

Ανάλυση, Μοντελοποίηση και Προσομοίωση Αναζήτησης σε Σύνθετα Δυναμικά Δίκτυα

Η Διδακτορική Διατριβή

υποβάλλεται στην ορισθείσα

από τη Γενική Συνέλευση Ειδικής Σύγκλησης
του Τμήματος Μηχανικών Η/Υ και Πληροφορικής
Εξεταστική Επιτροπή

από την

Σπυριδούλα Μαργαρίτη

ως μέρος των υποχρεώσεων για την απόκτηση του

ΔΙΔΑΚΤΟΡΙΚΟΥ ΔΙΠΛΩΜΑΤΟΣ

Πανεπιστήμιο Ιωαννίνων

Μάιος 2017

Συμβουλευτική Επιτροπή:

- **Βασίλειος Δημακόπουλος**, Αναπληρωτής Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Ευαγγελία Πιτουρά**, Καθηγήτρια, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Παναγιώτα Φατούρου**, Αναπληρώτρια Καθηγήτρια, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Εξεταστική Επιτροπή:

- **Βασίλειος Δημακόπουλος**, Αναπληρωτής Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Ευαγγελία Πιτουρά**, Καθηγήτρια, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Παναγιώτα Φατούρου**, Αναπληρώτρια Καθηγήτρια, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
- **Νικόλαος Μαμουλής**, Αναπληρωτής Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Παναγιώτης Τσαπάρης**, Αναπληρωτής Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Κωνσταντίνος Μαγκούτης**, Επίκουρος Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Κωνσταντίνος Οικονόμου**, Αναπληρωτής Καθηγητής, Τμήμα Πληροφορικής, Ιόνιο Πανεπιστήμιο

ΑΦΙΕΡΩΣΗ

Στην οικογένειά μου.

ΕΥΧΑΡΙΣΤΙΕΣ

Φθάνοντας στο τέλος μιας μακρόχρονης προσπάθειας, θα ήθελα να εκφράσω την ειλικρινή μου ευγνωμοσύνη σε όσους συνέβαλαν και συνέδραμαν στην ολοκλήρωση της παρούσας διατριβής.

Πρωτίστως, θα ήθελα να ευχαριστήσω τον επιβλέποντα της διατριβής μου, Αναπληρωτή Καθηγητή κ. Βασίλειο Δημακόπουλο, για την αμέριστη υποστήριξη, την ερευνητική καθοδήγηση, την πολύτιμη συμπαράσταση και συνεργασία που μου παρείχε όλα αυτά τα χρόνια. Τον ευχαριστώ, για τα διορατικά σχόλια και την εποικοδομητική κριτική που έπαιξαν καθοριστικό ρόλο στην ολοκλήρωση αυτής της διατριβής. Ευχαριστώ επίσης, την Καθηγήτρια κ. Ευαγγελία Πιτουρά και την Αναπληρώτρια Καθηγήτρια κ. Παναγιώτα Φατούρου, για τη συμμετοχή τους στην συμβουλευτική επιτροπή, για τις εύστοχες παρατηρήσεις τους και για τον πολύτιμο χρόνο που μου διέθεσαν.

Επίσης, θα ήθελα να εκφράσω θερμές ευχαριστίες στα μέλη της εξεταστικής επιτροπής της διατριβής, κκ. Νικόλαο Μαμουλή, Παναγιώτη Τσαπάρα, Κωνσταντίνο Μαγκούτη και Κωνσταντίνο Οικονόμου για την τιμή που μου κάνουν να μετέχουν σε αυτή.

Τέλος, οφείλω ένα θερμό ευχαριστώ στους δικούς μου ανθρώπους, στο σύζυγό μου Δημήτρη και στα παιδιά μου Αγγελική και Κωνσταντίνα για τη συνεχή αγάπη, στήριξη και ενδυνάμωση, στα εύκολα και τα δύσκολα αυτής της διαδρομής.

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Σχημάτων	iv
Κατάλογος Πινάκων	ix
Περίληψη	x
Extended Abstract	xii
1 Εισαγωγή	1
1.1 Το πρόβλημα της αναζήτησης	1
1.2 Στόχοι της διατριβής	4
1.3 Συνεισφορά της διατριβής	5
1.4 Δομή της διατριβής	6
2 Στοιχεία σύνθετων δικτύων - επισκόπηση βιβλιογραφίας	8
2.1 Βασικές έννοιες	8
2.1.1 Μετρικές δικτύων	10
2.2 Σύνθετα δίκτυα του πραγματικού κόσμου	13
2.2.1 Δίκτυα φυσικής υποδομής	13
2.2.2 Δίκτυα διάθεσης πληροφορίας και πόρων	14
2.2.3 Κοινωνικά δίκτυα	16
2.2.4 Φυσικά δίκτυα	17
2.3 Μοντέλα σύνθετων δικτύων	17
2.3.1 Στατικά μοντέλα	18
2.3.2 Δυναμικά μοντέλα	21
2.4 Δυναμικές διαδικασίες σε σύνθετα δίκτυα	24
2.4.1 Αναζήτηση	24
2.4.2 Άλλες διαδικασίες	24

3	Αναζήτηση σε σύνθετα δίκτυα	26
3.1	Γενικά	26
3.2	Επισκόπηση βιβλιογραφίας	29
3.3	Στρατηγικές αναζήτησης: λύσεις και προκλήσεις	33
3.3.1	Αποτίμηση αναζήτησης	34
3.3.2	Κριτήρια τερματισμού	36
3.4	Βελτίωση αναζήτησης	36
3.4.1	Αλλαγή τοπολογίας	37
3.4.2	Αντιγραφή (Replication)	38
3.5	Σύνοψη	40
4	Αναζήτηση με πλημμύρα σε αδόμητα δίκτυα: το πρόβλημα της πλεονάζουσας πληροφορίας	41
4.1	Εισαγωγή	41
4.2	Αδόμητα δίκτυα $p2p$ & πλημμύρα	44
4.3	Γενικά όρια επί των αναπόφευκτων διπλότυπων μηνυμάτων	48
4.4	Μοντελοποίηση απόδοσης πλημμύρας με χρήση της θεωρίας παιγνίων	52
4.5	Μοντέλο εκτίμησης κάλυψης του δικτύου	54
4.6	Επεκτείνοντας το μοντέλο	57
4.6.1	Δίκτυα με δεδομένο συντελεστή ομαδοποίησης	57
4.6.2	Δίκτυα power-law	59
4.7	Σύνοψη	60
5	APF: Μια νέα πιθανοτική στρατηγική πλημμύρας	62
5.1	Εισαγωγή	62
5.2	Επισκόπηση βιβλιογραφίας	64
5.3	Μοντέλο και υποθέσεις	67
5.3.1	Το δίκτυο	67
5.3.2	Η διαδικασία αναζήτησης	67
5.4	APF: Μια νέα πιθανοτική στρατηγική αναζήτησης	68
5.4.1	Εκτίμηση κάλυψης του δικτύου	70
5.5	Αξιολόγηση	72
5.5.1	Τυχαίοι γράφοι	73
5.5.2	Εφαρμογές σε άλλα δίκτυα	78
5.6	Σύνοψη	82

6	Προσομοίωση πρωτοκόλλων αναζήτησης σε δυναμικά δίκτυα μεγάλης κλίμακας	83
6.1	Εισαγωγή	83
6.2	Επισκόπηση βιβλιογραφίας	85
6.3	Το δικό μας πλαίσιο προσομοίωσης: Armonia	90
6.3.1	Η τοπολογία	90
6.3.2	Οι πόροι και τα αντίγραφα τους	92
6.3.3	Το πρωτόκολλο αναζήτησης	94
6.3.4	Στατιστικά	96
6.4	Ενδεικτική μελέτη προσομοίωσης και συζήτηση	96
6.5	Συμπεράσματα	101
7	Αναζήτηση σε δυναμικά δίκτυα μεγάλης κλίμακας	103
7.1	Εισαγωγή	103
7.2	Επισκόπηση βιβλιογραφίας	106
7.3	Μεθοδολογία και μοντέλο συστήματος	107
7.3.1	Μέσος βαθμός κόμβων σε απόσταση t	110
7.4	Προσαρμοζόμενη Advanced Probabilistic Search (A2PF)	113
7.5	Αξιολόγηση	118
7.6	Σύνοψη	125
8	Συμπεράσματα και Μελλοντική Εργασία	127
8.1	Συμπεράσματα	127
8.2	Μελλοντική εργασία	129
	Βιβλιογραφία	132

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

2.1	Η μαθηματική θεώρηση μικρών δικτύων: (α) με απλό γράφο, (β) με κατευθυνόμενο γράφο και (γ) με απλό γράφο με παράλληλες ακμές (οι 7 γέφυρες του Königsberg ¹).	9
2.2	Κατανομές Poisson και scale-free.	12
2.3	Κατανομή scale free.	13
3.1	Οι βασικές συνιστώσες της αναζήτησης.	27
4.1	Παράδειγμα πλημμύρας σε ένα απλό δίκτυο και τα παραγόμενα μηνύματα.	45
4.2	(α) Πλήθος μηνυμάτων που παράγει η απλή πλημμύρα με ή χωρίς μηχανισμό ανίχνευσης διπλότυπων μηνυμάτων κατά μέσο όρο σε κάθε βήμα, (β) οι κόμβοι που προσπελάστηκαν από το μήνυμα-ερώτηση και τα παραγόμενα μηνύματα με χρήση DDM.	48
4.3	Αποφυγή διπλότυπων μηνυμάτων σε γειτονικούς κόμβους.	50
4.4	(α) Μεταδόσεις μηνυμάτων και αριθμός κόμβων που έχουν επισκεφτεί από το μήνυμα-ερώτηση της αναζήτησης (β) αναπόφευκτα διπλότυπα μηνύματα, σε ένα τακτικό δίκτυο με βαθμό 5 των 100.000 κόμβων.	51
4.5	Παίγνιο χρήστη και τυχαίου δικτύου: (α) απόδοση παικτών και σημείο ισορροπίας και (β) κόστος αναζήτησης σε διπλότυπα μηνύματα (dups) που αυξάνεται μετά το $t = 6$ βήματα που αντιστοιχεί στο σημείο ισορροπίας.	54
4.6	Παίγνιο χρήστη και δικτύου Barabasi: (α) απόδοση παικτών και σημείο ισορροπίας και (β) κόστος αναζήτησης σε διπλότυπα μηνύματα (dups) που αυξάνεται μετά το $t = 6$ βήματα που αντιστοιχεί στο σημείο ισορροπίας.	54

4.7	Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο τακτικό δίκτυο με βαθμό 5 των 100.000 κόμβων.	57
4.8	Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο τακτικό δίκτυο με βαθμό 6 των 100.000 κόμβων.	58
4.9	Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε τυχαίο δίκτυο των 100.000 κόμβων, με $\bar{d} = 7$	58
4.10	Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε τυχαίο δίκτυο 100.000 κόμβων, με $\bar{d} = 7$ και συντελεστή ομαδοποίησης $cc=0.09$	59
4.11	Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο 100.000 κόμβων, με $\bar{d} = 7$ και συντελεστή ομαδοποίησης $cc=0.14$	60
4.12	Οι κόμβοι που έχει επισκεφθεί η αναζήτηση (α) και διπλότυπα μηνύματα (β) σε ένα δίκτυο με κατανομή power-law (μοντέλο barabasi) των 100.000 κόμβων, με $m = 2$	61
5.1	Κάλυψη δικτύου σε ένα τυχαίο τακτικό δίκτυο με βαθμό 6 των 100.000 κόμβων.	73
5.2	Πιθανότητα προώθησης για διαφορετικές τιμές δημοτικότητας ενός πόρου σε ένα τυχαίο τακτικό δίκτυο με βαθμό 6 των 100.000 κόμβων.	74
5.3	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο 100.000 κόμβων με αντιγραφή proportional και με κυμαινόμενη δημοτικότητα πόρου ($0.0005 \leq q \leq 0.21948$).	75
5.4	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο ($\bar{d} = 5$) 100.000 κόμβων με $q = 0.0005$	76
5.5	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο 6-τακτικό δίκτυο 100.000 κόμβων και $q = 0.0003$	77
5.6	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο 100.000 κόμβων, $\bar{d} = 7$ και συντελεστή ομαδοποίησης $cc = 0.14$	77

5.7	(α) Διπλά μηνύματα κατά την διάρκεια αναζήτησης σε δίκτυο BA 100.000 κόμβων, $\bar{d} = 4$ και $q = 0.0001$. (β) Μέσος βαθμός κόμβων σε απόσταση t από τον αρχικό κόμβο σε διάφορες τοπολογίες.	79
5.8	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα δίκτυο BA power-law 100.000 κόμβων με $q = 0.001$ αντίγραφα.	80
5.9	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα BA power-law δίκτυο 100.000 κόμβων και $q = 0.0005$	80
5.10	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε στιγμιότυπο gnutella των 26.518 κόμβων και $q = 0.0037$ αντίγραφα.	81
5.11	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε στιγμιότυπο του slashdot των 82.168 κόμβων και $q = 0.00006$ αντίγραφα. . . .	81
6.1	Ένας κόμβος του δικτύου που αντιστοιχεί σε μια κορυφή του γράφου.	88
6.2	Ένα στιγμιότυπο μιας προσομοίωσης με την Armonia: το δίκτυο OM-NeT++ (αριστερά) και η αντίστοιχη τοπολογία (δεξιά).	89
6.3	Η δομή του κάθε κόμβου, όπως υλοποιείται από το Armonia.	92
6.4	Η χρήση των δομικών στοιχείων protocol και resource στην Armonia.	92
6.5	Armonia: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα στατικό δίκτυο ER 25.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0012$	98
6.6	Armonia: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα στατικό δίκτυο BA 10.000 κόμβων, με $m = 2$ και $q = 0.0015$	99
6.7	Armonia: Ένα δείγμα ενεργών κόμβων για τα πρώτα 100sec της προσομοίωσης με το Armonia (αριστερά) και του αριθμού των μηνυμάτων (δεξιά) που παράγει η πλημμύρα σε ένα δυναμικό δίκτυο ER με 5.000 αρχικούς κόμβους.	101
6.8	Armonia: Ένα δείγμα ενεργών κόμβων για τα πρώτα 100sec της προσομοίωσης με το Armonia (αριστερά) και του αριθμού των μηνυμάτων (δεξιά) που παράγει η πλημμύρα σε ένα δυναμικό δίκτυο BA με 10.000 αρχικούς κόμβους· στη γραφική παράσταση απεικονίζονται ο συνολικός αριθμός των μηνυμάτων και των διπλότυπων μηνυμάτων.	102

7.1	Μεταβολή πλήθους κόμβων σε δίκτυο Barabasi με churn, με διάφορους συνδυασμούς διάρκειας ζωής και αφίξεων για αρχικό πλήθος κόμβων 10.000.	109
7.2	Ο μέσος βαθμός των κόμβων σε απόσταση ακριβώς t βημάτων, όπως υπολογίζεται σε συλλογές δεδομένων από πραγματικά δίκτυα. Αντίστοιχες μετρήσεις για συνθετικά δίκτυα βρίσκονται στο Σχ. 5.7. . . .	112
7.3	Η δομή του δικτύου όπως την αντιλαμβάνεται ο κόμβος u από την πληροφορία που συλλέγει από τα εισερχόμενα μηνύματα-ερωτήσεις q_m	113
7.4	Ο μέσος βαθμός των κόμβων σε απόσταση t , όπως υπολογίζεται από διαφορετικούς κόμβους ενός δυναμικού δικτύου ER με αρχικό πλήθος 1.000 κόμβων.	117
7.5	Αριθμητικός μέσος όρος των εκτιμήσεων των κόμβων από το Σχ. 7.4 (καμπύλη estimated) και πραγματικός μέσος βαθμός όπως μετρήθηκε για το ίδιο δυναμικό δίκτυο ER (καμπύλη measured).	117
7.6	Ο μέσος βαθμός των κόμβων σε απόσταση t σε δίκτυο Barabasi 10.000 κόμβων χωρίς churn.	118
7.7	Αριθμητικός μέσος όρος των εκτιμήσεων των κόμβων από το Σχ. 7.6 (καμπύλη estimated) και πραγματικός μέσος βαθμός όπως μετρήθηκε για το ίδιο δυναμικό δίκτυο BA (καμπύλη measured).	118
7.8	Ο μέσος βαθμός των κόμβων σε απόσταση t σε δίκτυο Barabasi με αρχικό πλήθος 1.000 κόμβων.	119
7.9	Αριθμητικός μέσος όρος των εκτιμήσεων των κόμβων από το Σχ. 7.8 (καμπύλη estimated) και πραγματικός μέσος βαθμός όπως μετρήθηκε για το ίδιο δυναμικό δίκτυο BA (καμπύλη measured).	119
7.10	Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο ER 30.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0006$ και με ρυθμό churn 14,5%.	121
7.11	Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο ER 30.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0006$ και με ρυθμό churn 45%.	121
7.12	Συγκριτικά αποτελέσματα αναζήτησης με υψηλό και χαμηλό ρυθμό churn (14,5% και 45% αντίστοιχα) σε δυναμικό δίκτυο ER με αρχικό πλήθος 30.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0006$	122

7.13	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα δίκτυο BA power-law με σταθερό πλήθος 30.000 κόμβων με $m = 2$, $q = 0.0016$.	123
7.14	Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) στο πραγματικό δίκτυο email-Enron 36.692 κόμβων και 183.831 ακμών, $q = 0.0054$ αντίγραφα.	123
7.15	Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο BA power-law 20.000 κόμβων, με $m = 2$ και $q = 0.005$ και με ρυθμό churn 14%.	124
7.16	Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο BA power-law 20.000 κόμβων, με $m = 2$ και $q = 0.005$ και με ρυθμό churn 45%.	124
7.17	Συγκριτικά αποτελέσματα αναζήτησης με υψηλό και χαμηλό ρυθμό churn (14, 5% και 45% αντίστοιχα) σε δυναμικό δίκτυο BA power-law με αρχικό πλήθος 20.000 κόμβων, με $m = 2$ και $q = 0.005$	125

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

5.1	Παράμετροι προσομοίωσης τυχαίου δίκτυο ER 100.000 κόμβων. . . .	75
5.2	Παράμετροι προσομοίωσης άλλων δικτύων	79
6.1	Παράμετροι προσομοίωσης για στατικά δίκτυα	98
6.2	Παράμετροι προσομοίωσης για δυναμικά δίκτυα	100
7.1	Συνοπτικός πίνακας συμβόλων που χρησιμοποιούνται σε αυτή την ενότητα.	105

ΠΕΡΙΛΗΨΗ

Σπυριδούλα Β. Μαργαρίτη, Δ.Δ., Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Μάιος 2017.

Ανάλυση, Μοντελοποίηση και Προσομοίωση Αναζήτησης σε Σύνθετα Δυναμικά Δίκτυα.

Επιβλέπων: Βασίλειος Δημακόπουλος, Αναπληρωτής Καθηγητής.

Στον πραγματικό κόσμο συναντάμε πληθώρα σύνθετων συστημάτων, που προέρχονται από διαφορετικούς τομείς (της επιστήμης, της φύσης, της τεχνολογίας, αλλά και της κοινωνίας), σε διάφορες κλίμακες και με ποικίλους σχηματισμούς. Τα συστήματα αυτά αναφύονται και εξελίσσονται ακολουθώντας απλούς κανόνες που καθορίζονται από τις αλληλεπιδράσεις μεταξύ των οντοτήτων που τα απαρτίζουν. Οι εγγενείς διαδικασίες σχηματισμού ή/και μεταβολής τους, οι δυναμικές λειτουργίες που εκτυλίσσονται σε αυτά και ο τεράστιος όγκος πληροφορίας που απαιτείται για τον προσδιορισμό τους, τα καθιστούν ιδιαίτερα πολύπλοκα. Για την περιγραφή τους, την αναπαραγωγή τους και την ποσοτικοποίησή τους χρησιμοποιούμε εργαλεία από την επιστήμη των δικτύων, τα οποία μας επιτρέπουν να κατανοήσουμε και να προβλέψουμε την συμπεριφορά τους αλλά και τη συμπεριφορά των δυναμικών διαδικασιών που εξελίσσονται σε αυτά, όπως η αναζήτηση.

Η αναζήτηση είναι ένα από τα σημαντικότερα ζητήματα στα σύνθετα δίκτυα με πρακτική εφαρμογή σε διάφορα πλαίσια, όπως, για παράδειγμα, την εύρεση ενός στόχου ή την αναζήτηση ενός προσώπου σε ένα κοινωνικό δίκτυο ή τον εντοπισμό ενός πόρου σε ένα δίκτυο ομότιμων (peer-to-peer) ή την ανάκτηση των δεδομένων που αποθηκεύονται σε μεγάλες ποσότητες στους κόμβους των δικτύων ή την εύρεση ενός δρομολογητή σε ένα δίκτυο επικοινωνίας. Η αναζήτηση είναι μια δυναμική διαδικασία που εξελίσσεται σε ένα δίκτυο μεταδίδοντας μηνύματα από κόμβο σε κόμβο. Όταν υπάρχει συνολική γνώση του δικτύου, η αναζήτηση στρέφεται στην εύρεση της καταλληλότερης διαδρομής με βάση κάποιο κριτήριο (π.χ. κόστος με-

τάβασης σε hops). Αντίθετα, η παντελής έλλειψη πληροφορίας σχετικά με τη δομή του δικτύου, έχει ως αποτέλεσμα η διαδικασία αναζήτησης να διεξάγεται τυφλά ή να βασίζεται μόνο σε τοπική πληροφορία.

Στόχος αυτής της διατριβής είναι η σχεδίαση και ανάπτυξη μοντέλων και τεχνικών που θα συμβάλουν σε δύο κατευθύνσεις: α) να διερευνήσουμε τη διαδικασία της αναζήτησης αναλυτικά και πειραματικά και β) να εξετάσουμε την εφαρμοσιμότητά τους σε σύνθετα δίκτυα. Εστιάζουμε σε αδόμητα δίκτυα, όπου οι συνδέσεις μεταξύ των κόμβων είναι αποτέλεσμα κανόνων τυχειότητας, στατικά ή δυναμικά και οι τεχνικές αναζήτησης βασίζονται στην πλημμύρα.

Αρχικά, εξετάζουμε και αναλύουμε τη συμπεριφορά της αναζήτησης και παρέχουμε γενικά όρια και προσεγγιστικά μοντέλα για την εκτίμηση της μη αναγκαίας παραγόμενης κίνησης (διπλότυπα μηνύματα) και την κάλυψη του δικτύου. Θέτοντας ως πρωταρχικό στόχο την εξάλειψη της πλεονάζουσας πληροφορίας που οφείλονται στην τοπολογία του δικτύου και στην ύπαρξη πολλαπλών μονοπατιών μεταξύ των κόμβων, προτείνουμε δύο πιθανοτικές στρατηγικές αναζήτησης που αξιοποιούν χαρακτηριστικά και μετρικές του δικτύου, όπως για παράδειγμα, το πλήθος των συνδέσεων του κάθε κόμβου (βαθμός) ή τον αριθμητικό μέσο των βαθμών. Η απόφαση ενός κόμβου για να μεταδώσει ένα μήνυμα-ερώτηση που έλαβε, βασίζεται στη δημοτικότητα των αντικειμένων και σε μια εκτίμηση για την κάλυψη του δικτύου. Ο κάθε κόμβος, πριν μεταδώσει, προσαρμόζει την πιθανότητα προώθησης, έτσι ώστε να μειώνονται τα διπλότυπα μηνύματα ενώ η πιθανότητα επιτυχίας παραμένει υψηλή.

Για να μελετήσουμε την αναζήτηση θέτουμε ένα πλαίσιο που περιλαμβάνει: την τοπολογία του δικτύου, τα προς αναζήτηση στοιχεία και τον μηχανισμό αναζήτησης. Σχεδιάζουμε και αναπτύσσουμε ένα πρωτότυπο, γενικού σκοπού προσομοιωτή, που τον ονομάσαμε Armonia. Παρέχει υλοποιημένα σύνολα, που έχουν προταθεί κατά καιρούς στη βιβλιογραφία, για διάφορα στατικά και δυναμικά μοντέλα δικτύων, διαφορετικές τεχνικές αναζήτησης και στρατηγικές εκχώρησης αντιγράφων.

Τέλος, χρησιμοποιούμε το πλαίσιο προσομοίωσης Armonia για την αξιολόγηση των προτεινόμενων τεχνικών. Εφαρμόζουμε τις στρατηγικές μας σε στατικά και δυναμικά δίκτυα που παράγουμε χρησιμοποιώντας είτε τυπικά μοντέλα δικτύων είτε διαθέσιμα σύνολα δεδομένων από πραγματικά δίκτυα. Συγκριτικά πειραματικά αποτελέσματα των προτεινόμενων τεχνικών μας με άλλες παρόμοιες στρατηγικές επιβεβαιώνουν την αποτελεσματικότητά τους.

EXTENDED ABSTRACT

Spiridoula V. Margariti, Ph.D., Department of Computer Science and Engineering,
University of Ioannina, Greece, May 2017.

Analysis, Modeling and Simulation of Search in Complex Dynamic Networks.

Advisor: Vassilios Dimakopoulos, Associate Professor.

Complex networks are present everywhere in the real world, coming from different fields (science, nature, technology, society), in various scales and with a variety of forms. These systems arise and evolve by following simple rules defined through the interactions between the participating entities. They are considered complex because of the complicated mechanics of their creation and evolution, because of the dynamic processes that take place within them and because of the huge amount of information required to identify them. Tools from network science are used to describe, reproduce and quantify them and also, to help us to understand and predict their behavior as well as the behavior of the dynamic processes that occur within them.

Search is one of the fundamental issues in complex networks with practical application in various contexts, such as finding a target or locating a person in a social network or discovering a desired resource. Search is a dynamic process that unfolds by transmitting messages from node to node. When there is global knowledge of the network, search is focused on finding the most appropriate route based on some criterion (eg. transition costs). In the absence of information about the network structure, search in these systems is a blind procedure, possibly using only local information.

The aim of this dissertation is to design and develop models and techniques that will contribute in two directions: a) towards the study of the search process both analytically and experimentally, and b) towards investigating their applicability in complex networks. We focus on static or dynamic unstructured networks where connections between nodes are governed by random rules, and on search mechanisms based on flooding.

We start by studying and analyzing the behavior of flooding with respect to unnecessarily produced traffic (duplicate messages) and provide simple bounds and approximate models to assess the associated overheads and network coverage. Our primary objective is the elimination of overheads that emanate from the overlay network topology itself and the multiplicity of paths among nodes. We then propose novel probabilistic search algorithms that exploit network features and metrics, such as the nodes degrees. The decision of a node to propagate a message (or not) is based on both the popularity of resources and an estimation of network coverage. Based on these parameters we adjust the forwarding probability at the time a node receives the query message so as to reduce the duplicate message overhead while maintaining a high probability of query success.

To study the search problem experimentally, we consider a comprehensive framework which includes the network topology, the resources and the searching strategies. We design and develop a novel, general-purpose simulator, called Armonia. It provides implementations for a wide suite of models and algorithms found in the literature related to network topology generation, search strategies and resources allocation and placement policies.

Finally, we use our simulation framework to evaluate our proposed algorithms. We apply our strategies in static and dynamic networks using well-known network models and data sets coming from real networks. The observed results confirm the effectiveness of our mechanisms and their superiority with respect to the other known strategies.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

-
- 1.1 Το πρόβλημα της αναζήτησης
 - 1.2 Στόχοι της διατριβής
 - 1.3 Συνεισφορά της διατριβής
 - 1.4 Δομή της διατριβής
-

1.1 Το πρόβλημα της αναζήτησης

Σύνθετα συστήματα είτε φυσικά είτε τεχνητά αναφύονται παντού, από τον επιστημονικό και τεχνολογικό χώρο ως το φυσικό και κοινωνικό κόσμο. Απαρτίζονται από ένα σύνολο δυναμικών οντοτήτων με διακριτή ταυτότητα και συμπεριφορά όπως άνθρωποι, πακέτα πληροφορίας, ενέργεια, κύτταρα, που αυτοδιοργανώνονται και αλληλεπιδρούν στενά μεταξύ τους. Είναι συχνά μεγάλης κλίμακας, χαρακτηρίζονται από αυθαίρετη δόμηση (τοπολογία) δυναμικά εξελισσόμενη στο χρόνο, ενώ κοινοί κανόνες διέπουν τις διεργασίες που λαμβάνουν χώρα σε αυτά (π.χ. αναζήτηση). Ονομάζονται σύνθετα (complex), καθώς είναι “απελπιστικά πολύπλοκα” [18] για δύο λόγους [63]: για τις ιδιότητες που αναδύονται από τις αλληλεπιδράσεις των μελών τους και για την ποσότητα των πληροφοριών που απαιτείται για να προσδιοριστούν.

Ο πιο απλός τρόπος αναπαράστασης αυτών των συστημάτων είναι ένας γράφος ή δίκτυο, ένα σύνολο δηλαδή από σημεία που συνδέονται μεταξύ τους με γραμμές ή ακμές. Το κάθε σημείο αποτελεί έναν κόμβο του δικτύου και μπορεί να είναι

οτιδήποτε: είτε ένα άτομο σε ένα κοινωνικό δίκτυο, είτε μια συσκευή σε ένα δίκτυο επικοινωνιών είτε ένα από αντικείμενο σε ένα ηλεκτρικό δίκτυο ή σε ένα οδικό δίκτυο είτε μια πρωτεΐνη σε ένα βιολογικό δίκτυο. Οι γραμμές που συνδέουν τους κόμβους απεικονίζουν τη μεταξύ τους σχέση όπως, για παράδειγμα, τη γνωριμία δύο ατόμων σε ένα κοινωνικό δίκτυο, την απευθείας σύνδεση μεταξύ δυο συσκευών σε ένα δίκτυο επικοινωνιών ή την παραπομπή σε ένα δίκτυο βιβλιογραφικών αναφορών. Το δίκτυο, ως περιγραφικός τρόπος αναπαράστασης συμβάλει όχι μόνο στην κατανόηση ενός σύνθετου συστήματος αλλά και στην παραπέρα μελέτη και διερεύνησή του σε δύο επίπεδα [61]: το πρώτο αφορά τη διασύνδεση των οντοτήτων και το δεύτερο την μεταξύ τους αλληλεπίδραση.

Στα σύνθετα συστήματα συντελείται πλήθος δραστηριοτήτων και παρέχεται πληθώρα υπηρεσιών. Για να μπορούν οι οντότητες να επωφεληθούν από αυτές τις υπηρεσίες πρέπει πρώτα να τις εντοπίσουν και να τις ανακτήσουν μέσω της αναζήτησης. Η αναζήτηση είναι η προσεκτική και διεξοδική εξέταση για την εύρεση ή τον εντοπισμό ενός στόχου, όπως για παράδειγμα η αναζήτηση ενός προσώπου σε ένα κοινωνικό δίκτυο ή ο εντοπισμός ενός πόρου σε ένα δίκτυο ομότιμων (peer-to-peer) ή εύρεση ενός δρομολογητή σε ένα δίκτυο επικοινωνίας. Όταν υπάρχει συνολική γνώση του δικτύου, η αναζήτηση στρέφεται στην εύρεση της καταλληλότερης διαδρομής με βάση κάποιο κριτήριο (π.χ. κόστος μετάβασης). Αντίθετα, η παντελής έλλειψη πληροφορίας έχει ως αποτέλεσμα η διαδικασία αναζήτησης να διεξάγεται τυφλά ή να βασίζεται μόνο σε τοπική πληροφορία.

Αποτελεί πραγματική πρόκληση να μελετηθεί η συμπεριφορά της αναζήτησης και να αποκαλυφθούν εκείνοι οι μηχανισμοί που συμβάλουν στην αποτελεσματική (εύρεση λύσης) και στην αποδοτική έκβασή της. Η παρούσα μελέτη επιχειρεί να δώσει απάντηση σε αυτή την πρόκληση μέσα από:

- **Ανάλυση:** Η ανάλυση των σύνθετων δικτύων, η τυπική περιγραφή και η απόδοση των χαρακτηριστικών τους βασίζεται στην θεωρία των γράφων [178]. Όμως, το μέγεθος των δικτύων, η αυθαίρετη δόμηση και η εμπειρική φύση τους [18] απαιτεί μια διαφορετική προσέγγιση. Σήμερα, είναι δυνατή η συλλογή και η επεξεργασία πληθώρας στατιστικών δεδομένων από τις αλληλεπιδράσεις που συμβαίνουν σε ένα δίκτυο.

Ο συνδυασμός της θεωρίας των γράφων και της στατιστικής μηχανικής αποτελεί εργαλείο που αποκαλύπτει τα δομικά μοτίβα και τις ιδιότητες του δικτύου

δίνοντας ουσιαστικά εικόνα και γνώση για το δίκτυο. Αυτό αλλάζει τον τρόπο κατανόησης, προσέγγισης και πρόβλεψης της συμπεριφοράς δυναμικών διεργασιών, όπως η αναζήτηση, που εκτυλίσσονται στο δίκτυο. Η ανάλυσή μας εστιάζει στην εύρεση στατιστικών ιδιοτήτων που χαρακτηρίζουν τη συμπεριφορά της αναζήτησης και που πιθανόν θα βοηθήσουν στον έλεγχό της.

- Μοντελοποίηση: Αρκετά μοντέλα έχουν προταθεί για την περιγραφή των δικτύων του πραγματικού κόσμου. Παρ' όλο που αυτή η περιγραφή είναι αφαιρετική—διαφορετικά θα απαιτούνταν ένας τεράστιος όγκος δεδομένων—υπάρχουν δομικές ιδιότητες που αναδεικνύονται [178], [63] και που χωρίς αυτές είναι δύσκολο να κατανοηθεί ο ακριβής τρόπος δόμησης των δικτύων. Για παράδειγμα, η δομή του διαδικτύου, η δομή ενός κοινωνικού δικτύου και η δομή ενός βιολογικού δικτύου παρουσιάζουν ομοιότητες.

Εδώ, χρησιμοποιούμε προτεινόμενα μοντέλα για τη δημιουργία τοπολογιών και αναπτύσσουμε μοντέλα που εξηγούν τη διαδικασία εξέλιξης της αναζήτησης, προσαρμόζονται σε πραγματικά δίκτυα και βοηθούν στην κατανόηση του μηχανισμού αναζήτησης και την εξερεύνηση της συμπεριφοράς της και των κανόνων/παραδοχών που συμβάλουν στην αποτελεσματική διεξαγωγή της. Επιπλέον, αξιοποιούμε τα μοντέλα για να εκτιμήσουμε και να προβλέψουμε τη συμπεριφορά της αναζήτησης. Τέλος, με βάση τα μοντέλα αυτά αναπτύσσουμε εξελιγμένους μηχανισμούς αναζήτησης που αποδεικνύεται ότι έχουν καλύτερη συμπεριφορά από άλλους γνωστούς μηχανισμούς.

- Προσομοίωση: Η προσομοίωση είναι απαραίτητη για δύο λόγους: α) είναι ένας τρόπος για την απόκτηση γνώσεων για τη συμπεριφορά της αναζήτησης, ένας τρόπος για να ελεγχθούν και να αξιολογηθούν δικές μας προτεινόμενες στρατηγικές και για να συγκριθούν τα παραγόμενα αποτελέσματα με άλλα σχετικά εμπειρικά ή βιβλιογραφικά αποτελέσματα, β) παρέχει τη δυνατότητα πραγματοποίησης μετρήσεων και πειραμάτων σε ευρύ φάσμα και σε συστήματα διαφορετικής κλίμακας. Αυτό είναι σχεδόν αδύνατο για ένα πραγματικό δίκτυο αφού δεν μπορεί εύκολα να προσπελαστεί και να γίνει διεξαγωγή πειραμάτων λόγω της έκτασης και του τεράστιου όγκου δεδομένων.

1.2 Στόχοι της διατριβής

Στην αυγή του 21ου αιώνα, αυξημένο και διεπιστημονικό είναι το ενδιαφέρον για τα σύνθετα δίκτυα. Κοινή επιδίωξη όλων, είναι να μελετήσουν το ίδιο το δίκτυο ή/και τις διαδικασίες που συμβαίνουν πάνω σε αυτό. Παρά τη διαφορετική προέλευση των σύνθετων δικτύων, η τοπολογία τους διέπεται από κοινούς κανόνες και αρχές. Η κατανόηση της τοπολογίας είναι ζωτικής σημασίας για την πρόβλεψη της συμπεριφοράς και τη βελτιστοποίηση των δυναμικών διαδικασιών. Σε αυτή τη διατριβή, στόχος μας είναι να διερευνήσουμε τη διαδικασία της αναζήτησης αναλυτικά και πειραματικά αλλά και να προτείνουμε αποδοτικούς αλγορίθμους ικανούς να εφαρμοστούν σε σύνθετα δίκτυα. Οι παραδοσιακές προσεγγίσεις, βασιζόμενες κυρίως στην πλημμύρα, προωθούν το μήνυμα-ερώτηση σε όλους τους κόμβους του δικτύου με αποτέλεσμα την παραγωγή ενός τεράστιου όγκου μηνυμάτων που κατακλύζει το δίκτυο, δρα ανασταλτικά στην απόδοση της αναζήτησης, σπαταλά τους πόρους του δικτύου και οδηγεί σε συμφόρηση.

Για να εξετάσουμε αυτά τα ζητήματα, θεωρούμε ένα αφαιρετικό μοντέλο που περιλαμβάνει τη δομή του δικτύου, τα περιεχόμενά του (πόρους) και τις αλληλεπιδράσεις τους. Εστιάζουμε σε δυο κατηγορίες δικτύων: σε αυτά που οι συνδέσεις μεταξύ των κόμβων τοποθετούνται τυχαία [62], και σε αυτά που ταυτοποιούνται από μια βασική αρχή οργάνωσης σχηματίζοντας δίκτυα με μικρό μέσο αριθμό συνδέσεων ανά κόμβο ενώ υπάρχει ένας μικρός αριθμός κόμβων με πολύ μεγάλο αριθμό συνδέσεων [19]. Στα δίκτυα αυτά, δεν υπάρχει καθολική πληροφόρηση και ο κάθε κόμβος γνωρίζει και αλληλεπιδρά μόνο με τους κόμβους με τους οποίους έχει απευθείας συνδέσεις. Αυτό είναι αρκετό, για να διεξαχθεί μια διαδικασία όπως η αναζήτηση που ξεκινώντας από έναν κόμβο και ακολουθώντας τις συνδέσεις εκτυλίσσεται στο αόρατο δίκτυο.

Αρχικά, εξετάζουμε γενικά την αναζήτηση με απλή πλημμύρα και στη συνέχεια επικεντρωνόμαστε στην πιθανοτική αναζήτηση. Σε αυτή την περίπτωση ουσιαστική πρόκληση αποτελεί η εύρεση (προσδιορισμός) της κατάλληλης πιθανότητας προώθησης του μηνύματος από κόμβο σε κόμβο, προκειμένου να αντεπεξέλθει στους περιορισμούς της πλημμύρας. Επιδίωξή μας είναι η παραγωγή μιας έκφρασης που θα εκτιμά την πιθανότητα προώθησης σε κάθε κόμβο αξιοποιώντας τις ιδιότητες του δικτύου και άλλες τοπικές πληροφορίες. Εστιάζουμε σε στατικές και δυναμικές τοπολογίες δικτύων και για να ολοκληρώσουμε την ανάλυσή μας υλοποιούμε

ένα πλαίσιο προσομοίωσης προκειμένου να εξετάσουμε, να συγκρίνουμε και να αξιολογήσουμε τις δικές μας και άλλες στρατηγικές αναζήτησης.

1.3 Συνεισφορά της διατριβής

Αυτή η διατριβή εστιάζει στο πρόβλημα της αναζήτησης σε σύνθετα δυναμικά και συνεισφέρει από διαφορετικά οπτικά πεδία καθώς: α) παρέχουμε μια συνολική θεώρηση των σύνθετων δικτύων και των μοντέλων που τα περιγράφουν προκειμένου να κατανοήσουμε την επίδραση της τοπολογίας τους στη διαδικασία της αναζήτησης, β) προτείνουμε τη δημιουργία μιας πρωτότυπης πιθανοτικής στρατηγικής αναζήτησης που αξιοποιεί τυπικές παραμέτρους του δικτύου (π.χ. πλήθος συνδέσεων ενός κόμβου) και δίνει ουσιαστική λύση στα προβλήματα της αναζήτησης γ) τροποποιούμε την προηγούμενη στρατηγική προκειμένου να προσαρμόζεται σε διαφορετικές κατηγορίες στατικών και δυναμικών δικτύων και γ) σχεδιάζουμε και υλοποιούμε έναν πρωτότυπο προσομοιωτή, που το ονομάζουμε *Armonia*, και χρησιμεύει στην αξιολόγηση των στρατηγικών αναζήτησης.

Παρά το μεγάλο ενδιαφέρον για την αναζήτηση, η προηγούμενη έρευνα εστιάζει σε στατικά δίκτυα και επιχειρεί να μετριάσει τα προβλήματα της αναζήτησης. Αντίθετα, εμείς εξετάζουμε τη συμπεριφορά της σε στατικά και σε δυναμικά δίκτυα με τυχαίους κανόνες δόμησης, εξηγούμε πως ακριβώς προκύπτουν τα προβλήματα, τα ποσοτικοποιούμε και προτείνουμε μηχανισμούς που σχεδόν εξαλείφουν κάποια από αυτά. Επιπλέον αναπτύξαμε ένα πρωτότυπο σύστημα προσομοίωσης, που σχεδιάστηκε για να υποστηρίξει τους χρήστες στη διεξαγωγή πειραμάτων. Παρέχει υλοποιημένα σύνολα από διαφορετικά μοντέλα δικτύων, πολιτικές εκχώρησης πόρων και διάφορες στρατηγικές αναζήτησης και διευκολύνει τη διεξαγωγή πειραμάτων, τη συλλογή και τη σύγκριση αποτελεσμάτων.

Συνοπτικά η συνεισφορά αυτής της διατριβής έχει ως εξής:

- υπολογίζουμε αναλυτικά όρια σε γενικά δίκτυα θεωρώντας τυφλή αναζήτηση που διεξάγεται σε διαδοχικά βήματα για:
 - το πλήθος των διαφορετικών κόμβων που λαμβάνουν το μήνυμα ερώτηση,
 - το πλήθος των διπλότυπων μηνυμάτων που παράγονται

σε κάθε βήμα,

- προτείνουμε ένα προσεγγιστικό μοντέλο που υπολογίζει το πλήθος των διαφορετικών κόμβων (που λαμβάνουν το μήνυμα-ερώτηση για πρώτη φορά), των μηνυμάτων (συνολικά μηνύματα) και των διπλότυπων μηνυμάτων σε κάθε βήμα της αναζήτησης.
- σχεδιάζουμε και αναπτύσσουμε έναν μηχανισμό πιθανοτικής αναζήτησης, τον οποίο ονομάζουμε APF, που βελτιώνει την απόδοσή της αναζήτησης καθώς: α) αποκόπτει τις περιττές διαδρομές της απλής πλημμύρας (overshooting), β) μειώνει αισθητά τον αριθμό των διπλότυπων μηνυμάτων, γ) προσαρμόζει την επέκτασή της στο δίκτυο,
- σχεδιάζουμε και αναπτύσσουμε ένα πρωτότυπο, γενικού σκοπού προσομοιωτή που παρέχει υλοποιημένα σύνολα για: διάφορα στατικά και δυναμικά μοντέλα δικτύων, διαφορετικές τεχνικές αναζήτησης και στρατηγικές εκχώρησης αντιγράφων.
- εισάγουμε μια νέα διάσταση στο μήνυμα-ερώτηση της αναζήτησης με την προσθήκη ενός πεδίου, το οποίο πληροφορεί για τον βαθμό του κόμβου που θέτει το ερώτημα και το αξιοποιούμε στην υλοποίηση ενός τοπικού μηχανισμού που “μαθαίνει” την τοπολογία του δικτύου από τα μηνύματα που λαμβάνει,
- βελτιώνουμε τη στρατηγική αναζήτησης APF κάνοντας χρήση του προαναφερμένου μηχανισμού ώστε να προσαρμόζεται σε δυναμικά ετερογενή δίκτυα.

1.4 Δομή της διατριβής

Η διατριβή οργανώνεται σε οκτώ κεφάλαια. Στα κεφάλαια 2 και 3 εισάγονται βασικές έννοιες και ζητήματα που αφορούν την ανάλυση και τη μοντελοποίηση της αναζήτησης. Στα υπόλοιπα κεφάλαια προτείνουμε τεχνικές και μοντέλα που δίνουν λύσεις στα σημαντικότερα ζητήματα της αναζήτησης. Συγκεκριμένα:

Στο 2ο κεφάλαιο παρουσιάζονται οι βασικές έννοιες από τη θεωρία των γραφών που χρησιμοποιούνται για να περιγράψουν χαρακτηριστικά του δικτύου και οι μετρικές για την ποσοτικοποίηση των στατιστικών ιδιοτήτων του. Επίσης δίνεται μια κατηγοριοποίηση των πραγματικών δικτύων, περιγράφονται τα κυριότερα μοντέλα που τα αναπαριστούν καθώς και οι δυναμικές διαδικασίες που συμβαίνουν σε

αυτά. Το 3ο κεφάλαιο εστιάζει στις στρατηγικές αναζήτησης που έχουν προταθεί, επισημαίνοντας τις προκλήσεις σχετικά με την απόδοση και τον έλεγχό τους. Επίσης συνοψίζονται προσεγγίσεις για τη βελτίωση της απόδοσής της, όπως αλλαγές στην τοπολογία του δικτύου και τεχνικές δημιουργίας αντιγράφων.

