μονοπάτι ανάμεσα στους κόμβους i και j , ή σε κάθε διαθέσιμο μονοπάτι ελάχιστου μήκους.

Για να υπολογίσουμε τη συνολική κεντρικότητα ενδιαμεσικότητας του κόμβου k , αθροίζουμε τη μερική ενδιαμεσικότητα (partial betweenness) για όλα τα μη διατεταγμένα ζεύγη των κόμβων $i \neq j \neq k$:

$$C_B(k) = \sum_{i \neq j \neq k} \sum_{i < j}^n b_{ij}(k) \quad 3.7$$

Το βασικό πρόβλημα της συγκεκριμένης κεντρικότητας είναι ότι προϋποθέτει και αυτή τον υπολογισμό του συνολικού αριθμού των συντομότερων μονοπατιών σε ένα δίκτυο. Ακόμη και με τον state-of-the-art αλγόριθμο που προτάθηκε από τον Brandes [5] για τον υπολογισμό της κεντρικότητας ενδιαμεσικότητας ενός κόμβου, ο οποίος κάνει χρήση αναδρομικών σχέσεων για τον υπολογισμό και την άθροιση της εξάρτησης ενός κόμβου-πηγής $s \in V$ πάνω σε ένα κόμβο $v \in V$, η συνολική πολυπλοκότητα για τον υπολογισμό της συγκεκριμένης μετρικής κεντρικότητας είναι $O(nm)$ για δίκτυα χωρίς βάρη και $O(nm + n^2 \log n)$ για δίκτυα με βάρη, όπου n ο αριθμός των κόμβων του δικτύου και m ο αριθμός των ακμών που υπάρχουν στο δίκτυο. Πολυπλοκότητα που καθιστά ιδιαίτερα χρονοβόρο, έως και αδύνατο, τον ακριβή υπολογισμό της κεντρικότητας ενός κόμβου με βάση τη συγκεκριμένη μετρική για πραγματικούς γράφους που φτάνουν σε μέγεθος ή και ξεπερνούν τις εκατοντάδες χιλιάδες κόμβων. Για το λόγο αυτό έχουν προταθεί διάφοροι προσεγγιστικοί αλγόριθμοι, όπως οι [6] και [7], ώστε να δώσουν μία εκτίμηση για την τιμή και την κατάταξη της κεντρικότητας ενδιαμεσικότητας ενός κόμβου σε υπολογιστικά προσιτό χρόνο.

Ο αλγόριθμος του Brandes [5] για τον υπολογισμό της κεντρικότητας ενδιαμεσικότητας, καθώς και παραλλαγές του για τον υπολογισμό άλλων τύπων κεντρικότητας που κάνουν χρήση των γεωδαισιακών μονοπατιών (αποτελούν δηλαδή παραλλαγές της κεντρικότητας ενδιαμεσικότητας) [8] θα αναλυθούν στο Κεφάλαιο 5.

3.3 Μοντέλα Αναπαράστασης και Δημιουργίας Συνθετικών

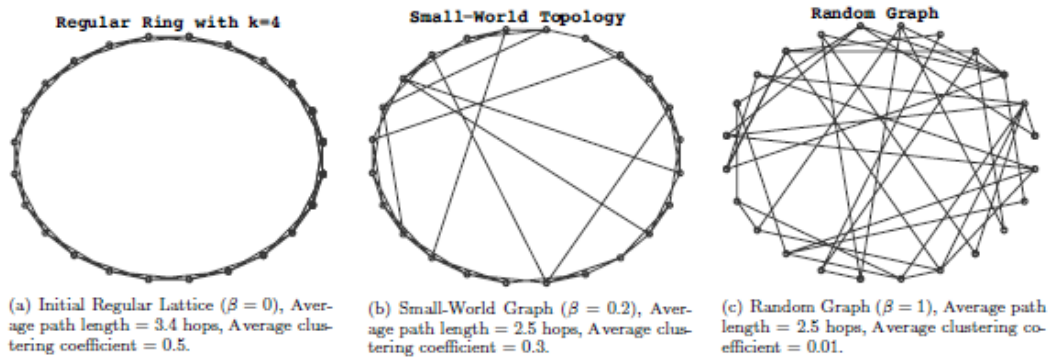
Δικτύων

Στην παρούσα ενότητα θα αναλυθούν συνοπτικά τρία αρκετά διαδεδομένα μοντέλα που έχουν προταθεί για την αναπαράσταση σύνθετων κοινωνικών δικτύων (δίκτυα μικρού κόσμου, ελεύθερης κλίμακας, τυχαίων γραφημάτων), καθώς και οι αντίστοιχοι αλγόριθμοι που στοχεύουν στη δημιουργία συνθετικών γραφημάτων που θα ακολουθούν τις θεωρητικές αρχές πάνω στις οποίες έχουν στηριχθεί τα τρία αυτά μοντέλα. Οι αλγόριθμοι που έχουν επιλεγεί για τη δημιουργία των συνθετικών αυτών γραφημάτων είναι το μοντέλο των Watts και Strogatz για τη δημιουργία συνθετικών δικτύων τύπου μικρού κόσμου, το μοντέλο της “Επιλεκτικής Προσκόλλησης” (Preferential Attachment) των Barabasi-Albert για τη δημιουργία συνθετικών δικτύων ελεύθερης κλίμακας, ενώ τα τυχαία γραφήματα και το μοντέλο των Erdős-Rényi για την παραγωγή των τυχαίων γραφημάτων αναλύθηκαν στο Κεφάλαιο 2.1.6. Και τα 3 αυτά μοντέλα χρησιμοποιήθηκαν στη διπλωματική εργασία για τη δημιουργία συνθετικών γραφημάτων για τον έλεγχο των αποτελεσμάτων.

3.3.1 Δίκτυα Μικρού Κόσμου

Η μελέτη της δομής των κοινωνικών δικτύων και των γραφημάτων που προκύπτουν από την αναπαράσταση της αλληλεπίδρασης των οντοτήτων που απαρτίζουν ένα κοινωνικό δίκτυο έριξε φως σε ορισμένα βασικά χαρακτηριστικά που ακολουθούν κατά βάση αυτά τα δίκτυα. Ένα βασικό χαρακτηριστικό τους είναι η υψηλή τιμή του συντελεστή ομαδοποίησης των κόμβων και το σύντομο κατά μέσο όρο μήκος των μονοπατιών που υπάρχουν στο δίκτυο. Θεωρούμε ως υψηλό έναν συντελεστή ομαδοποίησης όταν αυτός είναι σημαντικά μεγαλύτερος από την τιμή p , όπου p η πιθανότητα να συνδέονται δύο κόμβοι σε έναν τυχαίο γράφο (η οποία όπως εξηγήθηκε στην ενότητα 3.1.3 είναι και η τιμή του συντελεστή ομαδοποίησης σε τυχαίους γράφους). Τέτοια υψηλή τιμή του συντελεστή ομαδοποίησης προκύπτει σε ένα κανονικό πλέγμα (σχήμα 5. α) όπου κάθε κόμβος συνδέεται μόνο με τους γειτονικούς (τοπολογικά) του κόμβους με βάση ένα συγκεκριμένο, προαποφασισμένο τρόπο, που είναι ίδιος για όλους τους κόμβους. Επομένως, με βάση αυτό το πλέγμα είναι πολύ πιθανό οι γείτονες ενός κόμβου να είναι γείτονες και μεταξύ τους, οδηγώντας έτσι σε έναν πολύ υψηλό συντελεστή ομαδοποίησης για το συνολικό δίκτυο. Από την άλλη μεριά, ως χαμηλή τιμή για το μέσο μήκος μονοπατιών ενός γράφου θεωρούμε τους γράφους αυτούς για τους οποίους το μέσο μήκος μονοπατιού τους μεγαλώνει πιο αργά σε σχέση με τον αριθμό των κόμβων τους n από τη γραμμική σχέση, για παράδειγμα σαν $\log(n)$. Όπως είναι ήδη γνωστό οι τυχαίοι γράφοι χαρακτηρίζονται από μέσο μήκος μονοπατιού το οποίο μεγαλώνει λογαριθμικά σε σχέση με τον αριθμό n των κόμβων τους [1]. Οπότε με βάση αυτό το γεγονός, μπορούν να θεωρηθούν ως σημείο αναφοράς για συγκρίσεις σχετικά με την τιμή του μέσου μήκος μονοπατιού. Ένας γράφος ο οποίος έχει μέσο μήκος μονοπατιού συγκρίσιμο με το αντίστοιχο ενός τυχαίου γράφου με τον ίδιο αριθμό κόμβων και ακμών, τότε θεωρείται ότι έχει σύντομο μέσο μήκος μονοπατιού.

Ο όρος “μικρός κόσμος” (*small-world*) επομένως χρησιμοποιείται για να περιγράψει γράφους οι οποίοι δεν είναι ούτε κανονικοί, ούτε τυχαίοι. Πιο συγκεκριμένα, γράφους οι οποίοι έχουν υψηλό συντελεστή ομαδοποίησης, όπως τα κανονικά πλέγματα, και την ίδια ώρα χαμηλό μέσο μήκος μονοπατιού, όπως οι τυχαίοι γράφοι.



Σχήμα 5: Παρουσίαση του b -μοντέλου των Watts και Strogatz. Οι παράμετροι του δικτύου είναι $n = 24, k = 4$ [1]

Πέρα από τα μαθηματικά χαρακτηριστικά που αναλύθηκαν προηγουμένως, το φαινόμενο του μικρού-κόσμου μπορεί να περιγραφεί στην περίπτωση των κοινωνικών δικτύων από ξένους μεταξύ τους χρήστες που συνδέονται μέσω ενός μικρού αριθμού ατόμων.

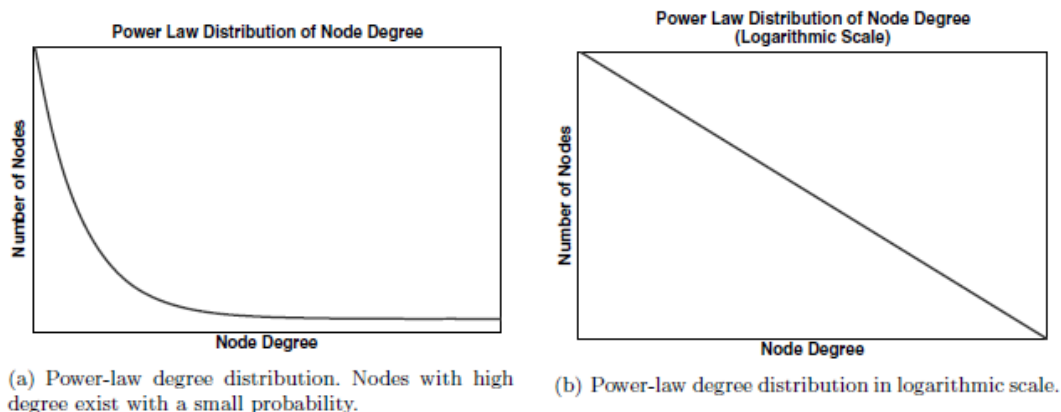
- **Το μοντέλο των Watts και Strogatz**

Το μοντέλο των Watts και Strogatz ξεκινάει από ένα γράφο με δομή κανονικού πλέγματος. Η αρχική αυτή δομή διασφαλίζει τον υψηλό βαθμό του συντελεστή ομαδοποίησης για τον γράφο που πρόκειται να παραχθεί. Έπειτα ένας κατάλληλος αριθμός “συντομεύσεων” (*shortcuts*) (δηλαδή ακμών μεταξύ δύο κόμβων που δεν είναι κοντινοί τοπολογικά στο πλέγμα) προστίθεται ώστε να εξασφαλιστεί επιπλέον και το σύντομο κατά μέσο όρο μήκος των μονοπατιών στο δίκτυο. Το πιο σύντομο μοντέλο που έχει προταθεί από τους Watts και Strogatz είναι το β – μοντέλο. Με βάση τον αλγόριθμο για το β – μοντέλο, ξεκινάμε από μία πλήρως καθορισμένη δομή πλέγματος δακτυλίου, όπου κάθε κόμβος i έχει k γείτονες, το σύνολο $\{i - k/2, \dots, i - 1, i + 1, \dots, i + k/2\}$, όπου χρησιμοποιούμε το σύμβολο $i - h$ για έναν κόμβο που βρίσκεται h θέσεις μακριά από τον i στην αντι-ωρολογιακή φορά και παρομοίως τον συμβολισμό $i + h$ στην ωρολογιακή φορά (βλ. Σχήμα 5(α)). Ο αλγόριθμος χρειάζεται $k/2$ γύρους για την εκτέλεση του, με κάθε έναν από αυτούς να αποτελείται από n υπο-γύρους. Σε έναν γύρο j , για κάθε έναν από τους υπο-γύρους, ένας κόμβος i του γράφου επιλέγεται με τη σειρά μαζί με την ακμή που τον συνδέει στον $i + j$ γείτονα του (με την ωρολογιακή φορά). Με πιθανότητα β η ακμή $(i, i + j)$ αναδιατάσσεται από τον κόμβο $i + j$ σε έναν τυχαία και ομοιόμορφα επιλεγμένο κόμβο του πλέγματος. Κατά τη διάρκεια των n υπο-γύρων επιλέγεται κάθε φορά ένας κόμβος (άρα συνολικά όλοι οι κόμβοι του γράφου) με ωρολογιακή φορά. Αμέσως μετά τον τερματισμό όλων των γύρων, όλες οι ακμές του γράφου

έχουν διασχιστεί ακριβώς μία φορά η κάθε μία. Στην περίπτωση που $\beta = 0$ ο κόμβος παραμένει ένα πλέγμα, ενώ όταν $\beta = 1$ ο αρχικός δακτύλιος τροποποιείται σε έναν απολύτως τυχαίο γράφο (η διαδικασία αυτή φαίνεται στο σχήμα 5).

3.3.2 Δίκτυα Ελεύθερης Κλίμακας

Ως δίκτυα ελεύθερης κλίμακας (*scale-free networks*) ονομάζουμε τα δίκτυα αυτά στα οποία υπάρχει μία έλλειψη κλιμάκωσης στην κατανομή του βαθμού των κόμβων του δικτύου. Συγκεκριμένα, η κατανομή του βαθμού των κόμβων αυτών των δικτύων ακολουθεί την κατανομή νόμου-δύναμης, όπου η πιθανότητα ένας κόμβος να αλληλεπιδρά με k ακόμη κόμβους είναι $P(k) \sim k^{-\gamma}$, όπου ο εκθέτης γ έχειδειχθεί ότι στα περισσότερα πραγματικά δίκτυα βρίσκεται μεταξύ των τιμών 2.1 και 4 [1]. Επιπρόσθετα, ένα ακόμη χαρακτηριστικό των δικτύων ελεύθερης κλίμακας είναι ότι οι κόμβοι που έχουν υψηλό βαθμό τείνουν να συνδέονται και μεταξύ τους [9]. Για παράδειγμα, στις ανθρώπινες κοινότητες, υπάρχουν συνήθως λίγοι άνθρωποι οι οποίοι είναι πολύ δημοφιλείς και γνωστοί στον περισσότερο κόσμο (όπως πολιτικοί, ηθοποιοί κ.α.) οι οποίοι επιπροσθέτως, όπως είναι φυσικό, γνωρίζονται και μεταξύ τους, ενώ στον αντίποδα οι περισσότεροι άνθρωποι είναι γνωστοί μόνο σε έναν πολύ μικρό κύκλο (οικογενειακό, επαγγελματικό κτλ.). Η κατανομή του βαθμού των κόμβων για τα δίκτυα ελεύθερης κλίμακας απεικονίζεται στο σχήμα 6, τόσο σε γραμμική κλίμακα όσο και σε λογαριθμική-λογαριθμική κλίμακα ($\log - \log$). Ιδιαίτερο χαρακτηριστικό των δικτύων ελεύθερης κλίμακας είναι ότι στην περίπτωση της $\log - \log$ κλίμακας η καμπύλη που απεικονίζει τη συσχέτιση μεταξύ βαθμού κόμβων και της κατανομής τους, φθίνει γραμμικά.



Σχήμα 6: Κατανομή του βαθμού κόμβων σε σύνθετα δίκτυα τύπου ελεύθερης κλίμακας

Πρέπει να τονιστεί ότι τα δίκτυα ελεύθερης κλίμακας παρουσιάζουν δύο σημαντικές ιδιότητες: είναι ιδιαίτερα ανθεκτικά σε περιστασιακά τυχαία σφάλματα, αλλά ιδιαίτερα

ευάλωτα σε συντονισμένες επιθέσεις [1]. Αυτό συμβαίνει γιατί σε ένα τυχαίο σφάλμα, ο κόμβος που θα τεθεί εκτός λειτουργίας κατά πάσα πιθανότητα θα παρουσιάζει μικρό αριθμό συνδέσεων με άλλους γείτονες, άρα δε θα οδηγήσει σε αδυναμία λειτουργίας του συνολικού δικτύου. Αντίθετα, μία οργανωμένη επίθεση σε έναν μικρό αριθμό κόμβων που όμως έχουν μεγάλο αριθμό συνδέσεων με γείτονες μπορεί να οδηγήσει, στην περίπτωση αδυναμίας λειτουργίας αυτών των κόμβων, σε πλήρη κατάρρευση του δικτύου.

Σε γενικές γραμμές, τα σύνθετα δίκτυα εξελίσσονται σύμφωνα με δύο βασικούς μηχανισμούς που δεν λαμβάνονται υπόψιν από τα μοντέλα τυχαίων γράφων. Το πρώτο είναι το μοντέλο της “ανάπτυξης” (*growth*) όπου οι νέοι κόμβοι τείνουν να συνδέονται σε υπάρχοντες κόμβους. Το άλλο, και πιο δημοφιλές μοντέλο, είναι το μοντέλο της “Επιλεκτικής Προσκόλλησης” (*preferential attachment*), όπου οι νέοι κόμβοι τείνουν να συνδέονται στους υπάρχοντες κόμβους με πιθανότητα ανάλογη στη δημοφιλία (δηλαδή των αριθμό συνδέσεων) των υπάρχοντων κόμβων. Πιο συγκεκριμένα, η πιθανότητα ένας κόμβος να συνδέεται σε έναν κόμβο που ήδη υπάρχει δεν είναι ομοιόμορφη, αλλά αντίθετως ευνοεί κόμβους με μεγαλύτερη συνδεσιμότητα ή βαθμό, δημιουργώντας έτσι δύο ομογενείς κατηγορίες κόμβων. Μία κατηγορία με μικρό αριθμό κόμβων με υψηλό βαθμό που λειτουργούν ως *κεντρικοί κόμβοι* (*hubs*) και μία άλλη κατηγορία που θα περιέχει την μεγάλη πλειοψηφία των κόμβων του δικτύου που έχουν μικρό βαθμό.

- **Το μοντέλο των Barabasi-Albert**

Σε αντίθεση με το μοντέλο Watts-Strogatz για τη δημιουργία συνθετικών γράφων τύπου Μικρού-Κόσμου, το μοντέλο που πρότειναν οι Barabasi-Albert για τη δημιουργία συνθετικών γράφων τύπου Ελεύθερης-Κλίμακας δεν στηρίζεται πάνω σε τροποποιήσεις μίας δεδομένης τοπολογίας, αλλά κατασκευάζει μία τοπολογία από την αρχή. Έτσι καλύπτεται και η ανάγκη να παραχθούν μοντέλα που θα καλύπτουν και τα δυναμικά χαρακτηριστικά που παρουσιάζουν τα κοινωνικά δίκτυα με την εξέλιξή τους στο πέρασμα του χρόνου (όπως η προσθήκη ή αφαίρεση κόμβων και ακμών). Το μοντέλο των Barabasi-Albert ακολουθεί ένα συνδυασμό των νόμων της “Επιλεκτικής Προσκόλλησης” (*preferential attachment*) και της “ανάπτυξης” (*growth*) τα οποία αναφέρθηκαν και προηγουμένως, για να δημιουργήσει γράφους συνδετικών δικτύων που θα ακολουθούν κατανομή νόμου-δύναμης για το βαθμό των κόμβων. Με αυτό τον τρόπο το δίκτυο που κατασκευάζεται συλλαμβάνει σε μεγάλο βαθμό τη δυναμική συμπεριφορά των πραγματικών δικτύων. Γενικότερα, μέσα από την παρακολούθηση και την στατιστική επεξεργασία, τα πραγματικά δίκτυα μεγαλώνουν σε μέγεθος με τον χρόνο, ενώ η πιθανότητα ένας κόμβος να συνδεθεί με έναν συγκεκριμένο κόμβο βασίζεται στη δημοφιλία (τον βαθμό) του κόμβου. Τέτοια παραδείγματα πραγματικών δικτύων παρουσιάζονται στις κοινωνικές

επιστήμες (γραφήματα συνεργασίας ηθοποιών ή συνεργασίας στη συγγραφή μίας επιστημονικής δημοσίευσης-μελέτης), στη μοριακή βιολογία (στο μεταβολισμό των κυττάρων και στα γενετικά ρυθμιστικά δίκτυα) και στο Internet (γραφήματα του διαδικτυακού ιστού, δηλαδή συνδέσεων μεταξύ ιστοσελίδων, γραφήματα τόσο στο επίπεδο router, όσο και στο επίπεδο αυτόνομων συστημάτων) [9].

Ο αλγόριθμος των **Barabasi-Albert** έχει ως εξής:

Αρχικά το δίκτυο αποτελείται από m_0 κόμβους. Κάθε νεοεισερχόμενος κόμβος συνδέεται στο δίκτυο μέσω m συνδέσεων με m κατάλληλους κόμβους (έναν για κάθε κόμβο), που ήδη υπάρχουν στο δίκτυο. Επομένως σε κάθε χρονοσχισμή t :

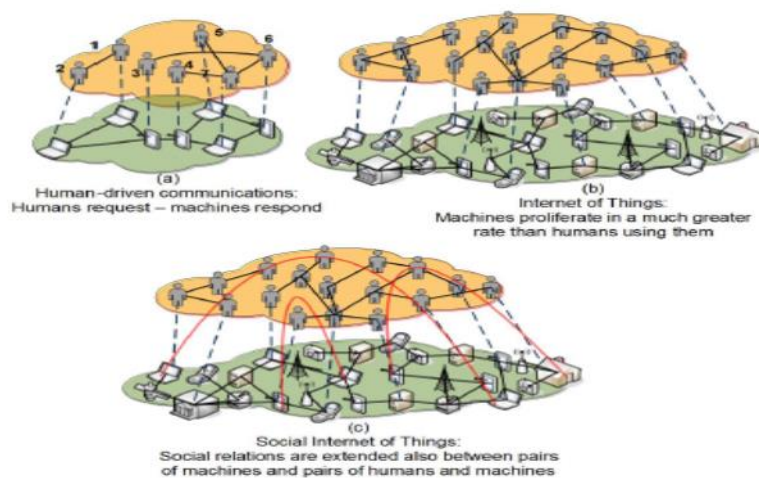
- Ένας νέος κόμβος εισέρχεται στο δίκτυο και συνδέεται με m υπάρχοντες κόμβους
- Κάθε ένας από τους m υπάρχοντες κόμβους επιλέγεται με βάση το μοντέλο της επιλεκτικής προσκόλλησης “preferential attachment”. Οπότε ο κόμβος i με βαθμό $k(i)$ επιλέγεται με πιθανότητα $P(k_i) = \frac{k_i(t)}{\sum_j k_j(t)}$
- Ο αριθμός των κόμβων στο δίκτυο ισούται με $N_t = M_0 + t$

Τα συνθετικά γραφήματα που κατασκευάζονται με βάση το μοντέλο των Barabasi-Albert, έχειδει ότι παρουσιάζουν σημαντικά υψηλότερο βαθμό του συντελεστή ομαδοποίησης και χαμηλότερο κατά μέσο όρο μήκος μονοπατιών απ' ότι ένας αντίστοιχος τυχαίος γράφος. Σε σχέση με τα δίκτυα Μικρού-Κόσμου, παρ' όλο που το μέσο μήκος μονοπατιού αυξάνεται λογαριθμικά σε σχέση με τον αριθμό των κόμβων στο δίκτυο, ο συντελεστής ομαδοποίησης μειώνεται με την αύξηση του αριθμού των κόμβων αντί να παραμένει σταθερός.

4 Ενσωμάτωση των Γράφων στην Υπερβολική Γεωμετρία

4.1 Ανάλυση Δεδομένων Μεγάλης Κλίμακας στον Υπερβολικό Χώρο Συντεταγμένων

Η ανάπτυξη των σύνθετων και κοινωνικών δικτύων έχει οδηγήσει στην παραγωγή και τη μεταφορά μεγάλου όγκου δεδομένων. Η αντίστοιχη περιοχή έρευνας και τεχνολογικής ανάπτυξης έχει ονομαστεί Ανάλυση Κοινωνικών Δικτύων. Τα δίκτυα επικοινωνιών και τα κοινωνικά δίκτυα έχουν αναπτυχθεί παράλληλα την τελευταία δεκαετία σε ένα σύνθετο ιεραρχικό σύστημα, το οποίο αναπτύσσεται μη-συμμετρικά με το χρόνο, όπως φαίνεται στο παρακάτω σχήμα:



Σχήμα 7: Συν-εξέλιξη δικτύων επικοινωνιών (σύνθετων)-κοινωνικών δικτύων [10]

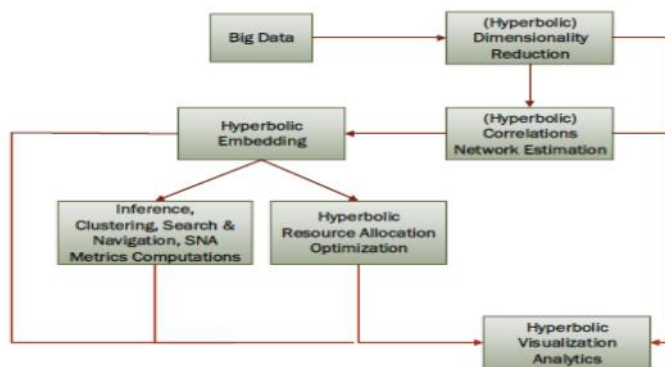
Στο Σχήμα 7 μπορούμε να διακρίνουμε τα δύο επίπεδα που δημιουργούνται στα σύνθετα δίκτυα και αλληλεπιδρούν μεταξύ τους. Το φυσικό επίπεδο (*physical layer*) είναι το επίπεδο εκείνο το οποίο αποτελείται από τη βασική δικτυακή υποδομή μετάδοσης της πληροφορίας και τις συνδέσεις σε επίπεδο υλικού μεταξύ διαφόρων συσκευών ή μηχανημάτων. Αξιοποιώντας το επίπεδο αυτό, και άρα πάνω από το επίπεδο αυτό, οι χρήστες ενός δικτύου επικοινωνιών ή ενός κοινωνικού δικτύου, δημιουργούν συνδέσεις μεταξύ τους (π.χ., επαφές skype, κατάλογος

e-mail κτλ.) οδηγώντας έτσι στην ανάπτυξη ενός ξεχωριστού επιπέδου το οποίο ονομάζουμε *κοινωνικό επίπεδο (social layer)*.