Στο 4ο κεφάλαιο παρουσιάζονται αναλυτικά τα προβλήματα της πλημμύρας και ένα μοντέλο που ποσοτικοποιεί κάποια από αυτά. Επίσης προτείνουμε ένα μοντέλο που κάνοντας χρήση της θεωρίας παιγνίων ελέγχει την εξάπλωση της πλημμύρας. Στο 5ο κεφάλαιο παρουσιάζεται η APF μια πρωτότυπη πιθανοτική στρατηγική αναζήτησης που αξιοποιεί τοπικά χαρακτηριστικά του δικτύου. Στο 6ο κεφάλαιο παρουσιάζεται το πρωτότυπο πλαίσιο προσομοίωσης Armonia και οι δυνατότητες του. Με τη βοήθειά του παράγονται ενδεικτικά συγκριτικά αποτελέσματα στρατηγικών αναζήτησης σε στατικά και δυναμικά δίκτυα. Στο 7ο κεφάλαιο εξετάζεται το πρόβλημα της αναζήτησης σε δυναμικά δίκτυα και αναπτύσσεται μια τροποποιημένη εκδοχή της APF που προσαρμόζεται σε δυναμικές τοπολογίες. Το 8ο και τελευταίο κεφάλαιο συνοψίζει τα αποτελέσματα της διατριβής και συζητά μελλοντική εργασία.

ΚΕΦΑΛΑΙΟ 2

ΣΤΟΙΧΕΙΑ ΣΥΝΘΕΤΩΝ ΔΙΚΤΥΩΝ - ΕΠΙΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ

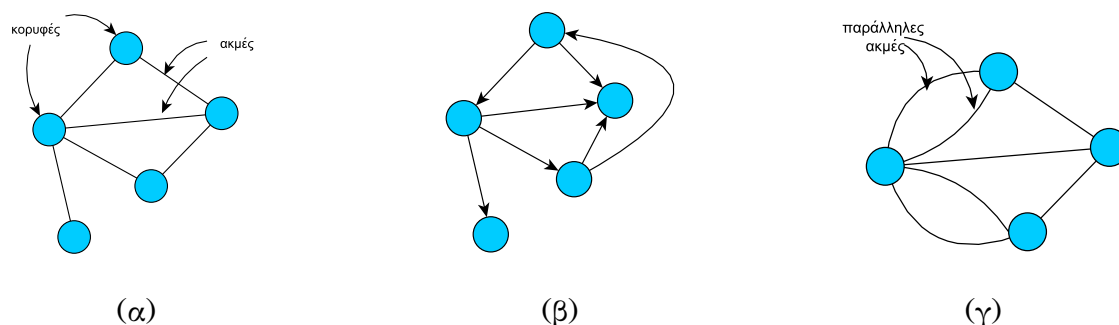
-
- 2.1 Βασικές έννοιες
 - 2.2 Σύνθετα δίκτυα του πραγματικού κόσμου
 - 2.3 Μοντέλα σύνθετων δικτύων
 - 2.4 Δυναμικές διαδικασίες σε σύνθετα δίκτυα
-

2.1 Βασικές έννοιες

Σε αυτή την ενότητα παρουσιάζονται βασικές έννοιες από τη θεωρία των γράφων, ικανές για την διατύπωση σχετικών όρων και χρήσιμες για την κατανόηση των ιδιοτήτων ενός σύνθετου δικτύου.

Ο γράφος αποτελεί τη μαθηματική θεώρηση για την αναπαράσταση ή τη μοντελοποίηση ενός σύνθετου δικτύου. Ως γράφος $G = (V, E)$ ορίζεται ένα πεπερασμένο μη κενό σύνολο V που περιέχει ως στοιχεία κορυφές ή κόμβους και ένα σύνολο E που περιέχει ζεύγη στοιχείων που σχετίζονται μεταξύ τους ή αλληλεπιδρούν. Κάθε στοιχείο $(u, v) \in E$, δηλαδή το μη διατεταγμένο ζεύγος των κορυφών u, v , καλείται ακμή και δηλώνει σχέση ή αλληλεπίδραση μεταξύ των κορυφών u, v .

Το πλήθος των κόμβων του συνόλου V συμβολίζεται με $N = |V|$ και το πλήθος των ακμών συμβολίζεται με $m = |E|$.



Σχήμα 2.1: Η μαθηματική θεώρηση μικρών δικτύων: (α) με απλό γράφο, (β) με κατευθυνόμενο γράφο και (γ) με απλό γράφο με παράλληλες ακμές (οι 7 γέφυρες του Königsberg¹).

Ένας γράφος λέγεται *μη κατευθυνόμενος* (undirected) αν τα ζεύγη των κορυφών που σχηματίζουν τις ακμές του δεν είναι διατεταγμένα, δηλαδή αν $(u, v) \in E$, τότε και $(v, u) \in E$. Ένας γράφος λέγεται *κατευθυνόμενος* (directed) αν οι ακμές του είναι διατεταγμένες ή προσανατολισμένες. *Πολλαπλές* (multiple) ή *παράλληλες* ακμές καλούνται οι ακμές που έχουν το ίδιο ζεύγος άκρων. *Βρόχος* (self loop) ονομάζεται η ακμή που αρχίζει και καταλήγει στην ίδια κορυφή, είναι δηλαδή της μορφής (u, u) .

Γειτονικοί (neighbors) είναι δύο κόμβοι $u, v \in V$ αν μεταξύ τους υπάρχει ακμή, δηλαδή αν $(u, v) \in E$.

Διαδρομή ή *μονοπάτι* (path) από τον κόμβο u στον κόμβο v είναι η ακολουθία από διαφορετικούς κόμβους που συνδέονται μεταξύ τους με διαφορετικές ακμές. Αν η διαδρομή αρχίζει και τελειώνει στον ίδιο κόμβο, είναι δηλαδή ένα κλειστό μονοπάτι λέγεται *κύκλος* (cycle). *Απόσταση* (distance) μεταξύ δύο κόμβων u, v καλείται το μήκος της συντομότερης διαδρομής που τους ενώνει. Δύο κόμβοι u, v είναι *συνδεδεμένοι* αν υπάρχει μονοπάτι μεταξύ u και v . Ένας γράφος G είναι *συνδεδεμένος* ή *συνεκτικός* (connected) αν κάθε ζεύγος κόμβων του γράφου είναι συνδεδεμένο. Αν κάθε ζεύγος κορυφών συνδέεται με μια ακμή τότε ο γράφος καλείται *πλήρης* (complete). Ένας πλήρης γράφος έχει $m = N(N - 1)/2$ ακμές. *Απλός* (simple) γράφος είναι ο γράφος που δεν έχει πολλαπλές ακμές ή βρόχους. *Δέντρο* (tree) είναι ένας συνδεδεμένος γράφος χωρίς κύκλους.

Τρίγωνο (triangle) είναι ένας πλήρης γράφος τριών κόμβων ή πιο απλά μια

¹Αποτελεί το γνωστό πρόβλημα του 18ου αιώνα για το αν μπορεί κανείς να διασχίσει τις 7 γέφυρες της πόλης χωρίς να περάσει δύο φορές από την ίδια γέφυρα. Την απάντηση έδωσε ο διάσημος μαθηματικός Euler.

τριάδα κόμβων που είναι γειτονικοί μεταξύ τους. Μια κορυφή συνδεδεμένη με άλλες δυο κορυφές σχηματίζει μια τριάδα (triple).

Βαθμός (degree) κόμβου: Σε έναν μη κατευθυνόμενο γράφο G καλείται βαθμός του κόμβου u και συμβολίζεται με d_u το πλήθος των ακμών που έχουν άκρο τον κόμβο u , δηλαδή $d_u = \#\{u \in V : (u, v) \in E\}$. Στα σύνθετα δίκτυα, ο βαθμός του κόμβου εκφράζει το πλήθος των συνδέσεων με άλλους κόμβους όπως, για παράδειγμα, το πλήθος των φίλων σε ένα κοινωνικό δίκτυο. Στα κατευθυνόμενα δίκτυα γίνεται διάκριση μεταξύ του έσω-βαθμού και του έξω-βαθμού. Ο έσω-βαθμός d_{in} είναι ο αριθμός των ακμών που καταλήγουν σε έναν κόμβο u . Ο έξω-βαθμός d_{out} εκφράζει τον αριθμό των ακμών που αναχωρούν από έναν κόμβο u . Ο μέσος βαθμός του γράφου G ορίζεται ως

$$\bar{d} = \frac{1}{N} \sum_{u \in V} d_u = \frac{2m}{N} \quad (2.1)$$

Αν όλοι οι κόμβοι του γράφου G έχουν τον ίδιο βαθμό d τότε ο γράφος ονομάζεται τακτικός (regular) και συμβολίζεται ως d -τακτικός.

2.1.1 Μετρικές δικτύων

Τα σύνθετα δίκτυα παρουσιάζουν ιδιαίτερα γνωρίσματα που έχουν να κάνουν με τις συσχετίσεις, τις δυναμικές αλληλεπιδράσεις ή τις συνεργασίες [200] μεταξύ των οντοτήτων που τα απαρτίζουν. Καθώς όμως, είναι σχεδόν αδύνατο η ανάλυση, η εξέταση και η ανασύνθεση αυτών των συστημάτων να βασιστεί στις τοπικές ιδιότητες των οντοτήτων, η έρευνα στράφηκε στη χρήση στατιστικών μετρικών [47], [63], [119]. Σήμερα υπάρχει η δυνατότητα για συλλογή και χειρισμό συνόλων δεδομένων από τέτοια δίκτυα, ικανά για να εκφράσουν και να ερμηνεύσουν τα ιδιαίτερα χαρακτηριστικά τους που αφορούν την τοπολογία ή την εξέλιξη τους. Οι κύριες υπάρχουσες μετρικές είναι η κατανομή βαθμού, ο συντελεστής ομαδοποίησης και δυναμική κίνηση των οντοτήτων (churn).

Κατανομή Βαθμών (Degree Distribution): Σε έναν μη κατευθυνόμενο γράφο G η κατανομή των βαθμών p_d ορίζεται ως το ποσοστό των κόμβων που έχουν βαθμό d ή ισοδύναμα, η τιμή της p_d εκφράζει την πιθανότητα ένας ομοιόμορφα επιλεγμένος κόμβος να έχει βαθμό d .

Η κατανομή των βαθμών χαρακτηρίζει τη δομή (τοπολογία) των σύνθετων δικτύων και επηρεάζει τη συμπεριφορά των δυναμικών διεργασιών, όπως για παράδειγμα η αναζήτηση, που διεξάγονται σε αυτά. Συνήθως, η κατανομή βαθμών απει-

κονίζεται σε μια γραφική παράσταση της p_d συναρτήσει του βαθμού d (Σχ. 2.2). Εναλλακτικά, αντί της p_d μπορεί να χρησιμοποιηθεί η συχνότητα των κόμβων που έχουν βαθμό d .

Στην πρόσφατη βιβλιογραφία διακρίνονται δύο κύριες κατηγορίες δικτύων που η τοπολογία τους περιγράφεται από την κατανομή βαθμών. Στην πρώτη κατηγορία ανήκουν τα δίκτυα που η κατανομή των βαθμών τους είναι κατανομή *Poisson*. Διαισθητικά αυτό σημαίνει ότι οι κόμβοι έχουν κατά μέσο όρο τον ίδιο αριθμό συνδέσεων. Στη δεύτερη κατηγορία εντάσσονται τα δίκτυα άνευ κλίμακας (*scale free*) με τυπικό παράδειγμα αυτό της κατανομής *power-law*. Πολλά δίκτυα του πραγματικού κόσμου εμφανίζουν κατανομή *power-law* όπου οι περισσότεροι κόμβοι έχουν μικρό αριθμό συνδέσεων και υπάρχουν λίγοι κόμβοι με μεγάλο αριθμό συνδέσεων.

Κατανομή Poisson: Έστω p η πιθανότητα δύο κόμβοι να συνδέονται μεταξύ τους / να γειτονεύουν. Τότε, η συνολική πιθανότητα ένας κόμβος να συνδέεται με ακριβώς d από τους υπόλοιπους $N - 1$ κόμβους είναι

$$p_d = \binom{N-1}{d} p^d (1-p)^{N-1-d} \quad (2.2)$$

ακολουθεί δηλαδή τη διωνυμική κατανομή. Για μεγάλες τιμές του N , η p_d μπορεί να προσεγγισθεί από τη συνάρτηση πυκνότητας πιθανότητας της κατανομής *Poisson*

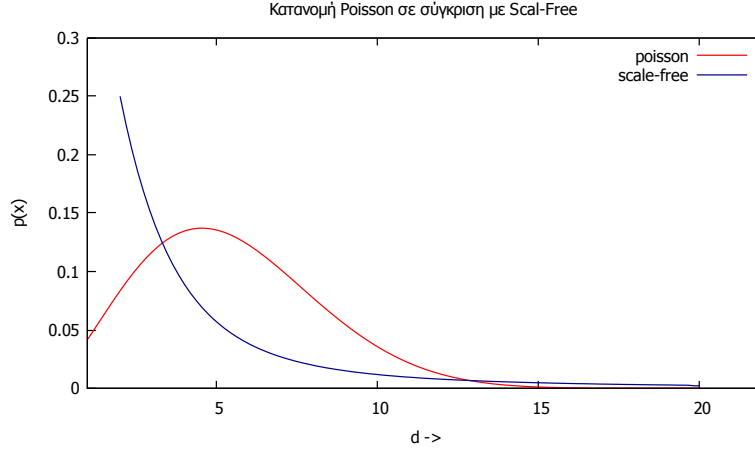
$$p_d = \frac{e^{-c} c^d}{d!} \quad (2.3)$$

όπου $c = Np$.

Κατανομή power-law: Μια κατανομή με συνάρτηση πυκνότητας πιθανότητας της μορφής

$$p_x = Cx^{-\gamma} \quad (2.4)$$

ονομάζεται κατανομή *power-law* [148], [139], [39], όπου η σταθερά γ αποτελεί τον εκθέτη της κατανομής *power-law* και η σταθερά C είναι παράγοντας κανονικοποίησης. Η κατανομή *power-law* είναι ιδιαίτερα γνωστή, αλλά και συνηθισμένη, στον πραγματικό κόσμο [39] και βρίσκει εφαρμογή μόνο για τιμές μεγαλύτερες μιας ελάχιστης τιμής (x_{min}) καθώς αποκλίνει για $x \rightarrow 0$. Όπως αναφέρθηκε και παραπάνω, η συγκεκριμένη κατανομή επιτρέπει σε πολύ λίγους κόμβους να έχουν πολύ μεγάλο βαθμό, δημιουργώντας ισχυρούς κεντρικούς κόμβους (*hubs*), ενώ οι υπόλοιποι και περισσότεροι έχουν πολύ μικρότερο βαθμό που συσσωρεύεται στην ουρά της κατανομής (*heavy tail*). Η κατανομή p_x , γραφικά αναπαρίσταται ως ευθεία με κλίση $-\gamma$



Σχήμα 2.2: Κατανομές Poisson και scale-free.

σε συνάρτηση με το βαθμό x και χρησιμοποιώντας λογαριθμική κλίμακα, δηλαδή εκφράζεται με τη μορφή

$$\ln p_x = -\gamma \ln x + C \quad (2.5)$$

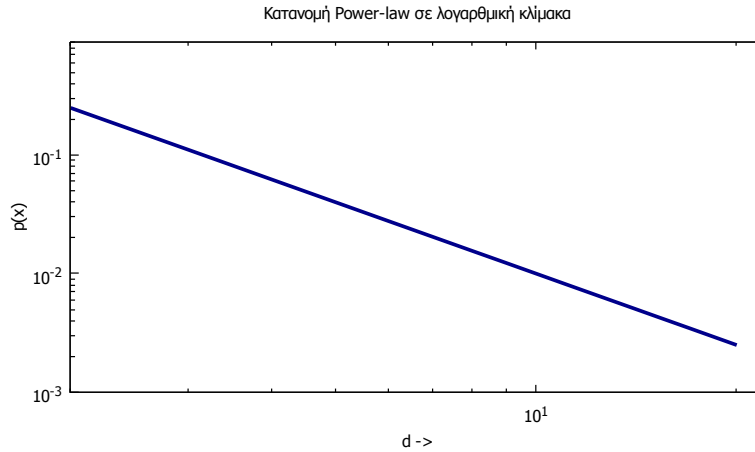
Μια τέτοια κατανομή απεικονίζεται στο Σχ. 2.3. Εκτός από το γνώρισμα της “βαριάς ουράς”, η κατανομή power-law έχει ένα ακόμη ιδιαίτερο χαρακτηριστικό: είναι *άνευ κλίμακας* (*scale-free*) που σημαίνει ότι η κατανομή παραμένει αμετάβλητη ακόμη και αν μεταβληθεί η κλίμακά της ή η κλίμακα του συστήματος. Για παράδειγμα, αν μεταβληθεί η ανεξάρτητη μεταβλητή x κατά τον παράγοντα α η μορφή της κατανομής θα αλλάξει μόνο κατά τον πολλαπλασιαστικό παράγοντα, δηλαδή $p(\alpha x) = \alpha^{-\gamma} p(x)$. Στο Σχ. 2.2 απεικονίζονται γραφικά οι δύο κατανομές Poisson και power-law σε γραμμική κλίμακα.

Συντελεστής Ομαδοποίησης (Clustering Coefficient): Είναι ένα συνηθισμένο χαρακτηριστικό που συναντάται σε πολλά πραγματικά δίκτυα και εκφράζει την ιδιότητα δύο γειτονικοί κόμβοι του κόμβου u να είναι επίσης γειτονικοί μεταξύ τους [144]. Ο συντελεστής ομαδοποίησης δίνεται από τη σχέση [144]:

$$cc = \frac{3 \times N_{\Delta}}{\sum_v \binom{d_v}{2}} = \frac{\# \text{ κύκλων μήκους } 3}{\# \text{ μονοπατιών μήκους } 2} \quad (2.6)$$

όπου d_v είναι ο βαθμός του κόμβου v και N_{Δ} είναι ο αριθμός των τριγώνων που εμφανίζονται στο δίκτυο.

Churn Rate: Churn είναι η συνεχόμενη διαδικασία των κόμβων να εισέρχονται και να εξέρχονται στο δίκτυο [162]. *Churn Rate* ορίζεται το ποσοστό των κόμβων



Σχήμα 2.3: Κατανομή scale free.

που εξέρχονται / εισέρχονται στο δίκτυο σε μια συγκεκριμένη χρονική περίοδο σε σχέση με το συνολικό αριθμό των κόμβων.

2.2 Σύνθετα δίκτυα του πραγματικού κόσμου

Πληθώρα σύνθετων δικτύων υπάρχουν τόσο στον τεχνολογικό όσο και στον φυσικό σύγχρονο κόσμο. Χάρη στη διαθέσιμη υπολογιστική ισχύ, σήμερα είναι δυνατή η συλλογή και ο χειρισμός συνόλων δεδομένων από αυτά τα δίκτυα που συνήθως είναι μεγάλης κλίμακας, ώστε να αναλυθούν και να εξαχθούν οι δομικές και λειτουργικές τους ιδιότητες. Τα περισσότερα γνωστά και μελετημένα πραγματικά σύνθετα δίκτυα παρουσιάζονται παρακάτω ταξινομημένα ως προς το σκοπό που εξυπηρετούν.

2.2.1 Δίκτυα φυσικής υποδομής

Τα δίκτυα αυτά βασίζονται σε φυσική υποδομή, έχουν φυσικές συνδέσεις και αποτελούν δημιούργημα του ανθρώπου για να καλύψει ανάγκες μεταφοράς και διανομής υλικών και άυλων αγαθών (ενέργεια, δεδομένα, εμπορεύματα, κ.ά.). Αν και οι ρίζες τους ανάγονται στο μακρινό παρελθόν, εξακολουθούν να υπάρχουν και να δημιουργούνται καινούρια όπως για παράδειγμα τα δίκτυα υπολογιστών. Τα δίκτυα φυσικής υποδομής γνώρισαν μεγάλη ανάπτυξη τον τελευταίο αιώνα και σε κάποιες περιπτώσεις μάλιστα αποτελούν τη βάση για την ανάπτυξη υπερκείμενων εικονικών δικτύων.

Χαρακτηριστικό τέτοιο παράδειγμα αποτελεί το διαδίκτυο (Internet). Το διαδίκτυο, ένα δίκτυο παγκόσμιας εμβέλειας, απαρτίζεται από υπολογιστές και δικτυακές συσκευές που διασυνδέονται σε φυσικό επίπεδο μεταξύ τους με χαλκό, οπτική ίνα ή και ασύρματα (με φυσικές συνδέσεις είτε ενσύρματα είτε ασύρματα). Είναι ένα δίκτυο πολύ μεγάλης κλίμακας, αυτοδιοικούμενο, συνεχώς εξελισσόμενο και αναπτυσσόμενο και με εσωτερική ετερογένεια, συνεπώς δύσκολο να μελετηθεί. Για τη μελέτη του προτάθηκαν δύο προσεγγίσεις. Η πρώτη, θεωρεί ως κόμβους τους δρομολογητές (routers) και ως ακμές τις μεταξύ τους φυσικές διασυνδέσεις. Η δεύτερη προσέγγιση, θεωρεί ως μια ολότητα το αυτόνομο σύστημα (ΑΣ), δηλαδή το σύνολο των υπολογιστών με συγκεκριμένη διαχείριση και όμοια πολιτική δρομολόγησης. Για το γράφο αναπαράστασης του διαδικτύου ο κόμβος είναι το αυτόνομο σύστημα και ακμή η φυσική σύνδεση μεταξύ τουλάχιστον δύο δρομολογητών που ανήκουν σε διαφορετικά ΑΣ.

Η μελέτη της αναπαράστασης του διαδικτύου ως δίκτυο αποτέλεσε το επίκεντρο πολλών εργασιών που βασίστηκαν κυρίως σε εμπειρικά δεδομένα που συλλέχθηκαν από το διαδίκτυο και διαπίστωσαν την ύπαρξη βασικών χαρακτηριστικών που παρατηρούνται στα σύνθετα δίκτυα. Οι μελέτες [65], [189] έδειξαν ότι οι βαθμοί των κόμβων ακολουθούν την κατανομή power-law. Οι Pastor-Satorras και Vespignani [158] αναφέρονται στην ύπαρξη κοινοτήτων και "μικρών κόσμων" ενώ στην [7] καταγράφεται ο υψηλός συντελεστής ομαδοποίησης και η ύπαρξη σύντομων μονοπατιών.

Στα δίκτυα υποδομής εντάσσονται επίσης και τα δίκτυα μεταφοράς ενέργειας, αγωγών, ατόμων [20] ή και ιών [16], το τηλεφωνικό δίκτυο [7], [147], δίκτυο ηλεκτρονικών κυκλωμάτων [147], τα δίκτυα παράδοσης και διανομής [145], [20], κ.ά.

2.2.2 Δίκτυα διάθεσης πληροφορίας και πόρων

Τα δίκτυα αυτής της κατηγορίας είναι κυρίως σύγχρονα και αναπτύσσονται πάνω στην υπάρχουσα υποδομή άλλων δικτύων. Σχηματίζονται από οντότητες, όπως υπολογιστές, χρήστες, ή τμήματα πληροφορίας και αλληλεπιδρούν μεταξύ τους σχηματίζοντας λογικές συνδέσεις. Στο δίκτυο του παγκόσμιου ιστού (world wide web) παρουσιάζονται οι ιστοσελίδες και πως συνδέονται μεταξύ τους, στα δίκτυα peer-to-peer (ή p2p) διαφαίνονται οι σχέσεις γειτνίασης μεταξύ των χρηστών και στις online κοινότητες η επικοινωνία ιδεών, ειδήσεων και γνώμων.

Το μεγαλύτερο και το περισσότερο γνωστό δίκτυο αυτής της κατηγορίας είναι ο παγκόσμιος ιστός. Οι κόμβοι είναι οι ιστοσελίδες και οι ακμές είναι οι υπερσυνδέσεις μεταξύ των σελίδων. Πολλές μελέτες εξετάζουν τον παγκόσμιο ιστό από διάφορες οπτικές και αποκαλύπτουν χαρακτηριστικές ιδιότητες των σύνθετων δικτύων. Η κατανομή των βαθμών είναι power-law με μικρό μέσο μήκος μονοπατιού [8], [110] και υψηλό βαθμό ομαδοποίησης [88]. Άλλες έρευνες εστίασαν στις ροές κίνησης ή στον βαθμό επισκεψιμότητας των ιστοσελίδων και απέδειξαν πως ακολουθούν κατανομή power-law [28].

Ένα ακόμη δίκτυο της ίδιας κατηγορίας που τράβηξε την προσοχή των μελετητών είναι το δίκτυο που σχηματίζεται από τα ιστολόγια (blogs). Οι κόμβοι είναι τα ιστολόγια και οι μεταξύ τους συνδέσεις αποτελούν τις ακμές. Οι ερευνητές, εκτός από την κατανομή βαθμών, εξετάζουν τα μοτίβα δόμησής τους, τη δημοτικότητα τους ή τη συχνότητα ανανέωσης της είδησης [2], [171], [118].

Ένας άλλος τύπος δικτύων αυτής της κατηγορίας είναι τα δίκτυα ομότιμων (ή δίκτυα p2p). Τα δίκτυα p2p αποτελούνται από μεγάλο αριθμό διασυνδεδεμένων υπολογιστών που παρέχουν υπηρεσίες και συμβάλλουν στην εκτέλεση σύνθετων ή δυναμικών διεργασιών (π.χ. αναζήτηση).

Τα δίκτυα p2p διαδόθηκαν ευρέως τη δεκαετία του '90 και έγιναν ιδιαίτερα δημοφιλή για τα πλεονεκτήματα και τις υπηρεσίες τους. Στα πλεονεκτήματα συγκαταλέγονται η εξοικονόμηση πόρων, η κλιμάκωση και η ανθεκτικότητα εξ αιτίας της κατανεμημένης φύσης τους [15], η ανεξαρτησία καθώς όλοι είναι ισότιμοι και δεν υπάρχει εποπτεύουσα αρχή και το ελάχιστο ή μηδαμινό κόστος πρόσβασης στο δίκτυο. Οι υπηρεσίες που παρέχουν περιλαμβάνουν διαμοίραση αρχείων (π.χ. Gnutella, BitTorrent), media streaming (π.χ. Napster, Spotify), τηλεφωνία (π.χ. Skype) ή ψηφιακό νόμισμα (π.χ. Bitcoin).

Τα δίκτυα p2p διακρίνονται σε δύο μεγάλες κατηγορίες ανάλογα με τον τρόπο σχηματισμού τους: *δομημένα* (structured) και *αδόμητα* (unstructured). Οι Lv κ.ά. [122] ταξινομούν τα δίκτυα p2p σε κεντριοποιημένα, αποκεντρωμένα-δομημένα και αποκεντρωμένα-αδόμητα ενώ οι Park κ.α. [156] διακρίνουν περαιτέρω τα αδόμητα συστήματα σε κεντριοποιημένα, υβριδικά και κατανεμημένα.

Στα δομημένα συστήματα οι συνδέσεις μεταξύ των κόμβων είναι σαφώς ορισμένες βάσει κανόνων όπως επίσης καθορισμένες είναι και οι θέσεις που θα τοποθετούν οι παρεχόμενοι πόροι. Με τον τρόπο αυτό επιτυγχάνεται αποτελεσματική αναζήτηση παρόλο που αυξάνεται η πολυπλοκότητα και το κόστος διατήρησης της δόμη-

σης σε δυναμικά περιβάλλοντα. Συνήθως για την διατήρηση της αυστηρής δομής και της αποθήκευσης των δεδομένων χρησιμοποιούνται κατακεντρωμένοι πίνακες κατακερματισμού (Distributed Hash Tables - DHTs). Παραδείγματα τέτοιων συστημάτων αποτελούν τα: CAN [161], Chord [179], Tapestry [206], Pastry [167] και Kademlia [134] που διαφέρουν όμως στον τρόπο δρομολόγησης των ερωτήσεων αναζήτησης.

Αντίθετα με τα δομημένα, στα αδόμητα συστήματα οι κόμβοι σχηματίζουν αυθαίρετα συνδέσεις μεταξύ τους και λειτουργούν με έναν πλήρως κατακεντρωμένο τρόπο. Στα κεντρικοποιημένα συστήματα διατηρείται ένα κεντρικό ευρετήριο με πληροφορία σχετική με τους κόμβους και τα αρχεία που διαθέτουν όπως, για παράδειγμα, στο Napster και στο BitTorrent. Η αναζήτηση χρησιμοποιεί τον κεντρικό κατάλογο και στη συνέχεια οι χρήστες ανταλλάσσουν αρχεία με απευθείας συνδέσεις μεταξύ τους. Τα μειονεκτήματα αυτής της προσέγγισης είναι το κεντρικό σημείο αποτυχίας και η αδυναμία κλιμάκωσης. Τα υβριδικά συστήματα υποστηρίζουν τόσο κεντρικοποιημένα (π.χ. με χρήση super nodes) όσο και κατακεντρωμένα (πρόσβαση και ανταλλαγή πόρων) οργάνωση και αναζήτηση (π.χ. kazza, Gnutella). Στα αποκεντρωμένα συστήματα, ο σχηματισμός του δικτύου είναι εντελώς αυθαίρετος, όπως αυθαίρετα γίνεται και η τοποθέτηση των πόρων ή η διαδικασία αναζήτησής (π.χ. Gnutella). Όπως έδειξαν σχετικές μελέτες, τα αποκεντρωμένα αδόμητα $p2p$ δίκτυα (ή απλά αδόμητα $p2p$) είναι άνευ κλίμακας και εμφανίζουν το φαινόμενο του μικρού κόσμου [18].

2.2.3 Κοινωνικά δίκτυα

Ένα κοινωνικό δίκτυο εκφράζει τις επαγγελματικές, φιλικές, οικογενειακές σχέσεις και δεσμούς μεταξύ ατόμων, ομάδων ή οργανισμών μιας κοινωνίας [147]. Ένα τέτοιο δίκτυο περιγράφει την κοινωνική δομή μεταξύ των μελών του, που αποτελούν τους κόμβους του δικτύου και τις μεταξύ τους συσχετίσεις που είναι οι ακμές του δικτύου. Πολλά από αυτά τα δίκτυα έχουν μελετηθεί διεξοδικά, όπως δίκτυα διαπροσωπικών σχέσεων [138], συνεργατικά δίκτυα, δίκτυα τηλεφωνικών κλήσεων, ανταλλαγής e-mails, βιβλιογραφικών αναφορών [147] και δίκτυα κοινωνικής δικτύωσης [96].

Οι μελέτες έδειξαν ιδιότητες όπως κατανομή βαθμών power-law, ύπαρξη κοινοτήτων και σύντομων μονοπατιών που επιτρέπουν την αποδοτική αναζήτηση συγκεκριμένων “στόχων” [61]. Όμως η επιτυχία της αναζήτησης βασίζεται κυρίως στους

ενδιάμεσους κόμβους [57].

2.2.4 Φυσικά δίκτυα

Η κατηγορία αυτή περιλαμβάνει πλήθος φυσικών δικτύων που σχετίζονται με την ύπαρξη, την εξέλιξη και τις διαδικασίες που συμβαίνουν στον φυσικό κόσμο και τους οργανισμούς. Κάθε ζωντανός οργανισμός αποτελεί πολύπλοκο σύστημα και σε κάθε βιολογικό σύστημα ένας τεράστιος αριθμός διεργασιών αλληλοδιαπλέκονται. Τα σύνθετα δίκτυα αποτελούν μονόδρομο για την μελέτη αυτών των συστημάτων. Η προσέγγιση γίνεται σε επίπεδο ειδών για την αναπαράσταση ενός οικοσυστήματος, ή σε κυτταρικό επίπεδο [20] για την αναπαράσταση χημικών αντιδράσεων, συνδέσεων μεταξύ κυττάρων (π.χ. εγκεφαλικά κύτταρα) γονιδιωμάτων και πρωτεϊνικών αλληλοεπιδράσεων. Στην ίδια κατηγορία δικτύων εντάσσεται και το τροφικό πλέγμα (food web). Το δίκτυο αυτό είναι κατευθυνόμενο και οι κορυφές του αντιστοιχούν στα είδη του οικοσυστήματος. Η κατευθυνόμενη ακμή από το ένα είδος στο άλλο δείχνει τη σχέση θηρευτή - θηράματος.

2.3 Μοντέλα σύνθετων δικτύων

Η επιρροή των ιδιαίτερων χαρακτηριστικών των πραγματικών δικτύων σε διάφορα φαινόμενα ή διαδικασίες αλλά και η ανάγκη για μελέτη και βαθύτερη κατανόηση αυτών των συστημάτων οδήγησε στην δημιουργία στατικών ή δυναμικών μοντέλων που αναπαράγουν τις ιδιότητες των πραγματικών δικτύων. Τα στατικά μοντέλα έχουν σταθερό αριθμό κόμβων από την αρχή της δημιουργίας τους και δεν μεταβάλλεται κατά τη διάρκεια της λειτουργίας τους, όπως για παράδειγμα το μοντέλο του τυχαίου γράφου των Erdős-Rényi [62] και το μοντέλο του μικρού κόσμου των Watts-Strogatz [197]. Οι συνδέσεις μεταξύ των κόμβων σχηματίζονται στατικά βάσει κανόνων τυχαιότητας. Τα δυναμικά μοντέλα μεταβάλλονται και εξελίσσονται με την πάροδο του χρόνου, όπως συμβαίνει και στα δίκτυα του πραγματικού κόσμου. Νέοι κόμβοι δημιουργούνται και προσαρτώνται σε ένα υπάρχον δίκτυο. Το δυναμικό μοντέλο είναι μια αλληλουχία στιγμιοτύπων του δικτύου σε κάθε χρονικό βήμα. Χαρακτηριστικό παράδειγμα αποτελεί το μοντέλο των Barabási-Albert [19].

2.3.1 Στατικά μοντέλα

Η πρώτη προσέγγιση για την μοντελοποίηση των δικτύων βασίστηκε στη θεωρία των τυχαίων γράφων. Η θεωρία των τυχαίων γράφων μελετά τις ιδιότητες ενός συνόλου γράφων. Ένας τυχαίος γράφος είναι το αποτέλεσμα μια τυχαίας διαδικασίας, για παράδειγμα μιας δειγματοληψίας από μια συλλογή γράφων. Αν και ο αριθμός των κόμβων του είναι σταθερός, οι συνδέσεις μεταξύ των κόμβων του περιγράφεται από κανόνες τυχαιότητας.

Δίκτυα Erdős - Rényi

Στα τέλη της δεκαετίας του 1950 δόθηκε ιδιαίτερη έμφαση στην μελέτη των τυχαίων γράφων και δυο σχεδόν όμοια μοντέλα εμφανίζονται: ο τυχαίος δυωνυμικός γράφος (random binomial graph) $G(N, p)$ [75] που αποτελείται από N κορυφές που συνδέονται ανά δύο μεταξύ τους με πιθανότητα p και ο ομοιόμορφα τυχαίος γράφος (uniform random graph) $G(N, m)$ [62], που έχει N κορυφές και m ακμές. Το μοντέλο $G(N, m)$ προτάθηκε από τους Paul Erdős και Alfréd Rényi και ουσιαστικά έδωσε νέες διαστάσεις στη θεωρία των τυχαίων γράφων. Για την δημιουργία ενός τυχαίου δικτύου, σε N αρχικά μεμονωμένους κόμβους τοποθετούνται m ακμές επιλεγμένες τυχαία και ομοιόμορφα από τις $N(N-1)/2$ δυνατές ακμές. Αντίστοιχα για την δημιουργία ενός τυχαίου δικτύου με το μοντέλο $G(N, p)$ τοποθετούνται N μεμονωμένοι κόμβοι και κάθε ζεύγος κόμβων συνδέεται με μια ακμή με ανεξάρτητη πιθανότητα p ή δεν συνδέεται με πιθανότητα $1-p$.

Και για τα δύο μοντέλα, τα παραγόμενα δίκτυα είναι μη κατευθυνόμενα, ενώ δεν επιτρέπουν την προσθήκη πολλαπλών ακμών. Για μεγάλες τιμές του N τα δύο μοντέλα είναι ισοδύναμα και στην βιβλιογραφία απαντώνται συνήθως σαν δίκτυα Erdős - Rényi (ER). Αν και διαφέρουν από τα δίκτυα του πραγματικού κόσμου και αδυνατούν να αποτυπώσουν με ακρίβεια τα χαρακτηριστικά τους [146], [18] είναι από τα περισσότερα μελετημένα δίκτυα καθώς αποκαλύπτουν ιδιότητες των δικτύων που υπολογίζονται κατά μέσο όρο από το σύνολο των γράφων που ανήκουν στην ίδια οικογένεια.

Παρακάτω παρουσιάζονται μερικές απλές ιδιότητες του μοντέλου ER για μεγάλες τιμές του N επισημαίνοντας ότι στα πραγματικά δίκτυα ο αριθμός των ακμών συνήθως δεν είναι σταθερός:

- Το μέσο πλήθος των ακμών m του γράφου $G(N, p)$ είναι: $m = pN(N-1)/2$.
- Ο μέσος βαθμός του γράφου είναι: $\bar{d} = 2m/N = (N-1)p$

- Το μοντέλο έχει κατανομή βαθμών (p_d) Poisson καθώς η εμφάνιση ή όχι των ακμών είναι ανεξάρτητη και η πιθανότητα ένας κόμβος να έχει βαθμό d δίνεται από την (2.2).
- Ο συντελεστής ομαδοποίησης του τυχαίου γράφου δίνεται από τη (2.6) και είναι

$$cc = \frac{\binom{N}{3} p^3}{\binom{N}{3} p^2} = p \quad (2.7)$$

αφού η πιθανότητα δυο κόμβοι να είναι γειτονικοί είναι p .

Αξίζει να σημειωθεί ότι οι δομικές ιδιότητες του δικτύου μεταβάλλονται ανάλογα με την τιμή της πιθανότητας p . Υπάρχει μια κρίσιμη τιμή κατωφλίου (threshold) που οδηγεί στην εμφάνιση ή στην απόκρυψη ορισμένων ιδιοτήτων του δικτύου [18], [146], [28]. Για παράδειγμα αν $p > \ln N/N$ [62] τότε το δίκτυο είναι συνδεδεμένο και αν $p < \ln N/N$ τότε το δίκτυο επιμερίζεται σε συνιστώσες που δεν επικοινωνούν μεταξύ τους. Οι ιδιότητες αυτές μπορούν να φανούν χρήσιμες και να αξιοποιηθούν στην αναζήτηση [153].

Οι τακτικοί τυχαίοι γράφοι είναι μια οικογένεια γράφων που ανήκει στο σύνολο των τυχαίων γράφων. Όλοι οι κόμβοι έχουν τον ίδιο αριθμό συνδέσεων, d , ενώ μπορεί να αποδειχθεί ότι ορισμένες ιδιότητες ισχύουν με μεγάλη βεβαιότητα, όπως για παράδειγμα η ύπαρξη γιγάντιας συνιστώσας. Τέτοια δίκτυα εμφανίζονται στην επιστήμη των υλικών και περιγράφουν τους δεσμούς μεταξύ των ατόμων των χημικών στοιχείων [18].

Γενικευμένα Τυχαία Δίκτυα

Το μοντέλο του τυχαίου γράφου είναι από τα περισσότερο μελετημένα, ενώ προτάθηκαν επεκτάσεις ή και γενικεύσεις του μοντέλου. Το γενικευμένο μοντέλο τυχαίου γράφου (Generalized Random Graph - GRG) [146], [140] εισάγει διαφορετική πιθανότητα για την κάθε ακμή σε αντιδιαστολή με την ομοιόμορφη πιθανότητα του ER μοντέλου και παράγει ετερογενή δίκτυα. Το μοντέλο διαμόρφωσης (configuration model) [147] παράγει δίκτυα με καθορισμένη ακολουθία βαθμών, γνωστή εκ των προτέρων (για παράδειγμα από μετρήσεις σε πραγματικά δίκτυα). Μια ειδική περίπτωση τυχαίων γράφων περιλαμβάνει τους γράφους που η δεδομένη ακολουθία βαθμών έχει κατανομή power-law. Το παραμετρικό μοντέλο του Aiello κ.ά. [5] θεωρεί ότι υπάρχουν y κόμβοι με βαθμό x (με $\log y = \alpha - \beta \log x$) και αντιστοιχεί ομοιόμορφη πιθανότητα σε όλους τους γράφους με βαθμό $y = e^\alpha / x^\beta$ (α, β παράμετροι του μοντέλου). Παρόμοιο είναι και το μοντέλο των Molloy–Reed [140] όπου σε

κάθε κόμβο αντιστοιχεί ένας αριθμός συνδέσεων από μια κατανομή power-law. Τα μοντέλα διαμόρφωσης παράγουν δίκτυα άνευ κλίμακας, όμως αυτά είναι στατικά καθώς τόσο ο αριθμός των κόμβων όσο και ο αριθμός των ακμών είναι προκαθορισμένες σταθερές.

Δίκτυα μικρών κόσμων

Οι Watts και Strogatz [197] πρότειναν ένα μοντέλο που συνδυάζει το χαρακτηριστικό των σύντομων διαδρομών του μοντέλου ER και το χαρακτηριστικό της ομαδοποίησης που εμφανίζεται στα κοινωνικά δίκτυα όπου τα άτομα τείνουν να σχηματίζουν ομάδες. Όπως διαπιστώθηκε [138] απαιτούνται έξι το πολύ διαδοχικά βήματα² (αλυσίδα ενδιάμεσων ατόμων) για να συνδεθούν (ή να επικοινωνήσουν) μεταξύ τους δύο τυχαία επιλεγμένοι πολίτες των Ηνωμένων Πολιτειών. Οι συγγραφείς του [197] ονόμασαν το μοντέλο τους μικρό - κόσμος (small-world) κατά αντιστοιχία του φαινομένου του μικρού κόσμου όπου ο όρος “μικρός” αναφέρεται στην εγγύτητα μεταξύ δύο ατόμων. Στην ίδια εργασία διαπίστωσαν ότι τα παραπάνω χαρακτηριστικά παρατηρούνται σε δίκτυα του πραγματικού κόσμου και αποτυπώνονται από το μοντέλο τους. Μελέτες που ακολούθησαν [7], [48] έδειξαν ότι η ιδιότητα του μικρού-κόσμου είναι “πανταχού παρούσα” στα πραγματικά δίκτυα. Παρότι το μοντέλο αποδίδει δύο βασικές ιδιότητες των πραγματικών δικτύων, δεν μπορεί να αποτυπώσει την κατανομή των βαθμών των κόμβων.

Για τη δημιουργία ενός τέτοιου δικτύου, αρχικά τοποθετούνται N κόμβοι. Ο κάθε κόμβος συνδέεται με τους k κοντινότερους γείτονές του. Στη συνέχεια, κάθε ακμή επανασυνδέεται [197] ή προστίθεται νέα ακμή [150] με πιθανότητα p . Με τον τρόπο αυτό επιτυγχάνεται η δημιουργία σύντομων διαδρομών μεταξύ των κόμβων του δικτύου. Η παράμετρος p του μοντέλου εξασφαλίζει τον σχηματισμό δικτύων που εκτείνονται από τακτικά ($p = 0$) έως τυχαία ($p = 1$). Βέβαια υπάρχει μια ενδιάμεση περιοχή τιμών της p που εξασφαλίζει τις επιθυμητές ιδιότητες, δηλαδή μικρό μήκος μονοπατιού και υψηλό συντελεστή ομαδοποίησης.

Ο Watts κ.ά. [196] εκμεταλλεύονται τη συχνότητα εμφάνισης της ιδιότητας του μικρού-κόσμου σε πολλές κατηγορίες δικτύων και προτείνουν ένα ιεραρχικό μοντέλο αναζήτησης που διευκολύνεται από την ύπαρξη κοινοτήτων και σύντομων διαδρομών. Αντίστοιχα, ο Kleinberg [97], [99] εξετάζει κατά πόσο τα ποιοτικά χαρακτηριστικά του μικρού-κόσμου σε συνδυασμό με τοπική πληροφορία μπορούν να αξιοποιηθούν για στοχευμένη αναζήτηση σε κατανεμημένο περιβάλλον.

²Η ιδιότητα αυτή είναι γνωστή και ως “six degeee separation” [138].

2.3.2 Δυναμικά μοντέλα

Τα στατικά μοντέλα (π.χ. *ER*) αδυνατούν να αποτυπώσουν τα πραγματικά δίκτυα (άνευ κλίμακας δίκτυα) καθώς υστερούν στην απόδοση σημαντικών ιδιοτήτων τους όπως για παράδειγμα η κατανομή βαθμών, ο συντελεστής ομαδοποίησης και οι σύντομες διαδρομές. Αυτό οφείλεται στις “κρυμμένες υποθέσεις” [18] ότι α) το μέγεθος του δικτύου N παραμένει σταθερό και β) ένας κόμβος επιλέγει τυχαία έναν άλλο κόμβο και αλληλεπιδρά ή συσχετίζεται μαζί του. Στην πλειονότητα τους όμως, τα πραγματικά δίκτυα συνεχώς εξελίσσονται καθώς προστίθενται ή αφαιρούνται κόμβοι και ακμές [19], [32], [115]. Άρα υπάρχει ανάγκη για ένα μοντέλο που δεν περιγράφει το δίκτυο, αλλά μοντελοποιεί την εξέλιξή του [18]. Στη συνέχεια παρουσιάζονται μοντέλα που ανταποκρίνονται σε αυτή την απαίτηση.

Uniform Attachment Model

Το μοντέλο της τυχαίας προσάρτησης [89] αποτελεί μια παραλλαγή του *ER* μοντέλου το οποίο λαμβάνει υπόψη την προσάρτηση ενός νέου κόμβου κάθε φορά. Αρχικά υπάρχουν $n+1$ κόμβοι πλήρως συνδεδεμένοι μεταξύ τους. Κάθε κόμβος που εισέρχεται στο δίκτυο δημιουργεί συνδέσεις με n τυχαία επιλεγμένους κόμβους που ανήκουν ήδη στο δίκτυο. Το μοντέλο προσεγγίζει μια κατανομή βαθμών που είναι παραλλαγή της εκθετικής κατανομής.

Preferential Attachment Model

Το 1995 ο Simon [173] μελετώντας συναρτήσεις κατανομής βασιζόμενος σε εμπειρικές παρατηρήσεις από κοινωνικά, βιολογικά και οικονομικά φαινόμενα διαπίστωσε ότι αυτές οι κατανομές έχουν κοινά χαρακτηριστικά. Για να εξηγήσει αυτή τη διαπίστωση, προτείνει μια κατηγορία στοχαστικών μοντέλων που οδηγούν σε κατανομή power-law. Ακολούθως, ο Price [54] διαπιστώνει αναλογικότητα μεταξύ του αριθμού των αναφορών μιας δημοσίευσης και των νέων αναφορών για τη συγκεκριμένη δημοσίευση καθώς ο “πλούσιος γίνεται πλουσιότερος” [54]. Μια νέα διάσταση δίνει το μοντέλο των Barabási-Albert [19] που εισάγοντας την έννοια “προτιμώμενη σύνδεση (preferential attachment)” επιχειρεί να εξηγήσει την προέλευση των δικτύων άνευ κλίμακας. Σύμφωνα με αυτό, ένας κόμβος που εισέρχεται στο δίκτυο επιλέγει ελεύθερα με ποιους άλλους κόμβους θα συνδεθεί, όμως δείχνει προτίμηση σε κόμβους με αυξημένο αριθμό συνδέσεων. Οι πολλές συνδέσεις ενός κόμβου οφείλονται σε ιδιαίτερες ιδιότητες που εμφανίζει. Για παράδειγμα η δημοτικότητα ή η αναγνωρισιμότητα ενός προσώπου, η επισκεψιμότητα ενός διαδικτυακού τόπου και

η συχνότητα εμφάνισης μιας βιβλιογραφικής αναφοράς.

Το μοντέλο των Barabási-Albert (BA) δημιουργεί ένα μη κατευθυνόμενο δίκτυο ως εξής:

- Σε κάθε χρονικό βήμα ένας νέος κόμβος εισέρχεται στο δίκτυο και επιλέγει ελεύθερα να συνδεθεί με τους m_0 πλήρως συνδεδεμένους κόμβους που υπάρχουν στο δίκτυο.
- Όμως “προτιμά” να συνδεθεί στον κόμβο i που έχει ήδη d_i συνδέσεις με πιθανότητα

$$\Pi(d_i) = \frac{d_i}{\sum_j d_j} \quad (2.8)$$

Το μοντέλο βασίζεται σε δυο μηχανισμούς: της ανάπτυξης και της προτιμώμενης σύνδεσης. Η απουσία ενός από τους δύο μηχανισμούς αποτυγχάνει να αναδείξει δίκτυα με την ιδιότητα άνευ κλίμακας [19] κατανομής βαθμών. Το μοντέλο BA παράγει δίκτυα με κατανομή βαθμών power-law με εκθέτη $\gamma = 3$.

Λόγω της απλότητάς του, το μοντέλο υστερεί στην περιγραφή χαρακτηριστικών των πραγματικών δικτύων [19], [18]: η κατανομή power-law έχει σταθερό εκθέτη ($\gamma = 3$) ενώ στα πραγματικά δίκτυα ο εκθέτης λαμβάνει και άλλες τιμές (μεταξύ 2 – 5), δεν παράγει κατευθυνόμενα δίκτυα, όπως είναι ο παγκόσμιος ιστός, δεν προβλέπει την επανασύνδεση ή την αφαίρεση συνδέσεων, γεγονός που ισχύει σε πραγματικά δίκτυα. Για ξεπεραστούν οι περιορισμοί του μοντέλου προτάθηκαν διάφορες τροποποιήσεις, όπως μια παραλλαγή για κατευθυνόμενα δίκτυα BA [105] και μια έκδοση με κυμαινόμενο εκθέτη για την κατανομή βαθμών [58], [105]. Το μοντέλο BA δεν είναι το μοναδικό για τη δημιουργία δικτύων άνευ κλίμακας.

Copying model

Σε αρκετές περιπτώσεις τα δίκτυα εμφανίζουν κατανομή power-law που ο μηχανισμός δημιουργίας της δεν μπορεί να εξηγηθεί από την κατά προτίμηση σύνδεση [147]. Οι Kleinberg [102] και Kumar [109], [108] οδηγούμενοι από την παρατήρηση πως οι δημιουργοί νέων ιστοσελίδων σχετικά με ένα θέμα τείνουν να αντιγράψουν συνδέσμους από άλλες σχετικές ιστοσελίδες, πρότειναν το μοντέλο αντιγραφής (copying model). Σε κάθε χρονική στιγμή που ένας κόμβος εισέρχεται στο δίκτυο, επιλέγει ομοιόμορφα τυχαία είτε να συνδεθεί με έναν κόμβο u (με πιθανότητα p), είτε να δημιουργήσει συνδέσεις με k γειτονικούς κόμβους του u (με πιθανότητα

$1 - p$). Αυτό έχει ως αποτέλεσμα την δημιουργία κατευθυνόμενων δικτύων με κατανομή βαθμών power-law και εκθέτη της κατανομής $\gamma = 1/(1 - p)$.

Forest Fire

Το μοντέλο Forest Fire [114] επιχειρεί να μιμηθεί τον πραγματικό κόσμο και τη συμπεριφορά που παρατηρείται για παράδειγμα στη δημιουργία μιας φιλίας ή στον τρόπο που γίνονται οι βιβλιογραφικές αναφορές. Όταν ένα νέο μέλος εισέρχεται σε μια ομάδα, διακρίνει ένα μέλος ως σημείο αναφοράς, συνδέεται μαζί του και διερευνά τη γειτονιά του για να δημιουργήσει νέες συνδέσεις. Επαναλαμβάνει την διαδικασία μέχρι να “κουραστεί”. Πιο τυπικά η λειτουργία του μοντέλου έχει ως εξής:

- ένας νέος κόμβος εισέρχεται στο δίκτυο και τυχαία επιλέγει έναν αρχικό κόμβο u και δημιουργεί μια σύνδεση μαζί του.
- επιλέγει τυχαία k γειτονικούς κόμβους του u από γεωμετρική κατανομή με μέση τιμή $p/(1 - p)$ και δημιουργεί συνδέσεις με τους k κόμβους. Αυτοί οι κόμβοι γίνονται ενεργοί, ενώ αντίστοιχα ο u παύει να είναι ενεργός.
- η διαδικασία του προηγούμενου βήματος επαναλαμβάνεται μέχρι να μην υπάρχει ενεργός κόμβος, οπότε και τερματίζεται.

Οι δημιουργοί του μοντέλου [114] θεωρούν ότι τώρα πετυχαίνουν όχι μόνο να εφαρμόσουν τα μοτίβα των προηγούμενων εμπειριών αλλά και να τα εμπλουτίσουν. Το μοντέλο της προτιμώμενης σύνδεσης οδηγεί σε κατανομές με βαριά ουρά, ενώ το μοντέλο αντιγραφής τείνει να δημιουργεί κοινότητες. Το μοντέλο Forest Fire εκτός από κατανομή βαριάς ουράς αποτυπώνει την συρρίκνωση της διαμέτρου και την πυκνότητα που εμφανίζεται στα πραγματικά δίκτυα.

Στατικά ή Δυναμικά Δίκτυα;

Στα στατικά δίκτυα τόσο το πλήθος των κόμβων όσο και ο αριθμός των συνδέσεων παραμένει σταθερός και δεν αλλάζει στο χρόνο. Στα δυναμικά δίκτυα και ο πληθυσμός και οι συνδέσεις μεταβάλλονται με την πάροδο του χρόνου. Τα στατικά δίκτυα είναι μια καλή προσέγγιση σε περιπτώσεις δικτύων που αυτά εξελίσσονται σχετικά αργά, όπως το διαδίκτυο [189]. Παράλληλα πλήθος άλλων εργασιών καταλήγουν πως η εξέλιξη στο χρόνο σε πολλά δίκτυα είναι τέτοια, που επηρεάζει τις διαδικασίες που εκτυλίσσονται σε αυτά.

2.4 Δυναμικές διαδικασίες σε σύνθετα δίκτυα

Το αυξημένο ενδιαφέρον και η ιδιαίτερη προσοχή σχετικά με τα σύνθετα δίκτυα πηγάζει από τις διάφορες φυσικές και δυναμικές αλληλεπιδράσεις που συμβαίνουν σε αυτά. Στα τέλη του 20ου αιώνα, οι ερευνητές συνειδητοποιούν την ανάγκη για συστηματική διερεύνηση της συμπεριφοράς και των χαρακτηριστικών των δυναμικών διεργασιών που εκτυλίσσονται σε συστήματα μεγάλης κλίμακας και υψηλής ετερογένειας. Η αναζήτηση, που είναι το αντικείμενο μελέτης αυτής της διατριβής, είναι μια τέτοια διαδικασία με πρακτικές επιπτώσεις σε πραγματικά δίκτυα.

2.4.1 Αναζήτηση

Σύμφωνα με τον Newman [145], η αναζήτηση είναι ο τρόπος για να γίνουν χρήσιμα τα δεδομένα που σε μεγάλες ποσότητες αποθηκεύονται στους κόμβους των δικτύων, ενώ ο Boccaletti [28] θεωρεί ότι είναι η ικανότητα εύρεσης σύντομων διαδρομών σε ένα πραγματικό δίκτυο. Η αναζήτηση είναι μια δυναμική διαδικασία που εξελίσσεται σε ένα δίκτυο μεταδίδοντας μηνύματα από κόμβο σε κόμβο. Αυτή η διαδικασία περιλαμβάνει τη δημιουργία του μηνύματος (μήνυμα-ερώτηση), την προώθησή του στο δίκτυο και την επεξεργασία του από τον κάθε κόμβο που το λαμβάνει ενώ έχει να διαχειριστεί μια σειρά από ζητήματα όπως την κλίμακα και την ετερογένεια του συστήματος, τη φυσική υποδομή (εύρος ζώνης, υπολογιστική ισχύς) και τον τεράστιο παραγόμενο όγκο μηνυμάτων. Αυτά και άλλα θέματα σχετικά με την αναζήτηση θα αναλυθούν στα επόμενα κεφάλαια, καθώς η αναζήτηση αποτελεί το αντικείμενο αυτής της διατριβής.

2.4.2 Άλλες διαδικασίες

Διάδοση

Η διάδοση είναι μια δυναμική διεργασία που συμβαίνει σε πολλές κατηγορίες δικτύων και αφορά τη μετάδοση που διευκολύνεται από τις απευθείας συνδέσεις μεταξύ των κόμβων και περιλαμβάνει τη μεταφορά ασθενειών, τη διάχυση ιδεών, φημών και γνώσεων, ιών υπολογιστών, την διαφήμιση εμπορικών προϊόντων, τη μεταφορά εμπιστοσύνης και αγαθών [190], [28], [157], [126], [18]. Αν και ο βασικός μηχανισμός του κάθε φαινομένου μπορεί να διαφέρει, η μετάδοση συμβαίνει από κόμβο σε κόμβο. Ο κάθε κόμβος που θα “μολυνθεί” μπορεί να μεταδώσει για ένα

συγκεκριμένο χρονικό διάστημα και να εξαπλώσει τη “μόλυνσή” του στους γειτονικούς του κόμβους που δεν έχουν μολυνθεί. Ο ρυθμός της διάδοσης εξαρτάται από το χρονικό διάστημα που κάποιος κόμβος μπορεί να μεταδώσει, την ευκολία με την οποία μπορεί να μεταδώσει και την ταχύτητα που ένας κόμβος μεταδίδει. Οι παράγοντες αυτοί εξαρτώνται από το είδος του προς μετάδοση αντικειμένου (π.χ. είδος ιού, πακέτο πληροφορίας ή γνώσης, φήμη) και τη “μολυσματικότητά” του αλλά και από τη δομή του δικτύου. Η εξάπλωση μπορεί να ελεγχθεί αν ο ρυθμός αναπαραγωγής δεν ξεπερνάει μια κρίσιμη τιμή (threshold). Αν και το όριο αυτό υπάρχει σε δίκτυα *ER* [147], στα άνευ κλίμακας δίκτυα τείνει στο μηδέν [157].

Ο στόχος των ερευνητών που μελέτησαν τη διάδοση στα δίκτυα είναι διπλός: α) να αποτρέψουν ή και να σταματήσουν την διάδοση όπως για παράδειγμα σε περιπτώσεις μολυσματικών ασθενειών και ιών υπολογιστών, β) να διευκολύνουν την εξάπλωση σε όσο το δυνατό μεγαλύτερο πληθυσμό σε περιπτώσεις διάδοσης γνώσης, ιδεών ή προβολής νέων προϊόντων [126]. Επιπλέον το θεωρητικό πλαίσιο και τα συλλεγόμενα δεδομένα σε πραγματικό χρόνο επιτρέπουν την ακριβή πρόβλεψη της εξέλιξης της διαδικασίας σε ένα δίκτυο [18].

Αντικείμενο μελέτης αποτελεί επίσης και η συμπεριφορά των κόμβων του δικτύου που πολλές φορές εμφανίζει συλλογικότητα ή συγχρονισμό [11], [197], [28]. Για παράδειγμα μια οντότητα επιλέγει συμπεριφορά (π.χ. επιλογή ενός προϊόντος) ώστε να είναι σε συμφωνία με το μεγαλύτερο μέρος των γειτονικών του κόμβων. Συλλογική συμπεριφορά παρατηρείται στις χημικές αντιδράσεις, στα οικολογικά συστήματα, στα φυσικά και κοινωνικά δίκτυα και για τη μελέτη της επιστρατεύεται η θεωρία των παιγνίων προσαρμοσμένη στις δικτυακές παραμέτρους [89].

Ένα δυναμικό φαινόμενο που εμφανίζεται στα σύνθετα δίκτυα συνήθως ως αποτέλεσμα της δυναμικής επικοινωνίας διαδικασιών είναι η συμφόρηση [28]. Η συμφόρηση συμβαίνει στο διαδίκτυο και στα δίκτυα μεταφοράς, σχετίζεται με τη δομή του δικτύου και με διαδικασίες που συμβαίνουν σε αυτό όπως για παράδειγμα η αναζήτηση [10].

ΚΕΦΑΛΑΙΟ 3

ΑΝΑΖΗΤΗΣΗ ΣΕ ΣΥΝΘΕΤΑ ΔΙΚΤΥΑ

3.1 Γενικά

3.2 Επισκόπηση βιβλιογραφίας

3.3 Στρατηγικές αναζήτησης: λύσεις και προκλήσεις

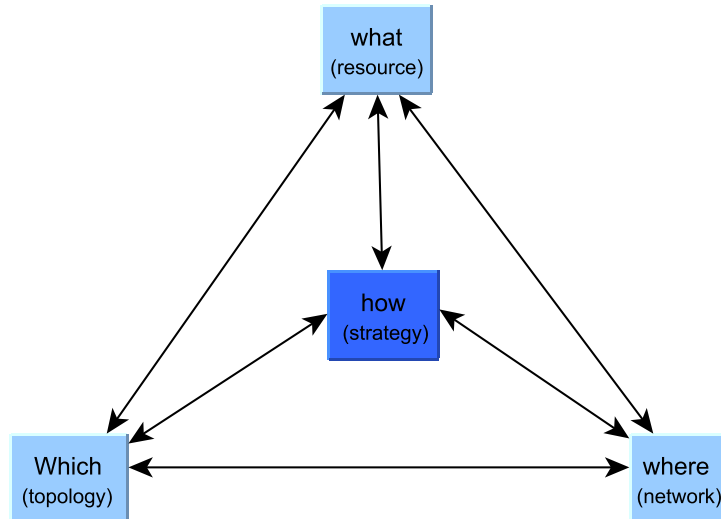
3.4 Βελτίωση αναζήτησης

3.5 Σύνοψη

3.1 Γενικά

Σε αυτή την ενότητα εξετάζονται οι τεχνικές αναζήτησης για την εύρεση ενός κόμβου ή για τον εντοπισμό ενός στοιχείου που αποθηκεύεται ή παρέχεται από έναν κόμβο του δικτύου. Αρχικά, γίνεται βιβλιογραφική επισκόπηση των μεθόδων και αξιολογούνται τα αποτελέσματά τους σε δύο κατευθύνσεις: την αποδοτική επίτευξη του στόχου (πλεονεκτήματα) και τα ζητήματα (μειονεκτήματα) που εισάγουν. Στη συνέχεια, μοντελοποιούμε τη διαδικασία, προσδιορίζουμε και ποσοτικοποιούμε παράγοντες που την επηρεάζουν, ενώ συζητάμε γενικές τεχνικές βελτίωσης.

Η αναζήτηση είναι μια διαδικασία που εκτυλίσσεται σε ένα δίκτυο και συνίσταται στη δημιουργία ενός μηνύματος που μεταδίδεται στο δίκτυο μέχρι να απαντηθεί ή να ικανοποιηθεί κάποια συνθήκη τερματισμού. Η διαδικασία αναζήτησης μπορεί να είναι στατική ή δυναμική. Στη στατική αναζήτηση ο κάθε κόμβος χρησιμοποιεί προκαθορισμένο και σταθερό τρόπο προώθησης της ερώτησης. Αντίθετα, στη δυναμική αναζήτηση ο κάθε κόμβος προσδιορίζει κάθε φορά, αν και που θα προωθηθεί



Σχήμα 3.1: Οι βασικές συνιστώσες της αναζήτησης.

το μήνυμα-ερώτηση.

Η μελέτη της αναζήτησης περιλαμβάνει τρεις διαφορετικές συνιστώσες: *που* (where) θα διεξαχθεί, δηλαδή σε ποιο πεδίο θα εφαρμοστεί (π.χ. web, κοινωνικά δίκτυα), *ποιό* (which) δικτυακό μοντέλο θα χρησιμοποιηθεί (π.χ. τυχαίο δίκτυο) και *τι* (what) θα αναζητηθεί, όπως για παράδειγμα ένας πόρος ή μια υπηρεσία που η διαθεσιμότητά τους μπορεί να ποικίλει. Το *πως* (how) θα εκτυλιχτεί η αναζήτηση, δηλαδή ποια στρατηγική θα ακολουθήσει εξαρτάται από τις παραπάνω συνιστώσες (τοπολογία, δίκτυο, πόροι/υπηρεσίες) όπως φαίνεται και στο Σχ. 3.1. Η στρατηγική αναζήτησης περιγράφει τη διαδικασία μετάδοσης ενός μηνύματος από έναν κόμβο σε έναν ή περισσότερους γειτονικούς του, ακολουθώντας ένα σύνολο κανόνων [158].

Είναι προφανές, πως η ικανότητα μιας στρατηγικής εξαρτάται από τη γνώση που διαθέτουν οι κόμβοι σχετικά με τη δομή του δικτύου. Η αναζήτηση μπορεί να είναι *τυφλή* (blind), όταν οι κόμβοι δεν διαθέτουν ή δεν διατηρούν πληροφορία σχετικά με τη θέση του αναζητούμενου στοιχείου, ή *πληροφορημένη* (informed) όταν οι κόμβοι διατηρούν σχετική πληροφορία [185], [199]. Στην πληροφορημένη αναζήτηση, είτε υπάρχει καθολική γνώση του δικτύου (όπως στα κεντριοποιημένα συστήματα π.χ. Napster), είτε οι κόμβοι διατηρούν μερική πληροφορία που καθοδηγεί την κατανεμμένη αναζήτηση [185], [199], [95]. Σε περίπτωση παντελούς έλλειψης καθολικής πληροφορίας για τη δόμηση του δικτύου, εφαρμόζονται δύο βασικοί μηχανισμοί: η *πλημμύρα* και ο *τυχαίος περιπατητής* που θα δούμε αναλυτικά στο κεφάλαιο 4. Σε πολλές παραλλαγές, οι παραπάνω μηχανισμοί, διευκολύνονται από τη γνώση που

διαθέτουν οι κόμβοι τοπικά.

Η ικανότητα εντοπισμού ενός κόμβου ή εύρεσης ενός αντικειμένου έχοντας μερική ή καθόλου γνώση για την θέση του, έχει ιδιαίτερη πρακτική σημασία για τα σύγχρονα πραγματικά δίκτυα. Για να δοθεί άμεσα απάντηση για την ύπαρξή ή την μη ύπαρξη του επιθυμητού στοιχείου στο δίκτυο, εφαρμόζονται διάφορες τεχνικές με πληθώρα ρυθμίσεων. Σε ένα πρωτοποριακό πείραμα, ο Milgram [138] διαπίστωσε πως τα άτομα βρίσκουν άλλα άγνωστα άτομα μέσω διαδρομών που περνούν από ενδιαμέσες άτομα και μάλιστα αυτές οι διαδρομές είναι σύντομες. Όπως επιβεβαιώθηκε και αργότερα [57] αυτές οι διαδρομές όχι μόνο υπάρχουν σε μεγάλα δίκτυα, αλλά μπορούν να εντοπιστούν χωρίς συνολική γνώση του δικτύου.

Το ζήτημα της αναζήτησης δεν αφορά μόνο τα κοινωνικά δίκτυα, αλλά είναι ουσιαστική λειτουργία και σε ένα ευρύ φάσμα δικτυακών εφαρμογών. Στον παγκόσμιο ιστό βελτιστοποιεί την εύρεση πληροφορίας που αποθηκεύεται στους διακομιστές. Στα δίκτυα p2p εξασφαλίζει την εύρεση πόρων ή αρχείων εφόσον υπάρχουν. Σε μια περιοχή ή σε μια πόλη χρησιμεύει στην εύρεση μιας διαδρομής ή στην ανακάλυψη οδών διάδοσης σημάτων στα κύτταρα ενός οργανισμού [28], [145], [33], [20]. Η προσπάθεια αναζήτησης εισβολέων σε ένα σύστημα [9] είχε ως κίνητρο την αναζήτηση χαμένων ή κρυμμένων ατόμων σε ένα σύστημα σπηλαίων. Η αναζήτηση έχει καθοριστικό ρόλο και στη δρομολόγηση πακέτων πληροφορίας ή στη δημιουργία μοντέλων κίνησης σε δίκτυα επικοινωνιών [38]. Μπορεί επίσης να χρησιμοποιηθεί για την εύρεση δικτυακών αρχιτεκτονικών που ελαχιστοποιούν την κυκλοφορική συμφόρηση [79].

Πέρα από την θεωρητική μελέτη της αναζήτησης, τα τελευταία χρόνια δόθηκε ιδιαίτερη έμφαση στην διερεύνησή της μέσω μοντελοποίησης και προσομοίωσης. Βασικός πυρήνας του μοντέλου είναι η τοπολογία του δικτύου που κωδικοποιεί και προβάλλει τις αλληλεπιδράσεις μεταξύ των συστατικών του [89], [145], [190]. Σήμερα, προκειμένου να αποκτηθεί μια πιο σφαιρική εικόνα, αξιοποιείται η δυνατότητα συλλογής και επεξεργασίας πληθώρας δεδομένων από τις αλληλεπιδράσεις που συμβαίνουν σε ένα δίκτυο. Αυτό αλλάζει τον τρόπο κατανόησης, προσέγγισης και συσχέτισης της συμπεριφοράς δυναμικών διεργασιών με την τοπολογία του δικτύου. Επιπλέον, τα συλλεγόμενα δεδομένα χρησιμοποιούνται για την επιβεβαίωση αναλυτικών αποτελεσμάτων, την απόκτηση χρήσιμων γνώσεων ή την ανακάλυψη κρυμμένων ιδιοτήτων. Τα δίκτυα του πραγματικού κόσμου εμφανίζουν τυπικά στατιστικά χαρακτηριστικά των σύνθετων δικτύων [7], [147], [28], [47], [45]. Παρ’

όλη την τυχαιότητα στις αλληλοσυσχετίσεις των κόμβων, οι αρχές οργάνωσης και δημιουργίας συνδέσεων οδηγούν σε δομές που επηρεάζουν σε μεγάλο βαθμό την εξέλιξη και την ανταπόκριση της αναζήτησης [157], [20]. Η δυσκολία της αναζήτησης είναι σύμφυτη με τη δομή του δικτύου [47] και ποσοτικοποιείται ως συνάρτηση των πληροφοριών που σχετίζονται με τον εντοπισμό ενός συγκεκριμένου στόχου [166].

3.2 Επισκόπηση βιβλιογραφίας

Το πρόβλημα της αναζήτησης απασχολεί την ερευνητική κοινότητα μέχρι και σήμερα, όχι μόνο για να κατανοήσει τη συμπεριφορά της, αλλά και για να αναλύσει τους παράγοντες που διευκολύνουν ή παρεμποδίζουν την αποτελεσματικότητά της. Όπως προαναφέρθηκε, το πρόβλημα της αναζήτησης είναι συνυφασμένο με το δίκτυο, τη δομή του, τη διαθεσιμότητα του περιεχομένου/πόρων και τη στρατηγική (μέθοδο) αναζήτησης.

Στα δομημένα δίκτυα, η διαδικασία της αναζήτησης είναι καθορισμένη από την αρχιτεκτονική του δικτύου, καθώς είναι γνωστή η θέση τόσο των κόμβων όσο και των δεδομένων, συνήθως από τον πίνακα κατακερματισμού που εφαρμόζουν (CAN [161], Tapestry [206], Chord [179], Pastry [167], Kademlia [134], και Viceroy [125]). Στα χωρικά δίκτυα, που όπως και στα δομημένα η θέση του κόμβου στόχου είναι γνωστή, η βέλτιστη αναζήτηση περιλαμβάνει την κατεύθυνση του ταξιδιού και τον αριθμό των συνδέσεων του γειτονικού κόμβου [182].

Όμως, πολλά πραγματικά δίκτυα είναι γενικά αδόμητα. Οι προτεινόμενες στρατηγικές, εφαρμόζοντας κλασσικούς αλγορίθμους αναζήτησης, προσπαθούν είτε να βρουν μια διαδρομή από έναν κόμβο σε έναν άλλο (*graph search*) είτε να διερευνήσουν συστηματικά όλες τις πιθανές διαδρομές για να φθάσουν στο αποτέλεσμα (*tree search*).

Το διάσημο πείραμα του Milgram [138] να παραδώσουν συνηθισμένοι άνθρωποι μηνύματα σε συγκεκριμένους ανθρώπους μέσα από ένα δίκτυο γνωριμιών οδήγησε σε δυο βασικά συμπεράσματα που παρατηρούνται στα κοινωνικά δίκτυα. Το πρώτο, είναι ότι υπάρχουν σύντομες διαδρομές ακόμη και σε πολύ μεγάλα σε μέγεθος δίκτυα. Το δεύτερο, είναι ότι οι άνθρωποι μπορούν να εντοπίσουν αυτές τις διαδρομές χρησιμοποιώντας μόνο τοπική πληροφορία, ή με άλλα λόγια το δίκτυο προσφέρεται

για αναζήτηση (*searcheable*).

Οι Watts και Strogatz [197] αποδίδουν αυτή τη συμπεριφορά στο φαινόμενο του “μικρού - κόσμου” και το μοντέλο που προτείνουν τοποθετεί σύντομες διαδρομές (*shortcuts*) μεταξύ δυο τυχαία επιλεγμένων κόμβων. Αργότερα ο Dodds κ.ά. [57] επιβεβαιώνουν τα αποτελέσματα του Milgram με ένα νέο πείραμα που βασίζεται στο διαδίκτυο και καταλήγουν πως το φαινόμενο του μικρού κόσμου από μόνο του δεν αρκεί για επιτυχή αναζήτηση, αλλά εξαρτάται από τα ατομικά κίνητρα (έναν κόμβος μπορεί να σταματήσει ή να προωθήσει μια αναζήτηση) και την τοπική γνώση.

Ο Kleinberg [101], [98], [99] προσεγγίζοντας θεωρητικά τις δυνατότητες αναζήτησης και παραφράζοντας την διατύπωση των Watts και Strogatz προτείνει τη μεταβολή της δομής του δικτύου με την προσθήκη σύντομων διαδρομών που κωδικοποιούν γεωγραφική πληροφορία για το υφιστάμενο δίκτυο. Η πιθανότητα να συνδεθούν δυο κόμβοι, τοποθετημένοι σε ένα διδιάστατο πλέγμα $d \times d$ με μια συντόμευση είναι ανάλογη του $r^{-\alpha}$, όπου r η απόσταση μεταξύ των κόμβων και α σταθερά, παράμετρος του μοντέλου. Σε αυτό το μοντέλο υπάρχει ένα σημείο απότομης μετάβασης από πλέγμα σε δίκτυο μικρού-κόσμου, για $\alpha = d$, όπου d η διάσταση του πλέγματος. Η αναζήτηση σε αυτή την περίπτωση εμφανίζει την καλύτερη απόδοση, ενώ τα δίκτυα αυτής της κατηγορίας καλούνται *searchable*.

Ο Watts κ.ά. [196] προτείνουν ένα ιεραρχικό μοντέλο βασιζόμενο στο μοντέλο του Kleinberg: τα άτομα οργανώνονται ιεραρχικά σε ομάδες σύμφωνα με την κοινωνική τους ταυτότητα. Βέβαια μια τέτοια ομαδική ιεράρχηση στα πραγματικά δίκτυα είναι δύσκολο να επιτευχθεί.

Στα πραγματικά άνευ κλίμακας δίκτυα, υπάρχουν κοινότητες (μικροί κόσμοι) [68], [117] που συνδέονται μεταξύ τους με σύντομες διαδρομές. Όμως, το ερώτημα είναι πως θα βρεθούν και θα χρησιμοποιηθούν αυτές οι διαδρομές από μια στρατηγική αναζήτησης [158], [3]. Οι Adamic κ.ά. [4] προτείνουν ως λύση τη χρήση τοπικής πληροφορίας που θα προδιαθέτει (*bias*) έναν τυχαίο περιπατητή προς τον κόμβο με το μεγαλύτερο αριθμό συνδέσεων, επιτυγχάνοντας έτσι καλύτερη απόδοση σε σχέση με τον εντελώς τυχαίο περιπατητή. Αν συμπεριληφθούν και επιπλέον τοπικές πληροφορίες, όπως η ταυτότητα και η συνεκτικότητα των γειτόνων, τότε η απόδοση βελτιώνεται ακόμη περισσότερο [3]. Ο βαθμός επέκτασης του δικτύου [123], ή η επιλογή επόμενου κόμβου από το σύνολο των γειτονικών που δεν έχουν ακόμη ερωτηθεί [201] φάνηκε και πειραματικά ότι διευκολύνουν την αναζήτηση και μειώνουν το κόστος σε δίκτυα άνευ κλίμακας. Αντίθετα, μικρή είναι η βελτίωση που

παρατηρείται σε τυχαία δίκτυα (Erdős-Rényi) [201], [3], [123].

Μπορεί το ζήτημα της αναζήτησης να απασχόλησε αρχικά τα κοινωνικά δίκτυα, η εξέλιξη όμως της τεχνολογίας έδωσε νέα ώθηση και κίνητρα για έρευνα. Δίκτυα όπως ο παγκόσμιος ιστός και τα δίκτυα p2p διαθέτουν υπηρεσίες, πόρους, πληροφορίες που αποθηκεύεται στους κόμβους τους. Για να έχουν πρακτική αξία και για να μπορέσουν οι χρήστες να αποκτήσουν πρόσβαση σε συγκεκριμένα στοιχεία από τον τεράστιο όγκο δεδομένων μετέρχονται διαδικασίες αναζήτησης.

Σχετικό ζήτημα μπορεί να θεωρηθεί και η αναζήτηση σελίδων στον παγκόσμιο ιστό χωρίς καθολική γνώση του δικτύου. Για τον εντοπισμό πληροφορίας στο διαδίκτυο αναπτύχθηκαν εξειδικευμένα εργαλεία αναζήτησης [127], [31], όμως πίσω από αυτά κρύβεται η εξαντλητική αναζήτηση και οι τυχαίοι περιπατητές. Για παράδειγμα, ένας χρήστης ξεκινώντας από μια τυχαία ιστοσελίδα ακολουθεί τους υπερσυνδέσμους ώστε να εντοπίσει την επιθυμητή πληροφορία. Εναλλακτικά θα υποβάλει ένα ερώτημα σε μια μηχανή αναζήτησης, που απαντά εξάγοντας αποτελέσματα από μια βάση δεδομένων που έχει συγκεντρώσει ύστερα από διεξοδική αναζήτηση (crawling).

Εκλεπτυσμένες προσεγγίσεις με ή χωρίς πληροφόρηση έχουν προταθεί, με σκοπό την επίτευξη αποδοτικών τεχνικών αναζήτησης σε δίκτυα p2p. Στις τυφλές (χωρίς πληροφόρηση) μεθόδους, κλασσικό παράδειγμα αποτελεί η στρατηγική της πλημμύρας που βασίζεται στον breadth-first search (BFS) αλγόριθμο. Ξεκινώντας από έναν κόμβο του δικτύου, η ερώτηση προωθείται σε όλους τους κόμβους. Αν και εγγυάται την εύρεση λύσης το πλήθος των μηνυμάτων που εισάγονται στο δίκτυο αυξάνεται εκθετικά με το χρόνο. Περιορίζοντας το βάθος του δέντρου αναζήτησης, η iterative deepening [199] στρατηγική δίνει μια μερική λύση στο πρόβλημα. Ο Akiba κ.ά. [6] θεωρούν, πως μπορούν να υπολογιστούν εκ των προτέρων όλες οι σύντομες διαδρομές σε ένα δίκτυο, εφαρμόζοντας BFS σε κάθε κόμβο. Η βελτίωση επιτυγχάνεται με την μείωση του χώρου αναζήτησης καθώς εφαρμόζεται τεχνική κλαδέματος κατά την διάρκεια της BFS αναζήτησης.

Εναλλακτικά εφαρμόζονται τυχαίοι αλγόριθμοι αναζήτησης όπως ο τυχαίος περιπατητής, και η πιθανοτική προώθηση. Ο τυχαίος περιπατητής, λόγω της στοχαστικής συμπεριφοράς του εφαρμόζεται ευρέως [76], [152], [205], [14] και με πολλές παραλλαγές. Στην βασική του έκδοση, επιλέγει τυχαία μια σύνδεση για να μεταβεί σε επόμενο κόμβο, εφόσον δεν βρει το αποτέλεσμα, σχηματίζοντας έτσι ένα τυχαίο μονοπάτι μέσα στο δίκτυο. Ύστερώντας ως προς το μέσο αριθμό βημάτων

(μεγαλύτερη απόσταση) σε σχέση με την πλημμύρα, υπερτερεί καθώς εισάγει πολύ λιγότερα μηνύματα. Ο αλγόριθμος της πιθανοτικής προώθησης εξετάζει για κάθε γειτονικό κόμβο, αν η τιμή ενός τυχαίου αριθμού $x \in [0, 1]$ είναι μεγαλύτερη από μια προκαθορισμένη τιμή κατωφλίου προκειμένου να διερευνήσει τον κόμβο αυτόν ή αλλιώς να τον αγνοήσει. Η πιθανοτική πλημμύρα προωθεί την ερώτηση μόνο σε ένα ποσοστό p_f γειτονικών κόμβων κάθε φορά. Αξιοποιώντας τοπική πληροφορία για την προσαρμογή της p_f προτάθηκαν ιδιαίτερα αποδοτικές στρατηγικές [95]. Η στρατηγική gossip, εμπνευσμένη από τη μορφή του σχολιασμού στα κοινωνικά δίκτυα, επιλέγει να προωθήσει την ερώτηση μόνο στους κόμβους που θεωρεί ότι δεν την έλαβαν ακόμη [74], [66].

Η πιθανοτική πλημμύρα απαντάται και σε στρατηγικές με πληροφόρηση. Οι intelligent-BFS [95] και directed-BFS [95] επιλέγουν να προωθήσουν το ερώτημα στους κόμβους που συγκέντρωσαν τις περισσότερες θετικές απαντήσεις για παρόμοιες ερωτήσεις στο πρόσφατο παρελθόν, σύμφωνα με στατιστικές παρατηρήσεις. Η υβριδική περιοδική πλημμύρα [209] επιχειρεί να βελτιώσει την απόδοση επιλέγοντας γειτονικούς κόμβους με βάση πολλαπλές μετρήσεις. Οι τυχαίοι περιπατητές με τοπική πληροφόρηση αποφέρουν καλύτερη απόδοση σε σχέση με τον απλό περιπατητή. Η τοπική γνώση καθοδηγεί τον περιπατητή προς τους κόμβους με τον μεγαλύτερο αριθμό συνδέσεων [4] ή προς έναν συγκεκριμένο κόμβο ανάλογα με τη δραστηριότητά του [159] ή εξετάζοντας το βάρος των συνδέσεων [205] ή το μέσο χρόνο επίσκεψης ενός κόμβου [152].

Μια άλλη κατηγορία στρατηγικών με πληροφόρηση, βασίζεται στην αξιολόγηση των απαντήσεων του πρόσφατου ιστορικού το οποίο διατηρούν και ενημερώνουν με σχετική πληροφορία. Στην Adaptive Probabilistic Search (APS) [184] η κατεύθυνση προώθησης εκτιμάται βάσει προηγούμενων θετικών ή αρνητικών απαντήσεων. Η [155] αποφεύγει να προωθεί τα μηνύματα στις ακμές που στατιστικά παρατηρούνται τα περισσότερα διπλότυπα μηνύματα. Routing indices [49], local indices [199] και query routing protocol [107], [165] είναι μερικά ακόμη παραδείγματα στρατηγικών βασιζόμενων σε ανατροφοδότηση από τις πρόσφατες αποκρίσεις.

Με στόχο την ενίσχυση της απόδοσης, άλλες στρατηγικές αξιοποιούν νέες δομές δεδομένων όπως τα φίλτρα bloom. Για παράδειγμα, στα φίλτρα bloom συνοψίζεται πληροφορία σχετικά με τη διαθεσιμότητα πόρων σε γειτονικούς κόμβους [163] ή κωδικοποιείται πιθανοτική πληροφορία που υποδηλώνει τη θέση του αντικειμένου [107].

3.3 Στρατηγικές αναζήτησης: λύσεις και προκλήσεις

Η διαδικασία της αναζήτησης σε σύνθετα δίκτυα περιλαμβάνει τα ακόλουθα βασικά στοιχεία:

- Πηγή (source (s)): Ο κόμβος που εισάγει την ερώτηση.
- Στόχος (target(t)): Ο κόμβος που αναζητείται ή αποθηκεύει ή παρέχει το αναζητούμενο στοιχείο. Μπορεί να μην είναι μοναδικός. Ανάλογα με την εφαρμογή, ενδιαφερόμαστε για την εύρεση είτε τουλάχιστον ενός είτε όλων.
- Στρατηγική (f_s): Ο αλγόριθμος που θα εφαρμοσθεί για τη δρομολόγηση των μηνυμάτων αναζήτησης.
- Χώρος/Πεδίο (S_s): Περιλαμβάνει το σύνολο των κόμβων που πρέπει να προσελαστούν στη χειρότερη περίπτωση, ώστε να δοθεί απάντηση και είναι $S_s \subseteq V$.

Ένας κόμβος s προσπαθεί να εντοπίσει έναν κόμβο-στόχο t εφαρμόζοντας μια στρατηγική f_s στο σύνολο ή σε μέρος των κόμβων S_s του δικτύου. Όπως αναφέρθηκε και στην Ενότητα 3.1, η στρατηγική αναζήτησης καθορίζει τη μετάβαση ή προώθηση ενός μηνύματος από κόμβο σε κόμβο και ο ρόλος της είναι καθοριστικός για την έκβαση της αναζήτησης: αν θα είναι επιτυχής ή όχι. Οι συνδέσεις (ακμές) που χρησιμοποιούνται για τη μετάβαση του μηνύματος-ερώτησης από έναν κόμβο u σε έναν γειτονικό του κόμβο v διακρίνονται σε¹: *ακμή δέντρου*, αν ο κόμβος v επισκέπτεται για πρώτη φορά διαμέσου της ακμής (u, v) , *πίσω ακμή*, η ακμή (v, u) , αν ο u είναι πρόγονος του v στο δέντρο αναζήτησης, *εμπρός ακμή*, η ακμή (v, u) , αν ο u είναι απόγονος του v στο δέντρο αναζήτησης, *διασυνδετική ακμή* κάθε άλλη ακμή. Το μήνυμα-ερώτηση μεταδίδεται σε μέρος ή στο σύνολο των κόμβων του δικτύου. Ορίζουμε ως *κάλυψη* του δικτύου τον αριθμό των διαφορετικών κόμβων που έχουν λάβει το μήνυμα-ερώτηση, μέχρι το ένα ορισμένο βήμα της διαδικασίας.

Μια στρατηγική αναζήτησης που μπορεί να καθοδηγείται από καθολική ή τοπική ή καθόλου πληροφορία, που μπορεί να εφαρμόζει μηχανισμούς ευρείας εκπομπής (broadcast) ή ανταλλαγής ιδιωτικών μηνυμάτων (message passing) που μπορεί να είναι στατική ή δυναμική αξιολογείται με βάση τα ακόλουθα ερωτήματα: *είναι καλή στρατηγική (αποτίμηση);* *πότε θα σταματά να εκτυλίσσεται (τερματισμός);*

¹η διάκριση γίνεται με βάση την αναζήτηση κατά πλάτος (BFS [30])

3.3.1 Αποτίμηση αναζήτησης

Για την εξαγωγή συμπερασμάτων για την καταλληλότητα και την αποδοτικότητα μιας στρατηγικής αναζήτησης κατασκευάζεται ένα γενικευμένο αφαιρετικό μοντέλο που βασίζεται σε παρατηρήσεις στη διάρκεια της εξέλιξής της. Το μοντέλο διέπεται από τις ακόλουθες παραδοχές:

- Η αναζήτηση εξαπλώνεται από κόμβο σε κόμβο ακολουθώντας τις μεταξύ τους συνδέσεις. Κάθε μετάβαση αποτελεί ένα βήμα και θεωρείται ότι γίνεται σε μια χρονική μονάδα.
- Κάθε κόμβος συνεργάζεται για να υποστηρίξει τη διαδικασία αναζήτησης. Υποθέτουμε δηλαδή, ότι δεν υπάρχουν κόμβοι με εγωιστική συμπεριφορά που δεν είναι διατεθειμένοι να διαβιβάσουν ένα μήνυμα.
- Ένας κόμβος παύει να προωθεί ένα ερώτημα μόνο όταν πληρούται τουλάχιστον μια συνθήκη τερματισμού.

Στόχος της κάθε στρατηγικής αναζήτησης είναι να αυξήσει την πιθανότητα εύρεσης του επιθυμητού στοιχείου επιλέγοντας από όλες τις δυνατές καταστάσεις, τις καλύτερες, εκείνες δηλαδή, που αυξάνουν την αποτελεσματικότητά της.