Το διασυνδεδεμένο φυσικό επίπεδο αναπτύσσεται τάξεις μεγέθους γρηγορότερα από το ρυθμό ανάπτυξης του υπερκείμενου κοινωνικού επιπέδου. Αυτό οδηγεί στην τεράστια παραγωγή δεδομένων και από τα δύο επίπεδα, για διαφορετικούς σκοπούς (π.χ., τη μεταφορά δεδομένων στο κατώτερο επίπεδο, δεδομένα ελέγχου και συνομιλίας στο υψηλότερο κτλ.). Ο χειρισμός όμως αυτών των δεδομένων, όπως είναι λογικό, καθίσταται δύσκολος και υπερβολικά χρονοβόρος (έως και αδύνατος για πραγματικό χρόνο). Για το λόγο αυτό έχουν αναπτυχθεί τεχνικές με χρήση του μετρικού χώρου της Υπερβολικής Γεωμετρίας (βασικά χαρακτηριστικά της οποίας θα αναλυθούν στην επόμενη ενότητα) για την αναπαράσταση και ανάλυση τέτοιων δεδομένων. Όπως έχει δειχθεί στη βιβλιογραφία (βλ. [10]) δίκτυα αυθαίρετα μεγάλου μεγέθους μπορούν να ενσωματωθούν σε Υπερβολικούς Χώρους μικρών διαστάσεων (ακόμη και δύο διαστάσεις αρκούν), χωρίς να θυσιάζεται σημαντική πληροφορία όσον αφορά τα δίκτυα επικοινωνίας (π.χ., δρομολόγηση) και τη δομή τους (π.χ., ιδιότητες ελεύθερης-κλίμακας).

Στο [10] παρουσιάζεται μία ροή διαδικασιών για το χειρισμό Δεδομένων Μεγάλης Κλίμακας. Ξεκινάει με την απόκτηση δεδομένων και την επιλογή ενός κατάλληλου μοντέλου αναπαράστασής τους. Τα δεδομένα εισόδου έχουν αρχικά μία ακατέργαστη μορφή, όπως για παράδειγμα μία λίστα γειτνίασης ή έναν πίνακα γειτνίασης που θα απεικονίζει τις συνδέσεις των χρηστών ενός κοινωνικού δικτύου. Ακολουθεί η προ-επεξεργασία των δεδομένων με τη μείωση των διαστάσεων τους, είτε με παραδοσιακές τεχνικές ή κάνοντας χρήση των ιδιοτήτων υπερβολικής γεωμετρίας. Στη συνέχεια, η αναπαράσταση των δεδομένων θα οδηγήσει στην κατάλληλη μεθοδολογία για την ενσωμάτωσή τους στο χώρο της υπερβολικής γεωμετρίας. Η ενσωμάτωση δεδομένων αφορά την ανάθεση συντεταγμένων για τους κόμβους ενός δικτύου στον υπερβολικό μετρικό χώρο. Το προτεινόμενο πλαίσιο ενεργειών θα εκμεταλλευθεί την ευελιξία της υπερβολικής γεωμετρίας που βασίζεται σε μηχανισμούς για την οπτικοποίηση δεδομένων, έτσι ώστε να επιτρέψει την ολιστική και ταυτόχρονα εστιασμένη ανάλυση των δεδομένων. Η χρήση της υπερβολικής γεωμετρίας είναι ικανή να παρέχει εργαλεία οπτικοποίησης που συλλαμβάνουν ταυτόχρονα καθολικά μοτίβα (π.χ., ιεραρχία, κεντρικότητα κόμβων) και τοπικά χαρακτηριστικά των δικτύων.

Στην ενότητα 4.3 θα αναλυθούν τεχνικές για την ενσωμάτωση των κόμβων ενός δικτύου σε συντεταγμένες της υπερβολικής γεωμετρίας. Εκμεταλλευόμενοι τη συνάρτηση υπολογισμού της απόστασης στην υπερβολική γεωμετρία και τεχνικές άπληστης δρομολόγησης, αποτελεσματικοί υπολογισμοί μετρικών της Ανάλυσης Σύνθετων Δικτύων (όπως είναι η κεντρικότητα των κόμβων, ο υπολογισμός της οποίας γίνεται δύσκολος για μεγάλα σύνολα δεδομένων) θα μελετηθούν και θα προταθούν στη συνέχεια.



Σχήμα 8: Η προτεινόμενη ροή ενεργειών στο χώρο της υπερβολικής γεωμετρίας για ανάλυση δεδομένων μεγάλης κλίμακας σε σύνθετα και κοινωνικά δίκτυα [10]

4.2 Θεμελιώδεις Αρχές του Υπερβολικού Γεωμετρικού

Χώρου

Οι μη ευκλείδειες γεωμετρίες, όπως η υπερβολική γεωμετρία [10][11], γεννήθηκαν από τον προβληματισμό γύρω από το πέμπτο αξίωμα της Ευκλείδειας Γεωμετρίας και την τροποποίηση αυτού. Σύμφωνα με το αξίωμα αυτό, δοθείσης μίας ευθείας και ενός σημείου το οποίο δεν βρίσκεται πάνω σε αυτήν, υπάρχει ακριβώς μία ευθεία γραμμή που διέρχεται από το δοθέν σημείο η οποία να είναι παράλληλη στη δοθείσα ευθεία.

Ο n -διάστατος υπερβολικός χώρος, που δηλώνεται ως $\mathbb{H}^n$, είναι ένας n -διαστάσεων τοπολογικός χώρος του Riemann με αρνητική κυρτότητα c την οποία συχνά θεωρούμε ως σταθερή και ίση με $c = -1$ (όπως θα κάνουμε και στη συνέχεια). Υπάρχουν αρκετά μοντέλα που περιγράφουν τον Υπερβολικό Χώρο, όπως το μοντέλο του δίσκου του Poincaré (Poincaré disk model), το μοντέλο ημιεπιπέδου του Poincaré (Poincaré half-plane model), το Υπερβολοειδές μοντέλο (Hyperboloid model), το μοντέλο του Klein (Klein model), κτλ. Όλα αυτά τα μοντέλα έχει αποδειχθεί ότι είναι ισομετρικά (δηλαδή υπάρχει μία απεικόνιση η οποία διατηρεί τις αποστάσεις μεταξύ των μετρικών χώρων), οπότε κάθε δύο από αυτά μπορούν να συσχετιστούν με έναν μετασχηματισμό που θα διατηρεί όλες τις γεωμετρικές ιδιότητες (π.χ., απόσταση) του χώρου. Στη συνέχεια θα αναλυθούν περισσότερο το μοντέλο του δίσκου του Poincaré και το Υπερβολοειδές Μοντέλο, τα οποία ήταν και τα δύο μοντέλα που χρησιμοποιήθηκαν στην παρούσα εργασία, για την ενσωμάτωση των κόμβων ενός δικτύου στο χώρο συντεταγμένων της υπερβολικής γεωμετρίας.

Το Υπεβολοειδές μοντέλο, θεωρεί τον $\mathbb{H}^n$ υπερβολικό χώρο σαν ένα υπερβολοειδές στον $\mathbb{R}^{n+1} = \{(x_0, x_1, \dots, x_n) | x_i \in \mathbb{R}, \mathbb{I} = \{0, 1, \dots, n\}\}$ με $x_0, \dots, x_n$ τέτοια ώστε $x_0^2 - x_1^2 - \dots - x_n^2 = 1, x_0 > 0$. Οι Υπερβολικοί χώροι έχουν μία μετρική συνάρτηση απόστασης η οποία διαφέρει από τη γνωστή Ευκλείδεια απόσταση, ενώ διαφέρει επίσης ανάμεσα στα διάφορα μοντέλα υπερβολικών χώρων. Στην περίπτωση του Υπερβολοειδούς μοντέλου, για δύο σημεία $x = (x_0, \dots, x_n)$ $y = (y_0, \dots, y_n)$ η υπερβολική απόσταση δίνεται από τον τύπο:

$$\cosh d_H(x, y) = \sqrt{(1 + \|x\|^2)(1 + \|y\|^2)} - \langle x, y \rangle \quad 4.1$$

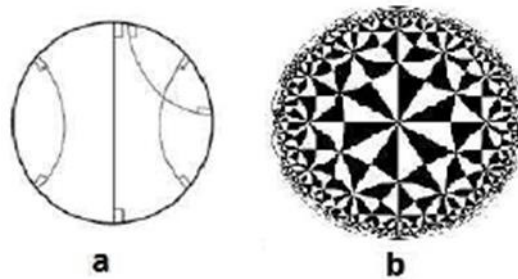
όπου $\|\cdot\|$ είναι η Ευκλείδεια νόρμα και το σύμβολο $\langle \cdot, \cdot \rangle$ αναπαριστά το εσωτερικό γινόμενο.

Το υπερβολοειδές μοντέλο μπορεί να χρησιμοποιηθεί για να κατασκευαστεί το μοντέλο του δίσκου του Poincaré, όπου το τελευταίο είναι μια προοπτική προβολή του προηγούμενου από την οπτική ($x_0 = -1, x_1 = 0, \dots, x_n = 0$), προβάλλοντας το πάνω ημιεπίπεδο του υπερβολοειδούς σε μία μοναδιαία σφαίρα στον $\mathbb{R}^n$ με κέντρο στο $x_0 = 0$.

Συγκεκριμένα, εστιάζοντας σε δύο διαστάσεις, ολόκληρος ο άπειρος γεωμετρικός χώρος μπορεί να αναπαρασταθεί μέσα σε έναν πεπερασμένο μοναδιαίο δίσκο $\mathbb{D} = \{z \in \mathbb{C} | \|z\| < 1\}$ του ευκλείδειου χώρου. Το μοντέλο αυτό είναι το 2-διάστατο μοντέλο του δίσκου του Poincaré. Η συνάρτηση για τον υπολογισμό της υπερβολικής απόστασης $d_{PD}(z_i, z_j)$, για δύο σημεία z_i, z_j , στο μοντέλο αυτό δίνεται από τον τύπο:

$$\cosh d_{PD}(z_i, z_j) = \frac{2z_i - z_j^2}{(1 - z_i^2)(1 - z_j^2)} + 1 \quad 4.2$$

Ο Ευκλείδειος κύκλος $\partial\mathbb{D} = \{z \in \mathbb{C} | \|z\| = 1\}$ είναι το όριο στο άπειρο για το μοντέλο του δίσκου του Poincaré. Επιπροσθέτως, σε αυτό το μοντέλο, η συντομότερη υπερβολική διαδρομή μεταξύ δύο κόμβων είναι είτε μέρος της διαμέτρου του $\mathbb{D}$, ή μέρος ενός ευκλείδειου κύκλου στο $\mathbb{D}$ κάθετου στο σύνορο $\partial\mathbb{D}$, όπως φαίνεται και στο σχήμα 9(α).



Σχήμα 9: Το μοντέλο του δίσκου του Poincaré (α) μαζί με τα συντομότερα μονοπάτια σε δύο διαστάσεις. Το (b) δείχνει τα διαγράμματα Voronoi του δίσκου του Poincaré για υπερβολικά τρίγωνα με ίσο εμβαδόν [10].

Πρέπει να σημειωθεί ότι οι καμπύλες αυτών των συντομότερων μονοπατιών διαφέρουν από τα τόξα που θα υπονοούνταν στην περίπτωση του Ευκλείδειου μέτρου.

Ένα αξιοσημείωτο πλεονέκτημα του υπερβολικού χώρου, σε σχέση με την εφαρμογή του στην ανάλυση δεδομένων μεγάλης κλίμακας, είναι η ιδιότητα της «εκθετικής κλιμάκωσης» σε σχέση με την ακτινική συντεταγμένη. Συγκεκριμένα, η περίμετρος C και το εμβαδόν A ενός κύκλου ακτίνας r στο 2-διάστατο μοντέλο του δίσκου του Poincaré, ακολουθούν τις παρακάτω σχέσεις [10]:

$$C(r) = 2\pi \sinh(r), \quad A(r) = 4\pi \sinh^2\left(\frac{r}{2}\right) \quad 4.3$$

Επομένως, για μικρές τιμές της ακτίνας r , για παράδειγμα γύρω από το κέντρο του δίσκου του Poincaré, ο υπερβολικός χώρος μοιάζει επίπεδος, ενώ για μεγαλύτερο r , τόσο η περίμετρος όσο και το εμβαδόν μεγαλώνουν εκθετικά με το r . Αυτή η εκθετική κλιμάκωση φαίνεται στο σχήμα 9(b), το οποίο απεικονίζει μία διαμέριση του χώρου του δίσκου του Poincaré σε υπερβολικά τρίγωνα ίσου εμβαδού. Τα τρίγωνα εμφανίζονται να μικραίνουν συνεχώς, καθώς πλησιάζουμε προς την περίμετρο του κύκλου με βάση την ευκλείδεια οπτική αναπαράστασης της τριγωνοποίησης, παρ' όλο που στην πραγματικότητα όλα έχουν το ίδιο εμβαδόν.

4.3 Ενσωμάτωση Δικτύων στον Υπερβολικό Χώρο

Στην παρούσα ενότητα εξετάζονται δύο τεχνικές ενσωμάτωσης που έχουν προταθεί και χρησιμοποιήθηκαν στη διπλωματική εργασία, οι οποίες αναθέτουν στους κόμβους ενός δικτύου συντεταγμένες στον Υπερβολικό Γεωμετρικό Χώρο. Ο ορισμός της ενσωμάτωσης ενός δικτύου στον χώρο συντεταγμένων της Υπερβολικής (ή της Ευκλείδειας) γεωμετρίας είναι ο ακόλουθος:

Ορισμός 1 [13]: Δοθέντος ενός συνδεδεμένου πεπερασμένου γράφου G με ένα σύνολο κόμβων V , μία ενσωμάτωση του G στον υπερβολικό χώρο $\mathbb{H}^d$, αντίστοιχα στον ευκλείδειο χώρο $\mathbb{R}^d$, είναι μία αντιστοίχιση $C(G): V \rightarrow \mathbb{H}^d$ (αντίστοιχα $C(G): V \rightarrow \mathbb{R}^d$) η οποία αναθέτει σε κάθε κόμβο $v \in V$ μία εικονική συντεταγμένη $C(v)$.

Η πρώτη από τις δύο τεχνικές ενσωμάτωσης που εξετάζονται είναι η ενσωμάτωση Rigel [12], η οποία αναθέτει στους κόμβους συντεταγμένες τέτοιες ώστε με προσεγγιστικό τρόπο να διατηρείται η γεωδαισιακή τους απόσταση στο πραγματικό δίκτυο και στην απόσταση δύο κόμβων με βάση τις συντεταγμένες τους στον Υπερβολικό Χώρο. Η τεχνική αυτή έχει το μειονέκτημα ότι σε συνθήκες *άπληστης δρομολόγησης* (*greedy routing*) (όταν, δηλαδή, κάθε φορά ένα πακέτο προωθείται προς τον κόμβο που μειώνει την απόσταση προς τον προορισμό)

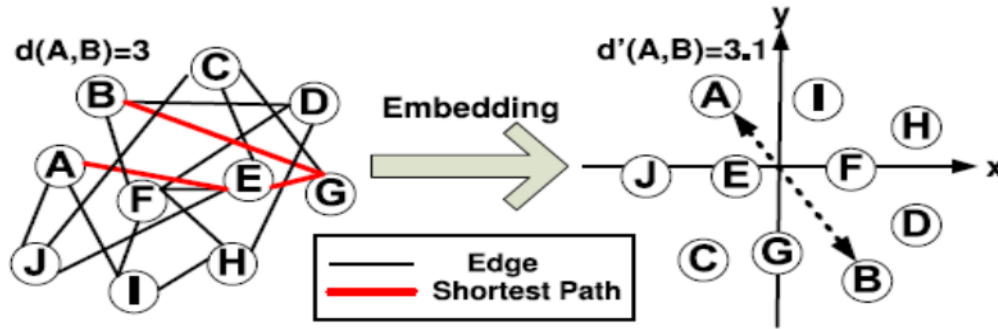
δεν εξασφαλίζει την αποφυγή κάποιου τοπικού ελαχίστου απόστασης (local minima of distance), δηλαδή μία κατάσταση στην οποία κατά τη διάρκεια της άπληστης δρομολόγησης μπορεί να βρεθούμε σε έναν κόμβο, ο οποίος δε θα έχει κάποιο γείτονα που να μειώνει την απόσταση προς τον προορισμό ακόμη κι αν υπάρχει μονοπάτι στο πραγματικό δίκτυο που διασυνδέει την πηγή με τον προορισμό. Βεβαίως, από τον ορισμό της ενσωμάτωσης δικτύων τηρείται η θεμελιώδης αρχή της Τριγωνικής Ανισότητας [14] (δηλαδή για τρία σημεία-κόμβους u, v, w με βάση και την υπερβολική τους απόσταση θα τηρείται ότι: $d(u, v) + d(v, w) \geq d(u, w)$), ενώ στη σχετική βιβλιογραφία έχειδειχθεί ότι σε δίκτυα ενσωματωμένα στον Υπερβολικό Χώρο μπορούμε να πετύχουμε ένα πολύ υψηλό ποσοστό επιτυχίας στο να βρούμε τον προορισμό, όταν η δρομολόγηση είναι άπληστη [15] [16]. Η δεύτερη τεχνική, η οποία θα αναφέρεται ως Άπληστη Ενσωμάτωση (Greedy Embedding) [13], μπορεί να μη διατηρεί (έστω σε έναν προσεγγιστικό βαθμό) τις πραγματικές αποστάσεις των κόμβων και στον Υπερβολικό Χώρο, εξασφαλίζει όμως την απουσία τέτοιων τοπικών ελαχίστων υπό την άπληστη δρομολόγηση, όπως αναφέρθηκε προηγουμένως, επιτυγχάνοντας 100% επιτυχία στην εύρεση μονοπατιού προς τον προορισμό.

4.3.1 Ενσωμάτωση Rigel

Η Ενσωμάτωση Rigel [12] στοχεύει στην ενσωμάτωση των κόμβων ενός δικτύου στον Υπερβολικό Χώρο με τέτοιο τρόπο ώστε η απόσταση μεταξύ των συντεταγμένων δύο κόμβων να είναι προσεγγιστικά κοντά στην αντίστοιχη απόσταση του γεωδαισιακού μονοπατιού που συνδέει τους δύο κόμβους στο πραγματικό δίκτυο. Στα σύγχρονα σύνθετα και κοινωνικά δίκτυα όπου ο αριθμός των κόμβων τους μπορεί να φτάνει τις εκατοντάδες χιλιάδες, ως και μερικά εκατομμύρια, μόνο ο υπολογισμός της απόστασης δύο κόμβων μπορεί να πάρει αρκετό χρόνο, από μερικά δευτερόλεπτα ως και μερικά λεπτά για υπερβολικά μεγάλα δίκτυα (όπως το Facebook με 700 εκατομμύρια κόμβους, το LinkedIn με 100 εκατομμύρια κόμβους κ.ο.κ.) χρησιμοποιώντας παραδοσιακούς αλγορίθμους όπως ο BFS [12]. Επομένως, προκύπτει η ανάγκη για τη χρήση προσεγγιστικών τεχνικών που να μπορούν να υπολογίσουν την απόσταση αυτή σε ελάχιστο χρόνο (κάποιων μικροδευτερολέπτων) με ικανοποιητική ακρίβεια. Άλλωστε, ο υπολογισμός της απόστασης δύο κόμβων έχει πρακτική εφαρμογή σε διάφορες περιπτώσεις. Σε μία αναζήτηση στον παγκόσμιο ιστό, η απόσταση ενός αρχικού σημείου από το οποίο γίνεται ένα ερώτημα από τη σχετική ιστοσελίδα μπορεί να είναι ένας σημαντικός παράγοντας της κατάταξης των αποτελεσμάτων. Στα κοινωνικά δίκτυα ένας χρήστης μπορεί να ενδιαφέρεται να προσθέσει άλλους χρήστες, ή να βρει περιεχόμενο από άλλους χρήστες οι οποίοι είναι κοντινοί του στον κοινωνικό γράφο.

Στη σχετική ενσωμάτωση, λοιπόν, προτείνεται η χρήση ενός Συστήματος Συντεταγμένων Γραφήματος (ΣΣΓ), δηλαδή ένα σύστημα στο οποίο σε κάθε κόμβο του

δικτύου, αναθέτονται με τη σειρά συντεταγμένες, οι οποίες υπολογίζονται ώστε να διατηρούν προσεγγιστικά τις πραγματικές αποστάσεις μεταξύ των κόμβων του δικτύου. Μία προσέγγιση η οποία αν και συνοδεύεται με το μειονέκτημα ενός αρχικού υπολογιστικού κόστους για την ενσωμάτωση των κόμβων στον υπερβολικό χώρο (το οποίο κλιμακώνεται γραμμικά σε σχέση με το μέγεθος του γραφήματος $O(n)$), στη συνέχεια μπορεί με σταθερό χρόνο $O(1)$ -ανεξάρτητο του μεγέθους του γράφου- να παρέχει μία καλή προσέγγιση της γεωδαισιακής απόστασης μεταξύ δύο κόμβων.



Σχήμα 10: Ένα παράδειγμα ενσωμάτωσης ενός γράφου στον Ευκλείδειο Χώρο. Για παράδειγμα, το συντομότερο μονοπάτι μεταξύ των κόμβων A και B είναι 3 στο γράφημα (αριστερά) και η Ευκλείδεια απόσταση των συντεταγμένων τους είναι 3.1 (δεξιά) [12]

Όπως έχει αναφερθεί, για την ενσωμάτωση των δικτύων επιλέχθηκε η χρήση συντεταγμένων στον υπερβολικό χώρο, καθώς πέρα από τις βιβλιογραφικές αναφορές που αναφέρθηκαν προηγουμένως, και η εμπειρική μελέτη και σύγκριση κατέδειξε πως η χρήση συντεταγμένων στον υπερβολικό χώρο έχει σαν αποτέλεσμα μικρότερα σφάλματα παραμόρφωσης σε σχέση με το να γινόταν η ενσωμάτωση στον Ευκλείδειο ή Σφαιρικό Χώρο.

Για την ενσωμάτωση των κόμβων σε ένα ΣΣΓ και τον υπολογισμό της απόστασης μεταξύ τους χρησιμοποιήθηκε το μοντέλο του Υπερβολοειδούς (που αναλύθηκε στην ενότητα 4.2) για δύο πρακτικούς λόγους: Πρώτον, ο υπολογισμός της απόστασης μεταξύ δύο σημείων σε αυτό το μοντέλο είναι υπολογιστικά πιο απλός σε σχέση με τα υπόλοιπα μοντέλα. Δεύτερον, η πολυπλοκότητα του υπολογισμού της απόστασης δύο σημείων είναι ανεξάρτητη από την κυρτότητα του χώρου. Η απόσταση δύο σημείων x και y στο συγκεκριμένο σύστημα, για κυρτότητα c σε n -διαστάσεις είναι:

$$\delta(x, y) = \text{arcosh} \left(\sqrt{\left(1 + \sum_{i=1}^n x_i^2\right) \left(1 + \sum_{i=1}^n y_i^2\right)} - \sum_{i=1}^n x_i y_i \right) \cdot |c| \quad 4.4$$

- **Διαδικασία Ενσωμάτωσης Δικτύου στον Υπερβολικό Χώρο Συντεταγμένων**

Η ενσωμάτωση αρχικά ξεκινάει με την επιλογή l κόμβων του δικτύου ως σημείων αναφοράς (landmarks), όπου $l \ll n$ και n ο συνολικός αριθμός κόμβων του δικτύου. Για κάθε έναν από αυτούς τους l κόμβους υπολογίζονται οι αποστάσεις τους από όλους τους υπόλοιπους κόμβους του δικτύου με χρήση του αλγόριθμου BFS. Συγκεκριμένα, απαιτούνται σε αυτό το στάδιο l εκτελέσεις του αλγόριθμου BFS. Στη συνέχεια χρησιμοποιείται ένας γενικός αλγόριθμος βελτιστοποίησης για τον καθορισμό των συντεταγμένων αυτών των l κόμβων, έτσι ώστε οι αποστάσεις μεταξύ τους στον χώρο συντεταγμένων να είναι όσο πιο κοντά γίνεται στις αντίστοιχες αποστάσεις σε βήματα των μονοπατιών που τους συνδέουν στο γράφο. Αναφερόμαστε σε αυτό το στάδιο ως “στήριγμα” (bootstrapping). Από τη στιγμή που έχουν καθοριστεί οι συντεταγμένες των κόμβων-σημείων αναφοράς, υπολογίζουμε τις θέσεις όλων των υπόλοιπων κόμβων, έτσι ώστε η απόσταση κάθε κόμβου από όλους τους κόμβους αναφοράς στο χώρο συντεταγμένων να προσεγγίζει ικανοποιητικά την πραγματική απόσταση (σε βήματα) αυτών των κόμβων από τους κόμβους-σημεία αναφοράς στο γράφο.

Για να υπολογιστεί η θέση με βάση τις συντεταγμένες ενός κόμβου του γράφου που δεν είναι σημείο αναφοράς, επιλέγονται τυχαία k από τους l κόμβους αναφοράς και με βάση τη μέθοδο Simplex μπορεί να υπολογιστούν οι βέλτιστες συντεταγμένες που ελαχιστοποιούν την απόκλιση των αποστάσεων μεταξύ του κόμβου και των k κόμβων αναφοράς ανάμεσα στο χώρο συντεταγμένων και στην πραγματική απόστασή τους στο δίκτυο με βάση των αριθμό βημάτων που τους χωρίζουν.