Ένα ευρέως χρησιμοποιούμενο μέτρο για την ποσοτικοποίηση της αποτελεσματικότητας είναι ο μέσος αριθμός βημάτων (δηλαδή το μήκος της διαδρομής) που απαιτείται για την προσέγγιση μιας συγκεκριμένης τοποθεσίας [122], [172], [20], [26]. Το μέσο μήκος διαδρομής ή ο μέσος χρόνος μετάβασης από μια πηγή (s) σε ένα στόχο (t) εξαρτάται από τον τύπο, το μέγεθος, τον μέσο βαθμό του δικτύου και τη διαθεσιμότητα του περιεχομένου. Εναλλακτικά, ή και σε συνδυασμό με τον αριθμό βημάτων χρησιμοποιείται μια ακόμη μέτρηση για την αποτελεσματικότητα, ο μέσος αριθμός μηνυμάτων που μεταδίδονται στο δίκτυο. Ένα ακόμη μέτρο είναι το ποσοστό επιτυχίας [185] που εκφράζει το ποσοστό των ερωτημάτων από τα συνολικά που απαντήθηκαν επιτυχώς. Ωστόσο το ζήτημα της απόδοσης είναι ιδιαίτερα περίπλοκο καθώς μετριάζεται από ποικίλους παράγοντες όπως: το εύρος ζώνης, την ενέργεια, το φόρτο του δικτύου, τη δυνατότητα αποθήκευσης [122], [194] με αρκετά από αυτά να σχετίζονται άμεσα με το παραγόμενο πλήθος των μηνυμάτων.

Καθώς είναι σχεδόν αδύνατο να χρησιμοποιηθούν όλοι αυτοί οι παράγοντες για την αξιολόγηση της αναζήτησης, στην παρούσα μελέτη εξετάζονται τα κυριότερα από αυτά: ο μέσος αριθμός μηνυμάτων και το ποσοστό επιτυχίας σε συνάρτηση με

τον αριθμό βημάτων. Τα μηνύματα που παράγει η διαδικασία, αλλά δεν συμβάλλουν στην επιτυχία της αναζήτησης (πλεονάζοντα μηνύματα) προκύπτουν από:

1. την επαναληπτική αποστολή μηνυμάτων ερώτησης (π.χ. expanding ring),
2. από την υπερκάλυψη (overshooting) [92] του χώρου αναζήτησης ως αποτέλεσμα του εκτιμώμενου κριτηρίου τερματισμού (περισσότερες λεπτομέρειες δίνονται στην επόμενη ενότητα (3.3.2)),
3. τα διπλότυπα μηνύματα, τα μηνύματα δηλαδή που αφορούν την ίδια ερώτηση και φθάνουν πολλές φορές σε έναν κόμβο από διαφορετικές διαδρομές-μονοπάτια του δικτύου.

Η υπερκάλυψη οφείλεται στα πακέτα που φθάνουν σε έναν κόμβο και χωρίς να είναι διπλότυπα, δεν συμβάλλουν στην απάντηση της ερώτησης. Για παράδειγμα αυτό μπορεί να συμβεί όταν τα επιθυμητά αποτελέσματα έχουν εντοπισθεί, αλλά η αναζήτηση συνεχίζεται σε άλλα μονοπάτια (π.χ. πλημμύρα). Το πρόβλημα της πλεονάζουσας παραγωγής μηνυμάτων, αν και έχει επισημανθεί από μεγάλη μερίδα ερευνητών, δεν έχει μελετηθεί διεξοδικά. Το πλήθος διπλότυπων μηνυμάτων μπορεί να φθάσει έως και $N(d - 2)$ σε ένα δίκτυο με N κόμβους και μέσο βαθμό d [121], [155]. Η μείωσή τους μπορεί να επιτευχθεί με τεχνικές βασιζόμενες στην ανατροφοδότηση [155], στην ελεγχόμενη εξάπλωση στο χώρο αναζήτησης και στην αύξηση διαθεσιμότητας των πόρων [92].

Στα πλαίσια αυτής της διατριβής, η αξιολόγηση της απόδοσης προσδιορίζεται από τα ακόλουθα κριτήρια σε σχέση με την απόσταση σε βήματα που απαιτείται για την ολοκλήρωσή της:

- Το ποσοστό επιτυχίας (success rate)
- Το πλήθος των κόμβων που εντοπίζονται σε κάθε βήμα της αναζήτησης (n_t)
- Το πλήθος των συνολικών μηνυμάτων που στέλνονται σε κάθε βήμα της αναζήτησης (m_t)
- Το πλήθος των διπλότυπων μηνυμάτων που στέλνονται σε κάθε βήμα ($dups_t$).

3.3.2 Κριτήρια τερματισμού

Η αναζήτηση είναι μια επαναληπτική διαδικασία καθώς, ο κάθε ενδιαμέσος κόμβος που λαμβάνει ένα ερώτημα καλείται εφαρμόζοντας κάποια στρατηγική να προωθήσει το μήνυμα που έλαβε. Η διαδικασία τερματίζεται όταν ικανοποιείται κάποιο από τα κριτήρια τερματισμού, που μπορεί να εφαρμόζονται μεμονωμένα ή σε συνδυασμό:

- *κριτήριο ικανοποίησης*, όταν βρεθεί ικανοποιητικός αριθμός αποτελεσμάτων (π.χ. τυχαίος περιπατητής) [122].
- *μέγιστης απόστασης*, όπου η αναζήτηση σταματά μετά από έναν συγκεκριμένο αριθμό βημάτων. Είναι συνήθης πρακτική σε στρατηγικές βασιζόμενες στον τυχαίο περιπατητή, στην πλημμύρα ή σε συγκεκριμένες προτεινόμενες τεχνικές (π.χ. *iterative deepening* [199]). Μια παράμετρος *time-to-live (TTL)* προσαρτάται στο μήνυμα και σε κάθε βήμα μειώνεται. Με την εκπνοή της, όταν φθάσει δηλαδή στην τιμή μηδέν, σταματά η μετάδοση του μηνύματος.
- *τιμή κατωφλίου*, όπου εφαρμόζοντας κάποια ευρετική μέθοδο, αποτιμάται μια συνεχώς μειούμενη παραμετρική ή πιθανοτική συνάρτηση που στηρίζεται σε τοπική πληροφορία. Όταν η τιμή της, που επανεκτιμάται σε κάθε βήμα, πέσει κάτω από ένα όριο, η αναζήτηση σταματά.

Η επιλογή κριτηρίου τερματισμού έχει καθοριστική σημασία, αφού μπορεί να οδηγήσει σε καλύτερη συνολική απόδοση [122], [199], [92], [133].

3.4 Βελτίωση αναζήτησης

Για να είναι επιτυχημένο ένα σχήμα αναζήτησης απαιτεί τη χρήση όσο το δυνατό λιγότερου χώρου για αποθήκευση και την επεξεργασία και τη μετάδοση μικρού αριθμού μηνυμάτων ανά βήμα. Όμως, η σχεδίαση ενός τέτοιου μηχανισμού είναι δύσκολη διαδικασία για πολλούς λόγους, όπως η μεγάλη κλίμακα των σημερινών δικτύων, η ανάγκη για προσαρμογή σε ετερογενή (στατικά, δυναμικά) δίκτυα και ο χειρισμός των διπλότυπων μηνυμάτων. Στη συνέχεια παρουσιάζονται προσεγγίσεις που επιχειρούν να διαπραγματευθούν αυτά τα ζητήματα.

3.4.1 Αλλαγή τοπολογίας

Οι περισσότεροι ερευνητές επισημαίνουν το σημαντικό ρόλο της τοπολογίας τόσο στην απόδοση όσο και στην αποτελεσματικότητα της αναζήτησης [101], [98], [4] [99]. Οι επισημάνσεις αυτές ήταν το κίνητρο για την ανάπτυξη πρωτοκόλλων που μεταβάλουν την τοπολογία του δικτύου προκειμένου να διευκολύνεται η αναζήτηση. Για το σκοπό αυτό εκμεταλλεύονται διάφορα τοπολογικά χαρακτηριστικά, όπως για παράδειγμα, τον βαθμό, τον συντελεστή ομαδοποίησης και τη σχέση ομοιότητας (π.χ. κοινά ενδιαφέροντα). Οι προσεγγίσεις αυτές μπορούν να ταξινομηθούν σε τρεις διαφορετικές κατηγορίες [137]: υβριδικές, υπερκόμβων και ομαδοποίησης.

Οι υβριδικές τεχνικές συνδυάζουν δομημένες και αδόμητες αρχιτεκτονικές ώστε να βελτιωθεί η αναζήτηση (κλιμάκωση) και να μειωθεί η επιβάρυνση από διπλότυπα μηνύματα [80]. Στην δεύτερη κατηγορία εντάσσονται τεχνικές που εισάγουν κόμβους με διακεκριμένο ρόλο, τους λεγόμενους “υπερκόμβους”. Οι πρώιμες μελέτες κατέδειξαν την επίδραση των κόμβων με υψηλή συνδεσιμότητα στην πλοήγηση και την αναζήτηση σε ένα δίκτυο. Τέτοιοι κόμβοι, που είναι επιπλέον αξιόπιστοι μπορούν να επιλεγούν ως υπερκόμβοι για μια ομάδα κόμβων. Οι υπερκόμβοι συνδέονται μεταξύ τους και διαμεσολαβούν για να μεταφέρουν μηνύματα από την ομάδα τους στους άλλους υπερκόμβους και το αντίστροφο. Κλασσικό παράδειγμα εφαρμογής αποτελεί η Gnutella 0.6 [174].

Οι τεχνικές ομαδοποίησης έχουν ως στόχο τη δημιουργία κοινοτήτων από κόμβους με κοινά χαρακτηριστικά ή ενδιαφέροντα. Τέτοιες οργανώσεις, που βασίζονται στις προτιμήσεις των χρηστών, επιφέρουν σημαντικές βελτιώσεις στην αποτελεσματικότητα και αποδοτικότητα της αναζήτησης, μειώνοντας το φόρτο κίνησης και το χρόνο αναζήτησης [176], [50], [43], [86], [191].

Τέλος, υπάρχουν κάποιες άλλες προσεγγίσεις που δεν εντάσσονται σε καμία από τις παραπάνω κατηγορίες. Ο Cooper [46] θεωρεί ότι μια τοπολογία είναι βέλτιστη για την αναζήτηση, αν ο βαθμός των κόμβων είναι ανάλογος της τετραγωνικής ρίζας της δημοτικότητας των περιεχομένων του. Ο Hota κ.ά. [85] προτείνει στη δόμηση της τοπολογίας (π.χ. τυχαίου γράφου, άνευ κλίμακας) να προστίθενται σε κάθε κόμβο m επιπλέον γειτονικοί κόμβοι, επιλεγμένοι κατά προτίμηση από το x^2 -τεστ² ως μέτρο ομοιότητας. Εφαρμόζοντας και μια ευρετική μέθοδο αναζήτησης επιτυγχάνει

²Είναι ένα στατιστικό τεστ που χρησιμοποιείται για να εξετάσει αν η σχέση ανάμεσα σε δύο μεταβλητές είναι ή δεν είναι στατιστικά σημαντική

κατανομή φόρτου και μείωση του κόστους.

3.4.2 Αντιγραφή (Replication)

Η διατήρηση πολλαπλών αντιγράφων ενός στοιχείου δεδομένων (Αντιγραφή - *Replication*) σε πολλαπλές θέσεις (κόμβους) έχει ως στόχο τη βελτίωση της απόδοσης της αναζήτησης [181], [183], [5], [175]. Επιπλέον η αντιγραφή συμβάλει στη διαθεσιμότητα, την αξιοπιστία και την κλιμάκωση του συστήματος. Με τη δημιουργία αντιγράφων (*replicas*), ο φόρτος του συστήματος ως αποτέλεσμα της αναζήτησης όχι μόνο κατανέμεται στους κόμβους του δικτύου και αποφεύγεται η συμφόρηση, αλλά μπορεί να μειωθεί σε βέλτιστο βαθμό [181]. Η μείωση αυτή οφείλεται σε δύο λόγους: μειώνεται ο χρόνος αναζήτησης καθώς αντίγραφα βρίσκονται “πιο κοντά” στην πηγή της ερώτησης, μειώνεται η επιβάρυνση από την κίνηση και την επεξεργασία των ερωτήσεων καθώς ο εντοπισμός της απάντησης οδηγεί στον τερματισμό της διαδικασίας.

Η αντιγραφή δεδομένων έχει να επιλύσει σύνθετα ζητήματα, όπως, ποιος θα είναι αριθμός αντιγράφων και πως θα γίνεται η ενημέρωσή τους, που θα τοποθετηθούν/κατανεμηθούν τα αντίγραφα.

Η διατήρηση της συνέπειας των αντιγράφων αποτελεί σχεδιαστική πρόκληση για τα πρωτόκολλα διαχείρισης αντιγράφων. Τα πρωτόκολλα αυτά καλούνται να αποφασίσουν για το *που* θα γίνουν οι ενημερώσεις (*single-master* vs. *multi-master*), για το *πότε* θα γίνουν οι ενημερώσεις (*synchronous* vs. *asynchronous*) και για το *πως* θα γίνουν οι ενημερώσεις (*full* vs. *partial replication*) [78], [5].

Για να απαντήσουν στο δεύτερο ερώτημα αρκετοί ερευνητές εφαρμόζουν μια κατανομή αντιγράφων (*replica distribution*) από την οποία προκύπτει ο αριθμός των αντιγράφων. Ο Cohen κ.ά. [44] εξετάζει τρεις στρατηγικές που καθορίζουν τον αριθμό των αντιγράφων ώστε να μειωθεί η επιβάρυνση από τα μηνύματα της αναζήτησης: αντιγραφή *uniform*, *proportional* και *square-root*. Στην αντιγραφή *uniform* ο αριθμός των αντιγράφων είναι ίδιος για κάθε στοιχείο δεδομένων. Στην αντιγραφή *proportional* ο αριθμός των αντιγράφων είναι ανάλογος της δημοτικότητας του στοιχείου δεδομένων. Στην αντιγραφή *square-root* ο αριθμός των αντιγράφων είναι ανάλογος της τετραγωνικής ρίζας της πιθανότητας ζήτησης του πόρου. Όπως αποδείχθηκε [44], [122] οι δυο πρώτες στρατηγικές δημιουργίας αντιγράφων σε συνδυασμό με στρατηγικές αναζήτησης όπως τυχαία αναζήτηση, τυχαίος περιπατητής

και πλημμύρα έχουν την παρόμοια συμπεριφορά ως προς το μέγεθος/χώρο αναζήτησης. Όμως η αντιγραφή square-root επιτυγχάνει βέλτιστη κατανομή αντιγράφων και ελαχιστοποιεί τον αναμενόμενο χώρο αναζήτησης. Οι ίδιες μελέτες υποδεικνύουν ως καταλληλότερο δικτυακό μοντέλο για αναζήτηση το μοντέλο Erdős-Rényi. Για το μοντέλο Erdős-Rényi αποδείχθηκε [181] ότι η καλύτερη στρατηγική αντιγραφής είναι η proportional αν ο αντικειμενικός στόχος είναι η μείωση του χρόνου αναζήτησης όταν εφαρμόζεται η πλημμύρα.

Εμπνευσμένος από την square-root αντιγραφή ο Leontiadis κ.ά. [112] προτείνει την αντιγραφή Pull-Then-Push. Η δημιουργία αντιγράφων ανατίθεται στον κόμβο-πηγή της ερώτησης και αποτελείται από δυο φάσεις: στην πρώτη (pull phase) επιχειρείται ο εντοπισμός ενός στοιχείου δεδομένων. Η δεύτερη φάση (push phase) που ακολουθεί μια επιτυχή αναζήτηση μεταδίδει το στοιχείο δεδομένων σε άλλους κόμβους του δικτύου εξαναγκάζοντας έτσι τη δημιουργία αντιγράφων. Ο αριθμός των κόμβων που διερευνώνται στη φάση pull είναι ίδιος με τον αριθμό των κόμβων που διερευνώνται στην φάση push.

Για την τοποθέτηση /κατανομή των αντιγράφων στους κόμβους του δικτύου τρεις είναι οι κύριες τεχνικές που εφαρμόζονται[122]: τοποθέτηση owner, τοποθέτηση path replication και τοποθέτηση random replication. Η τοποθέτηση owner μετά από μια επιτυχή αναζήτηση ο κόμβος στόχος (κάτοχος) αντιγράφει το στοιχείο δεδομένων στον κόμβο πηγή. Ο αριθμός των αντιγράφων ενός πόρου, που προκύπτει από αυτή τη διαδικασία, είναι ανάλογος των αιτημάτων για τον πόρο αυτό. Η τοποθέτηση owner βρίσκει εφαρμογή σε συστήματα όπως η Gnutella.

Η τοποθέτηση path, μετά από μια επιτυχή αναζήτηση δημιουργεί αντίγραφα του στοιχείου δεδομένων σε όλους τους ενδιαμέσους κόμβους της διαδρομής μεταξύ των κόμβων πηγής και στόχου. Βέβαια αυτό έχει ως αποτέλεσμα οι κόμβοι με μεγάλο αριθμό συνδέσεων να υπερφορτώνονται. Η τοποθέτηση random τοποθετεί τα αντίγραφα τυχαία στο δίκτυο. Συγκεκριμένα, μετά από μια επιτυχή αναζήτηση δημιουργεί τόσα αντίγραφα όσο και το μήκος της διαδρομής από την πηγή στον στόχο και τα τοποθετεί σε τυχαίες θέσεις επιλέγοντας από τους κόμβους που επισκέφθηκαν οι περιπατητές.

Πειραματικά αποδείχθηκε [122] ότι οι δύο τελευταίες στρατηγικές διανομής αντιγράφων υπερτερούν έναντι της τοποθέτησης owner ως προς το χρόνο αναζήτησης (σε βήματα) και επιτυγχάνουν μικρότερες καθυστερήσεις. Η τοποθέτηση random επηρεάζεται από την τοπολογία του δικτύου και έχει καλύτερες επιδόσεις

σε σχέση την τοποθέτηση path. Όμως η υλοποίησή της είναι πιο δύσκολη. Για περισσότερες λεπτομέρειες ο αναγνώστης μπορεί να ανατρέξει στο κεφάλαιο 25 του βιβλίου [5].

3.5 Σύνοψη

Σε αυτό το κεφάλαιο εστιάζουμε στη θεωρητική ανάλυση της αναζήτησης κάνοντας μια ανασκόπηση στη βιβλιογραφία. Κοινή επιδίωξη όλων είναι η επεξεργασία και η μετάδοση μικρού αριθμού μηνυμάτων ανά βήμα της διαδικασίας, ενώ ταυτόχρονα η αναζήτηση καταλήγει επιτυχώς. Επίσης, εδώ παρουσιάζονται και επισημαίνονται ζητήματα που εγείρονται με το σημαντικότερο το πλήθος των μηνυμάτων που παράγει η διαδικασία και μεταδίδονται στο δίκτυο, ενώ συνοψίζονται τεχνικές για την βελτίωσή της όπως, η αλλαγή της τοπολογίας και η καθιέρωση μιας πολιτικής διατήρησης αντιγράφων.

ΚΕΦΑΛΑΙΟ 4

ΑΝΑΖΗΤΗΣΗ ΜΕ ΠΛΗΜΜΥΡΑ ΣΕ ΑΔΟΜΗΤΑ ΔΙΚΤΥΑ: ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΠΛΕΟΝΑΖΟΥΣΑΣ ΠΛΗΡΟΦΟΡΙΑΣ

4.1 Εισαγωγή

4.2 Αδόμητα δίκτυα $p2p$ & πλημμύρα

4.3 Γενικά όρια επί των αναπόφευκτων διπλότυπων μηνυμάτων

4.4 Μοντελοποίηση απόδοσης πλημμύρας με χρήση της θεωρίας παιγνίων

4.5 Μοντέλο εκτίμησης κάλυψης του δικτύου

4.6 Επεκτείνοντας το μοντέλο

4.7 Σύνοψη

4.1 Εισαγωγή

Το πρόβλημα της αναζήτησης στα αδόμητα δίκτυα, δεδομένης της παντελούς έλλειψης πληροφορίας για τη δομή του δικτύου, προσεγγίζεται με στρατηγικές όπως η πλημμύρα και οι παραλλαγές της. Στην πλημμύρα, ο κόμβος που ξεκινάει την αναζήτηση, στέλνει ένα μήνυμα-ερώτηση στους γειτονικούς του κόμβους. Κάθε ένας από αυτούς, που δεν διαθέτει ή δεν γνωρίζει την απάντηση, προωθεί το μήνυμα στους γείτονές του. Η διαδικασία επαναλαμβάνεται μέχρι την επιτυχή έκβαση της

αναζήτησης ή την εκπλήρωση κάποιας συνθήκης τερματισμού.

Σε αυτή την εργασία επικεντρωνόμαστε στη μελέτη του προβλήματος της αναζήτησης σε αδόμητα δίκτυα $p2p$. Σημειώνουμε όμως, ότι τα αποτελέσματα μπορούν να βρουν εφαρμογή σε πολλούς άλλους τομείς σύνθετων δικτύων.

Ένα από τα βασικότερα ζητήματα στα αδόμητα δίκτυα $p2p$, είναι ο εντοπισμός ενός επιθυμητού πόρου. Ο πόρος είναι μια πηγή εφοδιασμού ή υποστήριξης [187], όπως για παράδειγμα αρχεία, μνήμη, υπολογιστική ικανότητα (CPU), κανάλια επικοινωνίας ή και υπηρεσίες [137].

Καθώς στα αδόμητα δίκτυα $p2p$ δεν υπάρχει καμιά σχέση μεταξύ της τοπολογίας και της τοποθέτησης των πόρων, όλες οι μέθοδοι αναζήτησης είναι κατά βάση τυφλές. Η πλημμύρα είναι ευρέως διαδεδομένη και το πρωτόκολλο Gnutella [164] είναι το πιο χαρακτηριστικό παράδειγμα εφαρμογής της.

Όπως είναι γνωστό, η πλημμύρα, από τη φύση της, παράγει τεράστιο αριθμό μηνυμάτων που μεταδίδονται στο δίκτυο. Έτσι, η διαδικασία αναζήτησης είναι ταχύτατη, αλλά την ίδια στιγμή είναι αρκετά δαπανηρή και χωρίς δυνατότητα κλιμάκωσης σε μεγαλύτερα δίκτυα. Η επίτευξη μεγαλύτερης κάλυψης σε ένα δίκτυο, επιφέρει καλύτερη απόδοση για την αναζήτηση, αλλά το δίκτυο υπερφορτώνεται από την υπερβολική κίνηση μηνυμάτων όπως θα εξηγήσουμε παρακάτω. Υπάρχουν πολλές εργασίες που επισημαίνουν την υπερβολική κίνηση που παράγεται από την πλημμύρα [92], [208], [56], [122] καθώς και πολλές άλλες, που προτείνουν στρατηγικές για τη μείωση της. Μερικές πρόσφατες εργασίες βασίζονται στην πιθανοτική πλημμύρα, όπου ένας κόμβος προωθεί την ερώτηση-μήνυμα στους γειτονικούς του κόμβους με μια δεδομένη πιθανότητα, σε τυχαία ή γενικευμένα τυχαία δίκτυα [153], [73], [52]. Ο Gaeta κ.ά. [71], [72] δημιούργησε αναλυτικά μοντέλα για να διαπιστώσει την επίδραση της τοπολογίας και να συγκρίνει σχήματα αναζήτησης βασιζόμενα στην πλημμύρα. Επιπλέον προτείνει αποτελεσματικές στρατηγικές πλημμύρας βασιζόμενες σε ρυθμιζόμενες παραμέτρους. Ο Gkantsidis [77] πρότεινε το συνδυασμό της πλημμύρας και των τυχαίων περιπατητών, ενώ άλλοι προσπαθούν να περιορίσουν τον χώρο αναζήτησης με την καθοδήγηση της αναζήτησης σε ορισμένες ακμές ή ομάδες κόμβων [207], [204], [192], [91], [155] ή χρησιμοποιούν αλγοριθμικές τεχνικές για να εξαλείψουν την υπερβολική κίνηση [193], [120].

Παρά τον μεγάλο αριθμό εργασιών που επιχειρούν να μειώσουν την υπερβολική κίνηση στο δίκτυο που σχετίζεται με την πλημμύρα, σύμφωνα με όσα είναι γνωστά, δεν υπάρχει μελέτη που να την ποσοτικοποιεί. Αυτό ακριβώς είναι το θέμα αυτής

της ενότητας. Ενδιαφερόμαστε να γνωρίσουμε πόσο σημαντική είναι η κίνηση που παράγουν τα μηνύματα της πλημμύρας κατά την διάρκειά της. Το ερώτημα αυτό αποτελεί το κίνητρο για την κύρια συνεισφορά αυτής της ενότητας: μια πρωτότυπη ανάλυση για τον όγκο μηνυμάτων που παράγει η απλή πλημμύρα. Πιο συγκεκριμένα, παράγουμε απλά αλλά γενικά όρια ως προς τον αριθμό των περιττών μηνυμάτων που παράγονται, βασιζόμενοι στον συνολικό αριθμό των μεταδιδόμενων μηνυμάτων και στους βαθμούς των κόμβων του δικτύου. Επιπλέον, παρουσιάζουμε ένα διαισθητικό μοντέλο που προσεγγίζει τον αριθμό των διπλότυπων μηνυμάτων που παράγει η διαδικασία της πλημμύρας σε κάθε βήμα.

Η συνεισφορά μας, συνοπτικά έχει ως εξής:

- υπολογίζει πάνω και κάτω όριο για το πλήθος των διαφορετικών κόμβων που ανακαλύπτονται (λαμβάνουν το μήνυμα της ερώτησης) από τη διαδικασία της αναζήτησης,
- υπολογίζει άνω όριο για τα διπλότυπα μηνύματα σε κάθε βήμα της αναζήτησης,
- προτείνει ένα προσεγγιστικό μοντέλο που υπολογίζει το πλήθος των νέων κόμβων (που λαμβάνουν το μήνυμα-ερώτηση για πρώτη φορά), των μηνυμάτων (συνολικά μηνύματα) και των διπλότυπων μηνυμάτων σε κάθε βήμα της αναζήτησης.

Το υπόλοιπο αυτού του κεφαλαίου είναι οργανωμένο ως εξής: στην Ενότητα 4.2 παρουσιάζονται οι λεπτομέρειες του προβλήματος που μελετούμε. Στην ενότητα 4.3 παράγουμε γενικά όρια για τον αριθμό των περιττών μηνυμάτων που παράγονται από την πλημμύρα. Στην Ενότητα 4.5 παρουσιάζεται ένα προσεγγιστικό μοντέλο για τον αριθμό των διπλότυπων μηνυμάτων και επικυρώνεται με προσομοιώσεις. Στην Ενότητα 4.6 διερευνώνται πιθανές επεκτάσεις του μοντέλου. Τέλος στην Ενότητα 4.7 συνοψίζονται τα κύρια ευρήματα του κεφαλαίου.

Μαζί με την ανάλυση παρέχονται πειραματικά αποτελέσματα βασιζόμενα σε έναν λεπτομερή προσομοιωτή που αναπτύχθηκε για αυτόν το σκοπό. Ο προσομοιωτής, υλοποιημένος σε γλώσσα C, λειτουργεί σε επίπεδο μηνυμάτων και παράγει στατικές τοπολογίες τυχαίων γράφων. Οι υποστηριζόμενες τοπολογίες περιλαμβάνουν ομοιόμορφα τυχαίους γράφους (σύμφωνα με το μοντέλο Erdős-Rényi), τυχαίους τακτικούς γράφους [93] και τυχαίους γράφους με κατανομή βαθμών power-law βασιζόμενους στο μοντέλο Barabasi [7]. Επιπρόσθετα, ο προσομοιωτής έχει την

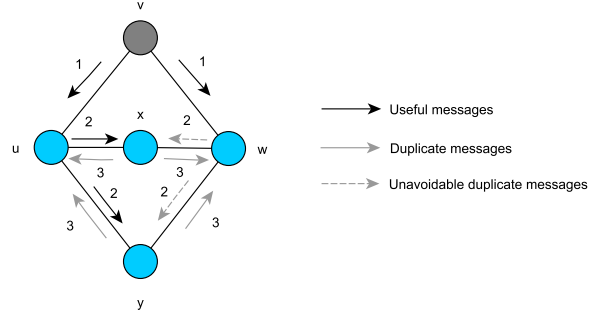
δυνατότητα να παράγει τοπολογίες με επιθυμητό συντελεστή ομαδοποίησης βασιζόμενος στο αλγόριθμο των Heath and Parikh [81] για τυχαία δίκτυα και στο [83] για δίκτυα άνευ κλίμακας. Ο προσομοιωτής έχει τη δυνατότητα να τοποθετεί R αντίγραφα εφαρμόζοντας αντιγραφή *uniform*.

Κάθε σενάριο προσομοίωσης αρχίζει με τη δημιουργία ενός στατικού δικτύου με δεδομένες παραμέτρους (π.χ. αριθμός κόμβων, πλήθος ακμών, μέσος βαθμός ή κατανομή βαθμών). Μια εκτέλεση θεωρεί έναν τυχαίο αρχικό κόμβο, ο οποίος ξεκινά τη διαδικασία αναζήτησης και χρησιμοποιεί την πλημμύρα για να προωθήσει το ερώτημα στους άλλους κόμβους του προσομοιούμενου δικτύου $p2p$. Οι ρυθμιζόμενες από τον χρήστη παράμετροι, επιτρέπουν τον περιορισμό της αναζήτησης σε συγκεκριμένη απόσταση από τον αρχικό κόμβο, όπως επίσης και την ανίχνευση / απαλοιφή διπλότυπων μηνυμάτων με βάση τη συζήτηση που ακολουθεί. Οι μετρήσεις είναι ο μέσος όρος πολλαπλών επαναλήψεων και περιλαμβάνουν τον αριθμό των συνολικά μεταδιδόμενων μηνυμάτων, τον αριθμό των διπλότυπων μηνυμάτων και το πλήθος των διαφορετικών κόμβων που έλαβαν το μήνυμα-ερώτηση.

4.2 Αδόμητα δίκτυα $p2p$ & πλημμύρα

Τα αδόμητα συστήματα $p2p$ είναι μεγάλης κλίμακας κατανεμημένα δίκτυα και χρησιμοποιούν μόνο τοπική πληροφορία για τη διεξαγωγή αποκεντρωμένων διαδικασιών όπως η αναζήτηση. Ως τοπική πληροφορία, εδώ, εννοείται η πληροφορία που έχει ο κάθε κόμβος για τους γειτονικούς του, για το πόσοι και ποιοι είναι. Θεωρούμε το δίκτυο $G = (V, E)$, με $|V| = N$ κόμβους και με ένα σύνολο ακμών E . Ο μέσος βαθμός του δικτύου δίνεται από τη σχέση 2.1.

Το Σχ. 4.1 απεικονίζει το πρόβλημα της πλημμύρας: ο κόμβος v πρόκειται να μεταδώσει ένα μήνυμα-ερώτηση και στους δύο γειτονικούς του κόμβους (u και w). Υποθέτουμε, ότι κάθε μετάδοση ενός μηνύματος πραγματοποιείται σε 1 χρονική μονάδα (βήμα). Στο σενάριο του Σχ. 4.1 ένα βέλος δηλώνει μια μετάδοση ενός μηνύματος και η ετικέτα του δείχνει το βήμα της διαδικασίας που συμβαίνει αυτή η μετάδοση. Οι κόμβοι x και y θα λάβουν το ίδιο μήνυμα δύο φορές στο δεύτερο βήμα, από διαφορετικά μονοπάτια. Καθώς δεν υπάρχει συντονισμός, στο τρίτο βήμα οι u και w θα λάβουν το ίδιο μήνυμα, άλλες δύο φορές. Αυτό οδηγεί σε μεγάλο αριθμό περιττών λήψεων του ίδιου μηνύματος (διπλότυπα μηνύματα). Τα διπλότυπα μη-



Σχήμα 4.1: Παράδειγμα πλημμύρας σε ένα απλό δίκτυο και τα παραγόμενα μηνύματα.

νύματα απεικονίζονται με βέλη χρώματος γκρι.

Για να ελεγχθεί η έκταση της πλημμύρας, συνήθως σε κάθε μήνυμα προσαρτάται μια παράμετρος *time-to-live* (TTL), η οποία μειώνεται σε κάθε βήμα της διαδικασίας. Η παραπέρα μετάδοση ενός μηνύματος σταματά, όταν η παράμετρος TTL φθάσει την τιμή μηδέν. Ο στόχος είναι να περιοριστεί ο χώρος αναζήτησης μέσα σε μια συγκεκριμένη απόσταση. Στο Σχ. 4.1 η τιμή 2 για την TTL παράμετρο μειώνει τα διπλότυπα μηνύματα που παράγονται στο τρίτο βήμα.

Για να γίνει το πρόβλημα πιο συγκεκριμένο, υποθέτουμε ένα τυχαίο τακτικό δίκτυο βαθμού d και την απλή πλημμύρα, όπου ένας κόμβος προωθεί κάθε μήνυμα που λαμβάνει σε όλους τους γειτονικούς του, εκτός από αυτόν από τον οποίο έλαβε το μήνυμα. Επίσης, θεωρούμε ότι κανένας κόμβος δεν γνωρίζει το αντικείμενο και συνεπώς όλοι προωθούν το μήνυμα. Αφού ο βαθμός του κάθε κόμβου v είναι d , κάθε παραλαβή οδηγεί σε $d - 1$ μεταδόσεις από τον κόμβο v . Αν υπάρχουν r_j λήψεις μηνυμάτων στο βήμα j , τότε θα υπάρξουν

$$r_{j+1} = (d - 1)r_j$$

μεταδόσεις μηνυμάτων στο βήμα $j + 1$. Η αναδρομή εύκολα δίνει

$$r_j = d(d - 1)^{j-1}, \quad j \geq 1,$$

αφού $r(1) = d$ για τον αρχικό κόμβο. Τότε, ο συνολικός αριθμός των μηνυμάτων που μεταδίδονται μέχρι τη χρονική στιγμή t ισούται με:

$$M_t = \sum_{j=1}^t r_j = d \frac{(d - 1)^t - 1}{d - 2} > (d - 1)^t, \quad (4.1)$$

που είναι εκθετικός στο βήμα του χρόνου.

Στο Σχ. 4.2(α) απεικονίζεται ο αριθμός μηνυμάτων που παράγεται σε σχέση με το βήμα της αναζήτησης με πλημμύρα σε ένα τακτικό δίκτυο βαθμού 5 των 100.000 κόμβων, σύμφωνα με τον αναλυτικό τύπο 4.1 και τα αποτελέσματα ενός σεναρίου προσομοίωσης (πάνω καμπύλες). Όπως φαίνεται και στη γραφική αναπαράσταση, τα αποτελέσματα του αναλυτικού τύπου και του σεναρίου προσομοίωσης συμπίπτουν. Με βάση το Σχ. 4.1, αυτό σημαίνει ότι ένας τεράστιος αριθμός μηνυμάτων επαναμεταδίδεται σε κόμβους που έχουν ήδη προσπελαστεί, δηλαδή είναι περιττά (unnecessary) διπλότυπα μηνύματα. Αντίστοιχα, στο Σχ. 4.2(β) απεικονίζεται ο συνολικός αριθμός των κόμβων που έλαβαν το μήνυμα-ερώτηση σε κάθε βήμα της αναζήτησης, για το ίδιο δίκτυο (καμπύλη με λευκούς κύκλους). Είναι απόλυτα σαφές, πως μετά από μερικά βήματα, ο αριθμός τέτοιων κόμβων, είναι δυσανάλογα μικρότερος από τον αριθμό των μεταδιδόμενων μηνυμάτων. Είναι προφανές πως αυτή η κατάσταση δεν είναι επιθυμητή και για τον λόγο αυτό πρέπει να εφαρμόζεται ένας μηχανισμός ανίχνευσης διπλότυπων μηνυμάτων ή *duplicate detection mechanism* (DDM) με σκοπό να περιορίσει τις περιττές μεταδόσεις μηνυμάτων. Ένας τέτοιος μηχανισμός χρησιμοποιείται και στη Gnutella όπου σε κάθε ερώτηση αντιστοιχίζεται ένα καθολικό μοναδικό αναγνωριστικό (*globally unique identifier, GUID*) [103]. Όταν ένας κόμβος λάβει ένα μήνυμα, αποθηκεύει το GUID του τοπικά (qcache) και το διατηρεί για κάποιο διάστημα. Αν ο κόμβος λάβει το ίδιο ερώτημα ξανά, (π.χ. με το ίδιο GUID), απλά το απορρίπτει, αποφεύγοντας περιττές μεταδόσεις.

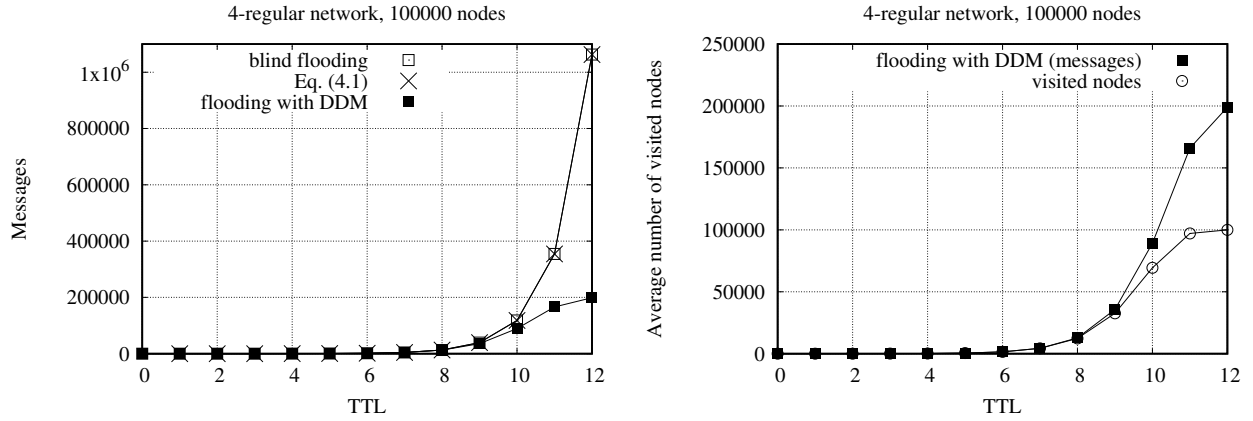
Αυτή η μέθοδος ανίχνευσης διπλότυπων μηνυμάτων μπορεί να απαιτήσει ένα μεγάλο ποσό μνήμης από τον κάθε κόμβο, ανάλογα με την κίνηση που διεκπεραιώνεται μέσα από αυτόν. Κόμβοι με υψηλή κίνηση, λαμβάνουν τεράστιο όγκο από ερωτήσεις ανά χρονική μονάδα και θα πρέπει έτσι είτε να διατηρούν μεγάλη qcache ή να διατηρούν τις εγγραφές της qcache για μικρότερο χρονικό διάστημα. Θα πρέπει επίσης να σημειωθεί ότι ο μηχανισμός δεν μπορεί να είναι τέλειος, υπό την έννοια ότι ένα διπλότυπο μήνυμα μπορεί να εμφανιστεί πολύ αργότερα από τη στιγμή που θα αφαιρεθεί το GUID από την qcache του κόμβου, π.χ. εξ αιτίας των καθυστερήσεων που μπορεί να υπάρχουν στο υφιστάμενο φυσικό δίκτυο. Παρά την απλότητα του μηχανισμού, αποδεικνύεται αρκετά ικανός να απαλείψει την πλειονότητα της περιττής πληροφορίας, όπως φαίνεται και στην κατώτερη καμπύλη στο Σχ. 4.2(α). Για αυτή την εκτέλεση, χρησιμοποιήθηκαν οι ίδιες παράμετροι προσομοίωσης όπως και για την πάνω καμπύλη, αλλά αυτή τη φορά κάθε κόμβος διατηρεί μια τέλεια

cache και εφαρμόζει τον DDM. Εύκολα διαπιστώνει κανείς, πως η συμπεριφορά της πλημμύρας παύει να ακολουθεί την εκθετική καμπύλη ως προς το πλήθος των μηνυμάτων.

Βέβαια, ακόμη και όταν χρησιμοποιείται ένας τέλειος DDM, υπάρχουν ορισμένα διπλότυπα μηνύματα που δεν μπορούν να αποφευχθούν. Ορίζουμε αυτά τα μηνύματα ως αναπόφευκτα διπλότυπα μηνύματα και αυτό το είδος διπλότυπων μηνυμάτων θα εξετάσουμε στις επόμενες ενότητες. Υποθέτουμε ότι ένας τέλειος DDM εφαρμόζεται τοπικά σε κάθε κόμβο. Με τον όρο τέλειο, όπως σημειώθηκε παραπάνω, εννοούμε ότι ο κόμβος ποτέ δεν ξεχνά ένα εισερχόμενο μήνυμα, έτσι ώστε αποφεύγει να προωθηθεί το ίδιο μήνυμα πολλές φορές. Με τον όρο τοπικά, εννοούμε ότι ο κάθε κόμβος δεν γνωρίζει ποια μηνύματα χειρίζονται οι γειτονικοί του κόμβοι.

Ορισμός 4.1. Όλα τα διπλότυπα μηνύματα που προκύπτουν, όταν κάθε κόμβος εφαρμόζει έναν τέλειο τοπικό μηχανισμό ανίχνευσης διπλότυπων μηνυμάτων καλούνται αναπόφευκτα διπλότυπα μηνύματα.

Αυτά αφορούν ίδια μηνύματα που φθάνουν στο ίδιο κόμβο από διαφορετικές διαδρομές. Σε αυτές τις περιπτώσεις, όλοι οι μηχανισμοί ανίχνευσης διπλότυπων μηνυμάτων μπορούν να σταματήσουν την διάδοσή τους μετά την άφιξή τους. Στο Σχ. 4.1 αποτυπώνεται ένα στιγμιότυπο αναζήτησης με πλημμύρα σε ένα μικρό δίκτυο, όπου ο κόμβος v θέτει ένα μήνυμα-ερώτημα. Η ερώτηση του κόμβου v δημιουργεί τρεις διαφορετικές κατηγορίες μηνυμάτων που εμφανίζονται συνολικά στα τρία πρώτα βήματα της διαδικασίας: χρήσιμα, διπλότυπα και αναπόφευκτα διπλότυπα μηνύματα. Τα χρήσιμα μηνύματα (μαύρα συνεχή βέλη) φθάνουν για πρώτη φορά σε έναν κόμβο μεταδίδοντας το μήνυμα-ερώτηση. Τα διπλότυπα μηνύματα (γκρι συνεχόμενα βέλη) φθάνουν σε έναν κόμβο, που έχει λάβει τουλάχιστον μια φορά το μήνυμα-ερώτηση από διαφορετικές διαδρομές. Τα αναπόφευκτα διπλότυπα μηνύματα φαίνονται στο Σχ. 4.1 με διακεκομμένα βέλη, και είναι αυτά, που ακόμη και αν εφαρμοσθεί μηχανισμός DDM, αυτά παραμένουν. Για παράδειγμα, στον κόμβο x θα φθάσει το ίδιο μήνυμα 2 φορές στο 2ο βήμα της διαδικασίας. Έστω ότι για πρώτη φορά θα φθάσει προωθημένο από τον κόμβο u . Όμως θα φθάσει και για δεύτερη φορά προωθημένο από τον κόμβο w , και αυτό είναι αναπόφευκτο διπλότυπο μήνυμα. Ακόμη και αν ο μηχανισμός DDM απαλείψει τα περισσότερα διπλότυπα μηνύματα, τα αναπόφευκτα διπλότυπα μηνύματα αντιπροσωπεύουν ένα μεγάλο



Σχήμα 4.2: (α) Πλήθος μηνυμάτων που παράγει η απλή πλημμύρα με ή χωρίς μηχανισμό ανίχνευσης διπλότυπων μηνυμάτων κατά μέσο όρο σε κάθε βήμα, (β) οι κόμβοι που προσπελάστηκαν από το μήνυμα-ερώτηση και τα παραγόμενα μηνύματα με χρήση DDM.

μέρος πλεονασμού όπως εξηγείται στο Σχ. 4.2(β). Η ανώτερη καμπύλη (με μαύρα τετράγωνα) είναι μια επανάληψη των μεταδιδόμενων μηνυμάτων που παρουσιάστηκαν στο Σχ. 4.2(α), με την υπόθεση ότι ένας τέλειος DDM εφαρμόζεται, ενώ η κατώτερη καμπύλη απεικονίζει τον αριθμό των κόμβων που έχουν προσπελαστεί, όπως περιγράφηκε παραπάνω. Σε κάθε βήμα της αναζήτησης, η διαφορά μεταξύ των δύο καμπυλών δίνει τα αναπόφευκτα διπλότυπα μηνύματα. Πρακτικά αυτό σημαίνει ότι το μήνυμα-ερώτηση ενός κόμβου u θα μεταδοθεί τουλάχιστον μια φορά σε όλες τις ακμές που οδηγούν σε κόμβους που βρίσκονται σε απόσταση $1, 2, \dots, t$ βήματα από τον u , χωρίς αυτό να είναι αναγκαίο, αλλά και χωρίς να υπάρχει προφανής τρόπος να αποτρέψουμε αυτή τη μετάδοση.

4.3 Γενικά όρια επί των αναπόφευκτων διπλότυπων μηνυμάτων

Σε αυτή την ενότητα, παρέχονται απλά γενικά όρια για αναζήτηση σε αυθαίρετα δίκτυα όπου υποθέτουμε ότι ένας τέλειος μηχανισμός ανίχνευσης διπλότυπων μηνυμάτων είναι σε ισχύ. Επιπρόσθετα του DDM, υποθέτουμε ότι οι γειτονικοί κόμβοι δεν ανταλλάσσουν ταυτόχρονα το ίδιο μήνυμα. Δηλαδή, αν οι κόμβοι v και u είναι γειτονικοί και συνδέονται με την ακμή e (όπως φαίνεται στο Σχ. 4.3) και λαμβάνουν το ίδιο μήνυμα-ερώτηση ταυτόχρονα από άλλους κόμβους, τότε μόνο ο ένας από

αυτούς θα χρησιμοποιήσει την ακμή e για την μετάδοση του μηνύματος, έτσι ώστε να αποφευχθεί η διπλή μετάδοση του ίδιου μηνύματος (αυτό μπορεί να εξασφαλισθεί, με την εφαρμογή ενός απλού πρωτοκόλλου χειραψίας για κάθε μετάδοση). Αυτή η ρύθμιση είναι απαραίτητη ώστε τα αποτελέσματα να είναι ανεξάρτητα από τις σχετικές ταχύτητες των κόμβων. Αν για παράδειγμα ο κόμβος u είναι ταχύτερος από τον κόμβο v ή λάβει το μήνυμα-ερώτηση λίγο νωρίτερα, τότε θα υπάρξει μόνο ένα διπλό μήνυμα (το οποίο είναι αναπόφευκτο) που θα μεταδοθεί προς τον κόμβο v : ο κόμβος v δεν θα μεταδώσει το ίδιο μήνυμα πίσω στον κόμβο u . Αν βέβαια και οι δυο κόμβοι στείλουν το μήνυμα ο ένας στον άλλον τυφλά, την ίδια χρονική στιγμή, τότε θα υπάρξουν δυο διπλότυπα μηνύματα (που είναι αναπόφευκτα) όπως φαίνεται στο Σχ. 4.3. Έτσι, ανάλογα με τη σχετική ταχύτητα του κάθε κόμβου, η ακμή e μπορεί να μεταδώσει μία ή δυο φορές το ίδιο μήνυμα. Με την παραπάνω υπόθεση, είναι βέβαιο ότι κάθε ακμή του δικτύου, θα χρησιμοποιηθεί το πολύ μία φορά για τη μετάδοση του μηνύματος. Με βάση τα παραπάνω, ενδιαφερόμαστε για τον αριθμό των αναπόφευκτων διπλότυπων μηνυμάτων.

Έστω ένα μήνυμα “εντοπισμού” ή “ping” που προωθείται σε ένα δίκτυο με την τεχνική της πλημμύρας με σκοπό να φτάσει σε όσο το δυνατό περισσότερους κόμβους. Υποθέτουμε ότι m είναι ο συνολικός αριθμός μηνυμάτων που μεταδίδονται στο δίκτυο. Τότε ισχύει το ακόλουθο λήμμα.

Λήμμα 4.1. *Αν μια αιτούμενη διαδικασία εντοπισμού (ανακάλυψης) παράγει m μηνύματα, τότε ο αριθμός των διαφορετικών εντοπιζόμενων κόμβων στο δίκτυο είναι $\sqrt{2m} \leq n \leq m + 1$.*

Απόδειξη. Υποθέτουμε ότι το αίτημα εντοπισμού επισκέπτεται n διαφορετικούς κόμβους, τον αρχικό κόμβο και άλλους $n - 1$ κόμβους. Προφανώς, αν δεν υπάρχουν διπλότυπα μηνύματα, κάθε μήνυμα φθάνει σε έναν νέο κόμβο και έτσι $n - 1 \leq m$, ή $n \leq m + 1$.

Επιπλέον, ο υπογράφος που αποτελείται από τους n κόμβους που έχουν ήδη προσπελαστεί, έχει το πολύ $n(n - 1)/2$ ακμές. Αφού εφαρμόζεται ο μηχανισμός ανίχνευσης διπλότυπων μηνυμάτων και βασιζόμενοι στην παραπάνω υπόθεση, καμιά ακμή δεν θα χρησιμοποιηθεί παραπάνω από μια φορά με σκοπό να προωθηθεί το αίτημα ανακάλυψης. Συνεπώς,

$$m \leq \frac{n(n - 1)}{2} \Rightarrow n \geq \sqrt{2m}.$$

□



Σχήμα 4.3: Αποφυγή διπλότυπων μηνυμάτων σε γειτονικούς κόμβους.

Λήμμα 4.2. Αν ο μέγιστος βαθμός ενός κόμβου είναι Δ και ένα αίτημα εντοπισμού παράγει m μηνύματα, ο αριθμός των διαφορετικών κόμβων που εντοπίζονται είναι $2m/\Delta \leq n \leq m + 1$.

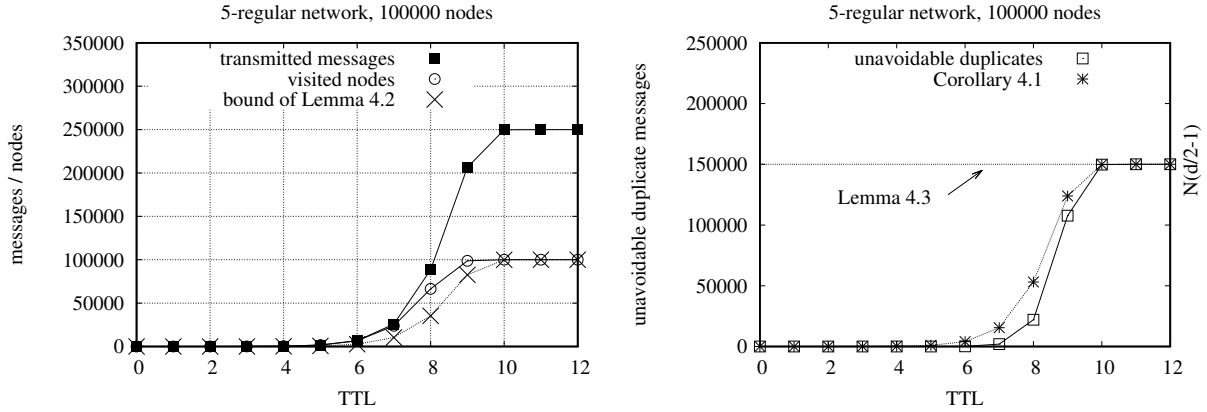
Απόδειξη. Η απόδειξη είναι όμοια με αυτή του Λήμματος 4.1, εκτός από το ότι στον υπογράφο που σχηματίζουν οι n εντοπισμένοι κόμβοι, υπάρχουν το πολύ $n\Delta/2$ ακμές αφού ο βαθμός του κάθε κόμβου είναι το πολύ Δ . $\square$

Πόρισμα 4.1. Αν το αίτημα εντοπισμού παράγει m μηνύματα, τότε ο αριθμός των αναπόφευκτων διπλότυπων μηνυμάτων είναι το πολύ $m + 1 - \max\{\sqrt{2m}, 2m/\Delta\}$.

Απόδειξη. Έστω ότι η διαδικασία της αναζήτησης παράγει m μηνύματα. Από αυτά μόνο τα $m - (n - 1) = m + 1 - n$ είναι τα αναπόφευκτα διπλότυπα μηνύματα. Από τα παραπάνω λήμματα προκύπτει ότι $n \geq \max\{\sqrt{2m}, 2m/\Delta\}$. $\square$

Δεδομένου του μηχανισμού ανίχνευσης διπλότυπων μηνυμάτων, και της υπόθεσης στην αρχή αυτής της ενότητας, είναι ξεκάθαρο ότι καμιά ακμή του γράφου δεν πρόκειται να χρησιμοποιηθεί παραπάνω από μια φορά από τη διαδικασία της πλημμύρας για τη μετάδοση ενός συγκεκριμένου μηνύματος. Αυτό σημαίνει ότι ο αριθμός των μεταδιδόμενων μηνυμάτων ποτέ δεν θα ξεπεράσει τον αριθμό των ακμών $|E|$, ή $\bar{d}N/2$, όπου $\bar{d}$ είναι ο μέσος βαθμός του κόμβου. Βέβαια, αν όλες οι ακμές του γράφου έχουν διασχιστεί, τότε όλοι οι κόμβοι του δικτύου θα έχουν προσπελαστεί. Το αποτέλεσμα είναι ότι με την συμπλήρωση των $\bar{d}N/2$ μηνυμάτων, ακριβώς N κόμβοι θα έχουν προσπελαστεί. Λαμβάνοντας υπόψη ότι τα $N - 1$ μηνύματα είναι αναγκαία για την προσπέλαση των N διαφορετικών κόμβων, το πολύ $\bar{d}N/2 - (N - 1)$ μηνύματα είναι διπλότυπα, οδηγώντας στο ακόλουθο λήμμα.

Λήμμα 4.3. Για κάθε αίτημα εντοπισμού, ο αριθμός των αναπόφευκτων διπλότυπων μηνυμάτων είναι το πολύ $N(\bar{d}/2 - 1) + 1$.



Σχήμα 4.4: (α) Μεταδόσεις μηνυμάτων και αριθμός κόμβων που έχουν επισκεφτεί από το μήνυμα-ερώτηση της αναζήτησης (β) αναπόφευκτα διπλότυπα μηνύματα, σε ένα τακτικό δίκτυο με βαθμό 5 των 100.000 κόμβων.

Στο Σχ. 4.4 απεικονίζονται τα όρια που παρουσιάστηκαν παραπάνω για ένα συγκεκριμένο τακτικό δίκτυο βαθμού 5 των 100.000 κόμβων. Η ενιαία γραμμή αναπαριστά τα δεδομένα από τα αποτελέσματα της προσομοίωσης ενώ η διακεκομμένη γραμμή τα παραγόμενα όρια από την παραπάνω ανάλυση. Συγκεκριμένα στο Σχ. 4.4(α) η ανώτερη καμπύλη είναι ο αριθμός των μεταδιδόμενων μηνυμάτων (m). Η κατώτερη καμπύλη αντιστοιχεί στο κατώτερο όριο του πλήθους των παραγόμενων μηνυμάτων, όπως διαπιστώνεται στο Λήμμα 4.2, το οποίο εφαρμόζεται σε αυτή την περίπτωση για $\Delta = 5$. Η μεσαία καμπύλη απεικονίζει τον πραγματικό αριθμό των κόμβων που προσπελάστηκαν στη διάρκεια της προσομοίωσης. Τα αποτελέσματα επιβεβαιώνουν την θεωρητική μας ανάλυση η οποία απεικονίζεται και γραφικά. Επίσης διαπιστώνουμε ότι η ποσότητα των διπλότυπων μηνυμάτων είναι υπολογίσιμη. Το Σχ. 4.4(β) απεικονίζει τα αναπόφευκτα διπλότυπα μηνύματα, όπου επιπλέον, το άνω όριο τους δίνεται από την οριζόντια γραμμή όπως παράγεται από το Λήμμα 4.3. Και εδώ, τα θεωρητικά αποτελέσματα επιβεβαιώνονται από τις πειραματικές μετρήσεις. Επισημαίνουμε ότι το άνω όριο των διπλότυπων μηνυμάτων σύμφωνα με το Λήμμα 4.3 παρατηρείται όταν η αναζήτηση προσπελάσει όλο το χώρο αναζήτησης. Αντίθετα το άνω όριο των διπλότυπων μηνυμάτων σύμφωνα με το Πρόρισμα 4.1 είναι πιο στενό και προσαρμοσμένο στο κάθε βήμα της αναζήτησης.

4.4 Μοντελοποίηση απόδοσης πλημμύρας με χρήση της θεωρίας παιγνίων

Η θεωρία των παιγνίων είναι ένα μαθηματικό εργαλείο που μπορεί να χρησιμοποιηθεί στη μελέτη και στη βελτιστοποίηση διεργασιών που εκτυλίσσονται στα σύνθετα δίκτυα. Μεταξύ τέτοιων διαδικασιών περιλαμβάνονται η δρομολόγηση πακέτων στο διαδίκτυο, η συμπεριφορά των πωλητών και των αγοραστών σε μια δημοπρασία [61] και η κατανομή πόρων [12] σε σύνθετα δίκτυα. Η θεωρία παιγνίων επιχειρεί να αναλύσει και να λύσει προβλήματα σχετικά με τη λήψη αποφάσεων σε περιπτώσεις συνεργασίας ή ανταγωνισμού. Η απόφαση μιας οντότητας που μετέχει στο παίγνιο δεν είναι μόνο μια επιλογή από το σύνολο των επιλογών, αλλά εξαρτάται και από τις επιλογές των άλλων οντοτήτων που αλληλεπιδρούν μαζί της. Σε ένα δίκτυο, η συμπεριφορά ενός μέλους είναι εξαρτώμενη από τη συμπεριφορά όλων των άλλων μελών του [61].

Σε αυτή την ενότητα προτείνουμε ένα μοντέλο, όπου η διαδικασία της αναζήτησης διαμορφώνεται ως ένα παίγνιο. Ένας κόμβος ξεκινάει την αναζήτηση με στόχο να επισκεφτεί όσο το δυνατό μεγαλύτερο αριθμό διαφορετικών κόμβων αφού έτσι αυξάνει την πιθανότητα της επιτυχίας. Το πλήθος των διαφορετικών κόμβων αυξάνεται όταν αυξάνεται ο αριθμός των βημάτων t της διαδικασίας. Όμως όσο αυξάνει το t αυξάνεται η κίνηση στο δίκτυο από τη μετάδοση πληθώρας μηνυμάτων, πολλά από τα οποία θα είναι διπλότυπα. Επομένως θα πρέπει να γίνει μια διαπραγμάτευση για τη μέγιστη τιμή t . Αν επιτευχθεί συμφωνία, τότε τα δύο μέρη θα λάβουν τις αντίστοιχες αποδόσεις, όπως τις ορίζουμε παρακάτω. Το παίγνιο ορίζεται ως ακολούθως:

- Στο παιχνίδι μετέχουν δυο ορθολογικοί παίκτες, ο κόμβος (χρήστης) και το δίκτυο.
- Η στρατηγική είναι ίδια και για τους δύο παίκτες που αποφασίζουν αν θα σταματήσει ή αν θα συνεχίσει η αναζήτηση.
- Η αμοιβή για τον κάθε παίκτη είναι η απόδοση του σε συνάρτηση με τα βήματα, όπως ορίζεται στη συνέχεια.

Η συνάρτηση απόδοσης για τον κάθε παίκτη ορίζεται ως εξής:

- *χρήστης*: $\alpha_u = \frac{N_t}{N-1}$, όπου N το πλήθος των κόμβων του δικτύου και N_t το πλήθος των διαφορετικών κόμβων που εντοπίστηκαν έως το βήμα t της

αναζήτησης.

- **δίκτυο:** $\alpha_n = \frac{M_t - \text{dups}_t}{M_t}$, όπου M_t το πλήθος των μηνυμάτων που στάλθηκαν έως το βήμα t και dups_t το κόστος από τα διπλότυπα μηνύματα έως το βήμα t . Ισχύει $N_t = M_t - \text{dups}_t$.

Αντικειμενικός σκοπός του χρήστη είναι να μεγιστοποιήσει τα κέρδη του, δηλαδή να επισκεφθεί περισσότερους κόμβους αυξάνοντας έτσι την πιθανότητα επιτυχίας, και του δικτύου να ελαχιστοποιήσει την ζημία του, δηλαδή τη σπατάλη των πόρων του από την επεξεργασία και τη μετάδοση διπλότυπων μηνυμάτων. Για να ποσοτικοποιήσουμε αυτούς τους δυο αντικρουόμενους στόχους ορίζουμε την συνάρτηση οφέλους ($utility(u)$) ως γινόμενο των δυο συνιστωσών:

$$u = \alpha_u \times \alpha_n$$

Για να μεγιστοποιηθεί η τιμή της u θα πρέπει να προβλεφθεί η τιμή του t έτσι ώστε ο χρήστης να μεγιστοποιεί το ελάχιστο κέρδος του και το δίκτυο να ελαχιστοποιεί τη μέγιστη ζημία του (κόστος). Υπό την προϋπόθεση $\alpha_u + \alpha_n \leq c$, (c σταθερά) το πρόβλημα έχει βέλτιστη λύση, την $\alpha_u = \alpha_n$. Το σημείο αυτό, είναι το σημείο ισορροπίας κατά Nash. Συνεπώς, από τη βέλτιστη λύση προκύπτει:

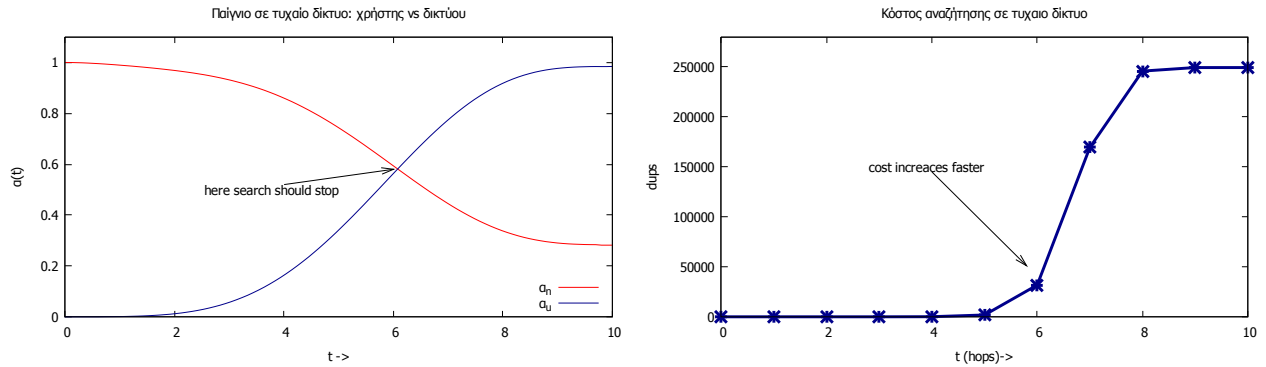
$$\alpha_u = \alpha_n \Leftrightarrow \frac{N_t}{N-1} = \frac{M_t - \text{dups}_t}{M_t} \Leftrightarrow M_t = N - 1$$

κάνοντας χρήση της $N_t = M_t - \text{dups}_t$.

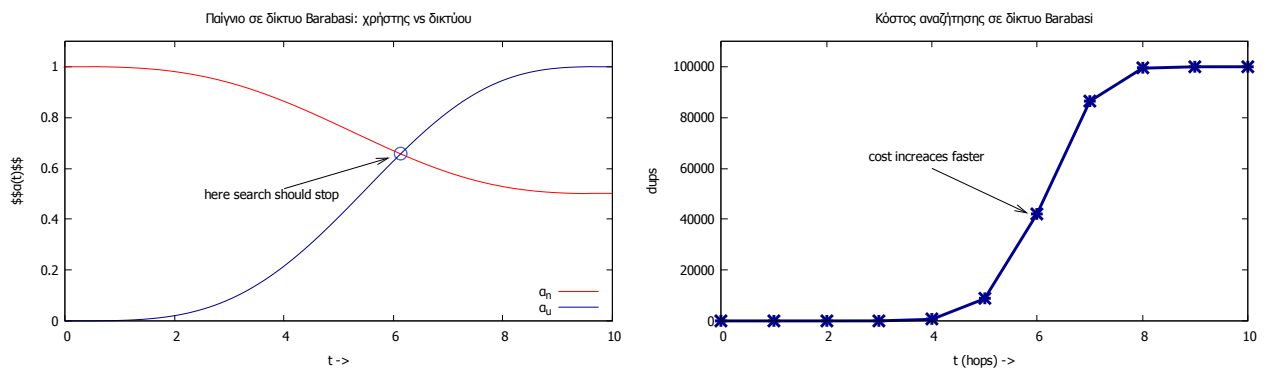
Η απόδοση και των δύο παικτών είναι συνάρτηση των βημάτων t της αναζήτησης. Δηλαδή θα πρέπει να προβλεφθεί ο αριθμός των βημάτων t που θα παράγει έως N μηνύματα ή σε κάθε βήμα να εξετάζεται αν έχουν σταλεί περισσότερα από $N - 1$ μηνύματα οπότε η αναζήτηση σταματά. Αλλιώς συνεχίζει στο επόμενο βήμα. Διαισθητικά αυτό σημαίνει ότι, από το σημείο ισορροπίας και μετά, η συνέχιση της αναζήτησης επιφέρει αυξημένο κόστος σε μηνύματα ενώ το όφελος (η πιθανότητα για θετικό αποτέλεσμα) μειώνεται αισθητά.

Για να επιβεβαιώσουμε τους ισχυρισμούς μας διεξάγουμε μια σειρά από πειραματικές μετρήσεις σε διάφορα δίκτυα Erdős-Rényi [62] και Barabási [19], με τη βοήθεια ενός προσομοιωτή που αναπτύχθηκε για τις ανάγκες αυτής της διατριβής και περιγράφεται στην Ενότητα 4.1.

Στα Σχ. 4.5, 4.6 αποτυπώνεται η παραπάνω ανάλυση σε δύο δίκτυα των 100.000 κόμβων, ένα σύμφωνα με το μοντέλο Erdős-Rényi μέσου βαθμού $\bar{d} = 7$ και ένα σύμφωνα με το μοντέλο Barabási μέσου βαθμού $\bar{d} = 4$. Όπως διαπιστώνουμε από τις



Σχήμα 4.5: Παίγνιο χρήστη και τυχαίου δικτύου: (α) απόδοση παικτών και σημείο ισορροπίας και (β) κόστος αναζήτησης σε διπλότυπα μηνύματα (dups) που αυξάνεται μετά το $t = 6$ βήματα που αντιστοιχεί στο σημείο ισορροπίας.



Σχήμα 4.6: Παίγνιο χρήστη και δικτύου Barabasi: (α) απόδοση παικτών και σημείο ισορροπίας και (β) κόστος αναζήτησης σε διπλότυπα μηνύματα (dups) που αυξάνεται μετά το $t = 6$ βήματα που αντιστοιχεί στο σημείο ισορροπίας.

γραφικές παραστάσεις των πειραματικών αποτελεσμάτων, το σημείο ισορροπίας προσδιορίζεται για $t = 6$. Αμέσως μετά, το πλήθος των διπλότυπων μηνυμάτων αυξάνεται απότομα (τουλάχιστον διπλασιάζεται). Αυτό αποτυπώνεται στην απόδοση του δικτύου που μειώνεται αισθητά, ιδιαίτερα στο δίκτυο Erdős-Rényi. Αυτό βέβαια οφείλεται στο διαφορετικό μέσο βαθμό (στο δίκτυο Erdős-Rényi εμφανίζονται περισσότερες ακμές που συμβάλουν στην αύξηση των διπλότυπων μηνυμάτων).

4.5 Μοντέλο εκτίμησης κάλυψης του δικτύου

Σε αυτή την ενότητα αναπτύσσουμε ένα προσεγγιστικό μοντέλο για να μελετήσουμε τη συμπεριφορά της αναζήτησης βασιζόμενη σε πλημμύρα, όταν ο μηχανισμός αντί-

χνευσης διπλότυπων μηνυμάτων είναι σε εφαρμογή. Θεωρούμε τυχαία δίκτυα, αποτελούμενα από N κόμβους και με μέσο βαθμό που δίνεται από $\bar{d}$, όπου οι μεταδόσεις μηνυμάτων έχουν την ίδια ταχύτητα, ίση με 1 χρονική μονάδα (ή βήμα). Υποθέτουμε ότι κάποιος κόμβος v έχει αρχίσει τη διαδικασία της πλημμύρας και έστω S_t το σύνολο των κόμβων που έχει επισκεφτεί η πλημμύρα, συμπεριλαμβανομένου και του κόμβου v , μέχρι το βήμα t . Με m_t συμβολίζουμε τον αναμενόμενο αριθμό των μεταδιδόμενων μηνυμάτων ακριβώς στο βήμα t και με n_t τον αναμενόμενο αριθμό των κόμβων που βρίσκονται σε απόσταση t από τον κόμβο v . Τέλος,

$$M_t = \sum_{i=1}^t m_i$$

και

$$N_t = \sum_{i=1}^t n_i$$

είναι ο συνολικός αριθμός των μηνυμάτων και ο συνολικός αριθμός των κόμβων που έχουν επισκεφθεί, αντίστοιχα, μέχρι και το βήμα t .

Ας δούμε τώρα τι συμβαίνει στο βήμα $t + 1$. Είναι σαφές ότι οι νέοι κόμβοι που προσπελάζονται στο βήμα $t + 1$ δίνονται από τη διαφορά των συνόλων S_{t+1} και S_t ή $(S_{t+1} \setminus S_t)$ και είναι οι κόμβοι που βρίσκονται ακριβώς σε απόσταση $t + 1$ από τον κόμβο πηγή, δηλαδή $n_{t+1} = |S_{t+1} \setminus S_t|$. Εξαιτίας του μηχανισμού ανίχνευσης διπλότυπων μηνυμάτων, οι μόνοι κόμβοι που μεταδίδουν νέα μηνύματα στο βήμα $t + 1$, θα είναι οι n_t κόμβοι που βρίσκονται σε απόσταση t από τον αρχικό κόμβο. Καθώς κανένας κόμβος δεν μεταδίδει το μήνυμα πίσω στον κόμβο που του το παρέδωσε, αυτοί οι κόμβοι θα μεταδίδουν μόνο σε νέους κόμβους σε απόσταση $t + 1$ χρησιμοποιώντας τις εμπρός ακμές ή σε κόμβους που έχουν ήδη προσπελαστεί σε απόσταση t μέσω των διασυνδετικών ακμών. Δεδομένου ότι ο μέσος βαθμός είναι $\bar{d}$, τότε θα υπάρχουν $(\bar{d} - 1)n_t$ μεταδόσεις στο βήμα $t + 1$.

Επειδή οι ακμές είναι ομοιόμορφα τυχαίες, οι προορισμοί των μηνυμάτων είναι επίσης ομοιόμορφα τυχαία κατανεμημένοι. Ως αποτέλεσμα, η πιθανότητα ένα μήνυμα να προορίζεται για έναν νέο κόμβο (έναν κόμβο που δεν έχει ήδη επισκεφτεί στο χρόνο t) δίνεται από τη σχέση $(N - N_t)/N$. Συνεπώς ο αριθμός των νέων κόμβων που ανακαλύπτονται ακριβώς στο βήμα $t + 1$ είναι

$$n_{t+1} = (\bar{d} - 1)n_t \left(1 - \frac{N_t}{N}\right). \quad (4.2)$$

Χρησιμοποιώντας παρόμοια επιχειρηματολογία μπορούμε να υπολογίσουμε και τον αριθμό των μεταδιδόμενων μηνυμάτων. Συγκεκριμένα, από τις $(\bar{d} - 1)n_t$ πιθανές

ακμές για την μετάδοση ενός μηνύματος από τους n_t κόμβους, μόνο ένα μέρος από τις $(|E| - |E_t|)/|E|$ θα χρησιμοποιηθεί από αυτούς

$$m_{t+1} = (\bar{d} - 1)n_t \left(1 - \frac{|E_t|}{|E|}\right), \quad (4.3)$$

όπου $|E_t|$ είναι ο αριθμός των ακμών που χρησιμοποιούνται μέχρι το βήμα t και $|E|$ είναι ο συνολικός αριθμός των ακμών στο δίκτυο. Προφανώς, $|E| = \bar{d}N/2$. Ο αριθμός των ακμών που χρησιμοποιούνται μέχρι το βήμα t προσεγγίζεται από

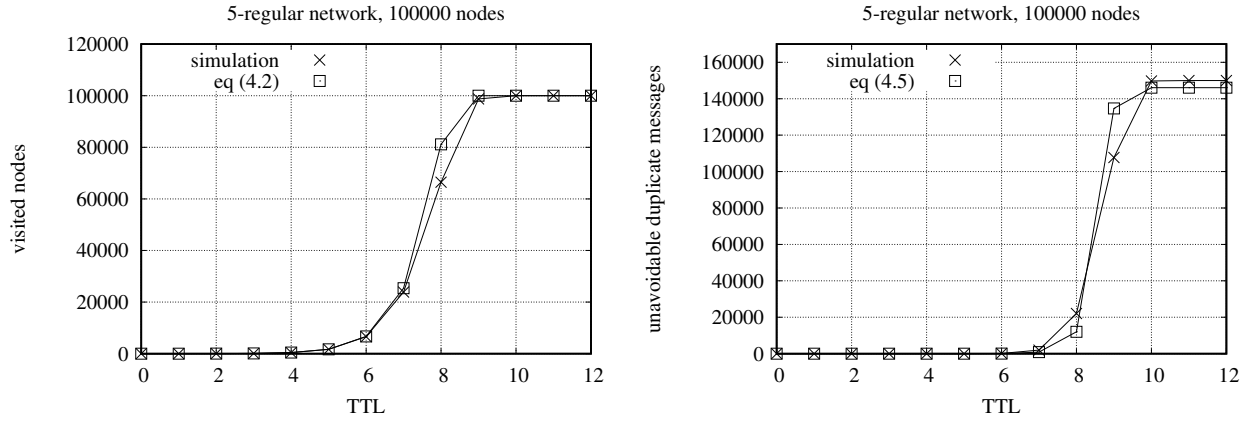
$$|E_t| = \frac{\bar{d}N_{t-1}}{2} + \frac{n_t}{2}. \quad (4.4)$$

Αυτό συμβαίνει επειδή, μέχρι το χρόνο t , οι κόμβοι στο σύνολο S_{t-1} έχουν χρησιμοποιήσει (για μεταδόσεις μηνυμάτων) όλες τις προσπίπτουσες ακμές, ο αριθμός των οποίων ισούται με $\bar{d}N_{t-1}$, διαιρούμενο με το 2 λόγω καταμέτρησης κάθε ακμής δύο φορές. Παρατηρούμε βέβαια ότι, στο t -ιοστο βήμα οι ακμές που οδηγούν σε νέους κόμβους, καταμετρούνται μόνο μια φορά, αφού οι νέοι κόμβοι δεν ανήκουν στο σύνολο των S_{t-1} . Αυτοί οι νέοι κόμβοι είναι n_t σε αριθμό και προσπελάζονται μέσω των n_t ακμών προσεγγιστικά, οδηγώντας έτσι στον διορθωτικό όρο $+n_t/2$. Επισημαίνουμε ότι αυτή είναι μια προσέγγιση, καθώς μπορεί να υπάρχουν δυο ή περισσότεροι κόμβοι σε απόσταση $t - 1$ γειτονικοί με έναν κόμβο σε απόσταση t . Βασιζόμενοι στα παραπάνω, μπορούμε να υπολογίσουμε τον αναμενόμενο αριθμό των διπλότυπων μηνυμάτων ως:

$$\text{dups}_{t+1} = m_{t+1} - n_{t+1}. \quad (4.5)$$

Σημειώνουμε ότι αυτή η εξίσωση (4.2) θυμίζει την διακριτή λογιστική εξίσωση που έχει χρησιμοποιηθεί στη μελέτη πληθυσμών. Οι Korhonen και Kurhinen [104] χρησιμοποίησαν την λογιστική συνάρτηση για να μοντελοποιήσουν τη διάδοση της πληροφορίας σε δίκτυα έξυπνων κινητών (mobile encounter network) υπό διάφορες παραδοχές, ενώ οι Oyama κ. ά. [154] προτείνουν τη λογιστική εξίσωση ως ένα θεωρητικό εργαλείο για να προσεγγίσουν πειραματικά αποτελέσματα για την διάδοση δεδομένων σε κινητά δίκτυα p2p.

Το μοντέλο των (4.2)–(4.5), αν και απλό, χρησιμεύει για μια πολύ αποτελεσματική προσέγγιση των κόμβων και των μηνυμάτων όπως επιβεβαιώνεται και από τα αποτελέσματα της προσομοίωσης. Στα Σχ. 4.7–4.9 δίνονται τα αποτελέσματα προσομοιώσεων και τα αποτελέσματα που προκύπτουν από το θεωρητικό μοντέλο. Στα Σχ. 4.7 και 4.8 εξετάζονται τυχαίοι τακτικοί γράφοι των 100.000 κόμβων με



Σχήμα 4.7: Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο τακτικό δίκτυο με βαθμό 5 των 100.000 κόμβων.

βαθμό 5 και 6 αντίστοιχα, ενώ στο Σχ. 4.9 το δίκτυο είναι τυχαίο (όχι τακτικό) με $\bar{d} = 7$.

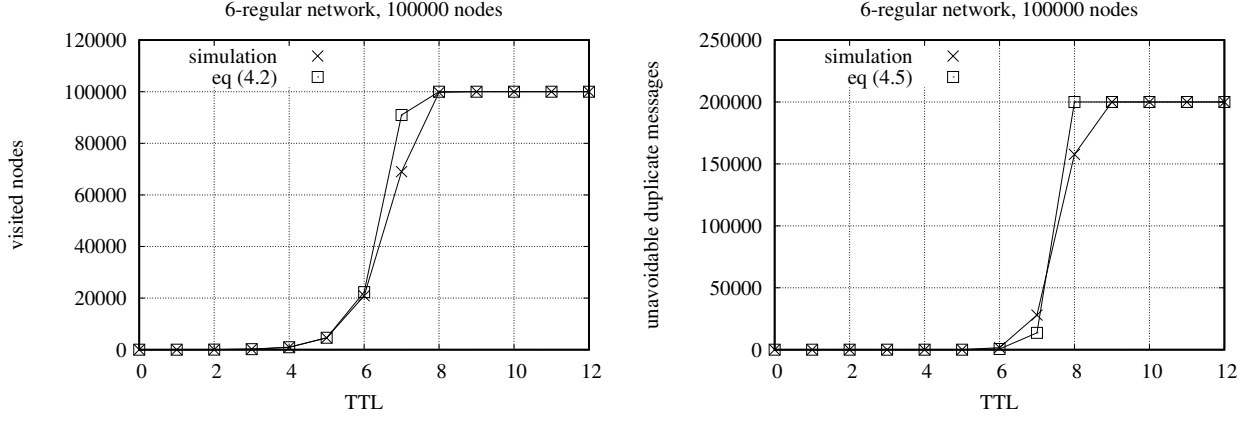
Οι γραφικές παραστάσεις επιβεβαιώνουν την ακρίβεια του μοντέλου. Οι περιορισμοί του, προέρχονται κυρίως από την απλούστευση της προσέγγισης που χρησιμοποιούμε στην (4.4) και από το γεγονός ότι ο καθολικός μέσος βαθμός του δικτύου ($\bar{d}$) που χρησιμοποιείται στον τύπο, είναι διαφορετικός από το μέσο βαθμό των κόμβων που έχουν επισκεφθεί σε κάθε βήμα. Παρόλα αυτά, το μοντέλο επιτυγχάνει να αποδώσει τη βασική δυναμική συμπεριφορά της διαδικασίας της πλημμύρας. Μάλιστα, σε ομογενή δίκτυα που ο βαθμός του κάθε κόμβου είναι περίπου ίσος με το μέσο βαθμό του δικτύου ($d \simeq \bar{d}$) το μοντέλο εμφανίζει ιδιαίτερη ακρίβεια.

4.6 Επεκτείνοντας το μοντέλο

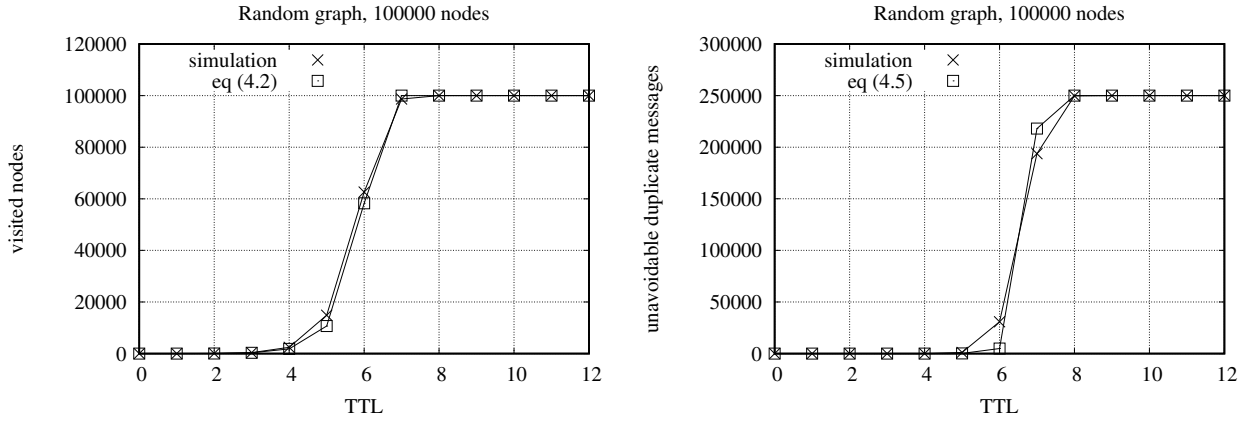
Σε αυτή την ενότητα διερευνούμε αν η εφαρμογή του απλού μοντέλου της ενότητας 4.5 μπορεί να επεκταθεί και σε άλλα δίκτυα πέρα από τα τυπικά τυχαία.

4.6.1 Δίκτυα με δεδομένο συντελεστή ομαδοποίησης

Ο συντελεστής ομαδοποίησης (cc) είναι ένα κοινό χαρακτηριστικό σε πολλά πραγματικά δίκτυα, όπως κοινωνικά, βιολογικά και τεχνολογικά [144] [169] και περιγράφεται από την (2.6). Εδώ εξετάζουμε πως το απλό μοντέλο της ενότητας 4.5 μπορεί



Σχήμα 4.8: Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο τακτικό δίκτυο με βαθμό 6 των 100.000 κόμβων.



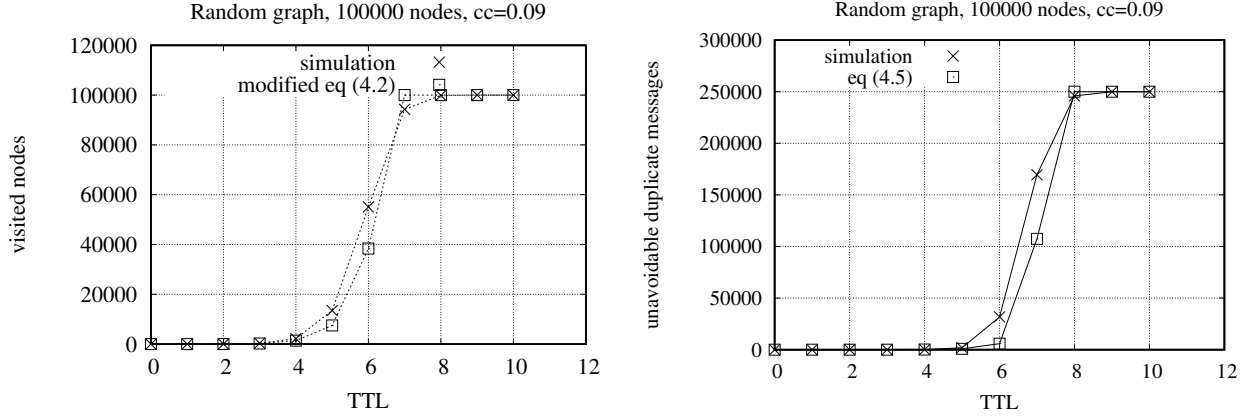
Σχήμα 4.9: Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε τυχαίο δίκτυο των 100.000 κόμβων, με $\bar{d} = 7$.

να επεκταθεί σε τυχαίους γράφους με ρυθμιζόμενο συντελεστή ομαδοποίησης. Οι γράφοι αυτοί μελετήθηκαν πρόσφατα από διάφορους ερευνητές [144] [81].

Αφού ο cc αναπαριστά την πιθανότητα δυο κόμβοι με κοινό γειτονικό κόμβο να είναι και μεταξύ τους γειτονικοί, μόνο ένα ποσοστό $1 - cc$ των ακμών ενός κόμβου που χρησιμοποιείται στο βήμα $t + 1$ θα οδηγήσει σε νέους κόμβους. Συνεπώς ο τύπος (4.2) τροποποιείται ως ακολούθως ώστε να συμπεριλάβει και τον συντελεστή ομαδοποίησης:

$$n_{t+1} = (\bar{d} - 1)n_t(1 - cc) \left(1 - \frac{N_t}{N}\right). \quad (4.6)$$

Ο προσομοιωτής επεκτάθηκε ώστε να παράγει τυχαίους γράφους με συγκεκριμένη κατανομή βαθμών και συντελεστή ομαδοποίησης σύμφωνα με τον αλγόριθμο



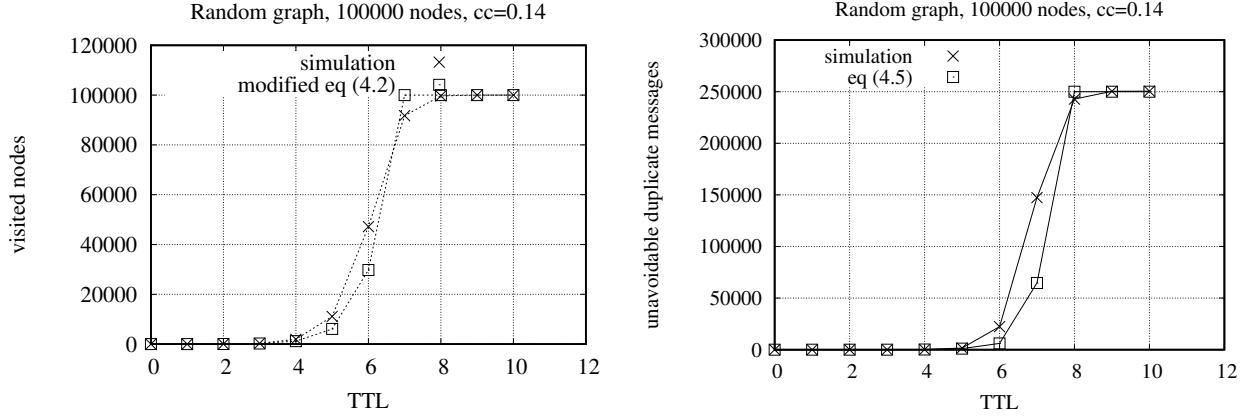
Σχήμα 4.10: Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε τυχαίο δίκτυο 100.000 κόμβων, με $\bar{d} = 7$ και συντελεστή ομαδοποίησης $cc=0.09$.

Heath and Parikh [81]. Διεξήχθησαν πειραματικές μετρήσεις με τη χρήση του ενημερωμένου μοντέλου χρησιμοποιώντας την (4.6) αντί της (4.2), για διάφορα δίκτυα με κατανομή βαθμών Poisson. Στα Σχ. 4.10 και 4.11 παρουσιάζονται τα αποτελέσματα για έναν τυχαίο γράφο με 100.000 κόμβους, μέσο βαθμό 7 και συντελεστή ομαδοποίησης 0.09 και 0.14, αντίστοιχα. Τα αποτελέσματα καταδεικνύουν ότι το τροποποιημένο μοντέλο επιτυγχάνει και ακολουθεί ικανοποιητικά τη συμπεριφορά της πλημμύρας σε αυτή την κατηγορία τοπολογιών.

4.6.2 Δίκτυα power-law

Το μοντέλο μας δεν είναι απευθείας εφαρμόσιμο σε αυτή την κατηγορία τοπολογιών. Επειδή η πλειονότητα των κόμβων έχει πολύ μικρό βαθμό και πολύ λίγοι κόμβοι έχουν πολύ μεγάλο αριθμό συνδέσεων, ο μέσος βαθμός του δικτύου ($\bar{d}$) είναι πολύ διαφορετικός από το μέσο βαθμό των κόμβων που επισκέπτεται η διαδικασία σε κάθε βήμα. Αν ο μέσος βαθμός, $\bar{d}_t$, σε απόσταση t από τον αρχικό κόμβο ήταν γνωστός, τότε το μοντέλο μας θα ήταν εύκολα εφαρμόσιμο με μόνη διαφορά τη χρήση του $\bar{d}_t$ αντί του $\bar{d}$ στις (4.2)–(4.5).