Η λογική πίσω από την επιλογή κόμβων-σημείων αναφοράς είναι ότι ο υπολογισμός της πραγματικής απόστασης μεταξύ δύο κόμβων για κάθε ζεύγος κόμβων σε ένα δίκτυο είναι μία ιδιαίτερα δαπανηρή διαδικασία την οποία στοχεύουμε να αποφύγουμε. Επομένως, για τον καθορισμό των συντεταγμένων δύο κόμβων θα απαιτούνταν n εκτελέσεις υπολογισμών BFS, κάνοντας έτσι τη συνολική πολυπλοκότητα του σταδίου «στηρίγματος» $O(nm)$, δηλαδή μεγαλύτερο και από το στάδιο της ενσωμάτωσης των κόμβων, το οποίο, επιπρόσθετα, θα είχε πολυπλοκότητα $O(n^2)$ αφού κάθε κόμβος θα ενσωματωνόταν με βάση όλους τους υπόλοιπους κόμβους του δικτύου και όχι μόνο με τους κόμβους αναφοράς. Χρησιμοποιώντας επομένως έναν πολύ μικρό- σε σχέση με το συνολικό- αριθμό κόμβων του δικτύου, μπορούμε να καθορίσουμε με ικανοποιητικό τρόπο τις συντεταγμένες των κόμβων του δικτύου σε σχέση με αυτούς τους l κόμβους, με πολύ μικρότερο υπολογιστικό κόστος.

Η επιλογή των κόμβων-σημείων αναφοράς έγινε επιλέγοντας τους κόμβους που έχουν τον υψηλότερο βαθμό στο γράφο, καθώς πέρα από το ότι είναι μία υπολογιστικά εύκολη διαδικασία, έχειδειχθεί ότι αποδίδει ικανοποιητικά σε σχέση με εναλλακτικές προσεγγίσεις [14]. Η κυριότητα για την ενσωμάτωση των δικτύων, που έγιναν χρησιμοποιώντας την

ενσωμάτωση Rigel, κατά τη διάρκεια της διπλωματικής εργασίας ορίστηκε ίση με -1 , καθώς κατά το σχετικό εμπειρικό έλεγχο [12] απέδωσε τα καλύτερα αποτελέσματα σε σχέση με άλλες τιμές.

Πρέπει να σημειωθεί ότι η ενσωμάτωση ενός δικτύου με την Rigel, μπορεί να πραγματοποιηθεί παράλληλα μειώνοντας το χρόνο που απαιτείται (και ο οποίος αυξάνεται γραμμικά με το μέγεθος του γραφήματος). Συγκεκριμένα, οι l υπολογισμοί BFS που χρειάζονται για τον υπολογισμό των αποστάσεων μεταξύ κάθε ενός από τους κόμβους αναφοράς και όλων των υπόλοιπων κόμβων στο δίκτυο μπορεί να πραγματοποιηθεί παράλληλα. Επίσης, κάθε κόμβος u που δεν είναι κόμβος αναφοράς, μπορεί να ενσωματωθεί παράλληλα και ανεξάρτητα από τους υπόλοιπους κόμβους του δικτύου [12].

4.3.2 Άπληστη Ενσωμάτωση

Η ύπαρξη επικοινωνιακών συσκευών που κάνουν χρήση του συστήματος GPS οδήγησε στην υιοθέτηση της άπληστης γεωμετρικής δρομολόγησης, που κάνει χρήση των πραγματικών συντεταγμένων μίας συσκευής, ώστε να γίνεται η δρομολόγηση μέσω συσκευών που είναι πιο κοντινές στον προορισμό μέσω της ευκλείδειας απόστασης. Η άπληστη δρομολόγηση που βασίζεται στις τοποθεσίες των κόμβων και τις Ευκλείδειες αποστάσεις μεταξύ τους έχει δειχθεί ότι μπορεί να έχει ένα υψηλό επίπεδο επιτυχίας, αλλά αποτυγχάνει όταν ένα πακέτο φτάνει σε έναν κόμβο που είναι κοντινότερος στον προορισμό από όλους τους άμεσους γείτονες του, ακόμη κι αν ένα τέτοιο μονοπάτι που συνδέει την πηγή με τον προορισμό δεν υπάρχει. Επομένως, γεννήθηκε η ανάγκη για συστήματα που θα αναθέτουν συνταγμένες στους κόμβους που θα εξασφαλίζουν 100% επιτυχία στην άπληστη δρομολόγηση.

Σε ένα πρωτοποριακό του έργο, ο R. Kleinberg [17] παρουσίασε μία κατασκευαστική απόδειξη με βάση την οποία δείχνει ότι κάθε πεπερασμένος, συνδεδεμένος και μη-κατευθυνόμενος γράφος έχει μία άπληστη ενσωμάτωση στις 2 διαστάσεις του υπερβολικού χώρου, δηλαδή, μία ενσωμάτωση που επιτρέπει μία επιτυχημένη άπληστη δρομολόγηση για κάθε ζεύγος πηγής-προορισμού. Ο γενικός αλγόριθμος του [17] μπορεί να βρει μία άπληστη ενσωμάτωση ενός άπειρου d -κανονικού δένδρου για κάθε ακέραιο $d \geq 3$. Για να ενσωματωθεί ένας πραγματικός γράφος G , πρώτα επιλέγεται ένα συνδεδετικό δένδρο T του G για να εξυπηρετήσει ως ένας ελάχιστος μη-κυκλικός υπόγραφος ο οποίος συνδέει όλους τους κόμβους και παρέχει ένα μοναδικό απλό μονοπάτι για κάθε ζεύγος κόμβων στο γράφο. Στη συνέχεια, καθορίζεται ο μέγιστος βαθμός d του T . Τελικά, οι κόμβοι του T ταυτίζονται με τους ενσωματωμένους κόμβους του d -κανονικού δένδρου, όπως αυτό κατασκευάζεται από τον αλγόριθμο ενσωμάτωσης. Με τον τρόπο αυτό ολοκληρώνεται η ενσωμάτωση του δένδρου T .

Είναι εύκολο να διαπιστωθεί ότι αυτή η ενσωμάτωση είναι επίσης και μία άπληστη ενσωμάτωση του γράφου G .

Η άπληστη ενσωμάτωση που προτείνεται από τους Cvetkovski και Crovella [13] βασίζεται στη δουλειά του Kleinberg [17] και επιπλέον παρέχει τη δυνατότητα για την κατασκευή μίας άπληστης ενσωμάτωσης γράφου η οποία υποστηρίζει την προσθήκη αυθαίρετου αριθμού κόμβων στην προϋπάρχουσα ενσωμάτωση χωρίς να απαιτούνται οποιεσδήποτε αλλαγές στις συντεταγμένες των κόμβων που είχαν ανατεθεί προηγουμένως ώστε να διατηρείται η ιδιότητα της 100% εξασφαλισμένης άπληστης δρομολόγησης.

Αρχικά πρέπει να διασαφηνιστούν κάποιοι ορισμοί σχετικοί με την ενσωμάτωση κόμβων, καθώς και οι αντίστοιχες κατασκευαστικές αποδείξεις που εξασφαλίζουν ότι μία ενσωμάτωση είναι άπληστη [13].

Ορισμός 2: Για δύο σημεία $v, w \in \mathbb{H}^d$, αντίστοιχα στον $\mathbb{R}^d$, η υπερβολική (ευκλείδεια) διχοτόμος του υπερβολικού (ευκλείδειου) τμήματος γραμμής που ορίζεται από τους v και w είναι η τοποθεσία των σημείων στον $\mathbb{H}^d$ ($\mathbb{R}^d$) τα οποία ισαπέχουν από τα v και w με βάση την υπερβολική (Ευκλείδεια) απόσταση τους.

Στον $\mathbb{R}^d$, η διχοτόμος είναι το Ευκλείδειο υπερεπίπεδο κάθετο στο τμήμα $[v, w]$ στη μέση του. Στον $\mathbb{H}^d$ η διχοτόμος είναι το υπερβολικό υπερεπίπεδο κάθετο στη μέση του τμήματος της υπερβολικής γραμμής που συνδέει τους v και w .

Λήμμα 1: Έστω ότι X είναι είτε το $\mathbb{R}^d$ ή το $\mathbb{H}^d$ και ρ είναι η αντίστοιχη συνάρτηση απόστασης. Έστω ότι τα v και w είναι διακριτά σημεία στο X και έστω ότι b είναι η διχοτόμος του τμήματος που συνδέει τους v και w . Τότε για κάθε $u \in X$ ισχύει ότι $\rho(v, u) < \rho(w, u)$ εάν και μόνο εάν τα σημεία v και u είναι στο ίδιο ημιεπίπεδο που ορίζεται από τη διχοτόμο b .

Απόδειξη [13]. Προκύπτει άμεσα από την τριγωνική ανισότητα όταν αυτή εφαρμόζεται στο τρίγωνο που ορίζεται από τα v, u και x , όπου x είναι η τομή της διχοτόμου b και του τμήματος που συνδέει τα u και w . Συγκεκριμένα, $\rho(v, u) < \rho(v, x) + \rho(u, x)$. Από τον ορισμό της διχοτόμου, $\rho(v, x) = \rho(w, x)$, και από τον ορισμό του x , $\rho(u, x) + \rho(x, w) = \rho(u, w)$. Συνδυάζοντας τα παραπάνω προκύπτει $\rho(v, u) < \rho(w, u)$.

Παρομοίως, για όλα τα $u \in X$ ισχύει ότι $\rho(w, u) < \rho(v, u)$, αν και μόνο αν, τα w και u είναι όλα στο ίδιο ημιεπίπεδο που ορίζεται από τη διχοτόμο b . □

Ορισμός 3: Για έναν γράφο $G(V, E)$ και την ενσωμάτωση του $C(G)$ στο $\mathbb{R}^d$, αντίστοιχα στο $\mathbb{H}^d$, έστω $e \in E$ μία ακμή που συνδέει τους κόμβους u και v . Μία

ενσωματωμένη ακμή του G είναι το Ευκλείδειο, αντίστοιχα Υπερβολικό, τμήμα γραμμής $C(e) = C(u, v)$ που συνδέει τα σημεία $C(u)$ και $C(v)$ στον $\mathbb{R}^d$, αντίστοιχα στον $\mathbb{H}^d$.

Λήμμα 2 (Άπληστη Ενσωμάτωση): Για έναν γράφο G με ενσωμάτωση $C(G)$, έστω T ένα συνδετικό δένδρο του G . Για κάθε ακμή $e \in T$, έστω ότι $b(e)$ είναι η κάθετη διχοτόμος της ενσωματωμένης ακμής $C(e)$. Τότε μία επαρκής συνθήκη για την C ώστε να είναι άπληστη ενσωμάτωση είναι ότι για κάθε $e \in T$, η $b(e)$ δεν τέμνει καμία ενσωματωμένη ακμή του T άλλη από την $C(e)$.

Απόδειξη [13]. Ας θεωρήσουμε μία ακμή $(u, v) \in T$. Για να αποδείξουμε το λήμμα, αρκεί να δείξουμε ότι το u έχει μία άπληστη διαδρομή για κάθε κόμβο s για τον οποίο τα μονοπάτια στο δένδρο από τον u έχουν ως επόμενο βήμα τον v . Ας θεωρήσουμε τη διχοτόμο b των (u, v) . Από τη στιγμή που η b δεν τέμνει καμία άλλη ακμή του T , τότε ο s πρέπει να είναι στο ίδιο ημιεπίπεδο, που ορίζεται από την b , με τον v (βλ. Λήμμα 1). Επομένως, $\rho(v, s) < \rho(u, s)$, έτσι ο u έχει ένα άπληστο επόμενο βήμα στον s , συγκεκριμένα τον v . Εφαρμόζοντας αυτή την αρχή σε κάθε ακμή στο μονοπάτι από τον u στον s επιβεβαιώνεται ότι η διαδρομή στο T (και επομένως στον G) από τον u στον s έχει μία μονότονα φθίνουσα απόσταση από τον προορισμό, είναι δηλαδή μία άπληστη διαδρομή. Κατά συνέπεια, η C είναι μία άπληστη ενσωμάτωση του G . $\square$

Η άπληστη ενσωμάτωση που προτείνεται στο [13] βασίζεται στην ανάθεση εικονικών συντεταγμένων στους κόμβους πάνω στο μοντέλο του δίσκου του Poincare που αναλύθηκε στην προηγούμενη ενότητα.

Στο μοντέλο του δίσκου του Poincare, τα μονοπάτια της υπερβολικής απόστασης αναπαρίστανται από τόξα του Ευκλείδειου κύκλου στο $\mathbb{D}$ τα οποία είναι κάθετα στο σύνορο $\partial\mathbb{D}$. Δύο διακριτά σημεία στο $\partial\mathbb{D}$ επομένως, καθορίζουν μία υπερβολική γραμμή στο $\mathbb{D}$. Για μία υπερβολική γραμμή στο $\mathbb{D}$ που καθορίζεται από δύο ιδανικά σημεία (δηλαδή σημεία στο άπειρο του $\mathbb{D}$), μας ενδιαφέρουν το κέντρο και η ακτίνα του Ευκλείδειου κύκλου στη Σφαίρα του Riemann $\mathbb{C} = \mathbb{C} \cup \infty$ που περιέχει τη συγκεκριμένη γραμμή. Είναι εύκολο ναδειχθεί ότι δοθέντος δύο ιδανικών σημείων $a = e^{i\alpha}$ και $b = e^{i\beta}$, το κέντρο του Ευκλείδειου κύκλου στο $\mathbb{C}$ που περιέχει την υπερβολική γραμμή η οποία έχει ως άκρα στο άπειρο τα σημεία a και b , και η αντίστοιχη ακτίνα δίνονται αντίστοιχα από τους τύπους:

$$c = 1/m^*, \quad R^2 = \frac{1}{|m|^2} - 1 \quad (4.5)$$

όπου m είναι το μέσο της Ευκλείδειας χορδής που ενώνει τα a και b , και m^* είναι ο συζυγής του m .

Δύο υπερβολικές γραμμές που δεν τέμνονται στον $\mathbb{D}$ λέγονται *παράλληλες*. Ειδικότερα, οι παράλληλες υπερβολικές γραμμές στο $\mathbb{D}$ που περιέχονται σε Ευκλείδειους κύκλους στο $\mathbb{C}$, οι οποίοι δεν τέμνονται μεταξύ τους, ορίζονται ως *υπερπαράλληλες*, για να μπορούν να διακριθούν από της παράλληλες υπερβολικές γραμμές που έχουν κοινό κάποια άκρο στο άπειρο.

- **Αλγόριθμος Δυναμικής Άπληστης Ενσωμάτωσης**

Όπως αναφέρθηκε προηγουμένως μία άπληστη ενσωμάτωση ενός δένδρου γράφου T το οποίο εκτείνεται πάνω σε έναν δοθέντα συνδεδεμένο γράφο G είναι επίσης μία άπληστη ενσωμάτωση του γράφου G . Επομένως, σαν ένα πρώτο βήμα, το δίκτυο κατασκευάζει ένα συνδετικό δένδρο T του γράφου G . Κάθε τύπος συνδετικού δένδρου μπορεί να χρησιμοποιηθεί για την άπληστη ενσωμάτωση. Ένας τύπος δένδρου κατάλληλος για κατανεμημένη κατασκευή είναι το *δένδρο ελάχιστου βάθους* (*minimal-depth tree*). Για να δημιουργήσουμε ένα δένδρο ελάχιστου βάθους, πρώτα οι κόμβοι του δικτύου ορίζουν έναν κόμβο ρίζα. Στη συνέχεια, κάθε κόμβος n επιλέγει, από τους N_n γείτονες του, ως πατέρα τον κόμβο ο οποίος έχει τη μικρότερη απόσταση σε βήματα από τον κόμβο-ρίζα. Εκτός από τον κόμβο-ρίζα r , κάθε κόμβος στο G θεωρείται επομένως ότι έχει ορίσει έναν κόμβο πατέρα για τον εαυτό του. Ο πατέρας ενός κόμβου n αναφέρεται ως p_n .

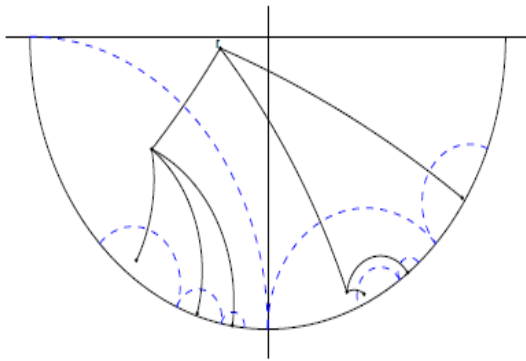
Διαδικασία Δυναμική Ενσωμάτωση

- 1) Αρχικοποίησε αναθέτοντας στον κόμβο ρίζα r του δένδρου:
 - (i) μία εικονική συντεταγμένη $C(r)$ στο υπερβολικό επίπεδο, και
 - (ii) τις γωνίες $\alpha_r = e^{i\alpha_r}$ και $\beta_r = e^{i\beta_r}$.
- 2) Για κάθε κόμβο $n \in G$:
 - α) Για τον πατέρα του p_n : (i) Στέλνει τα $C(p_n)$, $\alpha_n = \alpha_{p_n}$ και $\beta_n = (\alpha_{p_n} + \beta_{p_n}) / 2$ στον n , και (ii) ανανεώνει $\alpha_{p_n} := \beta_n$.
 - β) Ο κόμβος n : (i) υπολογίζει τα c και R σύμφωνα με την (4.5) με $\alpha_n = e^{i\alpha_n}$ και $\beta_n = e^{i\beta_n}$ και τις συντεταγμένες του

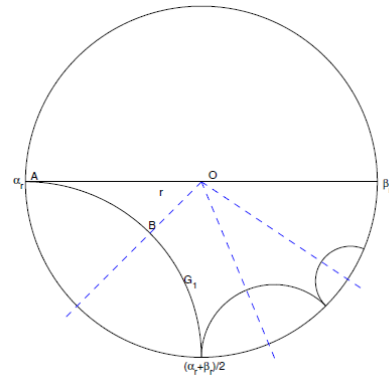
$$C(n) = \frac{R^2}{(C(p_n))^* - c^*} + c \quad (4.6)$$

και (ii) ανανεώνει $\alpha_n = (\alpha_n + \beta_n) / 2$

Στο στάδιο αρχικοποίησης του αλγορίθμου, οι τιμές των α_r και β_r ορίζουν τις πιθανές επιλογές για την τοποθεσία των συντεταγμένων του κόμβου-ρίζας $C(r)$. Οι συντεταγμένες $C(r)$ επιλέγονται από το εσωτερικό του υπερβολικού τριγώνου OAB , όπως αυτό ορίζεται από το γεωδαισιακό G_1 , τη διχοτόμο OB , και την ακτίνα OA που ορίζεται από το β_r όπως φαίνεται στο σχήμα 12. Με αυτή την επιλογή των αρχικών συνθηκών, η ανάθεση των εικονικών συντεταγμένων στους κόμβους του συνδετικού δένδρου T (και επομένως στο γράφο G) ανταποκρίνεται σε μία άπληστη ενσωμάτωση.



Σχήμα 11: Παράδειγμα μίας άπληστης ενσωμάτωσης ενός μη-κανονικού συνδετικού δένδρου στο μοντέλο του δίσκου Poincare [13]



Σχήμα 12: Κατάλληλη τοποθέτηση του κόμβου-ρίζας για μία άπληστη ενσωμάτωση [13]

Πρόταση 1: Εάν το $C(r)$ είναι ένα εσωτερικό σημείο του υπερβολικού τριγώνου OAB , όπως στο σχήμα 12, τότε η ενσωμάτωση $C(G)$ που προκύπτει από τον αλγόριθμο δυναμικής ενσωμάτωσης για έναν αυθαίρετο γράφο G με ένα συνδετικό δένδρο T είναι μία άπληστη ενσωμάτωση.

Απόδειξη [13]. Σύμφωνα με το Λήμμα 2 (άπληστης ενσωμάτωσης) που αναφέρθηκε προηγουμένως, αρκεί να δείξουμε ότι καμία διχοτόμος ακμής $e \in T$ ενσωματωμένη σε ένα υπερβολικό επίπεδο δεν τέμνει άλλη ακμή του ενσωματωμένου δένδρου. Ξεκινάμε παρατηρώντας διάφορες ιδιότητες για την παραπάνω διαδικασία δυναμικής ενσωμάτωσης.

Για έναν κόμβο $n \in T$, έστω G_n η υπερβολική γραμμή στο $\mathbb{D}$ που σχετίζεται με τον n , της οποίας τα σημεία στο άπειρο είναι τα α_n και β_n όπως αυτά ορίστηκαν στο βήμα 2α του αλγορίθμου, και ας συμβολίσουμε ως H_n την αντίστοιχη περιοχή του $\mathbb{D}$ που είναι φραγμένη από την G_n και περιέχει το σημείο $C(n)$. Οι εικονικές συντεταγμένες του κόμβου n που προέκυψαν από τη σχέση (2) είναι το είδωλο της τοποθεσίας του κόμβου-πατέρα $C(p_n)$ πάνω στην υπερβολική γραμμή G_n . Επομένως, το τμήμα της υπερβολικής γραμμής που ενώνει τα

$C(p_n)$ και $C(n)$ είναι η ενσωματωμένη ακμή $C(p_n, n)$ του T και η G_n είναι η κάθετη διχοτόμος σε αυτήν.

Απομένει να δείξουμε ότι για κάθε κόμβο $n \in T$, η G_n δεν τέμνει καμία ενσωματωμένη ακμή του T άλλη από την $C(p_n, n)$. Παρατηρούμε ότι ένα σημείο στον $\mathbb{D}$, το είδωλο του σε μία υπερβολική γραμμή και το κέντρο ενός Ευκλείδειου Κύκλου που περιέχει την υπερβολική γραμμή είναι συγγραμικά με βάση την ευκλείδεια λογική. Επομένως, για ένα σημείο P στο $\mathbb{D}$ και το είδωλο του από την υπερβολική γραμμή G_n πάντα θα βρίσκονται στο ίδιο μισό του υποχώρου H_n σε σχέση με την Ευκλείδεια διχοτόμο $b(G_n)$ του τόξου στο $\mathbb{D}$ που περιέχει την G_n . Από τη στιγμή που ένας κόμβος n και η υπερβολική γραμμή G_C που συνδέεται με έναν κόμβο-παιδί του n περιέχονται εκ κατασκευής σε αντίθετα μισά του H_n σε σχέση με την $b(G_n)$, συνεπάγεται ότι η ενσωματωμένη ακμή $C(p_n, n)$ και η G_C δεν τέμνονται. Τέλος, παρατηρούμε ότι εκ κατασκευής, για κάθε κόμβο $n \in T$, η υπερβολική γραμμή που περιέχει την ενσωματωμένη ακμή $C(p_n, n)$ είναι υπερπαράλληλη στην υπερβολική γραμμή που συνδέεται με κάθε αδερφό του n . Δηλαδή καμία ενσωματωμένη ακμή $C(p_n, n)$ δεν τέμνεται με την υπερβολική διχοτόμο οποιασδήποτε άλλης ενσωματωμένης ακμής του δένδρου T . Συνεπώς, η ενσωμάτωση $C(G)$ είναι μία άπληστη ενσωμάτωση. $\square$

Βασική επιπλέον προϋπόθεση για μία άπληστη ενσωμάτωση είναι το τρίγωνο OAB που ορίστηκε παραπάνω να έχει μη μηδενικό εμβαδό. Για να συμβεί αυτό αρκεί οι επιλεγθείσες τιμές α_r και β_r να ικανοποιούν τη σχέση $\beta_r - \alpha_r < 4\pi/3$. Στο σχήμα 12 απεικονίζεται η περίπτωση όπου $\beta_r - \alpha_r = \pi$.

5 Υπολογισμός Μετρικών Κεντρικότητας Με Χρήση Συντεταγμένων στην Υπερβολική Γεωμετρία για Δίκτυα Μεγάλης Κλίμακας

Στο κεφάλαιο 5.1 θα αναλυθεί ο αλγόριθμος του Brandes [5], ο οποίος είναι και ο αλγόριθμος με τη χαμηλότερη πολυπλοκότητα που έχει προταθεί στη βιβλιογραφία για τον ακριβή υπολογισμό της Κεντρικότητας Ενδιαμεσικότητας. Επίσης θα αναφερθούν παραλλαγές του συγκεκριμένου αλγορίθμου για τον υπολογισμό και άλλων μετρικών κεντρικότητας (όπως η Κεντρικότητα Φόρτου Κίνησης) [8]. Στη συνέχεια θα προταθεί μία νέα προσέγγιση για τον υπολογισμό της Κεντρικότητας Φόρτου Κίνησης μέσω της ενσωμάτωσης των δικτύων στον Υπερβολικό Χώρο, την οποία ονομάζουμε Υπερβολική Κεντρικότητα Φόρτου Κίνησης και η οποία παρέχει ικανοποιητικά προσεγγιστικά αποτελέσματα με χαμηλότερο υπολογιστικό κόστος σε σύγκριση με τον Υπολογισμό της Κεντρικότητας Φόρτου Κίνησης. Στο κεφάλαιο 5.2 θα παρουσιαστεί η επίδραση που έχει η αλλαγή του κόμβου-ρίζας του δένδρου που θα ενσωματωθεί στις υπερβολικές συντεταγμένες κάνοντας χρήση του αλγορίθμου της Άπληστης Ενσωμάτωσης [13] που αναλύθηκε στο προηγούμενο κεφάλαιο. Τα αποτελέσματα των παραπάνω μεθόδων, καθώς και συγκριτικά αποτελέσματα θα παρουσιαστούν στο Κεφάλαιο 6.