Για να επιβεβαιώσουμε τον ισχυρισμό μας, προσομοιώνουμε δίκτυα power-law χρησιμοποιώντας το μοντέλο των Barabási - Albert [7] και ακολουθούμε την παρακάτω διαδικασία. Αρχικά εφαρμόζουμε την πλημμύρα ώστε να εκτιμηθούν οι τιμές των $\bar{d}_t$ για κάθε βήμα t της διαδικασίας. Αμέσως μετά χρησιμοποιούμε αυτές τις εκτιμήσεις προκειμένου να εφαρμόσουμε το μοντέλο μας όπως περιγράφηκε



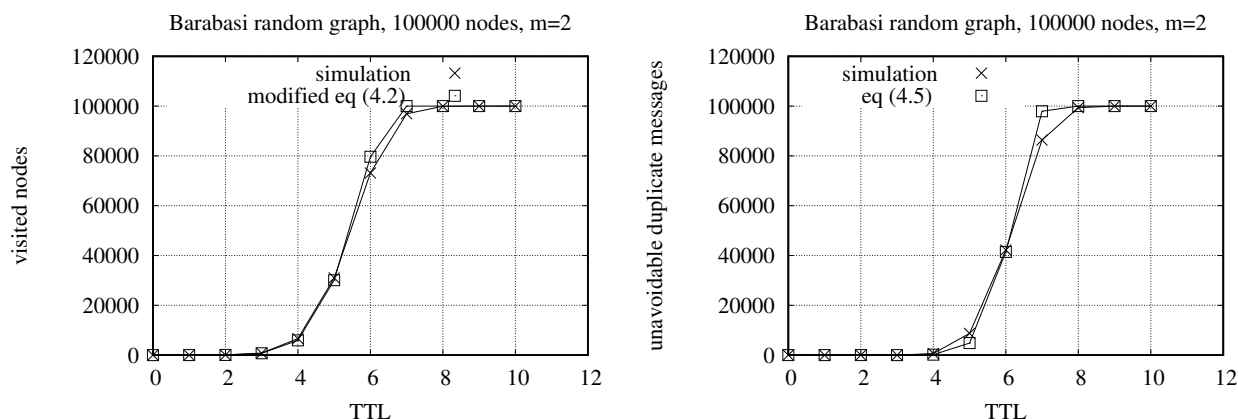
Σχήμα 4.11: Οι κόμβοι που έχει επισκεφθεί το μήνυμα ερώτηση της αναζήτησης (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο 100.000 κόμβων, με $\bar{d} = 7$ και συντελεστή ομαδοποίησης $cc=0.14$.

παραπάνω. Ενδεικτικά αποτελέσματα παρουσιάζονται στο Σχ. 4.12 για ένα δίκτυο 100.000 κόμβων και $m = 2$ (αριθμός συνδέσεων για κάθε νέο κόμβο) που επιβεβαιώνει την θεωρία μας. Από τα αποτελέσματα διαπιστώνουμε επίσης ότι η προσέγγιση του μέσου βαθμού $\bar{d}_t$ των κόμβων σε απόσταση t από έναν κόμβο u περιέχει ικανοποιητική πληροφορία για τη δόμηση του δικτύου. Με άλλα λόγια, και όπως διαπιστώνουμε και από τα αποτελέσματα που απεικονίζονται στο Σχ. 4.12 (α) όπου οι καμπύλες σχεδόν ταυτίζονται, η γνώση του μέσου βαθμού $\bar{d}_t$ των κόμβων σε απόσταση t μας επιτρέπει να προβλέψουμε την κάλυψη του δικτύου καθώς εκτυλίσσεται η διαδικασία της απλής πλημμύρας.

4.7 Σύνοψη

Σε αυτό το κεφάλαιο εξετάσαμε την πλημμύρα σε αδόμητα δίκτυα $p2p$. Αναλύθηκε η υπερβολική επιβάρυνση που συνδέεται με την τυφλή πλημμύρα, που δικαιολογεί την εφαρμογή μηχανισμού ανίχνευσης διπλότυπων μηνυμάτων. Ένας τέτοιος μηχανισμός είναι ικανός να μειώσει ουσιαστικά την περιττή επιβάρυνση, όμως δεν εξαλείφει τα “αναπόφευκτα” διπλότυπα μηνύματα, που οφείλονται στην τοπολογία του δικτύου και στην ύπαρξη πολλαπλών μονοπατιών μεταξύ των κόμβων.

Ανακαλύψαμε θεμελιώδη όρια για τον αριθμό αυτών των αναπόφευκτων διπλότυπων μηνυμάτων σε κάθε δίκτυο $p2p$, τα οποία δείξαμε και μέσω προσομοίωσης.



Σχήμα 4.12: Οι κόμβοι που έχει επισκεφθεί η αναζήτηση (α) και διπλότυπα μηνύματα (β) σε ένα δίκτυο με κατανομή power-law (μοντέλο barabasi) των 100.000 κόμβων, με $m = 2$

Χρησιμοποιώντας τη θεωρία παιγνίων προτείνουμε μια μέθοδο περιορισμού των διπλότυπων μηνυμάτων και βελτίωσης της απόδοσης της απλής πλημμύρας. Επιπλέον, αναπτύξαμε ένα απλό, αλλά αποτελεσματικό μοντέλο για την αποτύπωση της συμπεριφοράς της πλημμύρας σε κάθε βήμα της διαδικασίας, εφαρμόσιμο σε αδόμητα δίκτυα, όπως τα r2r που μοντελοποιούνται ως τυχαίοι γράφοι διαφόρων κατηγοριών.

ΚΕΦΑΛΑΙΟ 5

APF: ΜΙΑ ΝΕΑ ΠΙΘΑΝΟΤΙΚΗ ΣΤΡΑΤΗΓΙΚΗ ΠΛΗΜΜΥΡΑΣ

5.1 Εισαγωγή

5.2 Επισκόπηση βιβλιογραφίας

5.3 Μοντέλο και υποθέσεις

5.4 APF: Μια νέα πιθανοτική στρατηγική αναζήτησης

5.5 Αξιολόγηση

5.6 Σύνοψη

5.1 Εισαγωγή

Όπως παρουσιάστηκε στο Κεφ. 4, η πλημμύρα είναι η βασική στρατηγική αναζήτησης σε διάφορους τύπους δικτύων, που όμως, είναι υπεύθυνη για την υπερφόρτωση του δικτύου με μεγάλο αριθμό μηνυμάτων [56] τα οποία προκαλούν μια σειρά προβλημάτων όπως:

- κατασπατάληση πόρων του δικτύου,
- συμφόρηση.

Ο όρος υπερφόρτωση αναφέρεται στα διπλότυπα μηνύματα (3.3.1)) που εμφανίζονται στη διάρκεια της επαναπροώθησης, ειδικά σε δίκτυα με υψηλό βαθμό. Μελέτες

σε διάφορα πεδία [151], [122] κατέδειξαν τη σοβαρότητα του προβλήματος και τόνισαν ότι η επαναπροώθηση θα πρέπει να χρησιμοποιείται με προσοχή.

Η πιθανοτική πλημμύρα, η οποία ανήκει σε μια κατηγορία στρατηγικών αναζήτησης, προσπαθεί να μετριάσει τις αδυναμίες της απλής πλημμύρας. Αυτές οι στρατηγικές αξιοποιούν μια ιδιότητα ή μια εκτίμηση του δικτύου για να λάβουν μια απόφαση για το αν ένα μήνυμα θα διαβιβαστεί σε επόμενο κόμβο ή όχι. Πιο συγκεκριμένα, κάθε κόμβος που λαμβάνει ένα μήνυμα-ερώτηση, το προωθεί σε κάθε γειτονικό του κόμβο με κάποια πιθανότητα p_f , που ονομάζεται *πιθανότητα προώθησης (forwarding probability)*. Οι πιθανοτικές στρατηγικές, χρησιμοποιούν για την p_f τιμές είτε χωρίς καμία γνώση (τυφλές) (π.χ. [77], [122], [76] είτε από προκαθορισμένες σταθερές (π.χ. [95],[124]) είτε από ρυθμιζόμενες μεταβλητές (π.χ. [17], [51], [153], [70]) προκειμένου να αποφασίσουν ποια σύνδεση θα ακολουθήσουν κατά τη διαδικασία προώθησης του μηνύματος.

Σε αυτό το κεφάλαιο προτείνουμε μια πρωτότυπη πιθανοτική στρατηγική πλημμύρας που την καλούμε Advanced Probabilistic Flooding (APF). Στην μέθοδό μας, η πιθανότητα προώθησης είναι συνάρτηση α) της απόστασης από τον αρχικό κόμβο που εισάγει το μήνυμα-ερώτηση, β) τη δημοτικότητα του επιθυμητού πόρου, και γ) το πλήθος των συνδέσεων του κόμβου (βαθμός κόμβου). Η πιθανότητα προώθησης μειώνεται καθώς αυξάνεται η απόσταση από τον αρχικό κόμβο. Έτσι, η στρατηγική μας επιτυγχάνει να περιορίσει τα περιττά μονοπάτια της πλημμύρας (γνωστό και ως *overshooting*), να βελτιώσει την απόδοση της πλημμύρας και να ελαχιστοποιήσει το κόστος από τα πλεονάζοντα μηνύματα. Επιπλέον, καθώς η πιθανότητα προώθησης εξαρτάται από την δημοτικότητα του πόρου, η επέκταση της πλημμύρας είναι προσαρμοζόμενη και μεταβαλλόμενη ανάλογα με τον αναζητούμενο πόρο, γεγονός που λειτουργεί προσθετικά στην αποδοτικότητα του σχήματός μας. Για την επιβεβαίωση των αποτελεσμάτων, διεξήχθησαν λεπτομερείς πειραματικές προσομοιώσεις που αξιολογούν την απόδοση της προτεινόμενης στρατηγικής. Συμπληρωματικά, έγινε εξαντλητική σύγκριση με άλλες γνωστές πιθανοτικές στρατηγικές πλημμύρας και έδειξαν ότι, τα αποτελέσματα της προτεινόμενης μεθόδου υπερέχουν ως προς την απόδοση. Μάλιστα στις περισσότερες περιπτώσεις η ταχύτητα και το ποσοστό επιτυχίας είναι εφάμιλλα με αυτά της απλής πλημμύρας, ενώ παράλληλα εξαλείφονται τα διπλότυπα μηνύματα. Συνοπτικά η προσφορά της εργασίας μας έχει ως εξής:

- σχεδιάσαμε και αναπτύξαμε έναν μηχανισμό αναζήτησης που βελτιώνει την

απόδοσή της καθώς: α) αποκόπτει τις περιττές διαδρομές της απλής πλημμύρας (overshooting), β) μειώσει αισθητά τον αριθμό των διπλότυπων μηνυμάτων, γ) η επέκτασή της στο χώρο αναζήτησης προσαρμόζεται ανάλογα με την δημοτικότητα του αναζητούμενου πόρου,

- αξιολογούμε πειραματικά την αποδοτικότητα και την αποτελεσματικότητα της τεχνικής μας σε τοπολογίες από πραγματικά ή συνθετικά δεδομένα,
- παρέχουμε συγκριτικά αποτελέσματα με άλλες γνωστές στρατηγικές αναζήτησης για διάφορες τοπολογίες δικτύου.

Το υπόλοιπο αυτού του κεφαλαίου οργανώνεται ως εξής: στην επόμενη ενότητα γίνεται επισκόπηση σχετικών εργασιών. Στην Ενότητα 5.3 συνοψίζεται το μοντέλο του συστήματος καθώς και οι βασικές υποθέσεις που αφορούν τη διαδικασία αναζήτησης. Η Ενότητα 5.4 παρουσιάζει την προτεινόμενη πιθανοτική στρατηγική αναζήτησης και αναλύει τους απαιτούμενους υπολογισμούς. Η Ενότητα 5.5 εστιάζει στην αξιολόγηση της προτεινόμενης στρατηγικής μέσω πειραματικών προσομοιώσεων. Επιπρόσθετα, η στρατηγική μας επεκτείνεται σε δίκτυα με ρυθμιζόμενο συντελεστή ομαδοποίησης και συγκεκριμένες τοπολογίες, όπως τυχαία δίκτυα power-law. Τέλος η Ενότητα 5.6 συνοψίζει το κεφάλαιο.

5.2 Επισκόπηση βιβλιογραφίας

Ένας αριθμός στρατηγικών έχουν προταθεί για να βελτιώσουν την αποτελεσματικότητα της πλημμύρας. Μερικές από αυτές επιχειρούν να περιορίσουν την επέκταση της πλημμύρας με ντετερμινιστικό τρόπο, ενώ άλλες προσπαθούν να το πετύχουν με πιθανοτικό τρόπο. Τα ντετερμινιστικά σχήματα, όπως iterative deepening, local indices, και attenuated bloom filter χρησιμοποιούν μια εκτίμηση από προηγούμενες αναζητήσεις για να προωθήσουν την ερώτηση. Στη στρατηγική *local indices* [199], κάθε κόμβος διατηρεί συλλεγόμενους δείκτες προς τους πόρους άλλων κόμβων για μια προκαθορισμένη απόσταση. Αν ένα αίτημα αναζήτησης είναι έξω από τα όρια γνώσης του κόμβου, ο κόμβος εφαρμόζει αναζήτηση πλημμύρας. Στην *iterative deepening* [199], γνωστή επίσης και ως *expanding ring*, ο αρχικός κόμβος θέτει μια ακολουθία από αναζητήσεις πλημμύρας με αυξανόμενο βάθος (π.χ. τιμές TTL) μέχρι να βρεθεί ο επιθυμητός πόρος. Ένα attenuated bloom filter [163] βάθους h απο-

θηκεύεται σε κάθε κόμβο, και συνοψίζει πόρους που πιθανώς είναι διαθέσιμοι σε όλους τους κόμβους σε απόσταση h βήματα. Κατά τη διάρκεια της αναζήτησης, ο επόμενος γειτονικός κόμβος επιλέγεται από την εξέταση του attenuated bloom filter του τρέχοντα κόμβου.

Σε αυτό το κεφάλαιο, ενδιαφερόμαστε για τα πιθανοτικά σχήματα πλημμύρας. Οι πιθανοτικές στρατηγικές αναζήτησης μπορεί να εκμεταλλευτούν μια ιδιότητα ή μια εκτίμηση του δικτύου για να λάβουν μια απόφαση για το πότε μια ερώτηση πρέπει να προωθηθεί σε έναν άλλο κόμβο (με πιθανότητα p_f) ή όχι (με πιθανότητα $1 - p_f$). Μόλις ένας κόμβος λάβει ένα ερώτημα, το προωθεί σε ένα υποσύνολο από τους γειτονικούς του κόμβους με κάποια πιθανότητα προώθησης, p_f . Αυτές οι στρατηγικές μπορούν να ταξινομηθούν σύμφωνα με την φύση της πιθανότητας προώθησης. Μια κατηγορία πιθανοτικών στρατηγικών, που τις ονομάζουμε τυφλές, προωθούν το μήνυμα-ερώτηση σε ένα υποσύνολο γειτονικών κόμβων που επιλέγονται τυχαία. Οι τυχαίοι περιπατητές [122], [77], ανήκουν σε αυτή την κατηγορία. Στον τυπικό τυχαίο περιπατητή, όταν ένας κόμβος λάβει ένα μήνυμα και δεν διαθέτει τον επιθυμητό πόρο, προωθεί το μήνυμα ερώτημα σε έναν τυχαία επιλεγμένο γειτονικό κόμβο. Εναλλακτικά, μπορεί να χρησιμοποιηθούν πολλαπλοί περιπατητές παράλληλα για να εντοπίσουν τον επιθυμητό πόρο [122]. Η προσαρμογή των συνθηκών τερματισμού οδηγεί σε βελτίωση της απόδοσης ως προς τα διπλότυπα μηνύματα σε σχέση με την πλημμύρα. Αν και, οι τυχαίοι περιπατητές είναι γνωστοί για την απλότητά τους και την μικρή επιβάρυνση με μηνύματα, οδηγούν σε μεγαλύτερους χρόνους απόκρισης [77], [122], [76].

Οι εργασίες [95] και [124] παρουσιάζουν πιθανοτικές στρατηγικές που χρησιμοποιούν προκαθορισμένες τιμές για την πιθανότητα προώθησης. Στην εργασία [95] ένα μήνυμα-ερώτηση προωθείται σε σταθερό ποσοστό γειτονικών κόμβων (οι συγγραφείς χρησιμοποιούν μια προκαθορισμένη τιμή για την p_f , την $p_f = 0.5$) ενώ η τιμή $p_f = 0.6$ επιλέγεται από την Makino κ.ά. [124] για την ανταλλαγή πληροφοριών δρομολόγησης σε δίκτυα power-law. Σε συνδυασμό με τα χαρακτηριστικά του δικτύου, αυτές οι στρατηγικές μπορεί να μειώσουν την υπερφόρτωση μηνυμάτων συγκρινόμενες με την απλή πλημμύρα, αλλά μπορεί να υστερούν στο ποσοστό επιτυχίας καταλήγοντας σε χαμηλότερη συνολική απόδοση. Σε αντίθεση, η δική μας πιθανότητα προώθησης είναι ρυθμιζόμενη και προσαρμόσιμη.

Τέλος, μια τρίτη κατηγορία στρατηγικών, χρησιμοποιεί ρυθμιζόμενες τιμές για την πιθανότητα προώθησης. Σε αυτές τις στρατηγικές, η πιθανότητα προώθησης

είναι ίδια για όλους τους κόμβους [17], [95], [124], [51],[153] ή μεταβαλλόμενη [177], [203], [70]. Χρησιμοποιώντας την θεωρία διήθησης, οι Banaei-Kashani και Shahabi [17] εισάγουν την πιθανοτική πλημμύρα και πραγματεύονται πως η πιθανότητα προώθησης μπορεί να οριστεί σε μια κρίσιμη τιμή ώστε να εξαλείψει τα περιττά μηνύματα, αλλά χωρίς ρητές εγγυήσεις σχετικά με την πιθανότητα επιτυχίας. Σε δύο εργασίες [17], [51] η τιμή της p_f ρυθμίζεται έτσι ώστε το μήνυμα-ερώτηση να φθάσει σε όλους τους κόμβους με μεγάλη πιθανότητα ενώ ο Oikonomou κ.ά. [153] εξήγαγε μια σχέση μεταξύ της τιμής της p_f και του επιθυμητού επιπέδου κάλυψης του δικτύου. Θέτουν την p_f σε μια σταθερή τιμή μεταξύ 0.25 και 0.5. Σύμφωνα με τα αποτελέσματα της προσομοίωσης, μια κατάλληλη τιμή για την p_f είναι η 0.4, αλλά αυξάνει τον χρόνο τερματισμού. Ο Chrisostomo κ.ά. [52] ενδιαφέρεται για την p_f που θα είναι ικανή να οδηγήσει σε προσπέλαση όλων των κόμβων του δικτύου. Αυτή μπορεί να είναι γνωστή ή να ορίζεται σε μια τιμή κοντά στο 0.5.

Στη στρατηγική Adaptive Probabilistic Search (APS) [184], η πιθανότητα προώθησης των πολλαπλών περιπατητών υπολογίζεται χρησιμοποιώντας γνώση από παλιότερες αναζητήσεις. Παρόμοια, η στρατηγική SPUN [82] βασίζεται σε προηγούμενη γνώση που συλλέγει από τα επιτυχή μονοπάτια για να επιλέξει πιθανοτικά το καλύτερο υποσύνολο γειτονικών κόμβων και να προωθήσει την ερώτηση. Η Adaptive Resource-based Probabilistic Search (ARPS) [203] χρησιμοποιεί διαφορετικές πιθανότητες προώθησης, εξαρτώμενες από εκτιμήσεις για τη δημοτικότητα των πόρων. Επιπρόσθετα, όταν εφαρμόζεται σε δίκτυα power-law κάθε κόμβος προσαρμόζει την πιθανότητα προώθησης, ανάλογα με τον βαθμό του, ενώ στην [177], η πιθανότητα προώθησης εξαρτάται και από τον έσω-βαθμό και τον έξω-βαθμό του κάθε κόμβου. Κοντά στην προσέγγιση [177] βρίσκεται η γενικευμένη πιθανοτική πλημμύρα που μελετήθηκε στην εργασία [73]. Η p_f είναι μια συνάρτηση που λαμβάνει υπόψη την απόσταση από τον αρχικό κόμβο και τους βαθμούς προώθησης και λήψης των κόμβων. Τέλος, ο Gaeta κ.ά. [70] εξέτασαν την πιθανότητα προώθησης, η οποία μειώνεται εκθετικά ως προς την απόσταση από τον αρχικό κόμβο, σε αντιστάθμιση με τη μέγιστη επιτρεπτή τιμή της TTL

Στο δικό μας προτεινόμενο σχήμα APF η πιθανότητα προώθησης μειώνεται ανάλογα με την απόσταση από τον αρχικό κόμβο, αλλά σε αντίθεση με το [70] ο βαθμός μείωσης δεν είναι εκθετικός· αντιθέτως βασίζεται σε μια εκτίμηση του τρέχοντα κόμβου για την κάλυψη του δικτύου. Επιπρόσθετα, δεν βασίζεται μόνο στον βαθμό όπως στην [177] ή στην μειούμενη δημοτικότητα όπως στην [203]. Σε αντίθεση με τις πα-

ραπάνω εργασίες, που εξετάζουν μόνο κάποιες από τις εμπλεκόμενες παραμέτρους, η προτεινόμενη APF στρατηγική τις χρησιμοποιεί όλες.

5.3 Μοντέλο και υποθέσεις

5.3.1 Το δίκτυο

Σε αυτή την ενότητα εστιάζουμε στο σχεδιασμό και την αξιολόγηση τεχνικών αναζήτησης σε αδόμητα δίκτυα. Επιπλέον εξετάζουμε εάν και πως είναι δυνατό οι ιδιότητες της τοπολογίας μπορούν να συμβάλλουν στη διαχείριση της διαδικασίας αναζήτησης για την επίτευξη υψηλής απόδοσης και χαμηλής επιβάρυνσης (αναφορικά με το πλήθος των μηνυμάτων). Για να απαντηθεί το ερώτημα αυτό, εξετάζεται η αναζήτηση σε τυχαία δίκτυα που περιγράφονται και παράγονται από θεωρητικά μοντέλα. Ειδικότερα, η μελέτη επικεντρώνεται σε δύο πολύ γνωστές κατηγορίες δικτύων, τους τυχαίους γράφους ER και τους γράφους power-law. Για τα δίκτυα power-law χρησιμοποιείται το μοντέλο των Barabási-Albert (BA) [19].

Τέλος, εξετάζουμε τυχαίους γράφους με ρυθμιζόμενο συντελεστή ομαδοποίησης που περιγράφεται από την (2.6).

5.3.2 Η διαδικασία αναζήτησης

Σε ένα δίκτυο $p2p$, οι κόμβοι υποβάλλουν ερωτήσεις για τον εντοπισμό ενός αντίγραφου κάποιου πόρου για τον οποίο ενδιαφέρονται. Αν υπάρχουν $r \geq 1$ αντίγραφα ενός συγκεκριμένου πόρου τότε η δημοτικότητα του πόρου ορίζεται ως $q = r/N$, και εκφράζει το ποσοστό των κόμβων που τον διαθέτουν. Είναι ζωτικής σημασίας, οι στρατηγικές αναζήτησης σε αποκεντριοποιημένα αδόμητα συστήματα να είναι αποδοτικές και επιτυχείς. Με τον όρο αποδοτική εννοούμε ότι η παραγόμενη κίνηση από τη διάδοση της ερώτησης πάνω στο δίκτυο είναι μέσα σε λογικά όρια. Μια αναζήτηση θεωρούμε ότι είναι επιτυχής, αν ανακαλύψει τουλάχιστον ένα αντίγραφο του επιθυμητού πόρου.

Στην (απλή) πλημμύρα, ο κόμβος που θέτει το ερώτημα επικοινωνεί με όλους τους γειτονικούς του κόμβους. Κάθε γειτονικός, που δεν γνωρίζει για τον επιθυμητό πόρο προωθεί το μήνυμα-ερώτηση στους γειτονικούς του κόμβους, που με την σειρά τους προωθούν την ερώτηση στους δικούς τους γειτονικούς, επαναλαμβάνοντας τη

διαδικασία μέχρι, είτε να βρεθεί ο επιθυμητός πόρος είτε να ικανοποιηθεί κάποια συνθήκη τερματισμού.

Είναι ευρέως γνωστό ότι κατά τη διάρκεια αυτής της διαδικασίας ένα υπερβολικά μεγάλο ποσοστό παραγόμενης κίνησης, είναι περιττό, όπως αναφέρθηκε και στην Ενότητα 4.2. Για τη μείωση της υπερβολικής κίνησης υποθέτουμε ότι βρίσκεται σε ισχύ ένας μηχανισμός ανίχνευσης διπλότυπων μηνυμάτων *-(DDM)*.

Επιπλέον, εφαρμόζεται η χρήση της παραμέτρου TTL, που αποτελεί τυπική συνθήκη τερματισμού της διαδικασίας αναζήτησης. Σκοπός αυτής της ενότητας είναι να μειώσει παραπέρα την υπερφόρτωση από διπλότυπα μηνύματα. Όπως και σε άλλες στρατηγικές αναζήτησης που βασίζονται στην τυχειότητα, εισάγουμε έμμεσα μια επιπλέον συνθήκη τερματισμού: την πιθανότητα προώθησης, p_f . Ένας ενδιαμέσος κόμβος σταματά την παραπέρα προώθηση της ερώτησης στους γειτονικούς του κόμβους με πιθανότητα $1 - p_f$. Σε αντίθεση με άλλες στρατηγικές, αυτή η πιθανότητα δεν είναι σταθερή· η προώθηση είναι μια τοπική απόφαση του κάθε κόμβου, η οποία λαμβάνει υπόψη την απόσταση από τον αρχικό κόμβο, και μια εκτίμηση για το εάν η ερώτηση έχει ήδη απαντηθεί. Λεπτομέρειες παρουσιάζονται στις ακόλουθες ενότητες.

5.4 APF: Μια νέα πιθανοτική στρατηγική αναζήτησης

Θεωρούμε ότι ο κόμβος u έχει λάβει μήνυμα-ερώτηση για έναν συγκεκριμένο πόρο. Σύμφωνα με την πιθανοτική πλημμύρα, αν ο u δεν διαθέτει τον πόρο, αποφασίζει να προωθήσει το μήνυμα-ερώτηση στους γειτονικούς του κόμβους με πιθανότητα p_f ή να μην το προωθήσει (με πιθανότητα $1 - p_f$). Η κύρια ιδέα πίσω από τη στρατηγική μας είναι ότι το μήνυμα-ερώτηση μπορεί να προωθηθεί παραπέρα μόνο αν το ερώτημα δεν έχει απαντηθεί ακόμη· αυτό σημαίνει ότι η p_f θα πρέπει να είναι ίση ή μεγαλύτερη με την πιθανότητα ο ζητούμενος πόρος να μην έχει εντοπιστεί από την διαδικασία αναζήτησης μέχρι αυτό το σημείο (βήμα). Συνεπώς, η p_f θα πρέπει να είναι συνάρτηση του t , της απόστασης δηλαδή από τον αρχικό κόμβο ή ισοδύναμα του “βήματος” της αναζήτησης.

Ας υποθέσουμε ότι ο κόμβος u λαμβάνει την ερώτηση στο βήμα t , και ότι ως το βήμα t υπάρχουν N_t κόμβοι που έχουν λάβει την ερώτηση. Ο κόμβος u θα αποφασίσει αν θα προωθήσει το μήνυμα περαιτέρω, εκτιμώντας την πιθανότητα εντοπισμού

του πόρου μέχρι και το βήμα t . Ο αριθμός των κόμβων που δεν έχουν προσπελαστεί ακόμη δίνεται από τη σχέση $N - N_t$. Αν υπάρχουν r αντίγραφα από τον προς αναζήτηση πόρο, τότε, η πιθανότητα να μην έχει βρεθεί ακόμη ο πόρος ως το βήμα t , είναι ίση με την πιθανότητα όλα τα αντίγραφα να έχουν τοποθετηθεί μεταξύ των $N - N_t$ υπόλοιπων κόμβων (τα οποία είναι προφανώς r στον αριθμό, διαφορετικά ο πόρος έχει βρεθεί).

Επειδή στα αδόμετα συστήματα p2p δεν υπάρχει καμιά συσχέτιση μεταξύ της τοπολογίας και της τοποθέτησης των αντιγράφων των δεδομένων, μπορούμε να κάνουμε μια απλοποιημένη αλλά ακριβής παραδοχή ότι τα αντίγραφα είναι ανεξάρτητα τοποθετημένα. Ως αποτέλεσμα, η πιθανότητα ότι ένα αντίγραφο τοποθετήθηκε σε έναν από τους κόμβους που δεν έχουν προσπελαστεί ακόμη δίνεται από:

$$\frac{N - N_t}{N} = 1 - \frac{N_t}{N}.$$

Συνεπώς, η πιθανότητα ότι όλα τα r αντίγραφα είναι τοποθετημένα μεταξύ αυτών των κόμβων δίνεται από:

$$\left(1 - \frac{N_t}{N}\right)^r.$$

Αυτή είναι η πιθανότητα, ότι κανένας από τους μέχρι τώρα κόμβους που έχουν προσπελαστεί δεν είναι κάτοχος του πόρου. Σε αυτή την περίπτωση, ο κόμβος u θα πρέπει πράγματι να προωθήσει την ερώτηση και αυτό ακριβώς είναι που υλοποιεί ο προτεινόμενος APF αλγόριθμος.

Συνοψίζοντας την παραπάνω συζήτηση, στην APF κάθε κόμβος που λαμβάνει το μήνυμα-ερώτηση στο βήμα t , και δεν διαθέτει τον αιτούμενο πόρο, το προωθεί στους γείτονές του με πιθανότητα προώθησης

$$p_f(t) = \left(1 - \frac{N_t}{N}\right)^{qN}, \quad (5.1)$$

όπου $q = r/N$ είναι η δημοτικότητα του πόρου και N_t είναι ο αριθμός των διαφορετικών κόμβων που έχουν λάβει την ερώτηση μέχρι το βήμα t . Θέτουμε $p_f(0) = 1$, ώστε να εξαναγκάζεται ο αρχικός κόμβος να μεταδώσει την ερώτηση σε όλους τους γειτονικούς του κόμβους.

Ένας κόμβος, για να υπολογίσει την πιθανότητα προώθησης (5.1), πρέπει να γνωρίζει:

(α) τη δημοτικότητα του πόρου (q) και

(β) το συνολικό αριθμό των κόμβων που έχουν προσπελαστεί έως το βήμα t (N_t) (κάλυψη δικτύου).

Εξετάζουμε το ζήτημα της κάλυψης στην επόμενη ενότητα, ενώ σχετικά με τη δημοτικότητα των πόρων για τη συνέχεια υποθέτουμε ότι οι δημοτικότητες είναι γνωστές. Εναλλακτικά, αυτές μπορούν να εκτιμηθούν από κάθε κόμβο χρησιμοποιώντας τεχνικές παρακολούθησης της τοπικής κίνησης. Ο Zhang κ.ά. [203] χρησιμοποιεί ένα μηχανισμό για να εκτιμήσει τη δημοτικότητα των πόρων σε κάθε κόμβο, ο οποίος εκμεταλλεύεται την ανατροφοδότηση από προηγούμενες αναζητήσεις και ενημερώνει έναν τοπικό πίνακα με τις εκτιμώμενες δημοτικότητες. Μετά την προώθηση της ερώτησης, ο κόμβος μειώνει την εκτιμώμενη δημοτικότητα κατά έναν παράγοντα β . Θετικές απαντήσεις αυξάνουν τη δημοτικότητα του πόρου σύμφωνα με την απόσταση από τον ενδιαμέσο κόμβο έως τον κόμβο που διαθέτει τον πόρο. Οι Bisnik και Abouzeid [27] υποθέτουν ότι κάθε κόμβος διατηρεί έναν πίνακα δημοτικότητας για κάθε πόρο του συστήματος. Η εκτίμηση της δημοτικότητας των πόρων βασίζεται στην ανατροφοδότηση από πρόσφατες αναζητήσεις. Αν δεν υπάρχει σχετική γνώση ή μηχανισμός εκτίμησης ένας κόμβος μπορεί να θέσει μια μικρή τιμή για την q , ως μια συντηρητική εκτίμηση.

5.4.1 Εκτίμηση κάλυψης του δικτύου

Ας θεωρήσουμε τώρα την κάλυψη του δικτύου N_t , δηλαδή τον αριθμό των διαφορετικών κόμβων που έχουν λάβει το μήνυμα-ερώτηση, μέχρι και το βήμα t . Αν συμβολίσουμε ως n_i τον αριθμό των νέων κόμβων που προσπελάζονται στο βήμα i , $0 \leq i \leq t$, της διαδικασίας αναζήτησης, τότε είναι προφανές ότι:

$$N_t = \sum_{i=0}^t n_i, \quad (5.2)$$

όπου $N_0 = n_0 = 1$ καθώς ο αρχικός κόμβος θεωρείται πως γνωρίζει το ερώτημα ακριβώς πριν το πρώτο βήμα. Ως αποτέλεσμα, το πρόβλημά μας περιορίζεται στην εκτίμηση του αριθμού των νέων κόμβων που συναντώνται σε κάθε βήμα της διαδικασίας.

Μια προσέγγιση του μέσου αριθμού των κόμβων σε απόσταση i από έναν ορισμένο κόμβο δίνεται από τον Newman κ.ά. [149]. Συγκεκριμένα, αν p_d είναι η πιθανότητα ένας κόμβος να έχει βαθμό d , τότε η γεννήτρια συνάρτηση για την πιθανότητα

των βαθμών των κόμβων δίνεται από:

$$G_0(x) = \sum_{d=0}^{\infty} p_d x^d.$$

Η γεννήτρια συνάρτηση για τους βαθμούς των κόμβων που προσπελάζονται ακολουθώντας μια τυχαία ακμή συμβολίζεται με $G_1(x)$ και μπορεί να αποδειχθεί ότι $G_1(x) = G'_0(x)/G'_0(1)$, όπου G'_0 είναι η πρώτη παράγωγος της G_0 . Από αυτό εξάγεται ότι ο μέσος αριθμός των κόμβων που έχουν προσπελαστεί σε απόσταση i δίνεται από:

$$[G'_1(1)]^{i-1} G'_0(1).$$

Όπως σημειώνουν ο Chandra κ.ά.[36], [35], αυτή η προσέγγιση δεν αντιστοιχεί ακριβώς στον αριθμό των διαφορετικών κόμβων που προσπελάστηκαν από την πλημμύρα εξαιτίας των “διασυνδετικών” και “πίσω” ακμών που υπάρχουν στο δίκτυο. Οι διασυνδετικές/πίσω ακμές είναι συνδέσεις που οδηγούν στο σχηματισμό κύκλων. Οι συγγραφείς επιχειρούν να βελτιώσουν το μοντέλο με την εκτίμηση της πιθανότητας ύπαρξης διασυνδετικής ακμής όπως και της πιθανότητας ύπαρξης πίσω ακμής σε κάθε απόσταση i από έναν δεδομένο κόμβο.

Εμείς ακολουθούμε μια παρόμοια, αλλά απλούστερη προσέγγιση για την εκτίμηση του αριθμού των προσπελασμένων κόμβων σε απόσταση i βημάτων από τον αρχικό κόμβο που θέτει το ερώτημα, η οποία λαμβάνει επίσης υπ' όψη ότι με την πιθανοτική πλημμύρα δεν προσεγγίζονται όλοι οι πιθανοί γειτονικοί κόμβοι. Παρακάτω υποθέτουμε έναν τυχαίο γράφο, σύμφωνα με το τυχαίο γράφημα των Erdős and Rényi· ενώ αργότερα εξετάζουμε λεπτομερέστερα τυχαία δίκτυα power-law.

Έστω ότι n_i είναι το αναμενόμενο πλήθος των διαφορετικών κόμβων που έχουν προσεγγισθεί και που βρίσκονται σε απόσταση i από τον αρχικό κόμβο. Εξετάζουμε τη μετάδοση του μηνύματος κατά τη διάρκεια του βήματος i . Αφού υποθέσαμε ότι εφαρμόζεται μηχανισμός ανίχνευσης διπλότυπων μηνυμάτων, οι μόνοι κόμβοι που προωθούν την ερώτηση κατά τη διάρκεια αυτού του βήματος είναι οι n_{i-1} κόμβοι που βρίσκονται σε απόσταση $i-1$ από τον αρχικό κόμβο· οι κόμβοι που βρίσκονται σε μικρότερη απόσταση από $i-1$ έχουν ήδη μεταδώσει, οι κόμβοι που βρίσκονται σε μεγαλύτερη απόσταση από $i-1$ δεν μεταδίδουν αφού δεν έλαβαν ακόμη το μήνυμα-ερώτηση. Αφού ένας κόμβος δεν μπορεί να μεταδώσει στους προς τα πίσω γειτονικούς του, που τού παρέδωσαν το μήνυμα, οι κόμβοι που θα συνεχίσουν την αναζήτηση, μπορούν να μεταδώσουν μόνο σε νέους κόμβους σε απόσταση i . Δεδομένου ότι ο μέσος βαθμός των κόμβων του δικτύου είναι $\bar{d}$, τότε από τις $(\bar{d}-1)n_{i-1}$

μεταδόσεις μηνυμάτων, μόνο ένα ποσοστό από

$$1 - \frac{N_{i-1}}{N}$$

θα παραδοθεί σε νέους κόμβους. Αυτό συμβαίνει επειδή, από τους N πιθανούς προορισμούς ενός μηνύματος, οι N_{i-1} έχουν ήδη προσεγγιστεί στο βήμα $i - 1$. Λαμβάνοντας επίσης υπ' όψη ότι, λόγω της πιθανοτικής πλημμύρας, κάθε μια από τις μεταδόσεις των κόμβων σε απόσταση $i - 1$ συμβαίνει με πιθανότητα ίση με την πιθανότητα προώθησης $p_f(i - 1)$, παρατηρούμε ότι:

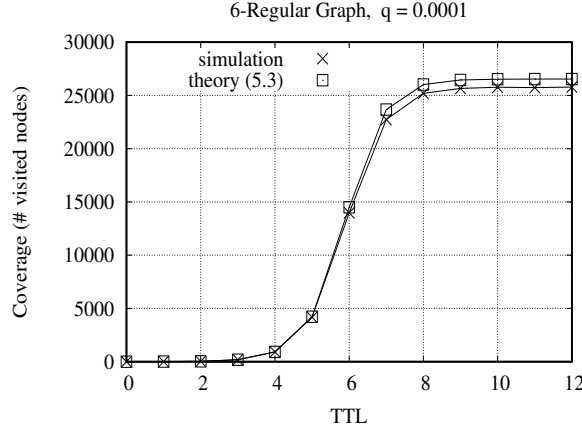
$$n_i = (\bar{d} - 1)n_{i-1} \left(1 - \frac{N_{i-1}}{N}\right) p_f(i - 1), \quad (5.3)$$

Συμπερασματικά, οι (5.1)–(5.3) αποτελούν ένα σύνολο αναδρομικών εξισώσεων που μπορούν να χρησιμοποιηθούν για να εκτιμηθεί η επιθυμητή πιθανότητα προώθησης σε κάθε βήμα της διαδικασίας αναζήτησης.

5.5 Αξιολόγηση

Σε αυτή την ενότητα, αξιολογείται και συγκρίνεται με παρόμοιες τεχνικές η προτεινόμενη στρατηγική μέσω πειραματικών προσομοιώσεων με τη βοήθεια του συστήματος προσομοίωσης που παρουσιάστηκε στην Ενότητα 4.1. Στον προσομοιωτή, που είχε τη δυνατότητα να τοποθετεί αντίγραφα πόρων σύμφωνα με την αντιγραφή *uniform*, προστέθηκε η δυνατότητα και για αντιγραφή *proportional* [44]. Όπως προαναφέρθηκε, στην αντιγραφή *uniform* κάθε πόρος έχει τον ίδιο αριθμό αντιγράφων ενώ στην αντιγραφή *proportional* ο αριθμός των αντιγράφων του κάθε πόρου είναι ανάλογος της ζήτησης για αυτόν τον πόρο. Έχει αποδειχτεί [44] ότι οι δύο πολιτικές είναι μεταξύ των στρατηγικών δημιουργίας αντιγράφων που οδηγούν σε χείριστη απόδοση αναζήτησης. Έχουν επιλεγεί για να δοκιμαστούν οι αλγόριθμοι αναζήτησης που εξετάζονται σε ακραίες καταστάσεις.

Μετά την κατασκευή της τοπολογίας, για κάθε προσομοίωση, επιλέγονται τυχαία κόμβοι του συστήματος και επισημαίνονται ως κάτοχοι του αντιγράφου του πόρου, σύμφωνα με τη δεδομένη πολιτική τοποθέτησης και τη δημοτικότητα των πόρων. Ακολούθως, ένας κόμβος, επιλεγμένος ομοιόμορφα τυχαία, εισάγει μια ερώτηση και ξεκινά η διαδικασία της αναζήτησης. Η αναζήτηση περιορίζεται από μια παράμετρο TTL, η οποία καθορίζει τον μέγιστο επιτρεπόμενο αριθμό βημάτων (μήκος μονοπατιού). Κατά τη διάρκεια της αναζήτησης συλλέγονται λεπτομερή στατιστικά, που



Σχήμα 5.1: Κάλυψη δικτύου σε ένα τυχαίο τακτικό δίκτυο με βαθμό 6 των 100.000 κόμβων.

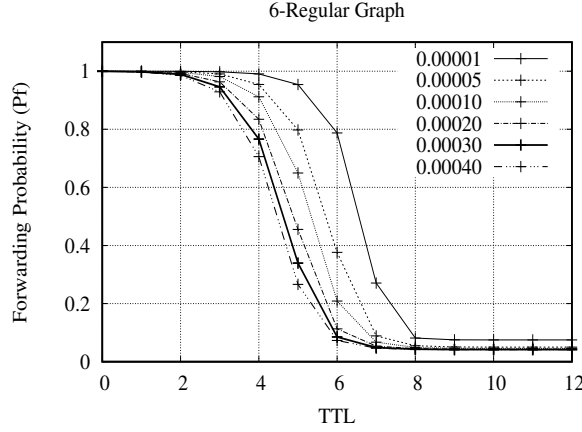
συμπεριλαμβάνουν τον συνολικό αριθμό μηνυμάτων, τον αριθμό των διπλότυπων μηνυμάτων, τον αριθμό των κόμβων που προσπελάστηκαν, κλπ. στη διάρκεια της αναζήτησης. Θεωρούμε ότι η αναζήτηση είναι επιτυχής, αν βρεθεί τουλάχιστον ένα αντίγραφο του επιθυμητού πόρου σε απόσταση t βήματα, διαφορετικά αυτή είναι ανεπιτυχής. Κάθε πειραματική μέτρηση επαναλαμβάνεται τουλάχιστον 1000 φορές και τα αποτελέσματα υπολογίζονται ως μέσοι όροι. Σε κάθε εκτέλεση της προσομοίωσης, το δίκτυο και η τοποθέτηση των πόρων παραμένει αμετάβλητη.

Για να κάνουμε μια συγκριτική μελέτη, ο προσομοιωτής συμπεριλαμβάνει υλοποιήσεις των πιο γνωστών πιθανοτικών στρατηγικών πλημμύρας, όπως παρουσιάστηκαν στην εισαγωγή. Εδώ συμπεριλαμβάνονται αποτελέσματα για αντιπροσωπευτικούς αλγόριθμους που περιλαμβάνουν την απλή πλημμύρα, τη modified BFS [95], την ARPS [204], σε σύγκριση με την προτεινόμενη APF.

5.5.1 Τυχαίοι γράφοι

Εξετάσαμε την εφαρμογή της APF σε έναν μεγάλο αριθμό από τυχαίους και τυχαίους τακτικούς γράφους. Σε έναν τυχαίο γράφο ο βαθμός των κορυφών μεταβάλλεται σύμφωνα με την πιθανότητα (p) ένταξης μιας ακμής στο δίκτυο, δίνοντας ένα μέσο βαθμό $\bar{d} = pN$.

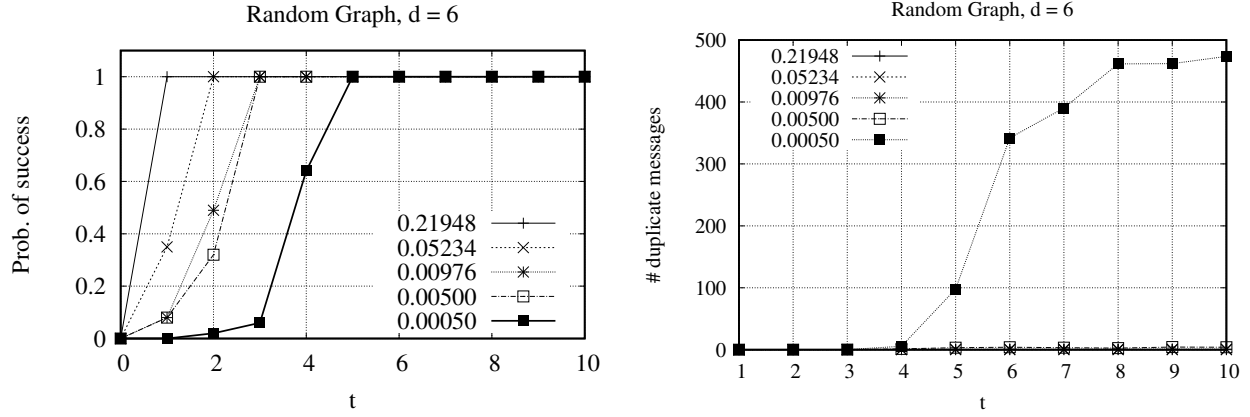
Πριν την παρουσίαση των συγκριτικών αποτελεσμάτων, δείχνουμε την αποτελεσματικότητα της μεθόδου μας. Στα Σχ. 5.1–5.2 εξετάστηκε ένα τακτικό δίκτυο με βαθμό 6. Για τη γραφική παράσταση στο Σχ. 5.1 θέτουμε ερωτήσεις για έναν πόρο



Σχήμα 5.2: Πιθανότητα προώθησης για διαφορετικές τιμές δημοτικότητας ενός πόρου σε ένα τυχαίο τακτικό δίκτυο με βαθμό 6 των 100.000 κόμβων.