5.1 Υπερβολική Κεντρικότητα Φόρτου για Σύνθετα Δίκτυα Μεγάλης Κλίμακας

Ο υπολογισμός μετρικών κεντρικότητας για σύνθετα δίκτυα μεγάλης κλίμακας είναι σήμερα μία σημαντική διεργασία, καθώς αποδίδει με ποσοτικό τρόπο τη σημασία ενός κόμβου στην κίνηση της πληροφορίας σε ένα δίκτυο. Ο αλγόριθμος αναφοράς σήμερα για τον υπολογισμό της Κεντρικότητας Ενδιαμεσικότητας είναι ο αλγόριθμος του Brandes [5], μέσα από τον οποίο προκύπτουν με μικρές τροποποιήσεις και διάφορες παραλλαγές του για τον υπολογισμό και άλλων μετρικών κεντρικότητας [8], με κοινό γνώρισμα ότι υιοθετούν ένα σχήμα δρομολόγησης με βάση τα συντομότερα μονοπάτια σε ένα δίκτυο. Ο υπολογισμός,

προσεγγιστικά, της κεντρικότητας των κόμβων σε ένα δίκτυο είναι μία ιδιαίτερα σημαντική διαδικασία, λόγω της συνεχούς αύξησης των δεδομένων και των δικτύων. Για το λόγο αυτό προτείνουμε την Υπερβολική Κεντρικότητα Φόρτου Κίνησης, η οποία εκμεταλλευόμενη την ενσωμάτωση των κόμβων στις υπερβολικές συντεταγμένες παρέχει μία αρκετά αποτελεσματική εκτίμηση για την κεντρικότητα των κόμβων σε μικρότερο υπολογιστικό χρόνο σε σχέση με την παραδοσιακή τεχνική για τον υπολογισμό της Κεντρικότητας Φόρτου Κίνησης.

5.1.1 Ο Αλγόριθμος του Brandes για την Κεντρικότητα Ενδιαμεσικότητας

Ο αλγόριθμος του Brandes [5] για τον υπολογισμό της Κεντρικότητας Ενδιαμεσικότητας παραμένει ως και σήμερα ο αλγόριθμος-αναφοράς (state-of-the-art) για τη συγκεκριμένη εργασία, καθώς είναι ο πιο αποτελεσματικός σε υπολογιστική πολυπλοκότητα με ασυμπτωτικό όριο $O(nm + n^2 \log n)$ για δίκτυα με βάρη και $O(nm)$ για δίκτυα χωρίς βάρη.

Αρχικά υπενθυμίζουμε ότι με $d_G(s, t)$ συμβολίζουμε το μήκος του συντομότερου μονοπατιού από τον κόμβο s στον κόμβο t για ένα γράφο G . Ενώ με σ_{st} συμβολίζουμε τον αριθμό των συντομότερων μονοπατιών από τον $s \in V$ στον $t \in V$, όπου $\sigma_{ss} = 1$, από σύμβαση. Σημαντική είναι και η ακόλουθη παρατήρηση:

Λήμμα 1 (Το κριτήριο του Bellman): Ένας κόμβος $v \in V$ βρίσκεται πάνω σε ένα συντομότερο μονοπάτι μεταξύ των κόμβων $s, t \in V$, αν και μόνο αν, $d_G(s, t) = d_G(s, v) + d_G(v, t)$.

Δοθέντων των αποστάσεων μεταξύ ζευγών κόμβων και του αριθμού των συντομότερων μονοπατιών μεταξύ αυτών, η εξάρτηση-ζεύγους (pair-dependency) $\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$ ενός ζεύγους κόμβων $s, t \in V$ σε έναν ενδιάμεσο κόμβο $v \in V$, δηλαδή ο λόγος των συντομότερων μονοπατιών μεταξύ των s και t στα οποία ο v είναι ενδιάμεσος, προς το συνολικό αριθμό μονοπατιών μεταξύ των s και t , δίνεται από τον τύπο:

$$\sigma_{st}(v) = \begin{cases} 0, & \text{εάν } d_G(s, t) < d_G(s, v) + d_G(v, t) \\ \sigma_{sv} \cdot \sigma_{vt}, & \text{διαφορετικά} \end{cases} \quad 5.1$$

Για να υπολογίσουμε την τιμή Κεντρικότητας Ενδιαμεσικότητας ενός κόμβου v , πρέπει να αθροίσουμε όλες τις εξαρτήσεις ζεύγους για όλα τα ζεύγη ως προς αυτόν τον κόμβο,

$$C_B(v) = \sum_{s \neq v \neq t \in V} \delta_{st}(v) \quad 5.2$$

Επομένως, ο υπολογισμός της Κεντρικότητας Ενδιαμεσικότητας παραδοσιακά αποτελείται από δύο βήματα:

1. Τον υπολογισμό του μήκους και του αριθμού των συντομότερων μονοπατιών ανάμεσα σε όλα τα ζεύγη κόμβων
2. Την άθροιση όλων των εξαρτήσεων-ζευγών

- **Υπολογισμός Αριθμού Συντομότερων Μονοπατιών**

Για να εκμεταλλευθεί την αραιή φύση των δεδομένων που αναπαριστώνται από έναν γράφο ο Αλγόριθμος του Brandes υπολογίζει το πλήθος των συντομότερων μονοπατιών χρησιμοποιώντας αλγόριθμους διάσχισης. Τόσο ο BFS για γράφους χωρίς βάρη, όσο και ο Dijkstra για γράφους με βάρη ξεκινούν από μία συγκεκριμένη πηγή $s \in V$ και σε κάθε βήμα προσθέτουν έναν από τους πλησιέστερους κόμβους στο σύνολο των ήδη επισκεφθέντων κόμβων με στόχο να βρουν συντομότερα μονοπάτια από την πηγή προς όλους τους άλλους κόμβους. Με αυτή τη διαδικασία υπολογίζονται όλα τα συντομότερα μονοπάτια που έχουν ως πηγή τον κόμβο s . Ορίζουμε το σύνολο των *προκάτοχων* (*predecessors*) ενός κόμβου v στα συντομότερα μονοπάτια από τον s ως εξής:

$$P_s(v) = \{u \in V : \{u, v\} \in E, d_G(s, v) = d_G(s, u) + \omega(u, v)\}$$

Όπου $\omega(u, v)$ είναι το βάρος της ακμής από τον κόμβο u στον κόμβο v . Σε περίπτωση που ο γράφος είναι χωρίς βάρη τότε $\omega(u, v) = 1$.

Λήμμα 2 (Συνδυαστική απαρίθμηση των συντομότερων-μονοπατιών): Για $s \neq v \in V$

$$\sigma_{sv} = \sum_{u \in P_s(v)} \sigma_{su} . \quad 5.3$$

Απόδειξη [5]. Από τη στιγμή που όλα τα βάρη ακμών είναι θετικά, η τελευταία ακμή ενός συντομότερου μονοπατιού από τον s στον v είναι μία ακμή $\{u, v\} \in E$, έτσι ώστε $d_G(s, u) < d_G(s, v)$. Εμφανώς, λοιπόν, ο αριθμός των συντομότερων μονοπατιών από τον s στον v που τελειώνουν με αυτή την ακμή ισούται με τον αριθμό των συντομότερων μονοπατιών από τον s στον u . Η ισότητα πλέον προκύπτει από το Λήμμα 1 (Κριτήριο του Belmann). $\square$

Τόσο ο αλγόριθμος του Dijkstra όσο και ο BFS μπορούν εύκολα να αξιοποιηθούν για να υπολογιστεί ο αριθμός των συντομότερων μονοπατιών σύμφωνα με το παραπάνω λήμμα.

Πρόταση 3 Δεδομένης μίας πηγής $s \in V$, τόσο το μήκος όσο και ο αριθμός όλων των συντομότερων μονοπατιών σε όλους τους άλλους γράφους μπορεί να υπολογιστεί με πολυπλοκότητα $O(m + n \log n)$ για γράφους με βάρη, και σε χρόνο $O(m)$ για γράφους χωρίς βάρη.

Κατά συνέπεια, ο αριθμός των συντομότερων μονοπατιών $\sigma_{st}, s, t \in V$, μπορεί να υπολογιστεί σε χρόνο $O(nm)$ για έναν γράφο χωρίς βάρη και σε χρόνο $O(nm + n^2 \log n)$ για γράφο με βάρη.

- **Άθροιση των εξαρτήσεων ζεύγων**

Για να ελαττωθεί η ανάγκη για μία εξαντλητική άθροιση όλων των ζεύγων-εξαρτήσεων (το οποίο θα είχε υπολογιστικό κόστος της τάξης του $O(n^3)$) ο αλγόριθμος του Brandes, αρχικά ορίζει την έννοια της εξάρτησης ενός κόμβου $s \in V$ σε κάθε απλό κόμβο $v \in V$, ως εξής:

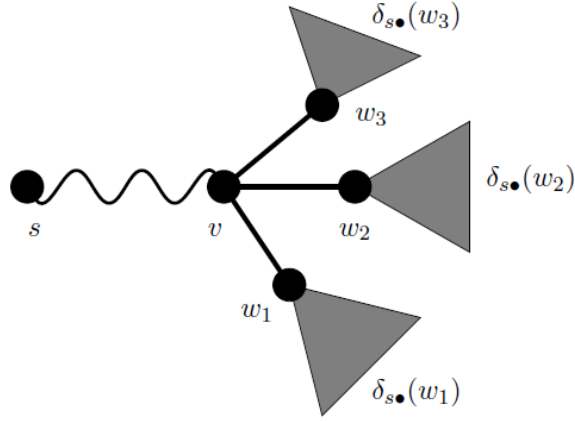
$$\delta_{s\cdot}(v) = \sum_{t \in V} \delta_{st}(v). \quad 5.4$$

Η κρίσιμη παρατήρηση που οδήγησε των Brandes στον αλγόριθμο αυτό είναι ότι τα μερικά αθροίσματα, όπως ορίζονται παραπάνω, υπακούουν σε μία αναδρομική σχέση.

Η σχέση αυτή είναι πιο εύκολο να αναγνωριστεί στην ακόλουθη ειδική περίπτωση:

Λήμμα 4: Εάν υπάρχει ακριβώς ένα συντομότερο μονοπάτι από τον $s \in V$ σε κάθε $t \in V$, η εξάρτηση από τον s σε κάθε $v \in V$ υπακούει στην ακόλουθη σχέση:

$$\delta_{s\cdot}(v) = \sum_{w: v \in P_s(w)} (1 + \delta_{s\cdot}(w)) \quad 5.5$$



Σχήμα 13: Με την υπόθεση του λήμματος 4, ένας κόμβος βρίσκεται ως ενδιάμεσος πάνω στα συντομότερα μονοπάτια προς τους διαδόχους του στο δένδρο των συντομότερων μονοπατιών από την πηγή [5]

Απόδειξη [5]: Η παραπάνω σχέση βασίζεται στην υπόθεση ότι οι κόμβοι και οι ακμές όλων των συντομότερων μονοπατιών που έχουν ως πηγή τον κόμβο s σχηματίζουν ένα δένδρο. Επομένως, ένας κόμβος v βρίσκεται ως ενδιάμεσος είτε σε όλα τα συντομότερα μονοπάτια μεταξύ των κόμβων s και κάποιων $t \in V$, είτε σε κανένα. Επομένως, η τιμή $\delta_{st}(v)$ ισούται είτε με 1 είτε με 0. Επιπλέον, ο v βρίσκεται «πάνω» σε όλα τα συντομότερα μονοπάτια προς τους κόμβους εκείνους για τους οποίους είναι ένας προκάτοχος και σε όλα τα συντομότερα μονοπάτια στα οποία αυτοί οι κόμβοι συμμετέχουν ως ενδιάμεσοι. $\square$

Θεώρημα 5: Η εξάρτηση των $s \in V$ σε κάθε $v \in V$ ακολουθεί την παρακάτω γενική αναδρομική σχέση:

$$\delta_s \cdot (v) = \sum_{w: v \in P_s(w)} \frac{\sigma_{sv}}{\sigma_{sw}} (1 + \delta_s(w)). \quad 5.6$$

Απόδειξη [5]. Υπενθυμίζεται ότι $\delta_{st}(v) > 0$ μόνο για εκείνους τους κόμβους $t \in V \setminus \{s\}$ για τους οποίους ο v βρίσκεται ως ενδιάμεσος σε τουλάχιστον ένα συντομότερο μονοπάτι από τον s στον t , ενώ παρατηρούμε ότι σε κάθε τέτοιο μονοπάτι υπάρχει ακριβώς μία ακμή $\{u, w\}$ με $v \in P_s(w)$.

Η έννοια της εξάρτησης-ζεύγους επεκτείνεται για να συμπεριλάβει μία ακμή $e \in E$ ορίζοντας ως $\delta_{st}(v, e) = \frac{\sigma_{st}(v, e)}{\sigma_{st}}$, όπου $\sigma_{st}(v, e)$ είναι ο αριθμός των συντομότερων μονοπατιών από τον s στον t ο οποίος περιέχει τόσο τον v και τον e . Τότε,

$$\delta_s(v) = \sum_{t \in V} \sum_{w: v \in P_s(w)} \delta_{st}(v, \{v, w\}) = \sum_{w: v \in P_s(w)} \sum_{t \in V} \delta_{st}(v, \{v, w\}).$$

Έστω τώρα w ένας κόμβος με $v \in P_s(w)$. Από τα σ_{sw} συντομότερα μονοπάτια από τον s στον w , σ_{sv} από αυτά πρώτα πηγαίνουν από τον s στον v και στη συνέχεια χρησιμοποιούν την ακμή $\{v, w\}$. Κατά συνέπεια, $\frac{\sigma_{sv}}{\sigma_{sw}} \cdot \sigma_{st}(w)$ συντομότερα μονοπάτια από τον s σε κάποιο $t \neq w$ περιέχουν τον v και την $\{v, w\}$. Έπεται, επομένως, ότι η εξάρτηση-ζεύγους του s και του t στον v και την $\{v, w\}$ είναι

$$\delta_{st}(v, \{v, w\}) = \begin{cases} \frac{\sigma_{sv}}{\sigma_{sw}}, & \text{εάν } t = w \\ \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \frac{\sigma_{st}(w)}{\sigma_{st}}, & \text{εάν } t \neq w \end{cases}$$

Εισάγοντας, λοιπόν, αυτή τη σχέση στην παραπάνω προκύπτει:

$$\begin{aligned} \sum_{w: v \in P_s(w)} \sum_{t \in V} \delta_{st}(v, \{v, w\}) &= \sum_{w: v \in P_s(w)} \left(\frac{\sigma_{sv}}{\sigma_{sw}} + \sum_{t \in V \setminus \{w\}} \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \frac{\sigma_{st}(w)}{\sigma_{st}} \right) \\ &= \sum_{w: v \in P_s(w)} \frac{\sigma_{sv}}{\sigma_{sw}} \cdot (1 + \delta_s(w)). \end{aligned} \tag{5.7}$$

□

Πρόταση 6: Δοθέντος ενός κατευθυνόμενου ακυκλικού γράφου των συντομότερων μονοπατιών από τον $s \in V$ στον G , οι εξαρτήσεις του s σε όλους τους άλλους κόμβους μπορούν να υπολογιστούν σε χρόνο $O(m)$ και χώρο $O(n + m)$.

Απόδειξη [5]. Διέσχισε τους κόμβους σε μη-αυξανόμενη σειρά σε σχέση με την απόσταση τους από τον s και η συσσώρευση των εξαρτήσεων εφαρμόζοντας το Θεώρημα 5. Χρειάζεται, επίσης, η αποθήκευση μίας εξάρτησης ανά κόμβο, και μία λίστα των προκατόχων του. Υπάρχει το πολύ ένα στοιχείο ανά ακμή σε κάποια από αυτές τις λίστες. □

Επομένως, σε σχέση με τον υπολογισμό της Κεντρικότητας Ενδιαμεσικότητας και την πολυπλοκότητα του αλγορίθμου του Brandes, προκύπτει το ακόλουθο θεώρημα:

Θεώρημα 7: Η Κεντρικότητα Ενδιαμεσικότητας μπορεί να υπολογιστεί σε χρόνο $O(nm + n^2 \log n)$ και χώρο $O(n + m)$ για γράφους με βάρη. Για γράφους χωρίς-βάρη, η χρονική πολυπλοκότητα μειώνεται σε $O(nm)$.

Ο ψευδοκώδικας που προκύπτει από τον αλγόριθμο του Brandes για τον υπολογισμό της Κεντρικότητας Ενδιαμεσικότητας για γράφους χωρίς βάρη δίνεται παρακάτω:

Algorithm 1: Betweenness centrality in unweighted graphs

```

1  $C_B[v] = 0; \quad v \in V$ 
2 for  $s \in V$  do
3    $S \leftarrow \text{empty stack};$ 
4    $P[w] \leftarrow \text{empty list}, w \in V;$ 
5    $\sigma[t] \leftarrow 0, t \in V; \sigma[s] \leftarrow 1;$ 
6    $Q \leftarrow \text{empty queue};$ 
7    $\text{enqueue } s \rightarrow Q;$ 
8   while  $Q$  not empty do
9      $\text{dequeue } v \leftarrow Q;$ 
10     $\text{push } v \rightarrow S;$ 
11    for each neighbor  $w$  of  $v$  do
12      //  $w$  found for the first time?
13      if  $d[w] < 0$  then
14         $\text{enqueue } w \rightarrow Q;$ 
15         $d[w] \leftarrow d[v] + 1;$ 
16      // shortest path to  $w$  via  $v$ ?
17      if  $d[w] = d[v] + 1$  then
18         $\sigma[w] \leftarrow \sigma[w] + \sigma[v];$ 
19         $\text{append } v \rightarrow P[w];$ 
20   $\delta[v] \leftarrow 0, v \in V;$ 
21  //  $S$  returns vertices in order of non-increasing distance from  $s$ 
22  while  $S$  not empty do
23     $\text{pop } w \leftarrow S;$ 
24    for  $v \in P[w]$  do
25       $\delta[v] \leftarrow \delta[v] + \frac{\sigma[v]}{\sigma[w]} \cdot (1 + \delta[w]);$ 
26    if  $w \neq s$  then
27       $C_B[w] \leftarrow C_B[w] + \delta[w]$ 

```

- **Κεντρικότητα Φόρτου Κίνησης**

Ο αλγόριθμος του Brandes για τον αποδοτικό υπολογισμό της Κεντρικότητας Ενδιαμεσικότητας μπορεί να αποτελέσει τη βάση, ώστε με μικρές τροποποιήσεις, να χρησιμοποιηθεί για τον υπολογισμό (με την ίδια χρονική πολυπλοκότητα) και άλλων μετρικών κεντρικότητας [8] οι οποίες εμπλέκουν την έννοια των συντομότερων μονοπατιών σε αριθμό βημάτων για τον υπολογισμό της κεντρικότητας των κόμβων ενός δικτύου.

Μία από αυτές τις μετρικές είναι και η *Κεντρικότητα Φόρτου Κίνησης* (Traffic Load Centrality).

Ορισμός Κεντρικότητας Φόρτου Κίνησης [8]: Κάθε κόμβος στέλνει μία μοναδιαία ποσότητα κάποιου αγαθού (κίνηση) σε έναν άλλο κόμβο. Αρχίζοντας από την αντίστοιχη κάθε φορά πηγή, αυτή η κίνηση στέλνεται πάντα στους γειτονικούς κόμβους που βρίσκονται πλησιέστερα στον προορισμό (π.χ., σε αριθμό βημάτων), και σε περίπτωση που υπάρχουν περισσότεροι του ενός τέτοιοι κόμβοι, η κίνηση μοιράζεται ισότιμα ανάμεσα τους. Η συνολική ποσότητα της κίνησης που διέρχεται από έναν κόμβο κατά τη διάρκεια όλων αυτών των ανταλλαγών καθορίζει το φόρτο του.

Algorithm 2: Traffic Load Centrality (TLC)

```

1   $TLC(v) = 0; \forall v \in V$ 
2  for each node  $s \in V$  do
3      % PART I: Computation of the next relays of a node in its
        shortest path towards  $s$ , where  $s$  is regarded as a destination
4       $S \leftarrow \text{empty stack}; Q \leftarrow \text{empty queue};$ 
5      %  $d(v)$ ,  $\sigma(v)$  length & # of shortest paths from  $s$  to  $v$ ;
6       $P[v] \leftarrow \text{empty list}; \sigma(v) = 0; d(v) = -1; \forall v \in V$ 
7       $\sigma(s) = 1; d(s) = 0; \text{enqueue } s \rightarrow Q;$ 
8      while  $Q$  not empty do
9           $\text{dequeue } v \leftarrow Q; \text{push } v \rightarrow S;$ 
10         for each neighbor  $w$  of  $v$  do
11             if  $d(w) < 0$  then
12                  $d(w) \leftarrow d(v) + 1; \text{enqueue } w \rightarrow Q;$ 
13             if  $d(w) = d(v) + 1$  then
14                  $\sigma(w) \leftarrow \sigma(w) + \sigma(v); \text{append } v \text{ in } P[w];$ 
15     % PART II: Accumulation of load dependencies ( $\delta$ ) and of TLC
        values
16      $\delta(v) = 1; \forall v \in V$ 
17     while  $S$  not empty do
18          $\text{pop } w \leftarrow S$ 
19         for  $v \in P[w]$  do
20              $\delta(v) = \delta(v) + \delta(w) \frac{1}{|P[w]|};$ 
21          $TLC(w) = TLC(w) + \delta(w);$ 

```

5.1.2 Υπερβολική Κεντρικότητα Φόρτου Κίνησης

Στα μελλοντικά σύνθετα δίκτυα (complex networks) που θα περιέχουν διάφορες ετερογενείς συσκευές και χρήστες, όλοι οι μηχανισμοί, όπως η ανάθεση πόρων και η δρομολόγηση, απαιτείται να μπορούν να λειτουργούν σε πολύ μεγάλη κλίμακα, αποδοτικά και με χαμηλή πολυπλοκότητα. Μία από τις πιο σημαντικές διαδικασίες σε δίκτυα τέτοιας κλίμακας θα πρέπει να είναι, επομένως, και η σωστή εκτίμηση για τους πιο σημαντικούς κόμβους που συμμετέχουν στη διαδικασία της δρομολόγησης, και η χρήση τέτοιων εκτιμήσεων για τη βελτίωση της δρομολόγησης, όπως και άλλων εφαρμογών που βασίζονται σε σχήματα δρομολόγησης.

Για το σκοπό αυτό προτείνεται η *Υπερβολική Κεντρικότητα Φόρτου Κίνησης* (Hyperbolic Traffic Load Centrality) (εφεξής ΥΚΦΚ) ως μία εναλλακτική προσέγγιση της Κεντρικότητας Φόρτου Κίνησης (ΚΦΚ) που αναφέραμε προηγουμένως. Για τον υπολογισμό της συγκεκριμένης μετρικής υποθέτουμε αυτή τη φορά ένα σχήμα άπληστης δρομολόγησης που κάνει χρήση συντεταγμένων στον υπερβολικό χώρο αντί για τη δρομολόγηση με χρήση συντομότερων μονοπατιών, όπως συνηθίζεται. Για την ανάθεση συντεταγμένων στον υπερβολικό χώρο στους κόμβους του δικτύου χρησιμοποιήσαμε την Ενσωμάτωση Rigel [12], η οποία επεξηγήθηκε στο προηγούμενο κεφάλαιο. Η επιλογή του συγκεκριμένου τύπου ενσωμάτωσης έγινε, καθώς πέρα από το ότι είναι κατάλληλη για δίκτυα μεγάλης κλίμακας, αναθέτει στους κόμβους συντεταγμένες τέτοιες ώστε οι αποστάσεις τους με βάση το μήκος των συντομότερων μονοπατιών να διατηρούνται κατά προσέγγιση και στον υπερβολικό χώρο. Έτσι, ο υπολογισμός κεντρικών μετρικότητας με χρήση αυτής της ενσωμάτωσης είναι πιο πιθανό να δώσει καλύτερα προσεγγιστικά αποτελέσματα. Το κίνητρο για την υιοθέτηση της ΥΚΦΚ είναι διπλό. Από τη μία μεριά ο υπολογισμός των συντομότερων μονοπατιών για μεγάλα κοινωνικά και σύνθετα δίκτυα μπορεί να γίνει εξαιρετικά δύσκολος, ενώ η άπληστη δρομολόγηση με βάση τις υπερβολικές συντεταγμένες απαιτεί μόνο τοπική γνώση σε κάθε κόμβο, δηλαδή τις υπερβολικές συντεταγμένες του γείτονά του και του προορισμού. Σε αντίθεση, δηλαδή, με τον υπολογισμό των συντομότερων μονοπατιών που απαιτείται από τα πριν για την ΚΦΚ. Δεύτερον, πέρα από την απλότητα της άπληστης δρομολόγησης με χρήση συντεταγμένων στον υπερβολικό χώρο, όπως ήδη αναφέρθηκε, βιβλιογραφικές αναφορές [15] [16] δείχνουν ότι η άπληστη δρομολόγηση στον υπερβολικό χώρο (κυρίως για δίκτυα ελεύθερης κλίμακας) οδηγεί σε άπληστα μονοπάτια που προσεγγιστικά έχουν το ίδιο μήκος με τα συντομότερα μονοπάτια. Μάλιστα τα δίκτυα ελεύθερης κλίμακας θεωρείται συχνά ότι έχουν μία «κρυμμένη» υπερβολική δομή [16].

Για τους σκοπούς της σχεδίασης του αλγορίθμου ΥΚΦΚ, εκμεταλλευτήκαμε την αλγοριθμική προσέγγιση της *Δρομολόγησης Κεντρικότητας Ενδιαμεσικότητας* (Routing Betweenness Centrality) [18] η οποία είναι μία γενίκευση της Κεντρικότητας

Ενδιαμεσικότητας για γενικούς σκοπούς δρομολόγησης. Συγκεκριμένα, υιοθετήσαμε τον αλγόριθμο *source-oblivious* για τη Δρομολόγηση Κεντρικότητας Ενδιαμεσικότητας του [18], καθώς κάθε σχήμα άπληστης δρομολόγησης γίνεται κάτω από καθεστώς άγνοιας-πηγής. Σαν αποτέλεσμα, η ΥΚΦΚ υπολογίζεται από τον παρακάτω αλγόριθμο:

Algorithm 3: Hyperbolic Traffic Load Centrality (HTLC)

```

1  $HTLC(v) = 0; \forall v \in V$ 
2 for each node  $s \in V$  do
3   % PART I: Sort all nodes in order of decreasing hyperbolic
   distance towards the destination  $s$ ,  $v_N = s$ 
4   Obtain  $S$  as  $S = \{v_1 \preceq v_2 \preceq \dots \preceq v_N\}$ , % “ $\preceq$ ” indicates the
   ordering of decreasing hyperbolic distance from  $s$ 
5   % PART II: Accumulation of load dependencies ( $\delta$ ) and of HTLC
   values
6    $\delta(v) = 1; \forall v \in V$ 
7   for  $i = 1 : N$  do
8     for all  $v_j \in N_G(i, s)$  (i.e., (greedy) neighbors of  $v_i$  in  $S$ ) do
9        $\delta(v_j) = \delta(v_j) + \delta(v_i) \frac{1}{|N_G(i, s)|}$ ;
10       $HTLC(v_i) = HTLC(v_i) + \delta(v_i)$ ;
11      remove  $v_i$  from  $S$ ;

```

Ο αλγόριθμος αποτελείται από έναν εξωτερικό βρόχο που επαναλαμβάνεται για όλους τους πιθανούς προορισμούς. Με $|N_G(i, s)|$ συμβολίζουμε το πλήθος των άπληστων γειτόνων του κόμβου v_i για τον προορισμό s , όπου ως άπληστο γείτονα ενός κόμβου θεωρούμε τον κόμβο ο οποίος μειώνει την απόσταση στην υπερβολική γεωμετρία προς τον προορισμό, και με V το σύνολο των κόμβων. Μέσα σε αυτό το βρόχο, το Μέρος 1 ταξινομεί όλους τους κόμβους με βάση τη φθίνουσα υπερβολική απόσταση από τον αντίστοιχο προορισμό. Αυτό αποσκοπεί στο να εξεταστούν οι κόμβοι με τη σωστή σειρά όταν αθροίζονται οι εξαρτήσεις φόρτου στη Μέρος 2. Συγκεκριμένα, στο Μέρος 2 υπολογίζεται η συνολική ποσότητα του φόρτου που διέρχεται από έναν κόμβο και κατευθύνεται προς τον αντίστοιχο προορισμό όπως αυτός ορίζεται από τον εξωτερικό βρόχο. Επιπλέον, από τη στιγμή που η δρομολόγηση είναι άπληστη και γίνεται με βάση τις συντεταγμένες στον υπερβολικό χώρο των κόμβων, όταν εξετάζουμε τους κόμβους με βάση μία φθίνουσα κατάταξή τους σε σχέση με την απόστασή τους από τον προορισμό, ένας κόμβος ο οποίος έχει ήδη εξεταστεί δεν μπορεί να χρησιμοποιηθεί ξανά από έναν επόμενο κόμβο και επομένως μπορεί να αφαιρεθεί από τη λίστα S των εναπομείναντων κόμβων.

Για λόγους σύγκρισης, ο αλγόριθμος της ΥΚΦΚ θα συγκριθεί τόσο αλγοριθμικά, όσο και ως προς την αποτελεσματικότητά του και το χρόνο εκτέλεσης του με τον αλγόριθμο του

Brandes για τον υπολογισμό της ΚΦΚ που παρουσιάστηκε στην προηγούμενη ενότητα. Οι δύο αλγόριθμοι διαφέρουν κατά βάση στο Μέρος 1. Πρέπει να επισημάνουμε ξανά ότι η Ενσωμάτωση Rigel δεν εξασφαλίζει 100% επιτυχία στην άπληστη δρομολόγηση στον Υπερβολικό Χώρο. Επομένως, θα μπορούσε να οδηγήσει σε απουσία άπληστων μονοπατιών για μερικά ζεύγη κόμβων σε μία τοπολογία. Παρ' όλα αυτά, το πρόβλημα είναι λιγότερο έντονο σε σύνθετα δίκτυα ελεύθερης-κλίμακας σύμφωνα με τα [19] [15]. Όπως θα δείξουμε στα πειραματικά αποτελέσματα που θα παρουσιαστούν στο επόμενο κεφάλαιο, πέρα από την καταλληλότητα της ΥΚΦΚ για τη μέτρηση της Κεντρικότητας Φόρτου των κόμβων σε περιπτώσεις δρομολόγησης που πραγματοποιείται κάτω από άπληστες μεθόδους, είναι αποδοτική στο να προσεγγίζει ικανοποιητικά και την ΚΦΚ.

Α. Ανάλυση Πολυπλοκότητας

Σε σχέση με την πολυπλοκότητα του Αλγορίθμου 3 για τον υπολογισμό της ΥΚΦΚ, το Μέρος 1 πραγματοποιεί ταξινόμηση όλων των κόμβων του δικτύου, οπότε έχει πολυπλοκότητα ίση με $O(N \log N)$ (για αλγόριθμους ταξινόμησης όπως η Quicksort ή η Heapsort). Το Μέρος 1 του αλγορίθμου 2 έχει πολυπλοκότητα ίση με $O(|E|)$, όπου με $|E|$ συμβολίζουμε το πλήθος των ακμών ενός γράφου. Εάν στον υπολογισμό της πολυπλοκότητας για αυτό το μέρος θεωρήσουμε και τον εξωτερικό βρόχο, οι πολυπλοκότητες για το Μέρος 1 γίνονται $O(N^2 \log N)$ και $O(N|E|)$ για την ΥΚΦΚ και την ΚΦΚ αντίστοιχα. Το Μέρος 2 επιβάλλει ίση πολυπλοκότητα και για τους δύο αλγόριθμους, η οποία ισούται με $O(N|E|)$ (εάν θεωρήσουμε επίσης τον εξωτερικό βρόχο). Επομένως, οι δύο αλγόριθμοι έχουν συγκρίσιμη πολυπλοκότητα με τη διαφορά ότι η ΥΚΦΚ βελτιώνει αυτήν για το Μέρος 1.

Συγκεκριμένα, σύμφωνα με το [20] ο αριθμός των ακμών ενός συνδεδεμένου γράφου $O(|E|)$ μπορεί να αναπαρασταθεί ως N^a (όπου a είναι ένας πραγματικός αριθμός ο οποίος βρίσκεται ανάμεσα στο 1 και το 2). Επομένως, έχουμε μία πολυπλοκότητα $O(NN^a)$ για το μέρος 1 της ΚΦΚ και $O(N^2 \log N)$ για το μέρος 1 της ΥΚΦΚ. Από τη στιγμή που $O(N^a) > O(N \log N)$, η πολυπλοκότητα του αλγορίθμου της ΥΚΦΚ για το Μέρος 1 βρίσκεται κάτω από το ασυμπτωτικό όριο που επιβάλλεται από την $O(|E|)$ της ΚΦΚ. Για πραγματικούς γράφους η τιμή του a εκτείνεται από 1.16 για δίκτυα αυτόνομων συστημάτων ως και 1.56 για δίκτυα αναφορών [20], κάνοντας αυτή την απόκλιση ακόμη πιο προφανή. Παράλληλα, ο αλγόριθμος για τον υπολογισμό της ΥΚΦΚ είναι λιγότερο απαιτητικός σε δομές δεδομένων και στην πρόσβαση και τροποποίηση αυτών.

Τα αριθμητικά αποτελέσματα που θα παρουσιαστούν στο Κεφάλαιο 6 επιβεβαιώνουν ότι η ΥΚΦΚ προσεγγίζει ικανοποιητικά την ΚΦΚ σε σχέση με την κατάταξη των κόμβων σε τοπολογίες ελεύθερης-κλίμακας, ενώ παράλληλα μειώνει σημαντικά τον υπολογιστικό χρόνο που απαιτείται. Πρέπει να σημειωθεί ότι ο αλγόριθμος για την ΥΚΦΚ έχει ένα αρχικό στάδιο

όπου αναθέτονται στους κόμβους των δικτύων υπερβολικές συντεταγμένες μέσω της Ενσωμάτωσης Rigel. Όπως θα δειχθεί στο Κεφάλαιο 6 αυτό το αρχικό στάδιο αυξάνει ελαφρώς τον υπολογιστικό χρόνο που απαιτείται για τον αλγόριθμο της ΥΚΦΚ, ακόμη κι έτσι όμως ο αλγόριθμος αυτός φέρνει καλύτερα αποτελέσματα σε σχέση με τον αλγόριθμο του Brandes για την ΚΦΚ. Όπως θα παρατηρήσουμε, ο χρόνος που απαιτείται για την ενσωμάτωση Rigel είναι αρκετά μικρότερος από τον αντίστοιχο του αλγορίθμου για τον υπολογισμό της ΥΚΦΚ, ιδιαίτερα για πυκνές τοπολογίες (μιας και η Ενσωμάτωση Rigel έχει πολυπλοκότητα γραμμική με τον αριθμό των κόμβων ενός δικτύου).

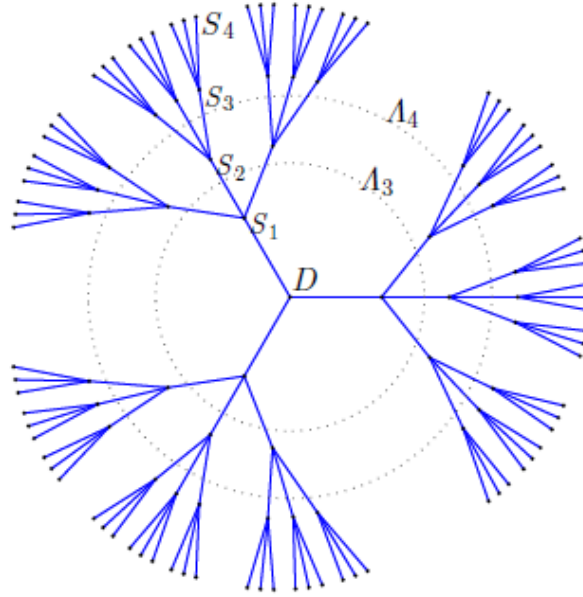
5.2 *Επίδραση της Επιλογής Δενδρικών Ακμών Κατά την*

Άπληστη-Ενσωμάτωση στο Φόρτο των Κόμβων

Όπως είδαμε στην ενότητα 4.3.2, κατά τη διαδικασία της άπληστης ενσωμάτωσης [13], επιλέγεται ένα συνδετικό δένδρο T του γράφου G το οποίο ενσωματώνεται στον υπερβολικό χώρο, δηλαδή οι κόμβοι του δένδρου T και άρα του γράφου G αποκτούν συντεταγμένες στον υπερβολικό χώρο, όπως ο τελευταίος μοντελοποιείται από το μοντέλο του δίσκου του Poincaré. Η επιλογή του συνδετικού δένδρου καθορίζει σημαντικά τη δρομολόγηση της πληροφορίας μέσα στο δίκτυο. Στην ακραία περίπτωση, όπου ο γράφος έχει ο ίδιος τη μορφή δένδρου ($G = T$), υπάρχει ακριβώς μία διαδρομή για κάθε ζεύγος κόμβων στο δίκτυο. Επομένως, τα άπληστα μονοπάτια που ορίζονται με την άπληστη δρομολόγηση που κάνει χρήση των υπερβολικών συντεταγμένων των κόμβων για την προώθηση των πακέτων ταυτίζονται με τα συντομότερα μονοπάτια του δικτύου σε αριθμό βημάτων. Ακριβώς το ίδιο συμβαίνει και στην περίπτωση, όπου ο γράφος G είναι ένας πλήρης γράφος, αφού σε αυτή την ιδανική περίπτωση κάθε ζεύγος κόμβων είναι και άμεσοι γείτονες, οπότε τόσο τα άπληστα όσο και τα συντομότερα μονοπάτια έχουν πάντοτε μήκος ενός βήματος. Σε κάθε ενδιάμεση περίπτωση, που είναι και η συνηθισμένη, δεν υπάρχει κατά κανόνα πλήρης ταύτιση μεταξύ των συντομότερων μονοπατιών και εκείνων που προκύπτουν από την άπληστη δρομολόγηση.

Θεωρούμε, λοιπόν, έναν συνδεδεμένο γράφο G οι κόμβοι του οποίου έχουν ενσωματωθεί με βάση τη διαδικασία της άπληστης ενσωμάτωσης. Η επιλογή ενός συνδετικού δένδρου T χωρίζει όλες τις ακμές του γράφου σε *δενδρικές και μη-δενδρικές (συντομεύσεις) ακμές (tree and non-tree (shortcut) edges)* [23]. Ένα άπληστο μονοπάτι τυπικά αποτελείται από κάποιες δενδρικές ακμές και κάποιες ακμές-συντομεύσεις. Διαισθητικά, οι ακμές συντομεύσεων μπορούν να οδηγήσουν σε άπληστη δρομολόγηση που θα παρεκκλίνει από τη δενδρική δομή και θα πλησιάζει σε αυτήν των συντομότερων μονοπατιών ενός γράφου.

Ως ένα παράδειγμα των παραπάνω, το σχήμα 14 απεικονίζει σχηματικά τα πρώτα $L = 5$ επίπεδα ενός κανονικού δένδρου με $d = 4$ το οποίο εκτείνεται πάνω σε έναν τυχαίο γράφο n κόμβων μαζί με τις τοποθεσίες αρκετών πιθανών προορισμών ενός πακέτου ($S_1 \dots S_4$), και έναν απλό προορισμό D . Ορίζοντας ως $\bar{d}$ το μέσο βαθμό των κόμβων του γράφου, μένουν κατά μέσο όρο $\bar{d} - d$ μη-δενδρικές ακμές προσπίπτουσες σε κάθε κόμβο (οι οποίες δε φαίνονται στο σχήμα).



Σχήμα 14: Ένα απλοποιημένο μοντέλο ενός d -κανονικού δένδρου ενός γράφου[23]

Η θέση ενός υποψήφιου κόμβου τον οποίο μπορούμε να προσεγγίσουμε ως επόμενο βήμα του μονοπατιού μέσω μίας ακμής-συντόμευσης όταν το πακέτο βρίσκεται στον κόμβο S_l στο επίπεδο l εμφανίζεται ως Λ_l στο σχήμα 14 και περιέχει κατά προσέγγιση $p_l = \# \Lambda_l = n/(d-1)^{-(L-l+1)}$ κόμβους. Η πιθανότητα να ακολουθηθεί μία μη-δενδρική ακμή ανάμεσα στον S_l και κάθε άλλο κόμβο θεωρείται ομοιόμορφη, και η πιθανότητα ότι μία ακμή-συντόμευσης που προσπίπτει στον S_l βρίσκεται μέσα στο Λ_l είναι κατά προσέγγιση $p_l = \# \Lambda_l/n = (d-1)^{L-l+1}$, όπου η πιθανότητα μίας τουλάχιστον χρήσιμης συντόμευσης είναι:

$$p_U = 1 - (1 - p_l)^{\delta-d} = 1 - (1 - (d-1)^{-(L-l+1)})^{\delta-d} \quad 5.8$$

Αυτό που είναι σημαντικό για εμάς να παρατηρήσουμε από τη σχέση 5.8 είναι ότι αυτή η πιθανότητα φθίνει εκθετικά, καθώς ένα πακέτο πλησιάζει προς τη ρίζα του δένδρου και ότι επίσης είναι μικρή για μέσες τιμές βαθμού κόμβων του γράφου G που είναι συγκρίσιμες με το βαθμό των κόμβων του συνδετικού δένδρου T .

Επομένως, αναμένουμε πως για κόμβους μικρού βαθμού το κλάσμα των χρησιμοποιούμενων ακμών συντομεύσεων προς το συνολικό αριθμό των ακμών που χρησιμοποιούνται κατά την άπληστη δρομολόγηση θα είναι μικρό. Όπως αναφέρεται και στο [23], πράγματι, οι κόμβοι μικρού βαθμού επιβάλουν συνολικά στο δίκτυο άπληστα μονοπάτια που κυρίως αποτελούνται από δενδρικές ακμές. Με άλλα λόγια, για κόμβους με βαθμό κοντινό στο μέσο βαθμό των δένδρων του δικτύου, οι ακμές-συντομεύσεις, παρ' όλου που είναι παρούσες, σπάνια χρησιμοποιούνται στην άπληστη δρομολόγηση. Για το λόγο αυτό, όσο περισσότερες οι ακμές εκείνες που με μεγάλη συχνότητα συναντώνται στη δρομολόγηση συντομότερων μονοπατιών, είναι και ακμές που θα χρησιμοποιηθούν στο συνδεδετικό δένδρο που ενσωματώνεται, τόσο πιο κοντά στα πραγματικά συντομότερα μονοπάτια οδηγείται η άπληστη δρομολόγηση στο δίκτυο. Στην πραγματικότητα, οι ακμές-συντομεύσεων χρησιμοποιούνται στην άπληστη δρομολόγηση κατά ένα ποσοστό που κινείται κοντά στο 15-30% (βλ. [23], fig.4) με το υπόλοιπο 70-85% της κίνησης να κυριαρχείται από δενδρικές ακμές.

Κάνοντας, λοιπόν, χρήση αυτής της παρατήρησης προτείνουμε ένα πλαίσιο ενεργειών που θα μπορούσε να ωφελήσει στη μεταφορά της συμφόρησης κίνησης από κόμβους υψηλού βαθμού σε άλλους κόμβους του δικτύου. Συγκεκριμένα, επιδιώκουμε το συνδεδετικό δένδρο που θα επιλεγεί να έχει όσο το δυνατόν μεγαλύτερο αριθμό ακμών που θα προσπίπτουν σε κόμβους χαμηλής κεντρικότητας και όσο το δυνατόν μικρότερο αριθμό δενδρικών ακμών που θα προσπίπτουν σε κόμβους υψηλής κεντρικότητας. Για να το επιτύχουμε αυτό κάνουμε χρήση της Άπληστης Ενσωμάτωσης [13] για την ανάθεση συντεταγμένων στον υπερβολικό χώρο στους κόμβους του δικτύου. Παράλληλα, χρησιμοποιούμε δένδρα ελάχιστου βάθους για να κατασκευάσουμε το συνδεδετικό δένδρο T του γράφου G . Η συνολική πολυπλοκότητα αυτής της διαδικασίας είναι γραμμική, αφού συγκεκριμένα χρειαζόμαστε χρόνο $O(m)$ για την κατασκευή του δένδρου ελάχιστου βάθους και $O(n)$ για την ενσωμάτωση του δένδρου αυτού (και κατ' επέκταση του γράφου G). Επομένως, η συνολική πολυπλοκότητα της διαδικασίας που προτείνουμε είναι γραμμική ως προς τον αριθμό των κόμβων και των ακμών του δικτύου $O(n + m)$ κάνοντας το συγκεκριμένο πλαίσιο-ενεργειών κατάλληλο για τη λειτουργία του πάνω σε πραγματικούς γράφους σύνθετων δικτύων.

Το συνδεδετικό δένδρο ελάχιστου βάθους επιλέγεται, καθώς (όπως εξηγήθηκε στην ενότητα 4.3.2) κάθε κόμβος επιλέγει ως πατέρα του τον κόμβο εκείνο ο οποίος απέχει την ελάχιστη απόσταση από τη ρίζα του δένδρου. Οπότε, ο κόμβος-ρίζα θα έχει δενδρικές-ακμές προς όλους τους γείτονες του, δηλαδή θα έχει την υψηλότερη δυνατή τιμή κεντρικότητας, αφού με μεγάλη πιθανότητα όλες του οι ακμές θα χρησιμοποιούνται ως ακμές της άπληστης δρομολόγησης. Παράλληλα, όσο απομακρυνόμαστε από αυτόν τόσο μειώνεται η πιθανότητα να υπάρχουν δενδρικές ακμές που θα προσπίπτουν σε κάποιον κόμβο, αφού κατά πάσα

πιθανότητα οι γείτονες ενός κόμβου θα έχουν σχηματίσει δενδρικές ακμές στο ακριβώς παραπάνω επίπεδο προς τη ρίζα του δένδρου.

Προκύπτει, επομένως, με βάση την επιλογή του συγκεκριμένου δένδρου μία άμεση συσχέτιση ανάμεσα στο επίπεδο του συνδετικού δένδρου που βρίσκεται ένας κόμβος, στον αριθμό των δενδρικών ακμών που προσπίπτουν σε αυτόν και συνακόλουθα στην κεντρικότητά του. Παραδείγματος χάριν, ο κόμβος 1 έχει 58 δενδρικές ακμές όταν είναι ρίζα του δένδρου (όσοι και οι γείτονες του), κατά μέσο αριθμό 50,4138 δενδρικές ακμές όταν βρίσκεται στο επίπεδο 2 του δένδρου και μόλις 2,3659 δενδρικές ακμές όταν βρίσκεται στο επίπεδο 3. Επομένως, μπορούμε να μετατρέψουμε το πρόβλημα της επιλογής δενδρικών ακμών για έναν κόμβο στο απλούστερο πρόβλημα της επιλογής κόμβου-ρίζας στο δένδρο και κατ' επέκταση της απόστασής του από τη ρίζα του δένδρου. Συγκεκριμένα, μπορούμε με μία μόνο εκτέλεση BFS από έναν κόμβο υψηλού βαθμού να βρούμε κάποιον κόμβο που απέχει μεγάλη απόσταση (ακόμη και τη μεγαλύτερη απόσταση) από τον κόμβο αυτόν. Θέτοντας έτσι, τον κόμβο αυτό ως κόμβο-ρίζα του συνδετικού δένδρου, θα έχουμε πετύχει εύκολα ο κόμβος υψηλής κεντρικότητας να ενσωματωθεί σε υψηλό επίπεδο του δένδρου και άρα να έχει μικρή πιθανότητα να προσπίπτουν σε αυτόν δενδρικές-ακμές.

Στην ενότητα 6.2 θα ακολουθήσει η πειραματική αξιολόγηση της παραπάνω διαδικασίας, αρχικά με βάση κάποια συγκεκριμένα παραδείγματα που θα αποδεικνύουν τη χρησιμότητά της και στη συνέχεια με τη χρήση στατιστικών συμπερασμάτων ώστε να δειχθεί σε πιο ευρεία κλίμακα η επίδραση της επιλογής του κόμβου-ρίζας στη μεταφορά της συμφόρησης κίνησης στο δίκτυο. Για το σκοπό αυτό θα εμπλακεί η Υπερβολική Κεντρικότητα Φόρτου Κίνησης που αναλύσαμε στο προηγούμενο κεφάλαιο, ώστε να δείξουμε την αλλαγή στην τιμή της Κεντρικότητας Φόρτου Κίνησης των κόμβων σε καθεστώς άπληστης δρομολόγησης. Η αξιολόγηση θα γίνει κάνοντας χρήση διαφορετικών επιλογών κόμβου-ρίζας των συνδετικών δένδρων και εξετάζοντας την επίδραση αυτή σε διάφορες τοπολογίες που θα ακολουθούν τη μορφή δικτύων που δείξαμε στην ενότητα 3.3.

6 Αποτελέσματα Υπολογισμού Μετρικών

Κεντρικότητας στην Υπερβολική Γεωμετρία

Στο κεφάλαιο αυτό θα παρουσιαστούν τα αριθμητικά αποτελέσματα των δύο τεχνικών που αναλύθηκαν στο Κεφάλαιο 5.

Στην Ενότητα 6.1 θα αξιολογηθεί η χρήση του Αλγόριθμου 3 για την Υπερβολική Κεντρικότητα Φόρτου Κίνησης και η απόδοση του θα αξιολογηθεί σε σύγκριση με τον υπολογιστικό χρόνο του Αλγόριθμου 2 για την Κεντρικότητα Φόρτου Κίνησης, αλλά και την ακρίβεια των προβλέψεων του στην κατάταξη των κόμβων σε σχέση με την κεντρικότητά τους. Για την υλοποίηση των 2 αλγορίθμων χρησιμοποιήθηκε το υπολογιστικό περιβάλλον MATLAB, ενώ χρησιμοποιήθηκαν ίδιες δομές δεδομένων και για τους 2 αλγορίθμους με σεβασμό στη διατήρηση της πολυπλοκότητας των πράξεων (π.χ., πρόσβαση σε στοιχεία, εισαγωγή, αφαίρεση κτλ.) που προτάθηκαν από τον αλγόριθμο του Brandes [8]. Για τα πειράματα χρησιμοποιήθηκε ένας κανονικός Προσωπικός Υπολογιστής με επεξεργαστή IntelCore i5 750 2.66Ghz, 5GB RAM και λειτουργικό σύστημα Windows 8.1 (64bit).

Στην Ενότητα 6.2 θα αξιολογηθεί η επίδραση της επιλογής κόμβου-ρίζας του συνδετικού δένδρου που θα ενσωματωθεί με την Άπληστη Ενσωμάτωση[13] και συνακόλουθα της κεντρικότητας των κόμβων (τόσο σε απόλυτη τιμή όσο και σε κατάταξη) σε σχέση με την απόσταση τους κάθε φορά από την ρίζα του δένδρου. Για την αξιολόγηση αυτή χρησιμοποιήθηκαν 3 συνθετικοί γράφοι 100 κόμβων και 500 μη-κατευθυνόμενων ακμών οι οποίοι ακολουθούν τα μοντέλα της ελεύθερης-κλίμακας, των τυχαίων γράφων των Erdős-Rényi και του μικρού κόσμου, αντίστοιχα. Ένα μειονέκτημα της Άπληστης Ενσωμάτωσης είναι ότι αναθέτει συντεταγμένες στους κόμβους ενός δικτύου πάνω στο μοντέλο του δίσκου του Poincare. Κατά συνέπεια, όταν στο δίκτυο υπάρχουν αρκετοί κόμβοι (>50) παρατηρείται το φαινόμενο πολλοί από αυτούς να τοποθετούνται αρκετά κοντά στο σύνορο στο άπειρο και, λόγω του φαινομένου που αναφέρθηκε στο Κεφάλαιο 4.2 και στις Εξισώσεις 4.3, ακόμη κι όταν οι αποστάσεις μεταξύ των κόμβων είναι μεγάλες, οι συντεταγμένες τους (σε σχέση με την Ευκλείδεια οπτική) είναι υπερβολικά κοντινές. Κατά συνέπεια, οι παραδοσιακές μεταβλητές σταθερής ακρίβειας (π.χ., 16 ή 32 bit) δεν επαρκούν για να καλύψουν την ακρίβεια που χρειάζεται στους υπολογισμούς. Επομένως, χρησιμοποιήσαμε τις βιβλιοθήκες GMP [24] και MPFR [25] της γλώσσας προγραμματισμού C ώστε να χρησιμοποιήσουμε μεταβλητές αυθαίρετης ακρίβειας (έως και 50.000 δεκαδικών ψηφίων) για να μπορέσουμε να πετύχουμε

την ενσωμάτωση των κόμβων και να υπολογίσουμε την μεταξύ τους απόσταση με την απαιτούμενη ακρίβεια. Πρέπει, φυσικά, να σημειώσουμε ότι η χρήση αριθμητικών αυθαίρετα μεγάλης ακρίβειας δεν γίνεται χωρίς κόστος σε υπολογιστικό χρόνο, καθώς πλέον οι αριθμητικές πράξεις πραγματοποιούνται σε επίπεδο λογισμικού και όχι στους καταχωρητές του επεξεργαστή.