με δημοτικότητα $q = 10^{-4}$. Οι καμπύλες δείχνουν των αριθμό των κόμβων που καλύπτονται, όπως παρατηρείται από την προσομοίωση, και όπως προβλέπεται από την (5.2), καταδεικνύοντας την ακρίβεια της εκτίμησης. Στο Σχ. 5.2 απεικονίζεται η πιθανότητα προώθησης για διάφορες τιμές της δημοτικότητας ενός πόρου. Όπως φαίνεται, όσο λιγότερο δημοφιλής είναι ένας πόρος τόσο μεγαλύτερη είναι η πιθανότητα προώθησης ώστε να εξαναγκάσει τη διάδοση της ερώτησης σε περισσότερους κόμβους. Μια άλλη σημαντική παρατήρηση είναι το γεγονός ότι το σχήμα της καμπύλης είναι ακριβώς αντίστροφο από το σχήμα της καμπύλης κάλυψης στο Σχ. 5.1. Η πιθανότητα προώθησης είναι υψηλή στα πρώτα βήματα και αργότερα μειώνεται· η μετάβαση εμφανίζεται σχεδόν στο ίδιο βήμα που η κάλυψη γίνεται υψηλή. Θεωρούμε ότι αυτό συμβάλλει σε σημαντικό βαθμό στην απόδοση της στρατηγικής μας.

Στο Σχ. 5.3, μετράται η αποδοτικότητα της APF χρησιμοποιώντας αντιγραφή proportional σε έναν τυχαίο γράφο $N = 100,000$ κόμβων και μέσου βαθμού ίσο με $\bar{d} = 6$. Υποθέτουμε ότι το πλήθος των ερωτήσεων για έναν πόρο j ακολουθεί κατανομή τύπου Zipf ($\propto j^{-\alpha}$). Επίσης οι δημοτικότητες των πόρων ακολουθούν κατανομή τύπου Zipf ($q_j \propto 1/j^{-\alpha}$), με δεδομένη τιμή για την παράμετρο α . Για παράδειγμα η τιμή της α τίθεται στο 0.8, ενώ οι προκύπτουσες δημοτικότητες πόρων κυμαίνονται από 0.22 έως 10^{-4} . Η γραφική παράσταση δείχνει ότι το πλήθος των διπλότυπων μηνυμάτων είναι αμελητέο ενώ επιτυγχάνεται υψηλό ποσοστό επιτυχίας. Ακόμη και αν η δημοτικότητα του πόρου είναι χαμηλή (π.χ. 0.0005), ο αριθμός των διπλότυπων μηνυμάτων εξακολουθεί να παραμένει σε χαμηλά επίπεδα. Ο λόγος είναι ότι κάθε κόμβος εκμεταλλεύεται τοπική πληροφορία επιπρόσθετα με την απόσταση από τον



Σχήμα 5.3: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο 100.000 κόμβων με αντιγραφή proportional και με κυμαινόμενη δημοτικότητα πόρου ($0.0005 \leq q \leq 0.21948$).

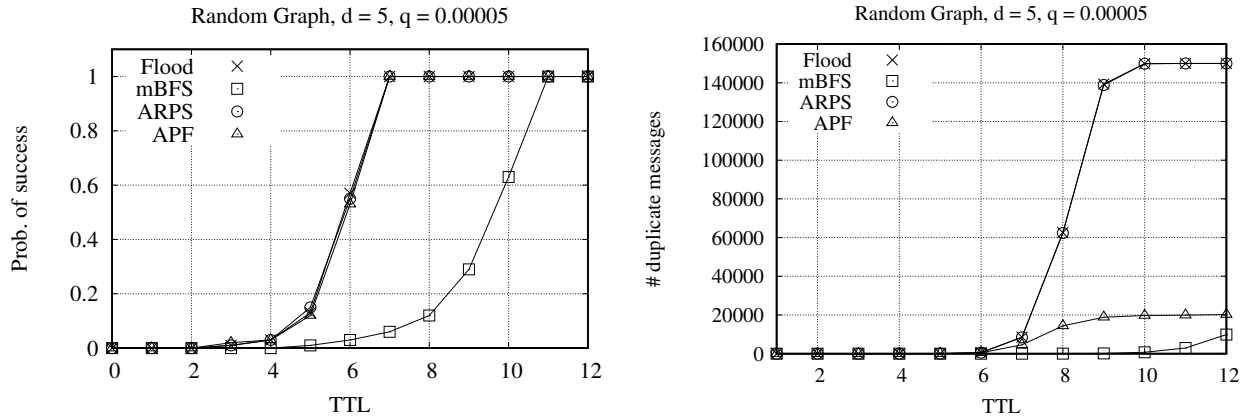
Πίνακας 5.1: Παράμετροι προσομοίωσης τυχαίου δίκτυο ER 100.000 κόμβων.

$\bar{d}$	q	flooding	mbfs	ARPS
5	5×10^{-5}	$p_f = 1$	$p_f = 0.5$	$p_f = 0.9$
6-regular	3×10^{-4}	$p_f = 1$	$p_f = 0.5$	$p_f = 0.8$
7 με $cc = 0.14$	5×10^{-5}	$p_f = 1$	$p_f = 0.5$	$p_f = 0.9$

αρχικό κόμβο.

Στη συνέχεια εφαρμόζεται αντιγραφή uniform καθώς, και η ομοιόμορφη και η αναλογική έχουν παρόμοια συμπεριφορά. Ο χώρος αναζήτησης είναι ανεξάρτητος από τον ρυθμό των ερωτήσεων [44] και η ομοιόμορφη κατανομή είναι πιο δίκαια ως προς την κατανομή των αντιγράφων. Ακολούθως, παρουσιάζονται αποτελέσματα με τυχαία και τυχαία τακτικά δίκτυα p_2p των $N = 100.000$ κόμβων. Ο μέσος βαθμός του δικτύου μεταβάλλεται από 5 έως 7, και η δημοτικότητα των πόρων κυμαίνεται από $5 \times 10^{-5} \leq q \leq 10^{-3}$, που σημαίνει ότι οι πόροι έχουν από 5 έως 100 αντίγραφα για το συγκεκριμένο μέγεθος δικτύου που εξετάζεται, δηλαδή όχι ιδιαίτερα δημοφιλή, ώστε να δοκιμάσουμε τη διαδικασία αναζήτησης σε όχι και ιδιαίτερα ευνοϊκές συνθήκες. Στον Πίνακα 5.1 συνοψίζονται οι παράμετροι προσομοίωσης.

Τα Σχ. 5.4–5.5 παρουσιάζουν την απόδοση της προτεινόμενης στρατηγικής σε σύγκριση με άλλες. Στο Σχ. 5.4 εξετάζουμε ένα τυχαίο γράφο με μέσο βαθμό κόμβων 5, όπου αναζητάμε σπάνιους πόρους με δημοτικότητα $q = 5 \times 10^{-5}$. Στις γραφικές παραστάσεις απεικονίζονται η πιθανότητα επιτυχίας και η ποσότητα των διπλό-

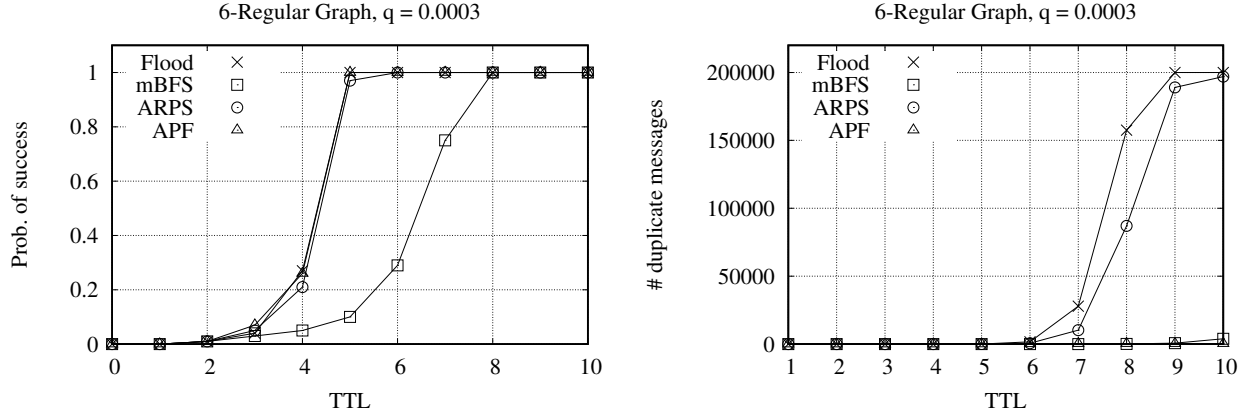


Σχήμα 5.4: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο ($\bar{d} = 5$) 100.000 κόμβων με $q = 0.0005$.

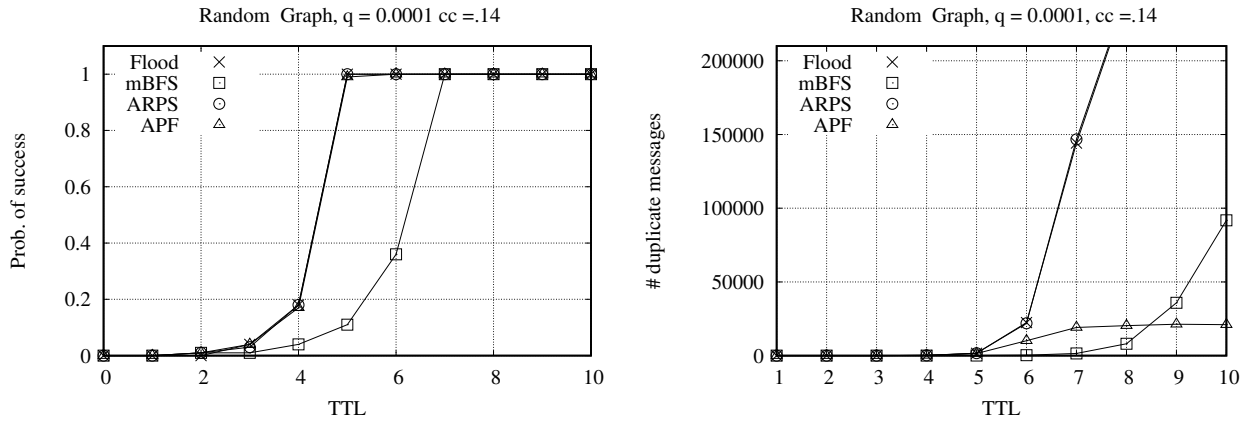
τυπων μηνυμάτων. Στο Σχ. 5.4(α) παρατηρούμε ότι οι περισσότερες στρατηγικές καταφέρνουν να επιτύχουν 100% πιθανότητα επιτυχίας σε απόσταση 7 βημάτων, όπου οι περισσότεροι κόμβοι έχουν προσπελαστεί. Την ίδια στιγμή, το Σχ. 5.4(β) δείχνει ότι αυτό επιτυγχάνεται με ελάχιστη υπερφόρτωση διπλότυπων μηνυμάτων για την APF. Η modified BFS έχει ακόμη μικρότερο κόστος διπλότυπων μηνυμάτων αλλά αυτό οφείλεται στο χαμηλότερο ποσοστό κάλυψης που οδηγεί σε χαμηλότερο ποσοστό επιτυχίας, όπως φαίνεται και στο Σχ. 5.4(α).

Στο Σχ. 5.5, παρουσιάζονται αποτελέσματα για μεγαλύτερη δημοτικότητα του αιτούμενου πόρου. Όπως διαπιστώνουμε, οι στρατηγικές της πλημμύρας, ARPS και APF έχουν υψηλότερο ποσοστό επιτυχίας σε αντίθεση με τη modified BFS που η επιτυχία εμφανίζεται σε επόμενα βήματα. Μάλιστα, το ποσοστό επιτυχίας της APF σχεδόν ταυτίζεται με αυτό της πλημμύρας. Επίσης παρατηρούμε ότι τα διπλότυπα μηνύματα για τις modified BFS και APF είναι σχεδόν αμελητέα. Βέβαια, ακόμη και σε αυτή την περίπτωση, η υπεροχή του σχήματός μας είναι προφανής, αφού συνδυάζει το υψηλό ποσοστό επιτυχίας με σχεδόν ανύπαρκτη υπερφόρτωση διπλότυπων μηνυμάτων, σε αντίθεση με τις άλλες μεθόδους.

Στη συνέχεια εξετάζουμε τυχαία δίκτυα με δεδομένο συντελεστή ομαδοποίησης. Επειδή ο συντελεστής ομαδοποίησης εκφράζει την πιθανότητα δύο κόμβοι με κοινό γειτονικό κόμβο, να είναι και μεταξύ τους γειτονικοί, ένας κόμβος χρησιμοποιεί μόνο ένα κλάσμα $1 - cc$ των ακμών του σε κάθε βήμα i που οδηγούν σε νέους (που δεν προσεγγίστηκαν ακόμη) κόμβους. Συνεπώς τροποποιούμε την (5.3) ως ακολούθως,



Σχήμα 5.5: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο 6-τακτικό δίκτυο 100.000 κόμβων και $q = 0.0003$.



Σχήμα 5.6: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα τυχαίο δίκτυο 100.000 κόμβων, $\bar{d} = 7$ και συντελεστή ομαδοποίησης $cc = 0.14$.

για να ληφθεί υπόψη ο δεδομένος συντελεστής ομαδοποίησης:

$$n_i = (\bar{d} - 1)n_{i-1}(1 - cc) \left(1 - \frac{N_{i-1}}{N}\right) p_f(i - 1), \quad (5.4)$$

Έτσι, για αυτόν τον τύπο δικτύων εφαρμόζουμε την (5.4) αντί της (5.3).

Στο Σχ. 5.6 παρουσιάζονται τα αποτελέσματα για έναν τυχαίο γράφο (με καταννομή βαθμών Poisson) με συντελεστή ομαδοποίησης $cc = 0.14$ ενώ στον Πίνακα 5.1 συνοψίζονται οι άλλες παράμετροι του δικτύου για αυτή την προσομοίωση. Τα αποτελέσματα για την πιθανότητα επιτυχίας (Σχ. 5.6(α)) και για τον αριθμό των διπλότυπων μηνυμάτων (Σχ. 5.6(β)) είναι σε συμφωνία με αυτά που παρατηρήθηκαν στα άλλα δίκτυα που εξετάστηκαν. Η APF, με την τροποποίηση (5.4), συμπεριφέρεται ιδιαίτερα αποτελεσματικά.

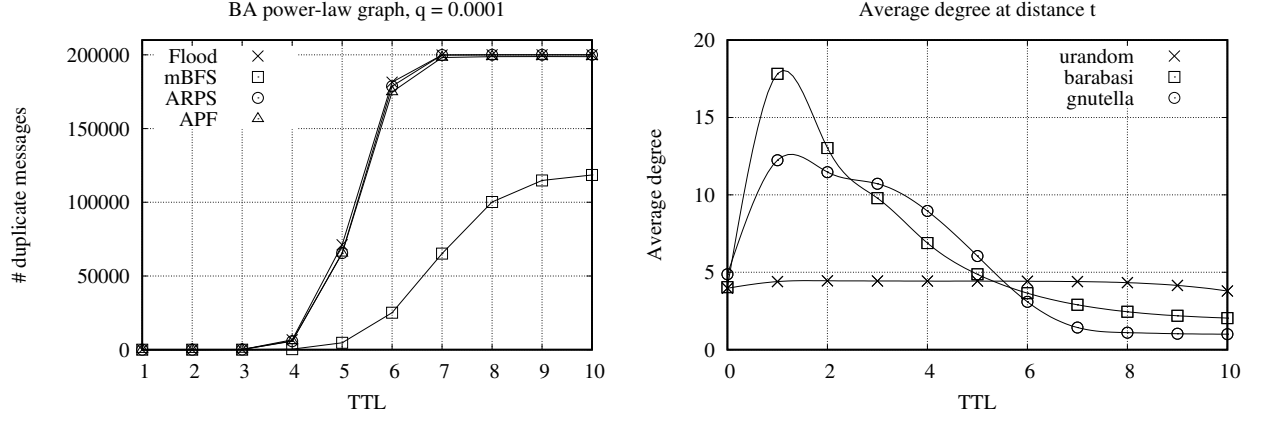
Συμπερασματικά, η συμπεριφορά της APF είναι εντυπωσιακή, τόσο ως προς το ποσοστό επιτυχίας όσο και ως προς το χειρισμό των διπλότυπων μηνυμάτων. Κατορθώνει να προσεγγίσει την απόδοση της απλής πλημμύρας χωρίς όμως το συνεπαγόμενο κόστος. Θυμίζουμε, ότι υποθέτουμε την εφαρμογή ενός ιδανικού μηχανισμού ανίχνευσης διπλότυπων μηνυμάτων όπως συζητήθηκε στις Ενότητες 4.2 και 5.3.2. Πραγματοποιήσαμε ακριβώς όμοια πειράματα με τον μηχανισμό DDM απενεργοποιημένο· ενώ η πιθανότητα επιτυχίας δεν επηρεάζεται, ο αριθμός των διπλότυπων μηνυμάτων αυξάνεται υπερβολικά σε όλους τους αλγορίθμους εκτός της APF τονίζοντας ακόμη περισσότερο τα οφέλη της δικής μας στρατηγικής.

5.5.2 Εφαρμογές σε άλλα δίκτυα

Σε αυτή την ενότητα εκτιμούμε την απόδοση της πρότασή μας σε άλλα δίκτυα, εκτός από τα ομοιόμορφα τυχαία. Εξετάζουμε τα δίκτυα power-law και συγκεκριμένα παράγουμε δίκτυα 100.000 κόμβων χρησιμοποιώντας το μοντέλο των Barabási-Albert [19], [7], αρχίζοντας με τρεις αρχικά συνδεδεμένους κόμβους και προσθέτοντας δύο ($m = 2$) συνδέσεις σε κάθε νεοεισερχόμενο κόμβο.

Εφαρμόζοντας τον αλγόριθμο APF πάνω σε αυτή την τοπολογία, παρατηρούμε ότι, ενώ η πιθανότητα επιτυχίας παραμένει σε υψηλά επίπεδα, ο αριθμός των διπλότυπων μηνυμάτων αυξάνεται σημαντικά. Αυτό απεικονίζεται στο Σχ. 5.7(α) για ένα δίκτυο με $\bar{d} = 4$, όπου η APF φαίνεται να μην έχει διαφορά από την απλή πλημμύρα. Αυτό οφείλεται στη δομή της τοπολογίας που οδηγεί σε μεταβαλλόμενο μέσο όρο βαθμών, $\bar{d}_t$, σε απόσταση t από τον αρχικό κόμβο, που είναι διαφορετικός από το μέσο βαθμό ($\bar{d}$) του συνολικού δικτύου. Στο Σχ. 5.7(β) περιγράφεται ο μέσος βαθμός σε απόσταση $t = 1, 2, \dots, 10$ από τον αρχικό κόμβο όπως παρατηρείται σε μια προσομοίωση για τρία διαφορετικά δίκτυα: έναν τυχαίο γράφο με $\bar{d} = 4$, έναν γράφο BA power-law με $\bar{d} = 4$ και ένα πραγματικό στιγμιότυπο Gnutella με $\bar{d} = 4.9$, που θα συζητηθεί παρακάτω. Σε αντίθεση με τον τυχαίο γράφο, για τον οποίο ο μέσος βαθμός των κόμβων σε απόσταση t από έναν κόμβο u παραμένει σχεδόν σταθερός, στα άλλα δύο δίκτυα αυτός μεταβάλλεται σημαντικά σε σχέση με το t .

Επειδή στα δίκτυα power-law η πλειονότητα των κόμβων έχει μικρό βαθμό και πολύ λίγοι κόμβοι έχουν μεγάλο αριθμό συνδέσεων, ο μέσος βαθμός του δικτύου είναι πολύ διαφορετικός από το μέσο βαθμό των κόμβων που έχουν επισκεφθεί



Σχήμα 5.7: (α) Διπλά μηνύματα κατά την διάρκεια αναζήτησης σε δίκτυο BA 100.000 κόμβων, $\bar{d} = 4$ και $q = 0.0001$. (β) Μέσος βαθμός κόμβων σε απόσταση t από τον αρχικό κόμβο σε διάφορες τοπολογίες.

Πίνακας 5.2: Παράμετροι προσομοίωσης άλλων δικτύων

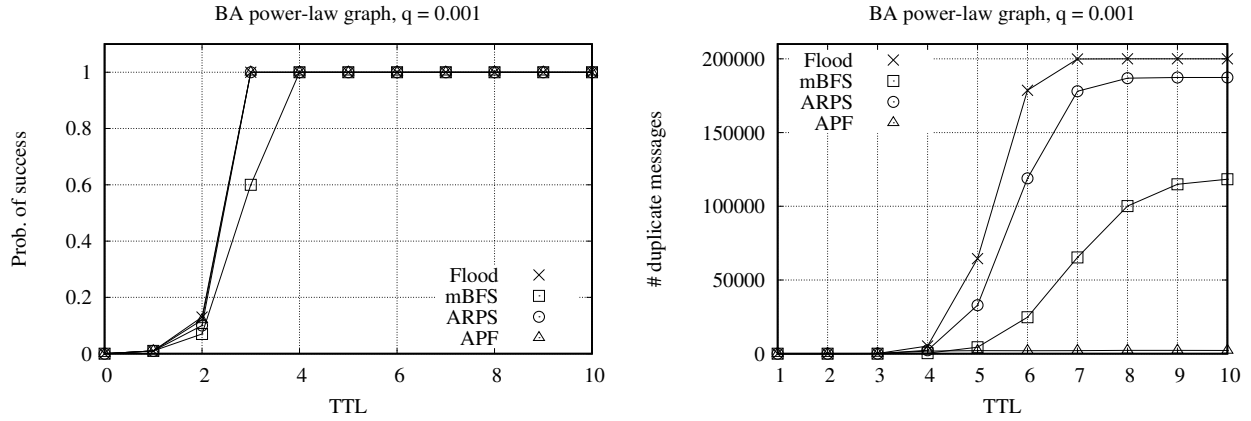
$\bar{d}$	graph	nodes	q	flooding	mBFS	ARPS
4	BA	100000	0.00005	$p_f = 1$	$p_f = 0.5$	$p_f = 0.9$
4	BA	100000	0.001	$p_f = 1$	$p_f = 0.5$	$p_f = 0.7$
4.9	gnutella	26518	0.0037	$p_f = 1$	$p_f = 0.5$	$p_f = 0.7$
23	slashdot	82168	0.00006	$p_f = 1$	$p_f = 0.5$	$p_f = 0.9$

σε κάθε βήμα της διαδικασίας αναζήτησης. Αν ο μέσος όρος $\bar{d}_t$, σε απόσταση t από τον αρχικό κόμβο ήταν γνωστός, τότε η στρατηγική μας θα ήταν εφαρμόσιμη χρησιμοποιώντας το $\bar{d}_t$ αντί του $\bar{d}$ του μοντέλου μας (5.3):

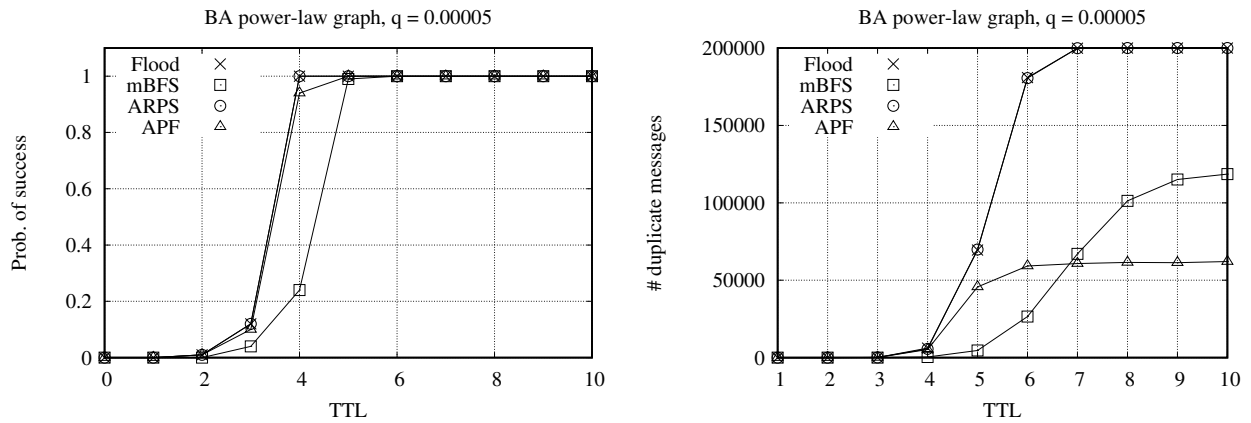
$$n_t = (\bar{d}_t - 1)n_{t-1} \left(1 - \frac{N_{t-1}}{N}\right) p_f(t-1), \quad (5.5)$$

Για να επιβεβαιώσουμε αυτόν τον ισχυρισμό, διεξάγουμε τα ακόλουθα πειράματα: καθορίζουμε το $\bar{d}_t$, από το μέσο βαθμό των κόμβων σε απόσταση t , όπως παρατηρείται με τη βοήθεια προσομοίωσης της απλής πλημμύρας. Έπειτα, χρησιμοποιούμε τις μετρούμενες ποσότητες και τις εφαρμόζουμε στην APF χρησιμοποιώντας την (5.5) αντί της (5.3). Τα αποτελέσματα παρουσιάζονται στα Σχ. 5.8–5.9. Οι παράμετροι προσομοίωσης συνοψίζονται στον Πίνακα 5.2.

Τα αποτελέσματα προσομοίωσης στα Σχ. 5.8–5.9 δείχνουν τη μείωση των διπλότυπων μηνυμάτων για αναζήτηση ενός πόρου με $r = 100$ και $r = 5$ αντίγραφα αντίστοιχα. Χρησιμοποιώντας γνώση για την τιμή του μέσου βαθμού σε σχέση με



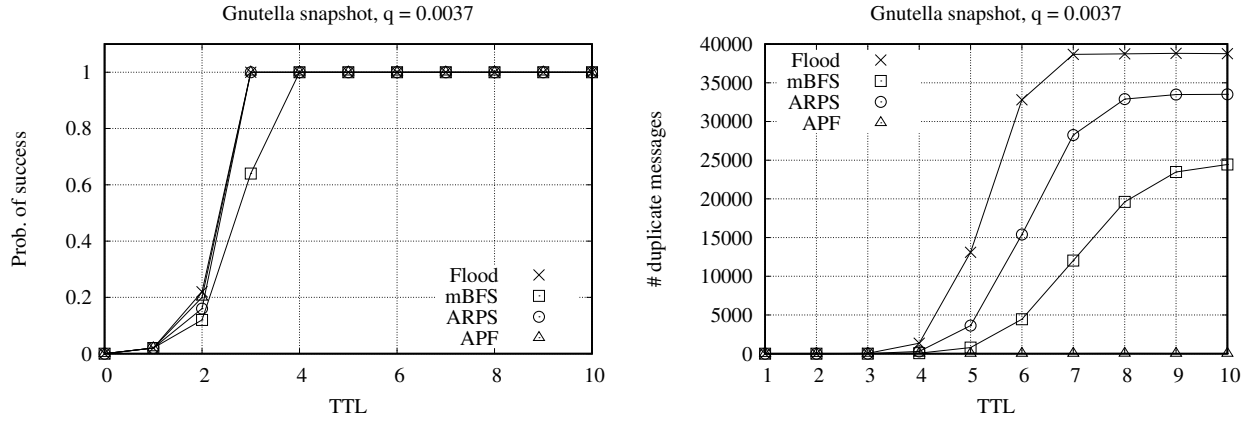
Σχήμα 5.8: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα δίκτυο BA power-law 100.000 κόμβων με $q = 0.001$ αντίγραφα.



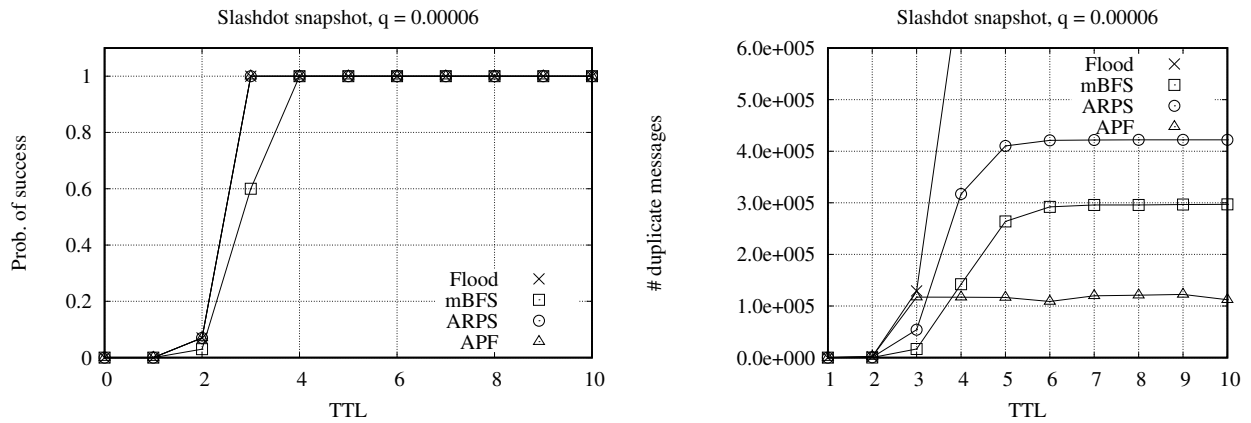
Σχήμα 5.9: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα BA power-law δίκτυο 100.000 κόμβων και $q = 0.00005$.

την απόσταση, η APF επιτυγχάνει ουσιαστική μείωση του κόστους αναζήτησης καθώς εξαλείφονται τα διπλότυπα μηνύματα, χωρίς όμως να επηρεαστεί ο χρόνος απόκρισης ή να μειωθεί το ποσοστό επιτυχίας.

Στα Σχ. 5.10–5.11 απεικονίζονται τα αποτελέσματα από την εφαρμογή της ίδιας ιδέας σε δύο “πραγματικού”-κόσμου δίκτυα. Συγκεκριμένα πρόκειται για ένα στιγμιότυπο από ένα σύστημα gnutella p2p και ένα στιγμιότυπο από το κοινωνικό δίκτυο slashdot όπως λήφθηκαν από το Stanford Network Analysis Platform [113]. Πιο συγκεκριμένα εξετάζουμε ένα δείγμα που λήφθηκε το 2002 από ένα δίκτυο gnutella p2p με 26.518 κόμβους και 65.369 ακμές (Σχ. 5.10) και ένα δεύτερο δείγμα το δίκτυο slashdot που λήφθηκε το Φεβρουάριο 2009 με 82.168 κόμβους (Σχ. 5.11). Οι υπόλοιπες παράμετροι προσομοίωσης συγκαταλέγονται στον Πίνακα 5.2. Όπως και



Σχήμα 5.10: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε στιγμιότυπο gnutella των 26.518 κόμβων και $q = 0.0037$ αντίγραφα.



Σχήμα 5.11: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε στιγμιότυπο του slashdot των 82.168 κόμβων και $q = 0.00006$ αντίγραφα.

παραπάνω, λαμβάνουμε την τιμή για $\bar{d}_t$ από την προσομοίωση με απλή πλημμύρα. Εκτελούμε όμοιες πειραματικές μετρήσεις όπως και στις παραπάνω παραγόμενες τοπολογίες. Τα συμπεράσματα παραμένουν ίδια: σε κάθε περίπτωση η προσέγγισή μας καταλήγει σε σημαντική μείωση των διπλότυπων μηνυμάτων. Στο δίκτυο gnutella τα διπλότυπα μηνύματα σχεδόν εξαλείφονται ενώ και στο δίκτυο slashdot είναι πολύ χαμηλά και σαφώς πολύ λιγότερα σε σχέση με αυτά που παράγουν οι άλλες τεχνικές. Όπως φαίνεται και στα σχήματα, αυτό δεν επηρεάζει την πιθανότητα επιτυχίας, η οποία διατηρείται ίδια όπως και στην απλή πλημμύρα.

5.6 Σύνοψη

Σε αυτό το κεφάλαιο παρουσιάσαμε την APF, μια πρωτότυπη πιθανοτική στρατηγική πλημμύρας που μπορεί να χρησιμοποιηθεί για την δρομολόγηση ερωτήσεων σε αδόμητα δίκτυα $p2p$. Κάθε κόμβος που λαμβάνει το μήνυμα-ερώτηση αποφασίζει τότε θα το προωθήσει περαιτέρω με πιθανότητα p_f η οποία βασίζεται σε μια εκτίμηση των κόμβων για την κάλυψη του δικτύου σε κάθε βήμα, και στην δημοτικότητα του αιτούμενου πόρου. Αυτό είναι σε αντίθεση με άλλες στρατηγικές που, είτε διατηρούν σταθερή την πιθανότητα προώθησης είτε την μειώνουν σύμφωνα με την απόσταση από τον αρχικό κόμβο. Αξιολογήσαμε πειραματικά την απόδοση της APF διεξάγοντας λεπτομερείς προσομοιώσεις, σε διάφορες τοπολογίες δικτύων που περιλαμβάνουν ομοιόμορφα τυχαίους γράφους, γράφους power-law και στιγμιότυπα γράφων από πραγματικά δίκτυα, με ή χωρίς συντελεστή ομαδοποίησης. Συγκρίνοντας τη με άλλες γνωστές προσεγγίσεις, δείξαμε ότι η στρατηγική μας οδηγεί σε υψηλή επιτυχία (σχεδόν όμοια με αυτή της απλής πλημμύρας) ενώ την ίδια στιγμή επιτυγχάνει πολύ χαμηλά επίπεδα διπλότυπων μηνυμάτων.

Η προσέγγισή μας βασίζεται σε δυο βασικές παραμέτρους του δικτύου: το μέσο βαθμό των κόμβων και την απόσταση από τον αρχικό κόμβο. Εδώ εκτιμήσαμε τις τιμές του μέσου βαθμού με τη βοήθεια πειραματικών μετρήσεων, και διαπιστώσαμε πειραματικά ότι αυτή η εκτίμηση είναι ουσιαστική και πολύτιμη για την απόδοση της αναζήτησης.

ΚΕΦΑΛΑΙΟ 6

ΠΡΟΣΟΜΟΙΩΣΗ ΠΡΩΤΟΚΟΛΛΩΝ ΑΝΑΖΗΤΗΣΗΣ ΣΕ ΔΥΝΑΜΙΚΑ ΔΙΚΤΥΑ ΜΕΓΑΛΗΣ ΚΛΙΜΑΚΑΣ

6.1 Εισαγωγή

6.2 Επισκόπηση βιβλιογραφίας

6.3 Το δικό μας πλαίσιο προσομοίωσης: Armonia

6.4 Ενδεικτική μελέτη προσομοίωσης και συζήτηση

6.5 Συμπεράσματα

6.1 Εισαγωγή

Η μοντελοποίηση και η μελέτη σύνθετων συστημάτων / δικτύων με αμιγώς αναλυτικές μεθόδους δεν είναι πάντα μια εύκολη προσέγγιση, πολλές φορές μάλιστα είναι αδύνατη. Υπάρχει επομένως ανάγκη για ένα πλαίσιο προσομοίωσης προκειμένου να αποκτηθούν γνώσεις για τη συμπεριφορά τους, να δοκιμαστούν και να αξιολογηθούν νέες ιδέες, να αναπαραχθούν αποτελέσματα, να συγκριθούν με εμπειρικά αποτελέσματα ή να επικυρωθούν ερευνητικές υποθέσεις. Από την άλλη πλευρά, μπορεί να είναι απλά αδύνατη η πρόσβαση και ο πειραματισμός με πραγματικά συστήματα, ενώ ένα πλαίσιο προσομοίωσης παρέχει όλες τις απαραίτητες λειτουργίες ώστε να είναι δυνατή η αλληλεπίδραση με ένα τέτοιο σύστημα. Οι ερευνητικές μελέτες πολύπλοκων συστημάτων στηρίζονται στην προσομοίωση για να κατανοηθεί η δομή, ο

σχηματισμός, οι επιπτώσεις πιθανών μεταβολών τους και κάθε εν εξελίξει διεργασία που συμβαίνει σε ένα πραγματικό σύστημα.

Ένα ρεαλιστικό πλαίσιο προσομοίωσης είναι βασικό για τρεις τουλάχιστον λόγους. Ο πρώτος είναι ότι μπορεί να αναπαράγει πραγματικά συστήματα και τις αλληλεπιδράσεις τους, για παράδειγμα, πως οι κόμβοι συνδέονται ή ποια είναι η σημαντικότητα ενός κόμβου σύμφωνα με τη θέση του στο δίκτυο. Ο δεύτερος είναι ότι μπορεί να χρησιμοποιηθεί για να σχεδιαστούν, να υλοποιηθούν, να αξιολογηθούν και να προταθούν νέες ιδέες ή πρωτόκολλα [22]. Ο τελευταίος λόγος είναι ότι μπορεί να χρησιμοποιηθεί για να επιβεβαιώσει αναλυτικά αποτελέσματα, να αποφέρει χρήσιμες γνώσεις ή να αποκαλύψει κρυμμένες ιδιότητες.

Ένας προσομοιωτής, λειτουργεί σε μια αφηρημένη εκδοχή, παρέχοντας ένα ρυθμιζόμενο περιβάλλον μέσω δομικών στοιχείων και παραμέτρων και αφαιρεί την πολυπλοκότητα της κατασκευής της τοπολογίας του δικτύου. Στη μελέτη της αναζήτησης αποτελεί καθιερωμένη πρακτική κάθε προτεινόμενος αλγόριθμος να συνοδεύεται και από “ειδικό” προσομοιωτή που χρησιμοποιείται για την επιβεβαίωση των αναλυτικών αποτελεσμάτων ή την πειραματική σύγκριση με παρόμοιες τεχνικές. Οι προσομοιωτές αυτοί έχουν υλοποιηθεί μόνο για υποστήριξη της συγκεκριμένης στρατηγικής και δεν μπορούν να χρησιμοποιηθούν ευρύτερα καθώς δεν καλύπτουν τις ανάγκες άλλων ερευνητών ([142], [143]).

Σε αυτό το κεφάλαιο προτείνουμε έναν πρωτότυπο προσομοιωτή, που τον ονομάζουμε *Armonia*. Από όσα γνωρίζουμε, αυτός είναι το πρώτο γενικού σκοπού εργαλείο προσομοίωσης για πρωτόκολλα τυχαία αναζήτησης σε αυθαίρετες τοπολογίες δικτύων. Εδώ σχεδιάζουμε και αναπτύσσουμε ένα πλαίσιο, που παρέχει διάφορα υλοποιημένα σύνολα στοχεύοντας να καλύψουμε τις ευρύτερες ανάγκες των χρηστών. Το *Armonia* είναι ευέλικτο εργαλείο που συνδυάζει ένα σύνολο λειτουργικών μονάδων και εκτελεί τρεις κύριες λειτουργίες: μοντελοποίηση (αυθαίρετων τοπολογιών), προσομοίωση (αναζήτηση και στρατηγικές τοποθέτησης αντιγράφων) και ανάλυση (συλλογή στατιστικών αποτελεσμάτων). Για να αξιοποιηθούν αυτά τα χαρακτηριστικά, ο χρήστης προσαρμόζει το πλαίσιο σύμφωνα με τις ανάγκες του, σε σχέση με την τοπολογία (η οποία μπορεί να είναι στατική ή δυναμική), την τοποθέτηση αντιγράφων και τη στρατηγική αναζήτησης. Υπάρχει ένας αριθμός από υλοποιημένα δομικά στοιχεία για την τοπολογία του δικτύου, την εκχώρηση αντιγράφων και τα πρωτόκολλα αναζήτησης, αλλά το πλαίσιο είναι επεκτάσιμο, επιτρέποντας την υλοποίηση νέων μοντέλων και πολιτικών που ενσωματώνονται ως πρόσθετα

δομικά στοιχεία. Η συνεισφορά μας, συνοπτικά έχει ως εξής:

- σχεδιάζουμε και αναπτύσσουμε ένα πρωτότυπο, γενικού σκοπού προσομοιωτή για προσομοιώσεις πραγματικών δικτύων, που εμφανίζουν δυναμικότητα και churn.
- υλοποιούμε ένα ευέλικτο πλαίσιο για τη διερεύνηση της συμπεριφοράς των πρωτοκόλλων αναζήτησης και των στρατηγικών εκχώρησης αντιγράφων σε σύνθετα δίκτυα.

Το υπόλοιπο αυτού του κεφαλαίου οργανώνεται ως ακολούθως: Η Ενότητα 6.2 συζητά την σχετική βιβλιογραφία και κάνει μια επισκόπηση των σχετικών εργαλείων. Στην Ενότητα 6.3 παρουσιάζεται το πλαίσιο προσομοίωσης Armonia και δίνεται λεπτομερής περιγραφή της αρχιτεκτονικής του. Στην Ενότητα 6.4 αναφέρονται τα κύρια ευρήματα και τα αποτελέσματα προσομοιώσεων. Τέλος, στην Ενότητα 6.5 συνοψίζεται το κεφάλαιο.

6.2 Επισκόπηση βιβλιογραφίας

Σημαντική προσπάθεια έχει καταβληθεί για την υλοποίηση ενός μεγάλου αριθμού εργαλείων προσομοίωσης για σύνθετα δικτυωμένα συστήματα. Διακρίνουμε τους προσομοιωτές στις ακόλουθες κατηγορίες:

- (α) μικρής διάρκειας ζωής, εξειδικευμένοι, που αποτελούν το πιο κοινό τύπο προσομοιωτών για προσωπική χρήση [142], [143]. Συνήθως υλοποιούνται από τους ίδιους συγγραφείς για την υποστήριξη των εργασιών τους,
- (β) λεπτομερείς, χαμηλού επιπέδου προσομοιωτές, που είναι γενικού σκοπού που εστιάζουν κυρίως στο φυσικό δίκτυο (όπως οι ευρέως γνωστοί NS-2, NS-3). Παρέχουν ουσιαστική υποστήριξη για προσομοιώσεις μέσων μεταφοράς (δικτύων), συσκευών και δικτυακών πρωτοκόλλων, και
- (γ) επιπέδου εφαρμογής, που είναι γενικού σκοπού προσομοιωτές, στοχεύοντας σε δίκτυα μεγάλης κλίμακας όπως τα δίκτυα p2p, τα κοινωνικά δίκτυα, τα βιολογικά δίκτυα.

Εδώ, ενδιαφερόμαστε για την τελευταία κατηγορία, και καταβάλλουμε προσπάθεια να υλοποιήσουμε ένα πλαίσιο, που μπορεί να διευκολύνει στην αναπαραγωγή

πραγματικών δικτύων, που εμφανίζουν δυναμικότητα και churn ενώ εξελίσσονται στο χρόνο. Στη βιβλιογραφία, για τα δίκτυα που μας ενδιαφέρουν, υπάρχουν πειραματικά αποτελέσματα κυρίως σε στατικά στιγμιότυπα αυτών των συστημάτων. Επιδίωξή μας είναι να μελετήσουμε περιβάλλοντα όπου οι κόμβοι εισέρχονται και εξέρχονται συνεχώς.

Οι σύντομης διάρκειας ζωής προσομοιωτές είναι εξειδικευμένοι και πολύ δύσκολο να επεκταθούν ή να συγκριθούν μεταξύ τους. Προφανώς, για την υλοποίησή τους συνεπάγεται καταβολή διπλής προσπάθειας, καθώς δύο ή περισσότερα άτομα παράγουν και χρησιμοποιούν ίδια ή παρόμοια υλοποίηση ατομικά.

Από τους χαμηλού επιπέδου προσομοιωτές, οι NS-2 και NS-3 είναι μεταξύ των πιο δημοφιλών. Ο NS-2 [136] είναι ένας προσομοιωτής διακριτών γεγονότων, γραμμένος σε γλώσσα C++ και OTcl. Ο NS-2, και ο διάδοχός του NS-3, υποστηρίζουν ερευνητικούς και εκπαιδευτικούς σκοπούς για δίκτυα και συστήματα διαδικτύου. Ο NS-3 [1] παρέχει δομικά στοιχεία για το πως λειτουργούν τα δίκτυα και πως εκτελούνται οι μεταδόσεις των πακέτων, και παρέχει μια λεπτομερή μηχανή ώστε οι χρήστες να διεξάγουν τις πειραματικές τους μετρήσεις.