6.1 Υπερβολική Κεντρικότητα Φόρτου Κίνησης με Χρήση της Ενσωμάτωσης Rigel

Σε αυτή την ενότητα, παρουσιάζονται ορισμένα αριθμητικά αποτελέσματα τα οποία συγκρίνουν από τη μία μεριά τον απαιτούμενο υπολογιστικό χρόνο της ΥΚΦΚ και της ΚΦΚ και από την άλλη μεριά την ακρίβεια της χρήσης της ΥΚΦΚ για τον προσεγγιστικό υπολογισμό της ΚΦΚ, με εστίαση σε γράφους που φέρουν την ιδιότητα της ελεύθερης-κλίμακας. Όπως αναφέρθηκε και στο Κεφάλαιο 5.1 η ΥΚΦΚ θεωρείται λιγότερο απαιτητική υπολογιστικά από την ΚΦΚ (βλέπε Κεφάλαιο 5.1.2.A) και σε συνδυασμό με την άπληστη δρομολόγηση πάνω σε συντεταγμένες του υπερβολικού χώρου (η οποία είναι επίσης υπολογιστικά αποτελεσματική) οδηγεί σε ένα κατάλληλο επιχειρησιακό πλαίσιο ενεργειών για δίκτυα μεγάλης κλίμακας.

A. ΣΥΝΘΕΤΙΚΑ ΔΙΚΤΥΑ

Ο Πίνακας 6.1 παρουσιάζει 6 συνθετικούς γράφους ελεύθερης-κλίμακας, παρέχοντας και τον τίτλο που χρησιμοποιούμε για να αναφερθούμε σε αυτούς, τον αριθμό των κόμβων και τον αριθμό των ακμών τους. Επίσης, ο ίδιος πίνακας παρουσιάζει και τον χρόνο σε δευτερόλεπτα για τον υπολογισμό της ΚΦΚ και της ΥΚΦΚ για κάθε δίκτυο. Από την στιγμή που ο αλγόριθμος της ΥΚΦΚ απαιτεί το προπαρασκευαστικό στάδιο της ανάθεσης συντεταγμένων στον υπερβολικό χώρο στους κόμβους του δικτύου χρησιμοποιώντας την ενσωμάτωση Rigel, ο χρόνος που απαιτείται για το στάδιο αυτό παρατίθεται στην τελευταία στήλη του Πίνακα 6.1 (όπου δε γίνεται χρήση παραλληλοποίησης [12]). Επομένως, ο συνολικός χρόνος που απαιτείται για τον υπολογισμό της ΥΚΦΚ προκύπτει από το άθροισμα των δύο τελευταίων στηλών του Πίνακα 6.1 για κάθε τοπολογία. Η τιμή που βρίσκεται στο εσωτερικό της παρένθεσης στην τελευταία στήλη αντιστοιχίζεται στον αριθμό των κόμβων-αναφοράς που χρησιμοποιήθηκαν για τον αλγόριθμο της Ενσωμάτωσης Rigel. Οι κόμβοι που χρησιμοποιήθηκαν ως κόμβοι αναφοράς ήταν οι κόμβοι με τον υψηλότερο βαθμό κάθε συνθετικού γράφου, παρ' όλα αυτά χρησιμοποιούμε περισσότερους κόμβους ως κόμβους-

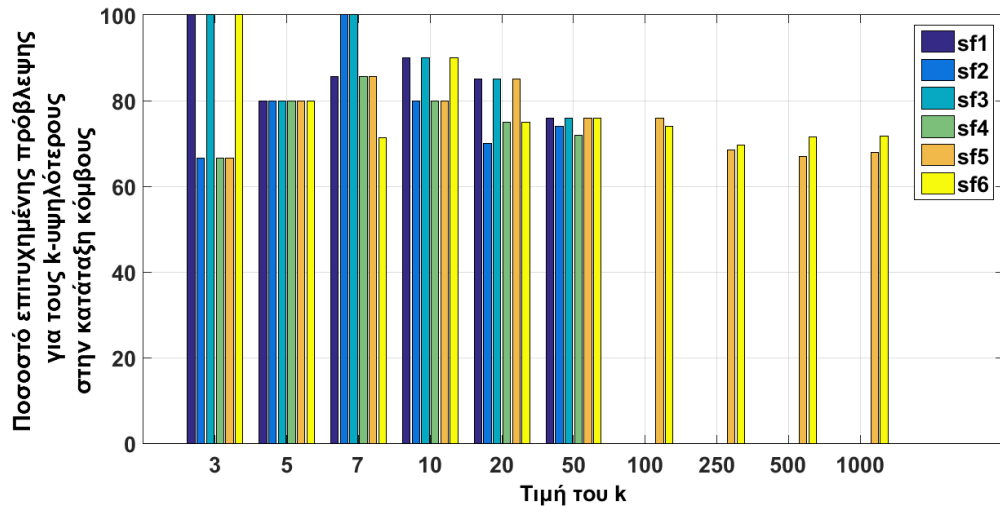
αναφοράς για μεγαλύτερες δικτυακές τοπολογίες (με βάση το πλήθος των κόμβων) προσπαθώντας να διατηρήσουμε την ακρίβεια της ενσωμάτωσης Rigel στην προσέγγιση των αποστάσεων (σε αριθμό βημάτων) μεταξύ των κόμβων ενός γράφου και στην αντίστοιχη απόσταση των κόμβων στον υπερβολικό χώρο.

ΠΙΝΑΚΑΣ 6.1

ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΣΥΝΘΕΤΙΚΩΝ ΓΡΑΦΩΝ & ΣΥΓΚΡΙΣΗ ΤΟΥ ΥΠΟΛΟΓΙΣΤΙΚΟΥ ΧΡΟΝΟΥ ΜΕΤΑΞΥ ΚΦΚ & ΥΚΦΚ

Ετικέτα Συνόλου Δεδομένων	# Κόμβων	# Ακμών	Χρόνος ΚΦΚ (sec)	Χρόνος ΥΚΦΚ (sec)	Χρόνος Rigel (sec) (# κόμβων- αναφοράς)
sf1	5000	29747*2	141.2	42.23	14 (8)
sf2	5000	92601*2	178.73	99.82	11 (8)
sf3	10000	59705*2	571.53	178.44	26 (12)
sf4	10000	187281*2	790.54	456.34	21 (12)
sf5	20000	119668*2	2281.75	737.42	56 (36)
sf6	20000	376522*2	3017.83	1978.51	93 (36)

Από τον Πίνακα 6.1, παρατηρούμε ότι η ΥΚΦΚ υπολογίζεται τουλάχιστον 1.5 φορές γρηγορότερα από την ΚΦΚ για πυκνούς γράφους και 2.5 φορές γρηγορότερα από την ΚΦΚ για λιγότερο πυκνούς γράφους. Οι πιο πυκνές τοπολογίες αυξάνουν τον χρόνο για τον υπολογισμό τόσο της ΥΚΦΚ, όσο και της ΚΦΚ. Συγκεκριμένα, για την ΥΚΦΚ, σε πυκνότερες τοπολογίες θα υπάρχουν περισσότερα άπληστα μονοπάτια τα οποία θα αυξάνουν τον υπολογιστικό χρόνο που χρειάζεται για το μέρος 2 του Αλγορίθμου για τον υπολογισμό της, παρ' όλο που για τον ίδιο αριθμό κόμβων δε θα υπάρχει επίδραση στο χρόνο κατάταξης του Μέρους 1. Ένας τρόπος για να αντιμετωπιστεί αυτό το πρόβλημα θα ήταν να περιοριστεί ο αριθμός των άπληστων γειτόνων θεωρώντας ως πιθανούς επόμενους στόχους για τη διάδοση της κίνησης (δηλαδή άπληστους γείτονες) μόνο στους γείτονες εκείνους ενός κόμβου για τους οποίους η απόσταση τους από τον προορισμό μειώνεται περισσότερο από ένα καθορισμένο κατώφλι.



Σχήμα 15: Συνθετικοί γράφοι ελεύθερης-κλίμακας: Ακρίβεια πρόβλεψης σε σύγκριση ΥΚΦΚ και ΚΦΚ

Το σχήμα 15 παρουσιάζει την ακρίβεια που επιτυγχάνεται στην προσέγγιση των τιμών της ΚΦΚ με τις τιμές της ΥΚΦΚ για τους συνθετικούς γράφους ελεύθερης-κλίμακας. Συγκεκριμένα, για κάθε τιμή k , η οποία αναπαριστά τους k – υψηλότερους κόμβους στην κατάταξη, υπολογίζουμε το ποσοστό της ταύτισης μεταξύ της κατάταξης των κόμβων με βάση την ΚΦΚ και την ΥΚΦΚ. Παρατηρούμε ότι για τα δίκτυα ελεύθερης-κλίμακας η ΥΚΦΚ επιτυγχάνει ένα ποσοστό ακρίβειας τουλάχιστον ίσο με 65%, το οποίο φτάνει ακόμη και το 100% σε αρκετές περιπτώσεις. Επίσης, παρατηρούμε ότι ο αριθμός των κόμβων ενός δικτύου δεν επηρεάζει την επιτευχθείσα ακρίβεια. Αυτή η συμπεριφορά είναι αναμενόμενη για δίκτυα ελεύθερης-κλίμακας τα οποία παρουσιάζουν μία «κρυμμένη» υπερβολική δομή ευνοώντας έτσι την άπληστη δρομολόγηση με χρήση συντεταγμένων στο χώρο της υπερβολικής γεωμετρίας. Επομένως, σε τέτοιες τοπολογίες η ΥΚΦΚ μπορεί να χρησιμοποιηθεί για να προσεγγίσει την ΚΦΚ και επομένως να αποφύγει την ανάγκη για υπολογισμό των συντομότερων μονοπατιών (δηλαδή να χρησιμοποιηθεί ο αλγόριθμος της ΥΚΦΚ αντί του αλγορίθμου της ΚΦΚ για να υπολογιστούν οι τιμές κεντρικότητας των κόμβων).

B. ΠΡΑΓΜΑΤΙΚΑ ΔΙΚΤΥΑ

Ανάλογα με τη διαδικασία που περιγράφηκε παραπάνω συγκρίσεις των αποτελεσμάτων και αξιολόγηση της επίδοσης του Αλγόριθμου 3 για τον υπολογισμό της ΥΚΦΚ έγινε και για σύνολα δεδομένων πραγματικών δικτύων τα οποία παρέχονται από το πανεπιστήμιο Stanford [21]. Όμοια με τα συνθετικά δίκτυα, ο Πίνακας 6.2 και το Σχήμα 16 παρέχουν παρόμοιες πληροφορίες με τον Πίνακα 6.1 και το Σχήμα 15, αντίστοιχα. Στον Πίνακα 6.2 η στήλη RMSE αναφέρεται στην ρίζα μέσης τετραγωνικής απόκλισης (root mean

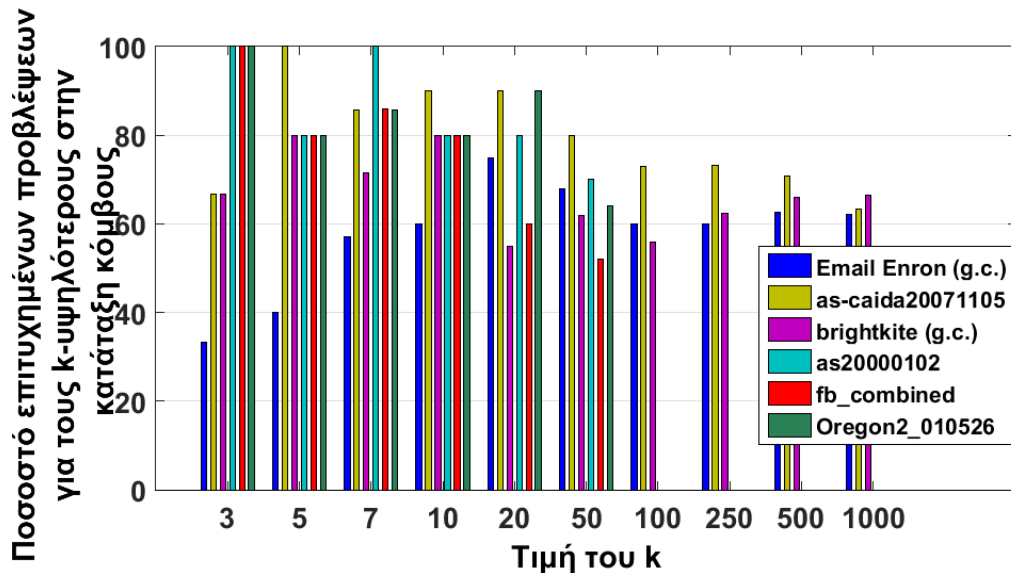
square error) η οποία αντιστοιχίζεται στο ταίριασμα (fitting) της κατανομής νόμου-δύναμης (power-law) στο αντίστοιχο σύνολο δεδομένων. Οι παρατηρήσεις που προκύπτουν είναι αντίστοιχες σε γενικές γραμμές με αυτές για τα συνθετικά δίκτυα ελεύθερης-κλίμακας. Συγκεκριμένα, χρησιμοποιώντας την ΥΚΦΚ αντί της ΚΦΚ οδηγούμαστε σε μία επιτάχυνση του υπολογιστικού χρόνου από 1.2 φορές (facebook_combined) έως και 4.1 φορές (AS-caida20071105). Ενώ η ΥΚΦΚ είναι η κατάλληλη κεντρική μετρικότητα για την περίπτωση της άπληστης δρομολόγησης από το Σχήμα 16 παρατηρούμε ότι προσεγγίζει επίσης ικανοποιητικά την κατάταξη των κόμβων που προκύπτει από την ΚΦΚ. Πιο συγκεκριμένα, για $k \geq 7$ η αντίστοιχη ακρίβεια είναι πάνω από 57% για όλα τα σύνολα δεδομένων. Είναι σημαντικό να σημειώσουμε εδώ ότι για όλα τα σύνολα δεδομένων του Πίνακα 6.2, εκτός από τον CA-HepTh, έχουμε μικρές τιμές της RMSE, δηλαδή παρουσιάζουν χαρακτηριστικά τοπολογιών ελεύθερης-κλίμακας. Για τα δεδομένα του γράφου CA-HepTh δεν υπάρχει καλή συσχέτιση μεταξύ της κατανομής του βαθμού κόμβων του συγκεκριμένου δικτύου και της κατανομής νόμου-δύναμης. Επομένως, για όλα τα σύνολα δεδομένων, εκτός του CA-HepTh αναμένεται υψηλή ακρίβεια όταν προσεγγίζουμε την ΚΦΚ με την ΥΚΦΚ, επειδή, όπως συζητήθηκε προηγουμένως, σε γράφους ελεύθερης-κλίμακας, τα άπληστα μονοπάτια πάνω σε υπερβολικές συντεταγμένες είναι παρόμοια με τα συντομότερα μονοπάτια με βάση των αριθμό βημάτων. Στον αντίποδα, για το σύνολο δεδομένων CA-HepTh δεν παρατηρούμε καλή ακρίβεια (π.χ. $0/k = 3, 0/5, 1/7, 3/10, 5/20, 19/50, 40/100$).

ΠΙΝΑΚΑΣ 6.2

ΠΕΡΙΓΡΑΦΗ ΠΡΑΓΜΑΤΙΚΩΝ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ & ΣΥΓΚΡΙΣΗ ΤΟΥ ΥΠΟΛΟΓΙΣΤΙΚΟΥ ΧΡΟΝΟΥ ΜΕΤΑΞΥ ΚΦΚ & ΥΚΦΚ. ΜΕ «Γ.Σ.» ΣΥΜΒΟΛΙΖΟΥΜΕ ΤΗ ΓΙΓΑΝΤΙΑΙΑ ΣΥΝΙΣΤΩΣΑ ΕΝΟΣ ΔΙΚΤΥΟΥ ΟΤΑΝ ΑΥΤΟ ΔΕΝ ΕΙΝΑΙ ΣΥΝΔΕΔΕΜΕΝΟ

Ετικέτα Συνόλου Δεδομένων [21]	# Κόμβων	# Ακμών/2	RMSE	Χρόνος ΚΦΚ (sec)	Χρόνος ΥΚΦΚ (sec)	Χρόνος Rigel (sec) (# κόμβων- αναφοράς)
as20000102	6474	25144	0.005859	198.34	41.74	15 (8)
Fb_combined	4039	88234	0.001594	112.9	78.35	12.6 (8)
Email Enron(Γ.Σ.)	33692	180811	0.002164	6048.95	2106.44	210 (50)
Oregon2_010 526	11461	32730	0.005414	645.35	153.76	27 (16)
as-caida 20071105	26475	106762	0.004704	3391.93	713.98	109 (50)

Brightkite(Γ.Σ)	56739	212945	0.000699	17016.45	4962.76	336 (50)
CA-HepTh	8638	24827	0.01354	382.61	93.9	21 (8)



Σχήμα 16: Γράφοι από πραγματικά σύνολα δεδομένων: Ποσοστό επιτυχίας στην πρόβλεψη των k -υψηλότερων στην κατάταξη κόμβων

6.2 Επίδραση Επιλογής Δενδρικών Ακμών στην Άπληστη

Ενσωμάτωση

Για την αξιολόγηση των αποτελεσμάτων μας χρησιμοποιήθηκαν, όπως αναφέρθηκε στην εισαγωγή του κεφαλαίου, 3 συνθετικοί γράφοι 100 κόμβων και 500 μη-κατευθυνόμενων ακμών που ακολουθούν τη δομή των δικτύων ελεύθερης κλίμακας, τυχαίων γράφων και μικρού κόσμου αντίστοιχα. Για να αξιολογηθεί η σημασία ενός κόμβου στη ροή της πληροφορίας χρησιμοποιήσαμε τη μετρική της Κεντρικότητας Φόρτου Κίνησης για δρομολόγηση που ακολουθεί τα συντομότερα μονοπάτια και την Υπερβολική Κεντρικότητα Φόρτου Κίνησης (βλ. Κεφάλαιο 5.1.2) που ορίζεται με χρήση της άπληστης δρομολόγησης.

Αρχικά θα παρουσιάσουμε ορισμένα παραδείγματα που αναδεικνύουν την επίδραση της επιλογής κόμβου ρίζας του συνδετικού δένδρου ελάχιστου βάθους στην κεντρικότητα των

κόμβων σε συγκεκριμένες περιπτώσεις, ενώ στη συνέχεια θα ακολουθήσει η παράθεση στατιστικών αποτελεσμάτων ώστε να προκύψει μία πιο συνολική εικόνα της μεθόδου.

Καταρχήν, όμως για να στοιχειοθετηθεί η επίδραση της επιλογής δενδρικών ακμών στην άπληστη δρομολόγηση αναφέρονται 2 σημαντικά στοιχεία που διαπιστώθηκαν:

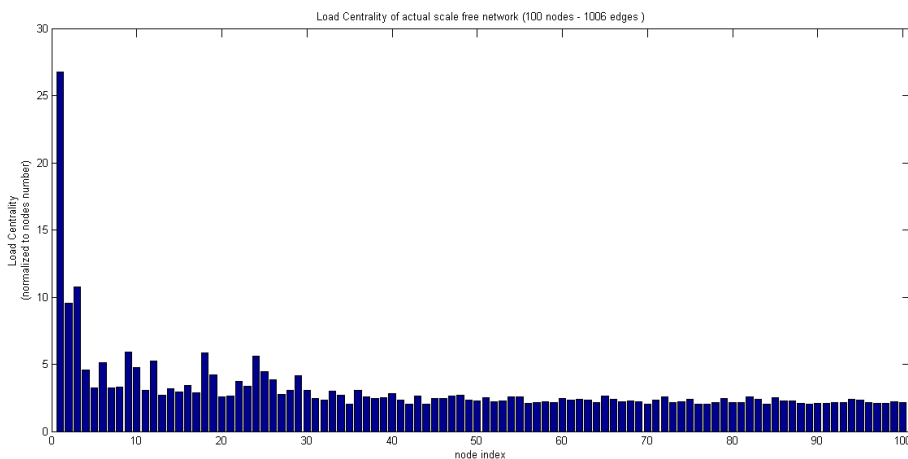
1. Ενσωματώνοντας το δένδρο μέγιστου βάρους, όπου αναθέτει βάρη στις ακμές ενός γράφου με βάση τη συχνότητά που αυτές συμμετέχουν στη δρομολόγηση συντομότερων αποστάσεων [23], παρατηρείται πως διαφορετικές επιλογές κόμβου-ρίζας δεν έχουν σημαντική επίδραση στην τιμή της ΥΚΦΚ των κόμβων. Επομένως, η επιλογή αυτή έχει νόημα –και επηρεάζει– την κεντρικότητα των κόμβων στην περίπτωση του δένδρου ελάχιστου βάθους για το λόγο που αναφέρθηκε προηγουμένως. Αντίθετα, για το δένδρο αυτό οι δενδρικές ακμές που χρησιμοποιούνται είναι ίδιες.
2. Αναφέρεται γενικώς πως για όλες τις επιλογές δένδρων που έγιναν στην παρούσα εργασία υπάρχει ευθεία ταύτιση μεταξύ του αριθμού των δενδρικών ακμών που προσπίπτουν σε έναν κόμβο και της τιμής της κεντρικότητάς του. Συγκεκριμένα, ανάμεσα στους 10 κόμβους με τη μεγαλύτερη τιμή προσπίπτοντων δενδρικών ακμών και τους 10 κόμβους με την υψηλότερη τιμή ΚΦΚ βρίσκουμε κοινούς κόμβους:
 - Στην περίπτωση του γράφου ελεύθερης-κλίμακας κατά μέσο όρο 8.57 κόμβους ανάμεσα στα δύο σύνολα (για κάθε επιλογή συνδεδετικού δένδρου) με τυπική απόκλιση 0.755, με τις αντίστοιχες τιμές να είναι 8.5 με 1.0632 για τον τυχαίο γράφο και 5.65 με 1.3870 για τον γράφο μικρού κόσμου.

A. ΠΑΡΑΔΕΙΓΜΑΤΑ

• Γράφος Ελεύθερης-Κλίμακας

Αρχικά στο *σχήμα 17* παρουσιάζεται η Κεντρικότητα Φόρτου Κίνησης (ΚΦΚ) των κόμβων του συνθετικού δικτύου που κατασκευάσαμε υπολογισμένη με τον Αλγόριθμο 2, αφού θεωρούμε δρομολόγηση συντομότερων μονοπατιών. Όπως παρατηρούμε υπάρχει ένας μικρός αριθμός κόμβων του δικτύου που έχει ουσιαστικά τον «έλεγχο» πάνω στη ροή της πληροφορίας, ενώ η συντριπτική πλειοψηφία των κόμβων παρουσιάζει μικρές τιμές κεντρικότητας.

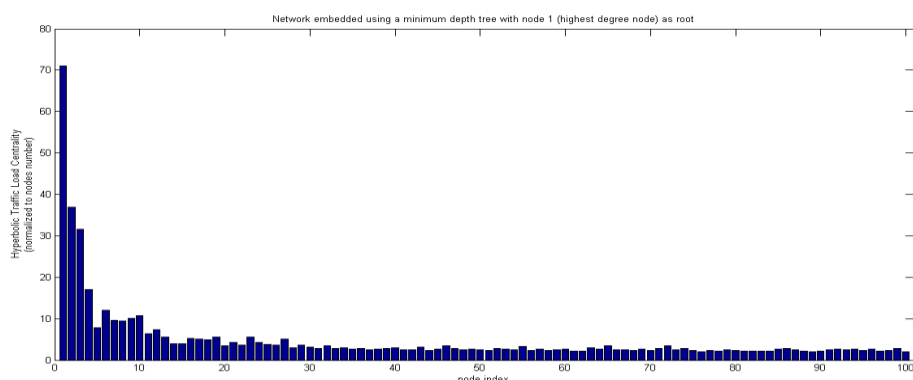
Πρέπει να σημειωθεί ότι στο επόμενο σχήμα, όπως και σε όλα τα σχήματα που θα ακολουθήσουν εκτός αν τονιστεί διαφορετικά, η τιμή της κεντρικότητας των κόμβων είναι κανονικοποιημένη με βάση τον αριθμό των κόμβων του δικτύου.



Σχήμα 17: ΚΦΚ των κόμβων του γράφου ελεύθερης κλίμακας

Στη συνέχεια παρέχουμε δύο παραδείγματα επιλογής κόμβων, τέτοια ώστε να αναπαρασταθεί με σαφήνεια η σημαντική επίδραση της επιλογής κόμβου-ρίζας και συνακόλουθα για τους λόγους που εξηγήθηκαν και δενδρικών ακμών στη μεταφορά της κίνησης του δικτύου, όπως θεωρητικά επεξηγήθηκε στο Κεφάλαιο 5.2. Σημειώνουμε ότι πλέον η δρομολόγηση γίνεται με βάση τα άπληστα μονοπάτια, οπότε η μετρική που χρησιμοποιείται για τον υπολογισμό της κεντρικότητας των κόμβων είναι η ΥΚΦΚ.

Αρχικά επιλέγουμε τον κόμβο 1 ως κόμβο-ρίζα του συνδετικού δένδρου (ελάχιστου βάθους) που κατασκευάζουμε για την ενσωμάτωσή του στον υπερβολικό χώρο. Ο συγκεκριμένος κόμβος είναι και ο κόμβος που παρουσιάζει την υψηλότερη τιμή κεντρικότητας στο δίκτυο, με σημαντική μάλιστα διαφορά από τους υπόλοιπους κόμβους. Όπως παρατηρούμε από το σχήμα 18 η κατάταξη των κόμβων δεν αλλάζει σημαντικά σε σχέση με αυτήν που παρατηρούμε στο σχήμα 17 για δρομολόγηση συντομότερων μονοπατιών. Συγκεκριμένα, 7/10 κόμβους από τους 10 πρώτους σε κατάταξη, με βάση την τιμή της ΚΦΚ, κόμβους βρίσκονται και πάλι ανάμεσα στους 10 κόμβους με την υψηλότερη τιμή ΥΚΦΚ. Αυτό μπορεί εύκολα να εξηγηθεί από το γεγονός ότι στους γράφους ελεύθερης-κλίμακας οι κόμβοι με τον υψηλότερο βαθμό δεν είναι μόνο και οι κόμβοι με την υψηλότερη τιμή κεντρικότητας, αλλά τείνουν επίσης να συνδέονται και μεταξύ τους σχηματίζοντας, έτσι, ένα στενά συνδεδεμένο

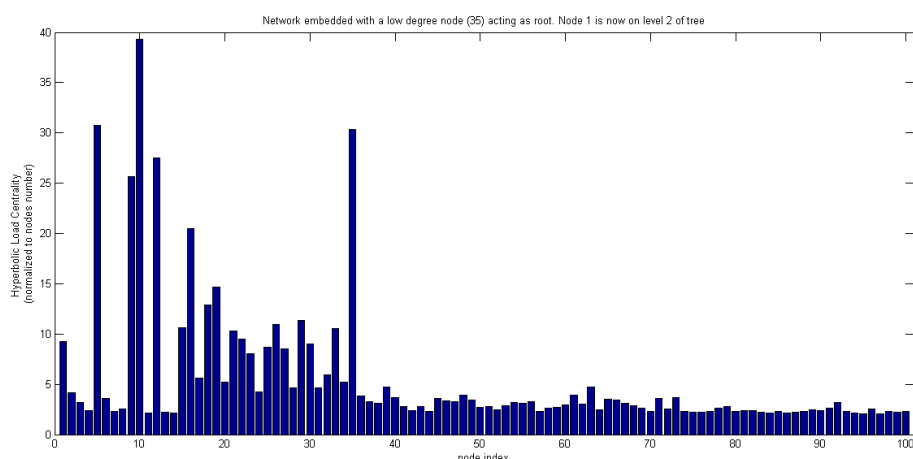


Σχήμα 18: ΥΚΦΚ των κόμβων με επιλογή του κόμβου 1 ως ρίζα του συνδετικού δένδρου

πυρήνα, μέσα από τον οποίο διέρχεται η συντριπτική πλειοψηφία των συντομότερων μονοπατιών του δικτύου [9]. Σαν αποτέλεσμα οι υπόλοιποι κόμβοι υψηλού βαθμού είναι

γείτονες του κόμβου-ρίζα και βρίσκονται και αυτοί σε χαμηλό επίπεδο στο δένδρο σχηματίζοντας μεγάλο αριθμό δενδρικών ακμών με τους υπόλοιπους κόμβους.

Στη συνέχεια, επιλέγουμε τον κόμβο 35 ως ρίζα του συνδετικού δένδρου. Το ενδιαφέρον στοιχείο αυτής της επιλογής είναι ότι πλέον ο κόμβος 1 έχει απόσταση 2 βημάτων από τον κόμβο 35 και επομένως βρίσκεται στο επίπεδο 3 του συνδετικού δένδρου που

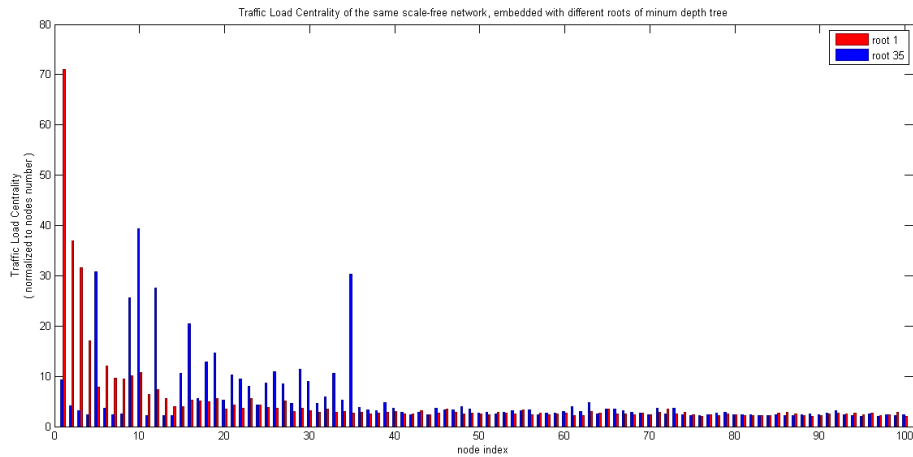


Σχήμα 19: ΥΚΦΚ των κόμβων με επιλογή του κόμβου 35 ως ρίζα του συνδετικού δένδρου

κατασκευάζουμε. Στο ίδιο επίπεδο του δένδρου βρίσκονται, επιπλέον, και οι κόμβοι 2 και 3, δηλαδή οι επόμενοι 2 κόμβοι στην κατάταξη κεντρικότητας των κόμβων. Για το λόγο αυτό, πλέον, οι κόμβοι αυτοί σχηματίζουν αντίστοιχα 4, 0 και 0 δενδρικές ακμές με άλλους κόμβους του δικτύου. Με αυτή την επιλογή προσπαθούμε να δημιουργήσουμε άπληστα μονοπάτια που θα αποκλίνουν από τον πυρήνα του δικτύου που σχηματίζεται από τους κόμβους υψηλής κεντρικότητας και να ακολουθήσουμε εναλλακτικά μονοπάτια διαμέσου άλλων κόμβων, όταν τέτοια μονοπάτια υφίστανται. Το σχήμα 19 επιβεβαιώνει αυτή την υπόθεση μας.

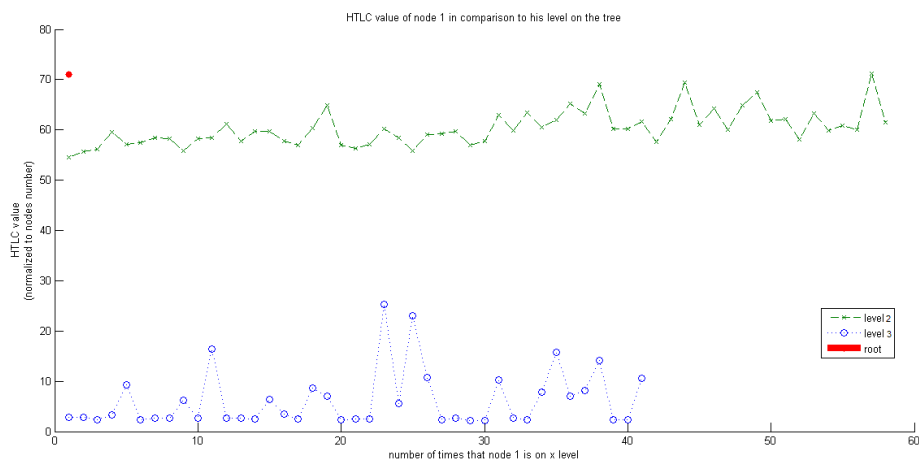
Πλέον, όχι μόνο οι κόμβοι 1,2,3 έχουν πολύ χαμηλή τιμή ΥΚΦΚ σε σύγκριση με την ΚΦΚ του πραγματικού δικτύου ή την ΥΚΦΚ των κόμβων με βάση την άπληστη δρομολόγηση που ακολουθείται όταν ο κόμβος 1 επιλέγεται ως δένδρο, αλλά η συνολική συμφόρηση στο δίκτυο μεταφέρεται από αυτούς τους κόμβους στο τμήμα εκείνο του δικτύου που αποτελείται από τον κόμβο 35 και τους γείτονες του (5,10,12,33), όπως επίσης και σε άλλους κόμβους που δεν συμμετέχουν στον πυρήνα του αρχικού δικτύου. Με αυτή την επιλογή, μόλις 4 κόμβοι, από τους αρχικά 10 πιο υψηλούς στην κατάταξη της ΚΦΚ του δικτύου κόμβους, παραμένουν ακόμη στους 10 πιο υψηλούς στην κατάταξη κόμβους της ΥΚΦΚ. Επιπροσθέτως, η τιμή της ΥΚΦΚ του κόμβου 1 μειώνεται κατά 86.94% σε σχέση με την τιμή του όταν βρισκόταν ως κόμβος-ρίζα στο συνδετικό δένδρο που ενσωματώθηκε στον υπερβολικό χώρο. Επίσης, πλέον βρίσκεται στην 15^η θέση της κατάταξης της ΥΚΦΚ.

Μία σύγκριση των τιμών της κεντρικότητας των κόμβων με βάση την ΥΚΦΚ σε αυτές τις 2 περιπτώσεις που εξετάστηκαν παραπάνω- δηλαδή όταν επιλέγεται αρχικά ο κόμβος 1 ως ρίζα του συνδετικού δένδρου και στη συνέχεια όταν επιλέγεται ο κόμβος 35- φαίνεται στο παρακάτω σχήμα:



Σχήμα 20: Σύγκριση μεταξύ της επιλογής των κόμβων 1 και 35 ως ρίζα του συνδετικού δένδρου

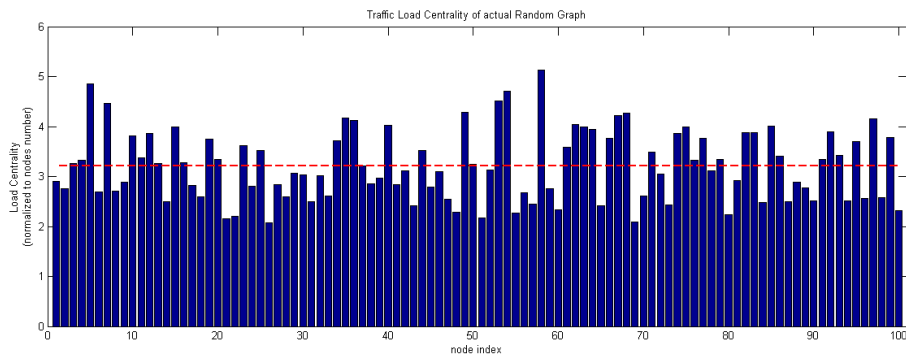
Η επίδραση της απόστασης ενός κόμβου από τη ρίζα του κόμβου στην τιμή της κεντρικότητάς του φαίνεται και στο παρακάτω σχήμα όπου απεικονίζεται η τιμή της ΥΚΦΚ του κόμβου 1 όταν αυτός είναι ρίζα του δένδρου (δηλαδή στο επίπεδο 1) και όταν βρίσκεται στα επίπεδα 2 και 3.



Σχήμα 21: ΥΚΦΚ του κόμβου 1 για διαφορετικά επίπεδά του στο δυνδετικό δένδρο

- **Τυχαίος Γράφος**

Στους τυχαίους γράφους η αθροιστική συνάρτηση κατανομής του βαθμού των κόμβων ακολουθεί την κατανομή Poisson [3]. Σαν αποτέλεσμα η τιμή της ΚΦΚ όλων των κόμβων βρίσκεται κοντά σε μία μέση τιμή, όπως μπορεί να παρατηρηθεί και στο επόμενο σχήμα, όπου η μέση τιμή της ΚΦΚ των κόμβων αναπαρίσταται με μία διακεκομμένη γραμμή.



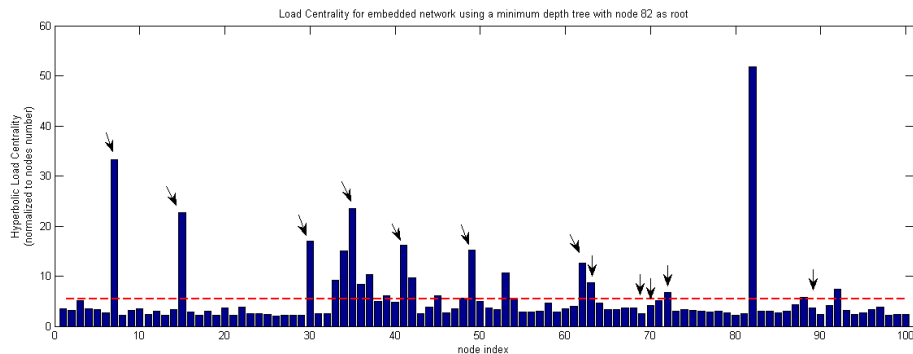
Σχήμα 22: Τιμή της ΚΦΚ των κόμβων ενός τυχαίου γράφου

Κατά συνέπεια, στην περίπτωση αυτή, δεν υπάρχει ένας μικρός αριθμός κόμβων που έχουν εξαιρετικά υψηλή τιμή κεντρικότητας σε σχέση με τη συντριπτική πλειοψηφία των κόμβων του δικτύου, όπως συμβαίνει στα δίκτυα ελεύθερης κλίμακας.

Λόγω της τυχαίας φύσης αυτών των γράφων επιλέγουμε τυχαία έναν κόμβο (θεωρώντας όλους τους κόμβους ισοπίθανους να επιλεγθούν), από το σύνολο όλων των κόμβων του δικτύου, ως κόμβο-ρίζα του συνδετικού δένδρου που θα χρησιμοποιηθεί για την ενσωμάτωση του δικτύου στον υπερβολικό χώρο. Αναπαριστούμε και πάλι την μέση τιμή της ΥΚΦΚ των κόμβων με μία διακεκομμένη γραμμή. Από το Σχήμα 23 μπορούμε να παρατηρήσουμε ότι, όταν η δρομολόγηση είναι άπληστη, υπάρχουν κόμβοι που παρουσιάζουν αρκετά υψηλότερη τιμή κεντρικότητας σε σχέση με τη μέση τιμή της ΥΚΦΚ. Συγκεκριμένα, ο κόμβος 82, που επιλέχθηκε ως ρίζα του δένδρου, και οι γείτονες του, παρουσιάζουν μία αρκετά υψηλότερη τιμή ΥΚΦΚ συγκριτικά με τους υπόλοιπους κόμβους του δικτύου. Οι κόμβοι αυτοί σημειώνονται στο σχήμα 23 με ένα βέλος, ώστε να μπορούμε να τους διακρίνουμε από τους υπόλοιπους κόμβους του δικτύου.

Θα πρέπει να τονίσουμε σε αυτό το σημείο ότι για την κατασκευή ενός δένδρου ελάχιστου βάθους, ένας κόμβος επιλέγει ως πατέρα του τον γείτονα του ο οποίος ελαχιστοποιεί την απόσταση από τον κόμβο-ρίζα σε σχέση με τους υπόλοιπους γείτονες. Στην αλγοριθμική υλοποίηση μας, σε περίπτωση που υπάρχουν περισσότεροι του ενός τέτοιοι κόμβοι, επιλέγεται ως πατέρας ο πρώτος κόμβος που συναντάται. Αυτός είναι ο λόγος για τον οποίο οι γείτονες του κόμβου-ρίζα στο σχήμα 23 που έχουν χαμηλότερη τιμή δείκτη (π.χ., οι κόμβοι 7,15,30,

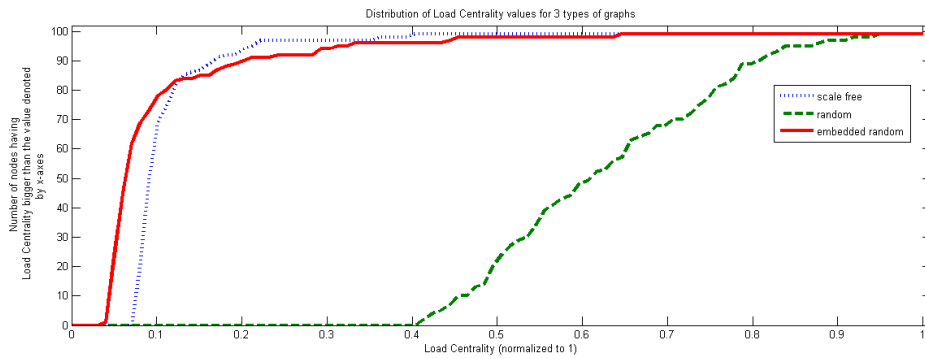
κ.ο.κ.) παρουσιάζουν υψηλότερη κεντρικότητα σε σχέση με κόμβους που έχουν υψηλή τιμή δείκτη (π.χ., 69,70,72,89), λόγω του ότι προσπίπτουν συχνότερα σε αυτούς τους κόμβους δενδρικές ακμές.



Σχήμα 23: ΥΚΦΚ των κόμβων του τυχαίου γράφου με τον κόμβο 82 ως ρίζα του δένδρου

Το γεγονός ότι πλέον λόγω του ότι οι κόμβοι που βρίσκονται σε χαμηλό επίπεδο στο συνδετικό δένδρο έχουν μεγαλύτερο αριθμό δενδρικών ακμών έναντι των υπολοίπων, ενώ ταυτόχρονα σε χαμηλό ύψος η πιθανότητα χρήσης κάποιας συντόμευσης έναντι δενδρικής ακμής ελαχιστοποιείται, έχει τη σημαντική επίδραση που παρατηρήσαμε και προηγούμενως: Συγκεκριμένα, ένας μικρός αριθμός των κόμβων παρουσιάζει πλέον σημαντικά υψηλότερη τιμή ΥΚΦΚ από τη συντριπτική πλειοψηφία των κόμβων του δικτύου.

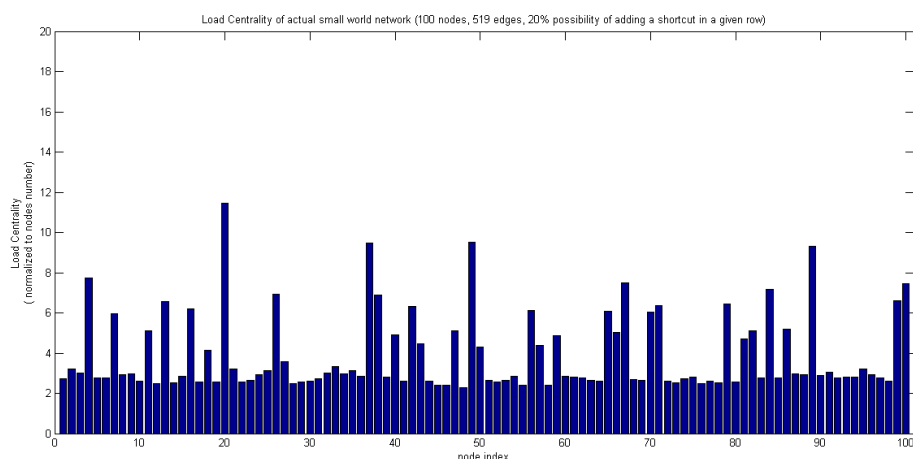
Πράγματι, σχεδιάζοντας την κατανομή των τιμών της ΥΚΦΚ ενός ενσωματωμένου τυχαίου γράφου και συγκρίνοντας αυτή την κατανομή με αυτές της ΚΦΚ του ίδιου τυχαίου γράφου και ενός γράφου ελεύθερης κλίμακας (του ίδιου μεγέθους και τάξης) παρατηρούμε ότι αυτή παρουσιάζει παρόμοια συμπεριφορά με τη δεύτερη παρά με την πρώτη. Σαν αποτέλεσμα, σε έναν τυχαίο γράφο μπορούμε να μεταφέρουμε τη συμμόρφωση της κίνησης σε ένα σύνολο κόμβων που βρίσκονται σε ένα χαμηλό επίπεδο του ενσωματωμένου δένδρου και να επιβάλλουμε ένα σχήμα δρομολόγησης το οποίο θα παρουσιάζει παρόμοια συμπεριφορά, σε σχέση με την κατανομή των τιμών της ΥΚΦΚ των κόμβων, με αυτήν ενός γράφου ελεύθερης-κλίμακας (φυσικά μέχρι ενός σημείου, μιας και η απόκλιση των τιμών κεντρικότητας και του βαθμού των κόμβων, λόγω της κατανομής νόμου δύναμης, σε έναν γράφο ελεύθερης κλίμακας είναι σημαντικά υψηλότερη από την αντίστοιχη ενός τυχαίου γράφου).



Σχήμα 24: Σύγκριση της κατανομής ΥΚΦΚ και ΚΦΚ

- **Γράφος Μικρού-Κόσμου**

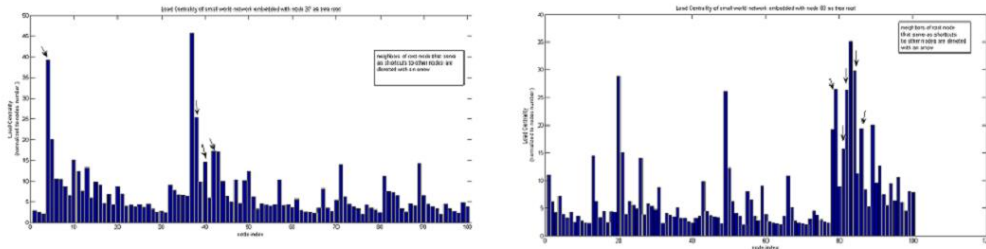
Ένας συνθετικός γράφος τύπου μικρού-κόσμου που κατασκευάζεται με βάση το μοντέλο των Watts-Strogatz (βλ. Κεφάλαιο 3.3.1) έχει μία σχεδόν ομοιόμορφη συνάρτηση κατανομής βαθμών. Συγκεκριμένα, οι κόμβοι αρχικά τακτοποιούνται σε έναν κανονικό δακτύλιο και ανατίθεται σε όλους ο ίδιος αριθμός γειτόνων (αυτοί που βρίσκονται τοπολογικά κοντά τους στο δακτύλιο). Ακμές-συντομεύσεων προστίθενται μεταξύ των δύο κόμβων με μία πιθανότητα b . Στην περίπτωση μας έχουμε θέσει την τιμή του b στο 0.2. Επομένως, ένας κόμβος έχει 20% πιθανότητα να έχει μία τουλάχιστον σύνδεση συντόμευσης με κάποιον άλλο κόμβο του δικτύου. Σαν αποτέλεσμα, οι κόμβοι οι οποίοι αποτελούν τα άκρα ακμών συντομεύσεων είναι σημαντικοί για τη ροή της πληροφορίας στο δίκτυο, όταν αυτή πραγματοποιείται μέσω δρομολόγησης συντομότερων μονοπατιών. Το Σχήμα 25 απεικονίζει την τιμή της ΚΦΚ του γράφου μικρής-κλίμακας που κατασκευάσαμε με το μοντέλο των Watts-Strogatz.



Σχήμα 25: Τιμή της ΚΦΚ των κόμβων του γράφου μικρού-κόσμου

Επομένως, μία άπληστη ενσωμάτωση και δρομολόγηση τέτοιου τύπου κόμβων έχει ως αποτέλεσμα τη μεταφορά της συμφόρησης στον κόμβο-ρίζα και τους γείτονες του, με τους

κόμβους που λειτουργούν ως άκρα συντομεύσεων (κατά τον ορισμό των Watts-Strogatz) να έχουν επίσης μία υψηλή τιμή ΥΚΦΚ σε σύγκριση με τους υπόλοιπους κόμβους. Παραδείγματα αυτού φαίνονται στο επόμενο 2 σχήματα, όπου οι κόμβοι 37 και 83 λειτουργούν ως ρίζες των δένδρων ελάχιστου βάθους. Πέρα από τους 10 τοπολογικά γειτονικούς κόμβους (άρα γειτονικούς και σε επίπεδο δείκτη) του κόμβου 37, οι κόμβοι 4 και 50 είναι επίσης γείτονες του μέσω ακμών συντομεύσεων. Ο κόμβος 83 δεν είναι άκρο σε καμία ακμή-συντόμευση.



Σχήμα 26: ΥΚΦΚ με τους κόμβους 37 και 83 ως ρίζες

B. ΣΤΑΤΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ

Στο σημείο αυτό θα παρουσιαστούν διάφορα στατιστικά αποτελέσματα που έχουν να κάνουν με την επίδραση του επιπέδου που βρίσκεται ένας κόμβος του δικτύου στο συνδεδετικό δένδρο που ενσωματώνεται στον υπερβολικό χώρο. Συγκεκριμένα, για έναν κόμβο θα παρουσιαστούν η μέση τιμή της ΥΚΦΚ και της κατάταξης του με βάση την τιμή αυτή σε σχέση και με τους υπόλοιπους κόμβους του δικτύου, καθώς και η τυπική απόκλιση των τιμών αυτών.

• Γράφος Ελεύθερης Κλίμακας

Για την παρουσίαση των στατιστικών αποτελεσμάτων επιλέξαμε 10 κόμβους του δικτύου, συγκεκριμένα τους 5 κόμβους με την υψηλότερη τιμή ΚΦΚ στο δίκτυο και τους 5 κόμβους με τη χαμηλότερη τιμή ΚΦΚ, όταν η δρομολόγηση γίνεται με βάση τα συντομότερα μονοπάτια. Η μέση τιμή (Μ.Τ.) και η τυπική απόκλιση (Τ.Α.) των τιμών της ΥΚΦΚ, σε σχέση με το επίπεδο που βρίσκεται κάθε φορά ο κόμβος στο συνδεδετικό δένδρο, για αυτούς τους 10 κόμβους υπολογίζονται χρησιμοποιώντας όλους τους 100 κόμβους του δικτύου ως κόμβους-ρίζες. Τα αποτελέσματα για την κατάταξη των κόμβων βρίσκονται μέσα σε παρένθεση.

ΠΙΝΑΚΑΣ 6.3

ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΕΛΕΥΘΕΡΗΣ ΚΛΙΜΑΚΑΣ ΜΕ ΤΗΝ ΥΨΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ

	Ρίζα (Επίπεδο 1)		Επίπεδο 2		Επίπεδο 3		Επίπεδο 4	
	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.

Κόμβος 1	71.0862 (1)	-	60.3472 (1)	3.5264 (0)	6.1942 (43.0976)	5.7155 (27.7469)	-	-
Κόμβος 3	51.6531 (2)	-	40.4042 (1.1492)	11.5027 (1.1492)	19.9555 (13.2131)	10.5479 (18.8061)	4.6208 (29)	1.4455 (9.8995)
Κόμβος 2	47.7139 (2)	-	40.5535 (1.871)	7.6035 (0.7184)	16.5343 (27.8088)	11.9950 (35.0476)	-	-
Κόμβος 9	38.4937 (2)	-	25.2137 (4.25)	13.8678 (2.1491)	9.3290 (15.0694)	4.7086 (16.4277)	6.4221 (24.8571)	2.703 (17.2572)
Κόμβος 18	36.6327 (2)	-	21.2630 (6.4762)	11.5555 (4.8437)	5.9883 (19.6667)	3.5075 (10.3991)	9.8708 (20.667)	5.5794 (20.0366)

ΠΙΝΑΚΑΣ 6.4

ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΕΛΕΥΘΕΡΗΣ
ΚΛΙΜΑΚΑΣ ΜΕ ΤΗΝ ΧΑΜΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ

	Ρίζα (Επίπεδο 1)		Επίπεδο 2		Επίπεδο 3		Επίπεδο 4	
	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.
Κόμβος 35	30.3483 (3)	-	2.2820 (77.25)	0.1377 (8.7321)	2.6503 (62.6098)	0.4530 (15.8238)	2.6108 (62.8678)	0.3568 (12.6294)
Κόμβος 70	21.5322 (5)	-	2.3001 (80.5)	0.2125 (14.1951)	2.0894 (87.0704)	0.2388 (9.125)	2.2867 (84.3636)	0.3857 (16.1733)
Κόμβος 42	27.2024 (3)	-	2.7044 (55.1666)	0.1746 (11.2679)	3.1120 (50.8088)	1.2232 (17.9428)	2.7795 (55.2857)	0.3800 (17.4841)
Κόμβος 44	14.1023 (6)	-	2.6462 (59.25)	0.3007 (16.8992)	2.2267 (81.6508)	0.3715 (11.7558)	2.3435 (79.9375)	0.2680 (11.4411)
Κόμβος 84	21.6504 (4)	-	2.1494 (85.5)	0.0771 (3)	2.1851 (85.5091)	0.078 (7.3783)	2.2281 (84.625)	0.0882 (7.9926)

Από τα παραπάνω αποτελέσματα επιβεβαιώνεται πλήρως η εξάρτηση της δυνατότητας ενός κόμβου να βρίσκεται ως ενδιάμεσος στα άπληστα μονοπάτια ενός δικτύου, από την απόσταση του από τον κόμβο-ρίζα στο δένδρο και συνακόλουθα από τον αριθμό των δενδρικών-ακμών που χρησιμοποιούνται. Ειδικότερα, για τους κόμβους που παρουσιάζουν την υψηλότερη τιμή της ΚΦΚ βλέπουμε πως καθώς απομακρυνόμαστε από τον κόμβο-ρίζα του δένδρου (άρα καθώς πηγαίνουμε σε υψηλότερο επίπεδο του δένδρου και έχουμε λιγότερες δενδρικές ακμές να προσπίπτουν στον κόμβο αυτό), τόσο μειώνεται η τιμή της ΥΚΦΚ και η τιμή της κατάταξης τους. Στους κόμβους με τη χαμηλότερη τιμή της ΚΦΚ παρατηρούμε πως οι κόμβοι αυτοί έχουν υψηλή τιμή ΥΚΦΚ μόνο όταν βρίσκονται ως κόμβοι-ρίζα στο δίκτυο. Το γεγονός αυτό οφείλεται στο ότι οι κόμβοι αυτοί έχουν, πλέον, δενδρικές ακμές με όλους τους γείτονες τους, με την πιθανότητα να χρησιμοποιούνται αυτές οι ακμές κατά την άπληστη δρομολόγηση να είναι η μέγιστη (αφού παράλληλα, η πιθανότητα χρήσης μη-δενδρικών ακμών ελαχιστοποιείται κοντά στη ρίζα του δένδρου). Από την άλλη μεριά, για τους κόμβους υψηλού βαθμού και κεντρικότητας δεν παρουσιάζεται τέτοια απότομη πτώση, καθώς ο υψηλός βαθμός συνδεσιμότητας τους, τους επιτρέπει να διατηρούν μεγαλύτερο αριθμό δενδρικών ακμών σε υψηλότερο επίπεδο του συνδετικού δένδρου, ενώ υπάρχει και η πιθανότητα να χρησιμοποιηθούν και ως ακμές-συντομεύσεων, όπως δείχνει η εξίσωση 5.8.

- **Τυχαίος Γράφος**

Για τον συνθετικό τυχαίο γράφο επιλέγουμε, τυχαία, 5 κόμβους για τους οποίους θα παρουσιάσουμε τα στατιστικά τους αποτελέσματα. Αυτή τη φορά θα χρησιμοποιήσουμε, τυχαία, 15 κόμβους ως κόμβους-ρίζες του δένδρου ελάχιστου βάθους.

ΠΙΝΑΚΑΣ 6.5

ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ 5 ΤΥΧΑΙΑ ΕΠΙΛΕΓΜΕΝΟΥΣ ΚΟΜΒΟΥΣ ΤΟΥ ΤΥΧΑΙΟΥ ΓΡΑΦΟΥ

	Ρίζα (Επίπεδο 1)		Επίπεδο 2		Επίπεδο 3		Επίπεδο 4	
	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.	Μ.Τ	Τ.Α.
Κόμβος 10	55.0195 (1)	-	32.9520 (2.5)	5.9786 (0.7071)	7.2536 (36.0909)	5.8815 (23.5561)	3.9322 (38)	-
Κόμβος 16	46.8590 (1)	-	21.53 (4.5)	2.7945 (0.7071)	2.7178 (72.8889)	0.2756 (9.5058)	3.1594 (56.3333)	0.1670 (7.0238)
Κόμβος 43	42.9554	-	8.4530	0.9813	2.7623	0.4994	2.45499	0.2671

	(1)		(15)	(4.2426)	(73.5714)	(15.5976)	(83.4)	(12.3410)
Κόμβος 64	56.4509	-	10.7517	5.7801	4.4502	1.5648	3.783	0.8699
	(1)		(15.5)	(9.1924)	(42.4444)	(23.3030)	(50.6667)	(12.5831)
Κόμβος 91	38.5984	-	4.4430	0.5803	3.4356	0.8884	3.1860	0.07030
	(1)		(36.3333)	(4.0415)	(56.2857)	(22.2165)	(57.35)	(19.8053)

Στους τυχαίους γράφους, βλέπουμε πάλι πως επιβεβαιώνεται η υπόθεση ότι όσο απομακρυνόμαστε από τη ρίζα του συνδετικού δένδρου, τόσο μειώνεται η τιμή (και η κατάταξη των κόμβων) της ΥΚΦΚ. Επιπρόσθετα, οι κόμβοι σε αυτό το επίπεδο έχουν μία κατανομή κόμβων που ακολουθεί την κατανομή Poisson. Επομένως, οι βαθμοί των κόμβων κινούνται γύρω από μία μέση τιμή. Οπότε, η επίδραση του ύψους των κόμβων στην ενσωμάτωση τους είναι κυρίαρχη και δεν επηρεάζεται τόσο (όπως στους γράφους ελεύθερης-κλίμακας) και από το βαθμό των κόμβων, αφού όλοι οι κόμβοι έχουν-σε γενικές γραμμές- παραπλήσιο αριθμό διαθέσιμων ακμών, δηλαδή ακμών οι οποίες δεν είναι δενδρικές ακμές. Παράλληλα, αυτή η αύξηση της τιμής της ΥΚΦΚ στο δίκτυο για τους κόμβους που επιλέγονται ως ρίζα του συνδετικού δένδρου έχει ως αποτέλεσμα την μεταβολή στην κατανομή της κεντρικότητας των κόμβων στην άπληστη δρομολόγηση, όπως είδαμε και στο σχήμα 24.

- **Γράφος Μικρού Κόσμου**

Για τον γράφο μικρού κόσμου ακολουθήσαμε την ίδια τακτική που ακολουθήσαμε και για τον γράφο ελεύθερης-κλίμακας. Συγκεκριμένα, επιλέγουμε να παρουσιάσουμε τα στατιστικά αποτελέσματα για τους 5 κόμβους με την υψηλότερη τιμή ΚΦΚ και τους 5 κόμβους με τη χαμηλότερη τιμή της ΚΦΚ. Αυτή τη φορά, επιλέγουμε τυχαία 20 διαφορετικούς κόμβους ως κόμβους-ρίζες του δένδρου ελάχιστου βάθους.

ΠΙΝΑΚΑΣ 6.6

ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΜΙΚΡΟΥ ΚΟΣΜΟΥ ΜΕ ΤΗΝ ΥΨΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ

	Ρίζα (Επίπεδο 1)		Ρίζα (Επίπεδο 2)		Ρίζα (Επίπεδο 3)		Ρίζα (Επίπεδο 4)		Ρίζα (Επίπεδο 5)	
	M.T.	T.A.	M.T.	T.A.	M.T.	T.A.	M.T.	T.A.	M.T.	T.A.
Κόμβος 20	49.9079	-	36.6924	7.0909	26.1421	4.2189	5.6712	2.3131	8,4834	2,7205
	(1)		(2,3333)	(0,5774)	(4,5000)	(2,4290)	(46,4286)	(19,9571)	(28,6667)	(11,5902)
Κόμβος 49	48.9275	-	32.5441	10.2893	11.4624	10.0788	10.1205	8.4314	4,9741	0,8937

	(1)		(2,8000)	(1,6432)	(39,3333)	(35,2288)	(36,7500)	(36,1605)	(52,0000)	(15,6205)
Κόμβος 37	45.7329	-	39.1505	0	12.9623	12.8766	11.1975	5.5837	5,5837	3,4370
	(1)		(2,0000)	(0,0000)	(36,0000)	(26,2932)	(38,6667)	(31,0075)	(52,0000)	(32,5269)
Κόμβος 89	55.8353	-	-	-	16.1156	6.1690	6.4253	9.502	9,5020	4,9093
	(1)				(14,9000)	(15,5167)	(42,0000)	(23,1517)	(26,0000)	(18,3848)
Κόμβος 4	42.8387	-	33.9974	4.5441	32.5149	0	3.9562	4.8388	4,8388	1,5970
	(1)		(2,3333)	(0,5774)	(2,0000)	(0)	(67,3636)	(19,8609)	(57,0000)	(19,9165)

ΠΙΝΑΚΑΣ 6.7

ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ 5 ΚΟΜΒΟΥΣ ΤΟΥ ΓΡΑΦΟΥ ΜΙΚΡΟΥ ΚΟΣΜΟΥ ΜΕ ΤΗ ΧΑΜΗΛΟΤΕΡΗ ΤΙΜΗ ΚΦΚ

	Ρίζα (Επίπεδο 1)		Ρίζα (Επίπεδο 2)		Ρίζα (Επίπεδο 3)		Ρίζα (Επίπεδο 4)		Ρίζα (Επίπεδο 5)	
	M.T.	T.A	M.T.	T.A	M.T.	T.A.	M.T.	T.A.	M.T.	T.A.
Κόμβος 58	35,5874	-	10,0333	1,5645	3,9658	0,8018	4,5261	1,0731	3,5859	0,6828
	(1)		(19,5000)	(3,5355)	(66,0000)	(15,0333)	(57,5000)	(11,5016)	(70,6667)	(8,3267)
Κόμβος 46	48.9275	-	8,9485	1,9644	3,8413	1,0482	3,7991	0,4557	3,6389	0,2450
	(1)		(21,7500)	(6,6521)	(69,0000)	(18,5068)	(63,5000)	(8,2158)	(72,0000)	(1,8257)
Κόμβος 55	34,3568	-	10,2470	0	6,1347	2,7254	6,1563	4,7039	6,1725	2,2718
	(3)		(20,0000)	(0)	(47,2500)	(29,6009)	(55,8000)	(36,7859)	(46,7500)	(25,7860)
Κόμβος 45	31,5894	-	8,1890	0,9132	4,4311	1,5761	2,9511	1,0704	3,9491	0,3068
	(1)		(24,7500)	(5,1881)	(60,6000)	(23,9019)	(80,6667)	(20,6171)	(65,0000)	(3,5590)
Κόμβος 48	30,4828	-	7,4990	2,2674	3,6537	1,4335	4,9858	3,4861	4,7785	0,1604
	(2)		(33,0000)	(12,7279)	(69,5000)	(27,9052)	(60,0000)	(33,8009)	(50,2500)	(4,0311)

Στους γράφους μικρού-κόσμου και πάλι η κατανομή του βαθμού των κόμβων κινείται γύρω από μία μέση τιμή. Επομένως, μπορούμε να εξάγουμε παρόμοια συμπεράσματα με αυτά της περίπτωσης τυχαίων γράφων, με τη διαφορά πως αυτή τη φορά οι κόμβοι που στην δρομολόγηση συντομότερων μονοπατιών παρουσιάζουν υψηλότερη τιμή ΚΦΚ, λόγω της φύσης τους ως άκρα των ακμών-συντομεύσεων (όπως αυτές ορίστηκαν από τον αλγόριθμο Watts-Strogatz) αποτελούν άκρα περισσότερων δενδρικών ακμών απ' ότι οι κόμβοι χαμηλής

ΚΦΚ. Για το λόγο αυτό και όταν βρίσκονται σε υψηλότερο, από τη ρίζα, επίπεδο στο δένδρο έχουν υψηλότερη τιμή από τους κόμβους της δεύτερης κατηγορίας.

- **Αλλαγή στην κατάταξη των κόμβων ανάλογα με την αλλαγή του κόμβου-ρίζας**

Τέλος, για να δειχθεί η αλλαγή στην κατάταξη των κόμβων με την αλλαγή του κόμβου-ρίζας, δημιουργήσαμε έναν άνω-τριγωνικό πίνακα, όπου για κάθε στοιχείο (i, j) του πίνακα σημειώσαμε τον αριθμό (από 0 έως 10) των κοινών κόμβων ανάμεσα στους 10 πρώτους σε κατάταξη ΥΚΦΚ κόμβους του δικτύου, όταν η ενσωμάτωση γίνεται χρησιμοποιώντας τον κόμβο i ως ρίζα του συνδετικού δένδρου, σε σύγκριση με την περίπτωση όπου η ενσωμάτωση γίνεται χρησιμοποιώντας τον κόμβο j ως ρίζα του συνδετικού δένδρου.

Για τον **γράφο ελεύθερης-κλίμακας** η σύγκριση έγινε χρησιμοποιώντας και τους 100 κόμβους ως ρίζες του συνδετικού δένδρου.

Η μέση τιμή του πίνακα ήταν 4.5709 με τυπική απόκλιση 2.0171.

Για τον **τυχαίο γράφο** η σύγκριση έγινε χρησιμοποιώντας τυχαία 15 κόμβους ως ρίζες.

Η μέση τιμή του πίνακα ήταν 1.6571 με τυπική απόκλιση 1.6320.

Τέλος, για τον **γράφο μικρού-κόσμου** η σύγκριση έγινε χρησιμοποιώντας τυχαία 25 κόμβους ως ρίζες.

Η μέση τιμή του πίνακα ήταν 1.9444 με τυπική απόκλιση 1.8134.

Και στις 3 περιπτώσεις παρατηρείται μία μικρή μέση τιμή με συγκριτικά μικρή τυπική απόκλιση, που υποδηλώνουν μία σταθερή αλλαγή της κατάταξης των κόμβων, ως προς την τιμή της ΥΚΦΚ τους. Το φαινόμενο είναι πιο έντονο στον τυχαίο γράφο και στο γράφο μικρού-κόσμου, λόγω και του πιο κατανεμημένου φόρτου κίνησης στο δίκτυο. Για τον γράφο ελεύθερης κλίμακας η μέση αυτή τιμή είναι υψηλότερη για τρεις λόγους. i) Λόγω της ύπαρξης ενός στενά συνδεδεμένου πυρήνα απ' όπου διέρχεται η κίνηση, με έναν μικρό αριθμό κόμβων να έχουν αρκετά υψηλότερη τιμή κεντρικότητας από τους υπόλοιπους κόμβους οδηγούμαστε ακόμη και σε περιπτώσεις όπου η μείωση της τιμής της κεντρικότητας ενός κόμβου κατά μεγάλο ποσοστό να μην αποτυπώνεται και σε αλλαγή στην κατάταξη των κόμβων. ii) Από το γεγονός ότι οι κεντρικοί αυτοί κόμβοι σπάνια βρίσκονται σε μεγάλη απόσταση από οποιονδήποτε κόμβο του δικτύου ώστε να βρίσκονται ταυτόχρονα σε υψηλό επίπεδο στο δένδρο που ενσωματώνεται στον υπερβολικό χώρο. iii) Οι κόμβοι μεγάλου βαθμού αυτών των δικτύων, ακόμη και όταν τοποθετούνται σε μεγάλη απόσταση από τη ρίζα και προσπίπτουν σε αυτούς λίγες δενδρικές-ακμές (σε σύγκριση με το βαθμό τους) μπορούν να συμμετέχουν στη δρομολόγηση μέσω ακμών-συντομεύσεων (βλ. Εξίσωση 5.8). Παρ' όλα αυτά ακόμη και σε

αυτό τον τύπο των δικτύων υπάρχει μεγάλη αλλαγή στην κατάταξη των κόμβων, ενώ σε περιπτώσεις όπως αυτή που δείχθηκε στο παράδειγμα του σχήματος 19 μπορούμε να αποσυμφορήσουμε πλήρως την κίνηση από τους πιο κεντρικούς κόμβους του δικτύου.

7 *Επίλογος*

Στο κεφάλαιο αυτό συνοψίζονται τα αποτελέσματα της διπλωματικής εργασίας, καθώς και τα συμπεράσματα που προέκυψαν από την χρήση της ενσωμάτωσης των κόμβων σύνθετων δικτύων στον υπερβολικό χώρο, για τον υπολογισμό μετρικών κεντρικότητας και την μεταβολή των τιμών κεντρικότητας των κόμβων ανάλογα με τη θέση τους στο συνδετικό δένδρο μίας άπληστης ενσωμάτωσης, κάτω από καθεστώς άπληστης δρομολόγησης. Τέλος, δίνονται κάποιες ιδέες και κατευθύνσεις για μελλοντική επέκταση της διπλωματικής εργασίας.

7.1 *Σύνοψη και Συμπεράσματα*

Στο πλαίσιο της διπλωματικής εργασίας, αρχικά επιλέχθηκε η ενσωμάτωση των κόμβων σύνθετων δικτύων στον υπερβολικό χώρο με χρήση της ενσωμάτωσης Rigel. Η ενσωμάτωση αυτή αναθέτει στους κόμβους του δικτύου συντεταγμένες στον υπερβολικό χώρο, τέτοιες ώστε η απόσταση των κόμβων στον υπερβολικό χώρο, να προσεγγίζει την απόσταση των κόμβων σε βήματα σε συνθήκες δρομολόγησης συντομότερων μονοπατιών. Αξιοποιώντας, το γεγονός ότι η γνώση των αποστάσεων των κόμβων μπορεί να υπολογιστεί σε σταθερό χρόνο (και άρα δεν απαιτείται η εξαντλητική εύρεση των συντομότερων μονοπατιών μεταξύ κάθε ζεύγους κόμβων) προτάθηκε η χρήση του αλγόριθμου της Υπερβολικής Κεντρικότητας Φόρτου Κίνησης, η οποία είναι μία παραλλαγή του αλγορίθμου για τον υπολογισμό της Κεντρικότητας Φόρτου Κίνησης. Ο αλγόριθμος αυτός μπορεί να υπολογίσει, σε μικρότερο υπολογιστικό χρόνο από τον 2^ο αλγόριθμο, την κεντρικότητα των κόμβων ενός δικτύου σε καθεστώς άπληστης δρομολόγησης. Παράλληλα, για γράφους που η δομή τους ακολουθεί το μοντέλο της ελεύθερης-κλίμακας, ο αλγόριθμος της Υπερβολικής Κεντρικότητας Φόρτου Κίνησης, μας παρέχει προσεγγιστικά πολύ κοντινά αποτελέσματα σε σχέση με τον αρχικό αλγόριθμο για τον υπολογισμό της Κεντρικότητας Φόρτου Κίνησης.

Στη συνέχεια, επιλέχθηκε για την ενσωμάτωση των κόμβων σύνθετων δικτύων, το μοντέλο της άπληστης ενσωμάτωσης. Στόχος της ενσωμάτωσης αυτή τη φορά δεν είναι η διατήρηση των αποστάσεων μεταξύ των κόμβων και στον υπερβολικό χώρο, αλλά η διασφάλιση 100% επιτυχίας στην εύρεση ενός προορισμού υπό καθεστώς άπληστης δρομολόγησης. Στο πλαίσιο της διπλωματικής εργασίας δείχθηκε ότι η θέση ενός κόμβου στο συνδετικό δένδρο που επιλέγεται για την ενσωμάτωση (σε σχέση με την απόσταση του από τη ρίζα του δένδρου του δένδρου ελάχιστου βάθους, γεγονός που επηρεάζει και τον αριθμό των προσπίπτοντως σε αυτόν δενδρικών ακμών) έχει σημαντική επίδραση στη ροή της πληροφορίας διαμέσου αυτού

του κόμβου και άρα στην κεντρικότητά του. Επομένως, προτάθηκε ένα πλαίσιο ενεργειών που μπορεί να χρησιμεύσει για την αποσυμφόρηση της κίνησης από συγκεκριμένους κόμβους και τη μεταφορά της προς άλλους κόμβους του δικτύου.

7.2 Μελλοντικές Επεκτάσεις

Το έργο της παρούσας διπλωματικής εργασίας μπορεί να επεκταθεί και στις εξής ακόλουθες ενδεικτικές, αλλά συγκεκριμένες, κατευθύνσεις:

- Να εξεταστεί η ικανότητα αποδοτικής χρήσης των συντεταγμένων των κόμβων σε υπερβολικό χώρο για τον υπολογισμό και άλλων μετρικών κεντρικότητας.
- Οι τοπολογίες που εξετάστηκαν παραπάνω είναι στατικές. Θα μπορούσε, επομένως, να εξεταστεί σε μελλοντική εργασία η δυνατότητα ανανέωσης, χωρίς να χρειάζεται υπολογισμός από την αρχή, των τιμών κεντρικότητας των κόμβων, καθώς κόμβοι προστίθενται ή αφαιρούνται σε ένα εξελισσόμενο δίκτυο.
- Με παρεμφερή τρόπο με την ανάθεση συντεταγμένων στους κόμβους, θα μπορούσε, πιθανώς, να χρησιμοποιηθεί και η διαμέριση Voronoi [26] για τον υπολογισμό συντομότερων μονοπατιών και αποστάσεων σε ένα γράφημα.
- Για τον τύπο της άπληστης ενσωμάτωσης, αξιοποιώντας την ισομετρία των διάφορων μοντέλων υπερβολικού χώρου, θα μπορούσε να εξεταστεί και η υιοθέτηση ενός άλλου μοντέλου του υπερβολικού χώρου για την ανάθεση συντεταγμένων στους κόμβους, ώστε να αποφευχθούν τα υπολογιστικά προβλήματα ακρίβειας που εμφανίστηκαν στην άπληστη ενσωμάτωση στο μοντέλο του δίσκου του Poincare.

Βιβλιογραφία

- [1] Karyotis, V., Stai, E., & Papavassiliou, S. (2013). *Evolutionary Dynamics of Complex Communications Networks*. CRC Press.
- [2] Diestel, R. (2012). *Graph theory*. Springer.
- [3] Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.
- [4] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *nature*, 393(6684), 440-442.
- [5] Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2), 163-177.
- [6] Brandes, U., & Pich, C. (2007). Centrality estimation in large networks. *International Journal of Bifurcation and Chaos*, 17(07), 2303-2318.
- [7] Geisberger, R., Sanders, P., & Schultes, D. (2008, January). Better approximation of betweenness centrality. In *Proceedings of the Meeting on Algorithm Engineering & Experiments* (pp. 90-100). Society for Industrial and Applied Mathematics.
- [8] Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2), 136-145.
- [9] Li, L., Alderson, D., Doyle, J. C., & Willinger, W. (2005). Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, 2(4), 431-523.
- [10] E. Stai, V. Karyotis, G.Katsinis, E.Tsiropoulou, S.Papavassiliou. A Hyperbolic Big Data Analytics Framework within Complex and Social Networks. (to appear)
- [11] Anderson, J. W. (2007). *Hyperbolic geometry. 2nd Edition*. Springer
- [12] Zhao, X., Sala, A., Zheng, H., & Zhao, B. Y. (2011, October). Efficient shortest paths on massive social graphs. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2011 7th International Conference on* (pp. 77-86). IEEE.
- [13] Cvetkovski, A., & Crovella, M. (2009, April). Hyperbolic embedding and routing for dynamic graphs. In *INFOCOM 2009, IEEE* (pp. 1647-1655). IEEE.

- [14] Zhao, X., Sala, A., Wilson, C., Zheng, H., & Zhao, B. Y. (2010). Orion: shortest path estimation for large social graphs. *networks*, 1, 5.
- [15] Papadopoulos, F., Krioukov, D., Bogua, M., & Vahdat, A. Greedy Forwarding in Dynamic Scale-Free Networks Embedded in Hyperbolic Metric Spaces. In *2010 Proceedings IEEE INFOCOM*.
- [16] Ban, X., Gao, J., & van de Rijt, A. (2010, January). Navigation in real-world complex networks through embedding in latent spaces. In *Proceedings of the Meeting on Algorithm Engineering & Experiments* (pp. 138-148). Society for Industrial and Applied Mathematics.
- [17] Kleinberg, R. (2007, May). Geographic routing using hyperbolic space. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE* (pp. 1902-1909). IEEE.
- [18] Dolev, S., Elovici, Y., & Puzis, R. (2010). Routing betweenness centrality. *Journal of the ACM (JACM)*, 57(4), 25.
- [19] Papadopoulos, F., Psomas, C., & Krioukov, D. (2015). Network mapping by replaying hyperbolic growth. *IEEE/ACM Transactions on Networking (TON)*, 23(1), 198-211.
- [20] Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005, August). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 177-187). ACM.
- [21] Leskovec, J., & Krevl, A. (2014). SNAP Datasets: Stanford large network dataset collection, June 2014. URL <http://snap.stanford.edu/data>.
- [22] Thomas H.. Cormen, Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (Vol. 6). Cambridge: MIT press.
- [23] Cvetkovski, A., & Crovella, M. (2013). On the choice of a spanning tree for greedy embedding of network graphs. *Networking Science*, 3(1-4), 2-12.
- [24] <https://gmplib.org/>
- [25] <http://www.mpfr.org/>
- [26] Honiden, S., Houle, M. E., Sommer, C., & Wolff, M. (2010). Approximate shortest path queries using Voronoi duals. In *Transactions on computational science IX* (pp. 28-53). Springer Berlin Heidelberg.