Οι προσομοιωτές επιπέδου εφαρμογής μπορεί να χρησιμοποιηθούν σε ενδεχομένως πολύπλοκες και ποικίλες συνθήκες, για να διαχειριστούν ένα ευρύ φάσμα αλληλεπιδράσεων σε ένα δίκτυο. Αυτοί οι προσομοιωτές αποτελούνται από δυο κύρια μέρη: α) τους κόμβους και β) το δίκτυο που διασυνδέει τους κόμβους μεταξύ τους. Ένας κόμβος συμπεριλαμβάνει τους πόρους (π.χ. αρχεία, μνήμη, ενέργεια). Το δίκτυο είναι συνήθως πολύ μεγάλο (απαρτίζεται από χιλιάδες κόμβους), στατικά ή δυναμικά όπου οι αλληλεπιδράσεις μεταξύ των κόμβων μεταβάλλονται και εξελίσσονται ραγδαία. Το δίκτυο χρησιμοποιείται για να μοντελοποιήσει με κάθε λεπτομέρεια τη δομή των πραγματικών συστημάτων. Οι PeerSim, SNAP, OMNeT++, και Oversim ανήκουν σε αυτή την κατηγορία προσομοιωτών.

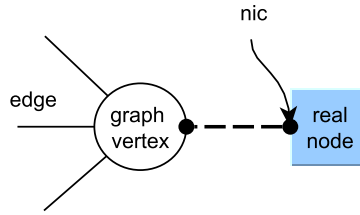
Ο PeerSim [141] είναι ένας προσομοιωτής για συστήματα p2p, βασισμένος σε Java με δύο μηχανές προσομοίωσης: μια cycle-based και μια event driven. Παρέχει υλοποιημένα μοντέλα όπως τα Pastry, Chord, Kademlia, Skpnet, Bittorrent, TMan και Cloudcast, ενώ ο χρήστης μπορεί να δημιουργήσει δικά του πρωτόκολλα. Η cycle-based μηχανή είναι κλιμακώσιμη καθώς αγνοεί τις λεπτομέρειες σε επίπεδο μεταφοράς που απαιτούνται για την προσομοίωση επικοινωνίας σε χαμηλό επίπεδο, αλλά έτσι παρέχονται λιγότερο ρεαλιστικά μοντέλα σε σχέση με την event driven μηχανή.

To Stanford Network Analysis Platform (SNAP) [113] είναι ένα σύστημα γενικού σκοπού, για την ανάλυση και τον χειρισμό μεγάλων δικτύων. Η βιβλιοθήκη του πυρήνα του SNAP είναι γραμμένη σε C++ και παρέχει διεπαφές προγραμματισμού εφαρμογών σε C++ και Python. Εύκολα επεκτείνεται σε μεγάλα δίκτυα με εκατοντάδες εκατομμυρίων κόμβους και δισεκατομμυρίων ακμών. Χειρίζεται αποδοτικά μεγάλους τακτικούς και τυχαίους γράφους, υπολογίζει δομικές τους ιδιότητες και υποστηρίζει χαρακτηριστικά των κόμβων και των ακμών. Όμως, σχεδιάστηκε για να υποστηρίξει μόνο την εξέλιξη της τοπολογίας των δικτύων και όχι διαδικασίες που συμβαίνουν πάνω σε αυτά.

Το OMNeT++ [188] είναι μια βιβλιοθήκη ανοικτού κώδικα και πλαίσιο προσομοίωσης, που χρησιμοποιείται ευρέως για την προσομοίωση πρωτοκόλλων επικοινωνίας και δικτύων. Αποτελείται από επιμέρους λειτουργικές μονάδες (modules), βασίζεται σε δομικά στοιχεία (components) και είναι γραμμένος σε C++. Χάρη στη γενική και ευέλικτη αρχιτεκτονική του χρησιμοποιείται κυρίως για την δημιουργία προσομοιωτών δικτύων σε διάφορα πεδία, όπως δίκτυα επικοινωνιών, σύνθετα συστήματα, δίκτυα ουρών ή αρχιτεκτονικές υλικού. Το δικό μας πλαίσιο προσομοίωσης βασίζεται στο OMNeT++.

Το Oversim [25] είναι ένα εργαλείο ανοικτού κώδικα υλοποιημένο σε C++ για προσομοίωση δικτύων p2p. Σχεδιάστηκε να εργάζεται πάνω από το περιβάλλον προσομοίωσης του OMNeT++. Ο προσομοιωτής παρέχει διάφορα μοντέλα για δομημένα (π.χ. Chord, Kademlia, Pastry) και αδόμητα (π.χ. GIA) συστήματα και πρωτόκολλα p2p. Ο χρήστης μπορεί να επεκτείνει το μοντέλο με την προσθήκη πρωτοκόλλων ή τοπολογιών δικτύου, όμως αυτή η διαδικασία είναι ιδιαίτερα απαιτητική, σύνθετη και χρονοβόρα.

Το SNAP υποστηρίζει μόνο μη δυναμικά δίκτυα, για παράδειγμα δεν καλύπτει την προσθήκη ή αφαίρεση κόμβων των μοντέλων μετά τον σχηματισμό της τοπολογίας. Από την άλλη πλευρά, το Oversim απαιτεί σημαντική προσπάθεια για να υλοποιηθεί με ακρίβεια ένα δικτυακό μοντέλο. Στο PeerSim, το δίκτυο είναι εντελώς αφηρημένο αποκλείοντας υλοποιήσεις που συμπεριλαμβάνουν πραγματικά χαρακτηριστικά (π.χ. latency σε φυσικό επίπεδο). Επίσης, το PeerSim δεν υποστηρίζει ικανοποιητικά τη συλλογή στατιστικών. Τα PeerSim και Oversim σχεδιάστηκαν ειδικά για την προσομοίωση δικτύων p2p ενώ το OMNeT++ είναι προσομοιωτής γενικής χρήσης. Επιπλέον, το OMNeT++ καλύπτει την ανάγκη για ακριβής προσομοίωση του δικτύου. Βασιζόμενοι στα παραπάνω, θεωρούμε ότι το OMNeT++ είναι



Σχήμα 6.1: Ένας κόμβος του δικτύου που αντιστοιχεί σε μια κορυφή του γράφου.

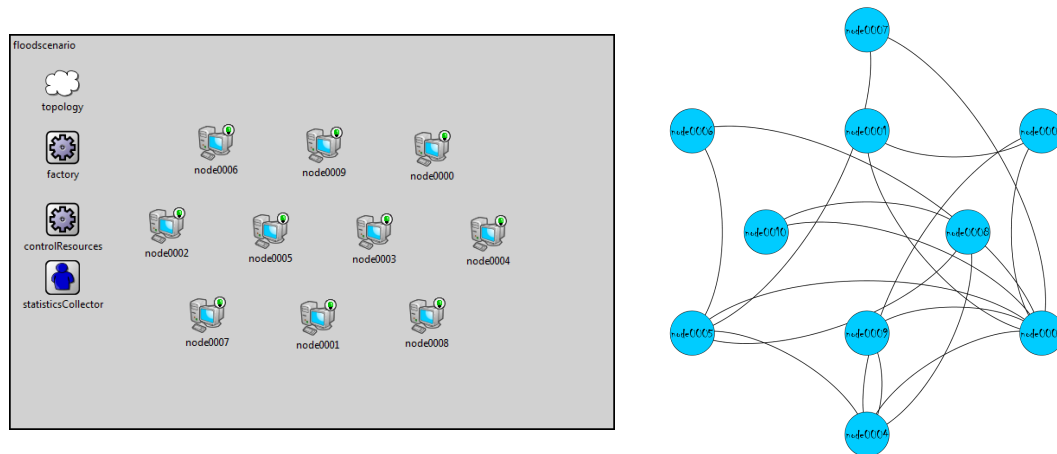
το πλέον κατάλληλο εργαλείο ως βάση για τους σκοπούς της εργασίας.

Μια άλλη ευέλικτη υποδομή για προσομοίωση σύνθετων δικτύων μεγάλης κλίμακας είναι το DRGSimLib toolbox [94] που επίσης υλοποιείται πάνω από το OM-NeT++. Λειτουργεί σε υψηλό επίπεδο αφάιρεσης, και οι αλληλεπιδράσεις μεταξύ των προσομοιούμενων αντικειμένων μιμούνται τα πραγματικά δίκτυα (επικοινωνιών). Το δίκτυο αναπαρίσταται ως γράφος και η τοπολογία του επιλέγεται από μια γεννήτρια διαφορετικών μοντέλων που λειτουργεί ως προσθήκη (plug-in). Παρακάτω δίνουμε περισσότερες λεπτομέρειες σχετικά με αυτή την βιβλιοθήκη, καθώς η τοπολογία της δικιάς μας υλοποίησης εμπνέεται εν μέρει από αυτή.

Οι κύριες λειτουργικές μονάδες της βιβλιοθήκης DRGSimLib είναι το *node*, το *nodeFactory*, και το *topology*. Το *node* είναι μια σύνθετη λειτουργική μονάδα που χρησιμοποιείται για να αναπαραστήσει κάθε προσομοιούμενο αντικείμενο, π.χ. τον κόμβο ενός δικτύου, όπως φαίνεται και στο Σχ. 6.1. Η λειτουργική μονάδα *nodeFactory* διαχειρίζεται τη δημιουργία και τη διαγραφή των κόμβων ενώ η λειτουργική μονάδα *topology* χειρίζεται τις συσχετίσεις του κόμβου ή τις διασυνδέσεις του δικτύου.

Η λειτουργική μονάδα *nodeFactory* λειτουργεί ανεξάρτητα από τα άλλα δομικά στοιχεία και διαχειρίζεται τον πληθυσμό των κόμβων στο σύστημα. Δημιουργεί και διαγράφει κόμβους σύμφωνα με τις προδιαγραφές για τη διάρκεια ζωής τους. Η λειτουργική μονάδα *node*, περιλαμβάνει ένα ουσιαστικό δομικό στοιχείο, το *topologyControlNic* (*nic*) που καταχωρεί ή καταργεί τον κόμβο από την τοπολογία του δικτύου. Το δομικό στοιχείο *nic* λειτουργεί ως μια διεπαφή μεταξύ του προσομοιούμενου αντικειμένου και μιας κορυφής του γράφου που αναπαριστά την τοπολογία του δικτύου. Είναι ενήμερο για το ποιοι είναι οι γειτονικοί του κόμβοι και για κάθε αλλαγή συνδέσεων που γίνεται τοπικά στη γειτονιά του.

Η λειτουργική μονάδα *topology* διαχειρίζεται τη δημιουργία και τη διατήρηση του γράφου που αναπαριστά το δικτυωμένο σύστημα. Αποτελείται από δύο δομικά



Σχήμα 6.2: Ένα στιγμιότυπο μιας προσομοίωσης με την Armonia: το δίκτυο OM-NeT++ (αριστερά) και η αντίστοιχη τοπολογία (δεξιά).

στοιχεία, τα: *topologyControl* και *generator*. Το δομικό στοιχείο *generator* λειτουργεί ως προσθήκη και χειρίζεται τον σχηματισμό του δικτύου. Το δομικό στοιχείο *topologyControl* επιβλέπει το γράφο και επικοινωνεί με το δομικό στοιχείο *generator* για προσθήκη ή διαγραφή κόμβων. Επίσης σε συνεργασία με το δομικό στοιχείο *nic*, διατηρεί τη δομή του γράφου. Η λειτουργική μονάδα *topology* είναι υπεύθυνη για την αυστηρή αντιστοιχία κάθε κορυφής του γράφου με κάθε κόμβο καθώς και για τη διαχείριση του *churn* των κόμβων.

Συνοψίζοντας, τα διαθέσιμα πακέτα προσομοίωσης σε επίπεδο εφαρμογής, παρέχουν μοντέλα μόνο για την τοπολογία του δικτύου και την εξέλιξή τους. Εξυπηρετούν σκοπούς όπως τη μελέτη της συμπεριφοράς της δόμησης και της εξέλιξης του συστήματος στο χρόνο και την άντληση γραφο-θεωρητικών μετρήσεων. Δεν υπάρχουν γνωστοί προσομοιωτές γενικού σκοπού που να επιτρέπουν πειραματικές μετρήσεις για αλγορίθμους αναζήτησης και στρατηγικές αντιγράφων, που να μπορούν να πραγματοποιηθούν αποδοτικά σε διάφορες τοπολογίες, παρόλο που αυτά τα προβλήματα έτυχαν ιδιαίτερης προσοχής στην ανοιχτή βιβλιογραφία. Ο προσομοιωτής μας, ο οποίος παρουσιάζεται στη συνέχεια, επιχειρεί να καλύψει αυτό το κενό.

6.3 Το δικό μας πλαίσιο προσομοίωσης: Armonia

Στόχος μας είναι ένα πλαίσιο προσομοίωσης που θα κάνει εύκολη τη μελέτη αυθαίρετων στρατηγικών αναζήτησης και εκχώρησης αντιγράφων σε αυθαίρετα σύνθετα δυναμικά δίκτυα. Ενώ οι άλλοι γνωστοί προσομοιωτές στοχεύουν στη μοντελοποίηση δικτύων μεγάλης κλίμακας, εμείς επιδιώκουμε να μοντελοποιήσουμε και τις σύνθετες αλληλεπιδράσεις πάνω σε αυτά τα δίκτυα, όπως η αναζήτηση ή η εκχώρηση αντιγράφων ενός πόρου. Ο προσομοιωτής μας είναι γενικής χρήσης και παρέχει:

- τη διαχείριση της τοπολογίας, δηλαδή τη δημιουργία των κόμβων και των συνδέσεων τους σύμφωνα με στατικά και δυναμικά μοντέλα δικτύων. Μπορεί να χρησιμοποιηθεί για να οριστούν αυθαίρετα τυχαίες τοπολογίες ενώ παρέχεται ένας αριθμός, από έτοιμες για χρήση τοπολογίες.
- τη διαχείριση των πόρων-αντιγράφων, παρέχοντας μια συλλογή από δομικά στοιχεία που ελέγχουν την δημιουργία και την εκχώρηση των πόρων.
- τη διαχείριση πρωτοκόλλων αναζήτησης για την υλοποίηση και την εφαρμογή προκαθορισμένων μεθόδων αναζήτησης σε διάφορα δίκτυα.
- τη διαχείριση των στατιστικών, που επιτρέπει τη συλλογή καθολικών ή τοπικών στατιστικών, ανάλογα με τις απαιτήσεις του χρήστη.

Το Armonia απαρτίζεται από τις λειτουργικές μονάδες topology, factory, controlResource και statisticCollector που παρουσιάζονται αναλυτικά στη συνέχεια. Στο Σχ. 6.2 απεικονίζεται ένα παράδειγμα από ένα απλό δίκτυο και τη δομή του. Περιλαμβάνει τους κόμβους, όπως παρουσιάζονται από το γραφικό περιβάλλον του OMNeT++, μαζί με την τοπολογία όπως οπτικοποιείται με το γνωστό εργαλείο Gephi [21].

6.3.1 Η τοπολογία

Όπως αναφέρθηκε παραπάνω, η βιβλιοθήκη DRGSimLib [94] αποτέλεσε την βάση για το τμήμα της τοπολογίας του Armonia. Τροποποιήσαμε αυτή τη βιβλιοθήκη σημαντικά, προσθέτοντας νέες λειτουργίες και κάνοντας διαρθρωτικές βελτιώσεις. Επίσης, συμπεριλάβαμε πρόσθετες δικτυακές τοπολογίες. Συγκεκριμένα, διατηρήσαμε το αφηρημένο δικτυακό πλαίσιο, που περιλαμβάνει τις λειτουργικές μονάδες

topology και nodeFactory για τη δημιουργία και διαγραφή των κόμβων και τη διαχείριση των συνδέσεων τους. Το topology αποτελείται από ένα δομικό στοιχείο για τη διαχείριση της τοπολογίας (*topologyControl*) και μια συλλογή από δικτυακά μοντέλα (*generator*). Στη λειτουργική μονάδα topology έγιναν επεκτάσεις και προσαρμογές. Συγκεκριμένα σε σχέση με τη βιβλιοθήκηDRGSimLib

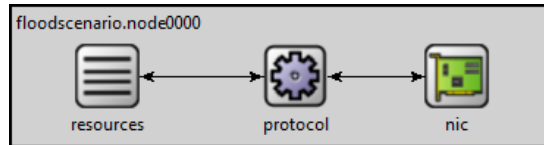
- Προστέθηκαν νέα δικτυακά μοντέλα.
- Δεν επιτρέπονται οι πολλαπλές ακμές ανάμεσα σε ένα ζεύγος κόμβων.
- Προστέθηκε η ιδιότητα της ομαδοποίησης στα δικτυακά μοντέλα.
- Το δομικό στοιχείο του κόμβου (*node*) επανασχεδιάστηκε και προστέθηκαν νέα χαρακτηριστικά.

Στο σύνολό τους, αυτές οι τροποποιήσεις προσθέτουν επιπλέον δυνατότητες και λειτουργίες στον προσομοιωτή.

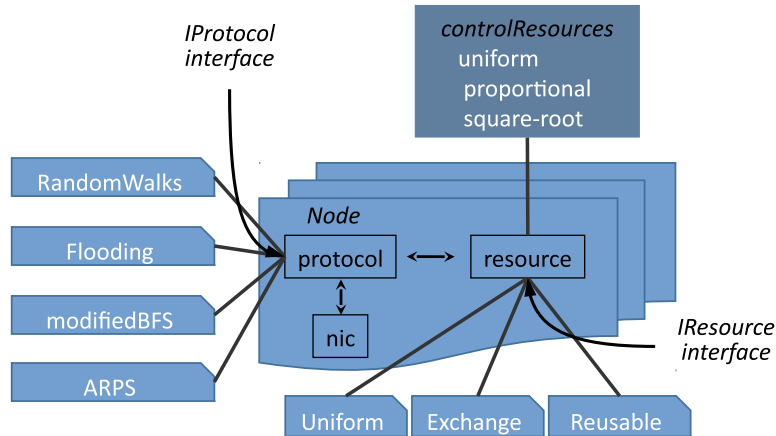
Νέα μοντέλα δικτύων. Οι τοπολογίες που υλοποιήθηκαν από τους συγγραφείς του [94] περιλαμβάνουν Barabási, FullyConnected, random, Tree και gnpRandom (μια τροποποιημένη έκδοση του δικτύου Erdős-Rényi). Εμείς έχουμε υλοποιήσει τον κλασσικό τυχαίο γράφο Erdős-Rényi [62] και το μοντέλο Caveman-Solaria που επιτρέπει την δημιουργία δικτύων που εμφανίζουν την ιδιότητα του μικρού κόσμου [195]. Επιπλέον υλοποιήσαμε μια νέα κατηγορία που παρέχει την δυνατότητα πρόσβασης και επεξεργασίας σε σύνολα δεδομένων από πραγματικά δίκτυα μεγάλης κλίμακας (όπως αυτά που περιλαμβάνονται στο SNAP [113]).

Πολλαπλές ακμές. Βάσει της επιλογής του χρήστη, οι συνδέσεις μεταξύ ενός ζεύγους κόμβων (u, v) μπορεί να περιοριστούν στο πολύ σε μια (ακμή). Με αυτό τον τρόπο, αποφεύγεται η τοποθέτηση πολλαπλών ακμών ανάμεσα σε ένα ζεύγος κόμβων κατά την δημιουργία ή την εξέλιξη του δικτύου, που μπορεί να επηρεάζει την ορθή λειτουργία ορισμένων πρωτοκόλλων αναζήτησης.

Clustering coefficient. Η ομαδοποίηση είναι ένα κοινό χαρακτηριστικό που εντοπίζεται σε πολλά πραγματικά δίκτυα και δίνεται από την (2.6). Η ιδιότητα της ομαδοποίησης είναι ένας κρίσιμος παράγοντας [169] και επηρεάζει τη λειτουργικότητα και τη δομή του δικτύου. Ο χρήστης μπορεί να την προσαρμόσει μέσω μιας



Σχήμα 6.3: Η δομή του κάθε κόμβου, όπως υλοποιείται από το Armonia.



Σχήμα 6.4: Η χρήση των δομικών στοιχείων protocol και resource στην Armonia.

παραμέτρου του συστήματος (*clusterCoefficient*) που προστέθηκε στη γεννήτρια της τοπολογίας. Θέτοντας την τιμή της στο μηδέν, η ιδιότητα αυτή απενεργοποιείται.

Το δομικό στοιχείο του κόμβου. Ο κόμβος αναπαριστά κάθε οντότητα, μαζί με τα χαρακτηριστικά της. Διατηρεί το αυθεντικό δομικό στοιχείο επικοινωνίας (*nic*). Αν και παραμένει το δομικό στοιχείο *protocol* οι λειτουργίες του επαναπροσδιορίζονται. Μια διεπαφή (*IProtocol*) συνδέει ως προσθήκη ένα από τα διαθέσιμα υλοποιημένα πρωτόκολλα αναζήτησης σύμφωνα με το εκάστοτε σενάριο εκτέλεσης. Επιπλέον προστέθηκε το δομικό στοιχείο *resource* για να διαχειριστεί τους πόρους που παρέχονται από τον κόμβο. Και εδώ, μια διεπαφή (*IResource*) συνδέει στο δομικό στοιχείο *resource* την επιλεγόμενη από το χρήστη πολιτική διαχείρισης αντιγράφων. Τα δομικά στοιχεία του κόμβου απεικονίζονται στο Σχ. 6.3. Το πρωτόκολλο αναζήτησης και η πολιτική διαχείρισης αντιγράφων που προσαρτώνται στο *protocol* και *resource* αντίστοιχα, μπορούν να οριστούν στο αρχείο διαμόρφωσης του προσομοιωτή.

6.3.2 Οι πόροι και τα αντίγραφά τους

Οι πόροι κατανέμονται σε όλο το δίκτυο, σε διάφορους κόμβους και μπορούν να προσπελαστούν από άλλους κόμβους. Σε ένα τέτοιο περιβάλλον εγείρονται σημα-

ντικά ζητήματα που διευθετούνται με:

- Το μοντέλο κατανομής πόρων (resource distribution) που καθορίζει πόσοι μοναδικοί πόροι είναι διαθέσιμοι στο σύστημα και πως αντιστοιχίζονται στους κόμβους.
- Το μοντέλο κατανομής αντιγράφων (replica distribution), που προσδιορίζει πόσα αντίγραφα του κάθε πόρου εμφανίζονται στο σύστημα.
- Το μοντέλο απόκτησης πόρων (resource acquisition), που εστιάζει στην απόκτηση των κατάλληλων πόρων που καλύπτουν τις ανάγκες του κόμβου.

Το δικό μας πλαίσιο διαχείρισης πόρων είναι αρκετά γενικό, επιτρέποντας στους ερευνητές να ενσωματώσουν τη δική τους πολιτική διαχείρισης πόρων.

Η λειτουργική μονάδα *controlResource* είναι μια απλή λειτουργική μονάδα OM-NeT++ και χρησιμοποιείται για να χειριστεί τους πόρους και την κατανομή των αντιγράφων. Υποθέτοντας ότι υπάρχουν m διαφορετικού τύπου πόροι στο σύστημα και κάθε πόρος i διατηρεί αντίγραφα σε r_i κόμβους, ο συνολικός αριθμός πόρων είναι $R = \sum_{i=1}^m r_i$. Υλοποιήσαμε τρία διαφορετικά στατικά σχήματα αντιγράφων, τα: uniform, proportional και square-root [122], ως ακολούθως:

- *Uniform*: ίδιος αριθμός αντιγράφων για όλους τους πόρους ($r_i = R/m$)
- *Proportional*: Ο αριθμός των αντιγράφων ενός πόρου i είναι ανάλογος της σχετικής δημοτικότητας, που συμβολίζεται ως q_i και ορίζεται ως το ποσοστό των ερωτήσεων που τίθενται για i^{th} πόρο ($r_i \propto q_i$).
- *Square-root*: Ο αριθμός των αντιγράφων ενός πόρου i είναι ανάλογος της τετραγωνικής ρίζας της σχετικής δημοτικότητας ($r_i \propto \sqrt{q_i}$).

Οι παράμετροι της λειτουργικής μονάδας *controlResource* περιλαμβάνουν το μέγιστο αριθμό των διαφορετικών πόρων που υπάρχουν στο σύστημα (*maxNumOfResources*) και το όνομα του πολιτικής κατανομής αντιγράφων (*distributionName*).

Ο αριθμός και η κατανομή των πόρων καθορίζεται κεντρικά. Σε κάθε κόμβο αντιστοιχίζεται και διατηρείται ένα σύνολο πόρων. Το δομικό στοιχείο *resource* του κόμβου χειρίζεται και την απόκτηση των πόρων. Στην αρχικοποίηση του κόμβου, η μονάδα *resource* αποκτά πόρους από τη λειτουργική μονάδα *controlResources*

εφαρμόζοντας πολιτική εκχώρησης *uniform*, όπου οι πόροι επιλέγονται ομοιόμορφα τυχαία από το σύνολο των διαθέσιμων πόρων.

Μετά τον σχηματισμό του δικτύου, με την έναρξη του *churn* οι κόμβοι αρχίζουν να εισέρχονται και να εξέρχονται από το δίκτυο, κάθε νεοεισερχόμενος κόμβος αποκτά πόρους σύμφωνα με μια από τις τρεις προκαθορισμένες πολιτικές: *uniform*, *reusable* και *exchange*. Στην πολιτική *uniform*, οι πόροι εξακολουθούν να επιλέγονται ομοιόμορφα τυχαία από το σύνολο των διαθέσιμων πόρων. Στην πολιτική *reusable*, πριν ένας κόμβος εγκαταλείψει το σύστημα, επιστρέφει τους πόρους του στη μονάδα *controlResources* για να ξαναχρησιμοποιηθούν. Στη συνέχεια, όταν ένας νέος κόμβος συνδέεται στο σύστημα, η μονάδα *controlResources* θα του εκχωρήσει πόρους από τους επαναχρησιμοποιούμενους, εφόσον υπάρχουν. Τέλος, αν εφαρμόζεται η πολιτική *exchange*, ο κόμβος ζητά πόρους από τους γειτονικούς του, επιλέγοντας ένα τυχαίο σύνολο και το αντιγράφει τοπικά. Οι παραπάνω πολιτικές μας επιτρέπουν να μοντελοποιήσουμε συμπεριφορές που απαντώνται σε πραγματικά συστήματα. Για παράδειγμα η πολιτική *exchange* χρησιμοποιείται σε συστήματα p2p (π.χ. στη Gia [37]) και σε εικονικές κοινότητες, ενώ η πολιτική *reusable* χρησιμοποιείται για τη μοντελοποίηση του σχήματος διαχείρισης πόρων στο Hadoop [198].

6.3.3 Το πρωτόκολλο αναζήτησης

Στο Armonia, το δομικό στοιχείο *protocol* είναι υπεύθυνο για την εφαρμογή των αλγορίθμων αναζήτησης στο δίκτυο. Είναι ενσωματωμένο στη μονάδα του κόμβου. Το κάθε υλοποιημένο πρωτόκολλο αναζήτησης, που κληρονομεί από μια βασική κλάση (*BaseProtocol*), συνδέεται στο δομικό στοιχείο *protocol* την ώρα της εκτέλεσης μέσω της διεπαφής *IProtocol*. Ο στόχος της κλάσης *BaseProtocol* είναι διπλός: πρώτον, διευκολύνει τη χρησιμοποίηση των διάφορων αλγορίθμων αναζήτησης, καθορίζοντας απλώς το όνομα της κλάσης του πρωτοκόλλου στο αρχείο ρυθμίσεων του προσομοιωτή, και δεύτερο, πραγματοποιεί όλες τις κοινές εργασίες (*BaseProtocol*) που απαιτούνται από την διαδικασία της αναζήτησης.

Η υποδομή μας περιλαμβάνει έναν μηχανισμό, που βοηθά στην αποφυγή της περαιτέρω προώθησης των ήδη απεσταλμένων μηνυμάτων. Στην πραγματικότητα εφαρμόζει έναν τοπικό μηχανισμό παρακολούθησης της κάθε σύνδεσης του κόμβου για να περιορίσει τις επαναμεταδόσεις σε αυτή την ακμή. Για να γνωρίζει ένας κόμβος, αν ένα συγκεκριμένο ερώτημα λήφθηκε και μεταδόθηκε στο παρελ-

θόν, διατηρεί μια δομή *pastQueriesBuffer* που αποθηκεύει προηγούμενες ερωτήσεις. Ο χρήστης μπορεί να περιορίσει την χωρητικότητα του *pastQueriesBuffer* μέσω της παραμέτρου *pastQueryCounter*. Επιπλέον, σημειώνει την ακμή από όπου έφτασε το μήνυμα και (με την βοήθεια ενός ρυθμιζόμενου flag) δεν την περιλαμβάνει στην ομάδα προώθησης. Τέλος, παρέχει επίσης ένα μηχανισμό για την επικοινωνία με το δομικό στοιχείο *resource* ώστε να χειρίζεται την εκχώρηση πόρων στο στάδιο της αρχικοποίησης των νέων κόμβων (π.χ. όταν εφαρμόζεται η πολιτική απόκτησης πόρων *exchange*).

Οι παραπάνω μηχανισμοί μπορούν να χρησιμοποιηθούν από κάθε πολιτική αναζήτησης και προαιρετικά μπορούν να παρακαμφθούν, αν αυτό απαιτείται. Η διαδικασία της αναζήτησης ολοκληρώνεται λαμβάνοντας υπόψη την αρχικοποίηση του ερωτήματος και την επιλογή της ομάδας προώθησης. Η όλη υποδομή απλοποιεί σημαντικά την υλοποίηση νέων σχημάτων αναζήτησης ενώ παρέχεται πλήθος από πλήρως υλοποιημένες στρατηγικές αναζήτησης:

- *Flooding*, η απλή πλημμύρα, η οποία είναι πιθανώς η πιο γνωστή και πιο γενική στρατηγική. Χρησιμοποιείται συχνά λόγω έλλειψης οποιασδήποτε πληροφορίας σχετικά με την τοποθέτηση των πόρων. Η πλημμύρα είναι γενικά γρήγορη, αλλά παράγει υπερβολικό αριθμό μηνυμάτων τα οποία τις περισσότερες φορές είναι διπλότυπα (περιττά).
- *Random walkers* [122, 77]. Στον τυπικό τυχαίο περιπατητή, όταν ένας κόμβος λάβει μια ερώτηση και εφόσον δεν διατηρεί τον επιθυμητό πόρο, προωθεί το μήνυμα σε έναν τυχαία επιλεγμένο γειτονικό του κόμβο. Εναλλακτικά, μπορούν να εφαρμοστούν πολλαπλοί περιπατητές που διαδίδουν την ερώτηση παράλληλα και εντοπίζουν τον πόρο γρηγορότερα, με κόστος τις περισσότερες μεταδόσεις μηνυμάτων.
- *Modified BFS* [95], όπου το μήνυμα ερώτημα προωθείται σε κάθε γειτονικό κόμβο με μια σταθερή πιθανότητα (*forwarding probability*) για να μειώσει τον αριθμό των περιττών μηνυμάτων της πλημμύρας.
- *ARPS* [204], που χρησιμοποιεί διαφορετικές πιθανότητες προώθησης για διαφορετικούς πόρους, ανάλογα με την εκτίμηση της δημοτικότητας του πόρου.
- *APF* [2], όπου η πιθανότητα προώθησης μεταβάλλεται σύμφωνα με την απόσταση από τον αρχικό κόμβο και την τοπική γνώση σχετικά με την δημοτικό-

τητα του πόρου την οποία την παρουσιάσαμε στο Κεφάλαιο 5.

6.3.4 Στατιστικά

Η συλλογή των στατιστικών εξυπηρετεί δύο σκοπούς: πρώτον, ο χρήστης μπορεί να λάβει μετρήσεις για το σύστημα συνολικά (καθολικά στατιστικά) και δεύτερο, ο κάθε κόμβος μπορεί να συλλέγει τοπικά αποτελέσματα (τοπικά στατιστικά) για να εκτιμήσει κάποια μέτρηση σχετική με το δίκτυο ή τους πόρους (π.χ. οι σχετικές δημοτικότητες των πόρων) που μπορεί να αλλάξουν τον τρόπο εξέλιξης των δυναμικών διεργασιών που συμβαίνουν στο δίκτυο.

Η λειτουργική μονάδα *statisticCollector* είναι μια απλή μονάδα του OMNeT++ και χρησιμοποιείται για να χειριστεί και να συναθροίσει στατιστικά από ολόκληρο το δίκτυο. Τα στατιστικά έχουν την ακόλουθη μορφή:

$$\langle \text{statName}, \text{variable1}, \text{variable2} \rangle,$$

όπου *statName* είναι το όνομα της μετρούμενης ποσότητας (π.χ. μηνύματα) και η *variable2* συναθροίζει και περιγράφει μετρήσεις όταν η *variable1* έχει μια συγκεκριμένη τιμή· για παράδειγμα, αν η *statName* χρησιμοποιείται για να μετρηθούν τα μηνύματα σε κάθε βήμα της διαδικασίας αναζήτησης, η *variable1* αποθηκεύει την τιμή της παραμέτρου TTL (δηλαδή το βήμα) και η *variable2* καταμετρά το συνολικό αριθμό μηνυμάτων για αυτό το συγκεκριμένο TTL.

Για τη συλλογή των τοπικών στατιστικών, κατασκευάσαμε την κλάση *localStatistics*, που περιλαμβάνεται στη μονάδα του κόμβου. Χρησιμοποιεί την ίδια μορφοποίηση, όπως και η μονάδα των καθολικών στατιστικών.

6.4 Ενδεικτική μελέτη προσομοίωσης και συζήτηση

Αξιολογήσαμε το Armonia, εφαρμόζοντας διάφορους συνδυασμούς αλγορίθμων αναζήτησης, τοπολογιών και πολιτικών διαχείρισης πόρων/αντιγράφων.

Για ένα σενάριο προσομοίωσης, ο χρήστης θα πρέπει μόνο να καθορίσει (μέσω ρυθμιζόμενων παραμέτρων): την επιθυμητή πολιτική αναζήτησης, τη δομή του δικτύου και τη στρατηγική των αντιγράφων. Όταν μια περίοδος προσομοίωσης ολοκληρωθεί, τα αποτελέσματα (καθολικά συλλεγόμενα στατιστικά) είναι διαθέσιμα και μπορούν να αξιολογηθούν.

Εδώ παρουσιάζεται ενδεικτικά ένα παράδειγμα προσομοίωσης που αφορά τη συγκριτική μελέτη της απόδοσης διάφορων στρατηγικών αναζήτησης σε διάφορες τοπολογίες. Εξετάζουμε την απόδοση χρησιμοποιώντας τρεις καταμετρήσεις:

- Η *πιθανότητα επιτυχίας* (probability of success), που μετρά την πιθανότητα μια ερώτηση να απαντηθεί, για δεδομένο όριο (βήματος) TTL. Μια ερώτηση θεωρείται επιτυχής αν εντοπίσει τουλάχιστον ένα αντίγραφο του πόρου που αναζητείται.
- Ο συνολικός αριθμός των μηνυμάτων (messages) που μεταδόθηκαν.
- Ο αριθμός των διπλότυπων μηνυμάτων (duplicate messages) που χρησιμεύει ως ένα μέτρο, για το πόσο αποτελεσματικά μια πολιτική αξιοποιεί τους πόρους του δικτύου.

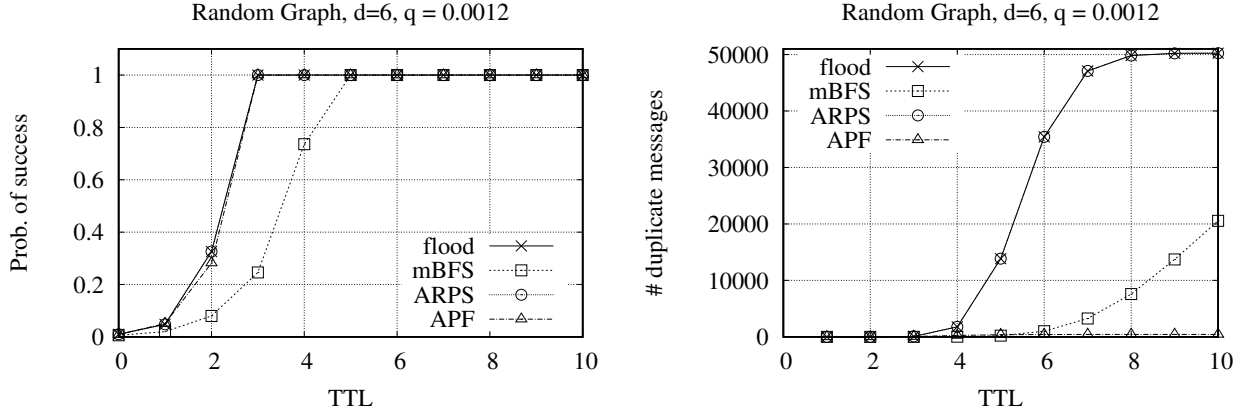
Συγκρίνουμε τέσσερις στρατηγικές αναζήτησης βασιζόμενες στην πλημμύρα, που είναι οι: απλή πλημμύρα (flood), modified BFS (mBFS), Adaptive Resource-based Probabilistic Search (ARPS) και APF τις οποίες παρουσιάσαμε στις Ενότητες 5.2 και 5.4. Όλες οι στρατηγικές περιορίζονται ως προς το μέγιστο αριθμό βημάτων που καθορίζεται από την παράμετρο TTL. Για τις τελευταίες τρεις στρατηγικές, η πιθανότητα προώθησης (p_f) ακολουθεί τις κατευθυντήριες γραμμές των συγγραφέων. Συγκεκριμένα, στην mBFS, η ερώτηση προωθείται στο 50% των γειτονικών κόμβων ($p_f = 0.5$), ενώ μια τιμή $p_f = 0.8$ χρησιμοποιείται στην ARPS. Η πλημμύρα με $p_f = 1$ επικοινωνεί ουσιαστικά με όλους τους κόμβους του δικτύου σε κάθε περίπτωση. Για την APF, η πιθανότητα προώθησης υπολογίζεται σε κάθε κόμβο [2]. Όλες οι πολιτικές αναζήτησης αξιοποιούν το μηχανισμό pastQueriesBuffer, με χωρητικότητα τουλάχιστον 500, για τη μείωση των διπλότυπων μηνυμάτων.

Η μελέτη εστιάζει σε δύο μοντέλα δικτύων:

- Το κλασσικό τυχαίο δίκτυο ER που σχηματίζεται από N κόμβους και κάθε ζεύγος κόμβων συνδέεται με μια ακμή με πιθανότητα $p \in (0, 1)$. Ο αναμενόμενος μέσος βαθμός ($\bar{d}$) δίνεται από $\bar{d} = pN$.
- Το μοντέλο Barabási (BA) που αρχικά έχει m_0 συνδεδεμένους κόμβους και κάθε νεοεισερχόμενος κόμβος συνδέεται με m κόμβους που υπάρχουν ήδη στο δίκτυο.

Πίνακας 6.1: Παράμετροι προσομοίωσης για στατικά δίκτυα

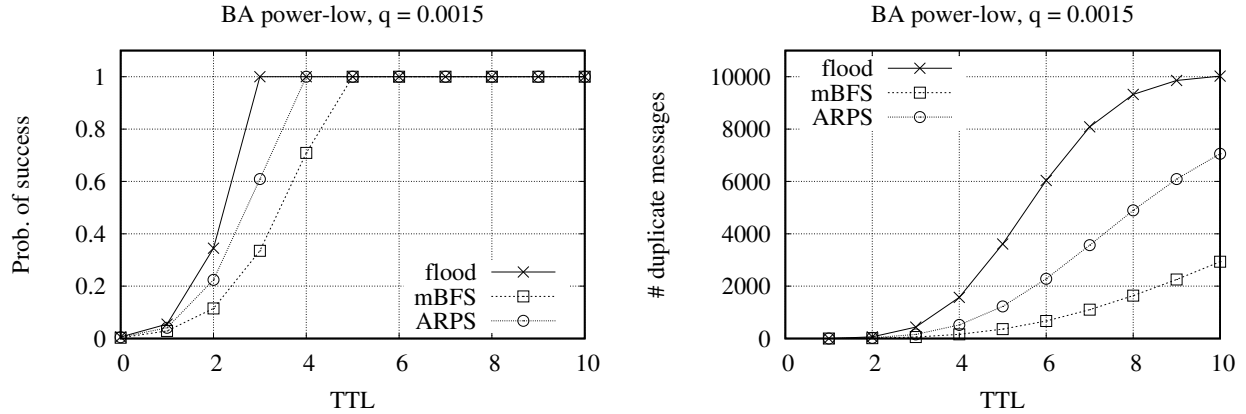
graph	parameter	nodes	r_i	flood	mBFS	ARPS
ER	$d = 6$	25000	30	$p_f = 1$	$p_f = 0.5$	$p_f = 0.8$
BA	$m = 2$	10000	15	$p_f = 1$	$p_f = 0.5$	$p_f = 0.8$



Σχήμα 6.5: Armonia: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα στατικό δίκτυο ER 25.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0012$.

Αρχικά εξετάζουμε στατικά δίκτυα, που σημαίνει ότι μετά τον σχηματισμό της τοπολογίας, κανένας κόμβος δεν εισέρχεται ή εξέρχεται από το δίκτυο. Εφαρμόζουμε ομοιόμορφη πολιτική εκχώρησης και κατανομής πόρων/αντιγράφων. Ένας προκαθορισμένος αριθμός αντιγράφων (r_i) για κάθε πόρο τοποθετείται ομοιόμορφα τυχαία στους κόμβους, όπου η τιμή r_i είναι ίδια για όλους τους πόρους i . Κάθε κόμβος μπορεί να υποβάλλει μια ερώτηση μέχρι ένας μέγιστος συνολικός αριθμός να συμπληρωθεί. Για αυτά τα πειράματα, ο συνολικός αριθμός των ερωτήσεων περιορίζεται στις 1000. Οι παράμετροι προσομοίωσης συνοψίζονται στον Πίνακα 6.1.

Τα καθολικά συλλεγόμενα αποτελέσματα (με τη βοήθεια της λειτουργικής μονάδας statisticCollector) της προσομοίωσης φαίνονται στα Σχ. 6.5–6.6 και μας παρέχουν άμεσα πληροφόρηση για την απόδοση των πρωτοκόλλων. Στις γραφικές παραστάσεις απεικονίζονται η πιθανότητα επιτυχίας και ο αριθμός των διπλότυπων μηνυμάτων. Στο τυχαίο δίκτυο *ER* για το συγκεκριμένο σενάριο εκτέλεσης, όλες οι στρατηγικές αναζήτησης εκτός από την mBFS, επιτυγχάνουν τον εντοπισμό του αιτούμενου πόρου από το τρίτο βήμα ($TTL \geq 3$). Η mBFS στρατηγική απαιτεί τουλάχιστον 5 βήματα. Η αποτελεσματικότητα των στρατηγικών εμφανίζεται ξε-



Σχήμα 6.6: Armonia: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα στατικό δίκτυο BA 10.000 κόμβων, με $m = 2$ και $q = 0.0015$.

κάθαρα με τον αριθμό των διπλότυπων μηνυμάτων. Ενώ η mBFS στρατηγική είναι πιο αργή στον εντοπισμό των πόρων, παράγει επίσης και μικρότερο αριθμό διπλών μηνυμάτων. Η ARPS φαίνεται να συμπεριφέρεται σχεδόν σαν την πλημμύρα ενώ η APF επιτυγχάνει τον χαμηλότερο αριθμό διπλότυπων μηνυμάτων και πιθανότητα επιτυχίας σχεδόν όμοια με αυτή της πλημμύρας.

Στο σενάριο προσομοίωσης σε δίκτυα BA δεν περιλαμβάνεται η APF στρατηγική, καθώς για την καλή της απόδοση απαιτείται γνώση του μέσου βαθμού των κόμβων σε δεδομένη απόσταση από τον αρχικό κόμβο που θέτει την ερώτηση. Από όσα γνωρίζουμε, δεν είναι γνωστή η αναλυτική εκτίμησή του σε δίκτυα power-law. Εκτενέστερη μελέτη για αυτό το ζήτημα γίνεται στο επόμενο κεφάλαιο. Σε αυτή την περίπτωση, βασιζόμενοι πάλι στα καθολικά συλλεγόμενα αποτελέσματα, διαπιστώνουμε ότι η mBFS είναι πιο αργή (επιτυχής μετά το 5ο βήμα) αλλά και πιο αποδοτική (λιγότερα διπλότυπα μηνύματα). Η ARPS χρειάζεται τουλάχιστον 4 βήματα για τον εντοπισμό του πόρου ενώ παράγει λιγότερα διπλότυπα μηνύματα συγκριτικά με την πλημμύρα.

Στη συνέχεια αξιοποιώντας τις δυνατότητες του Armonia εξετάζουμε δυναμικά δίκτυα με διάφορους ρυθμούς churn. Ένας κόμβος συμμετέχει στο σύστημα για ένα ορισμένο χρονικό διάστημα, που ονομάζεται *lifetime*, και είναι ο χρόνος που έχει παρέλθει από την πρώτη εμφάνιση του στο σύστημα μέχρι την οριστική του αποχώρηση. Για τη μοντελοποίηση της διάρκειας ζωής (*lifetime*) του κόμβου, μπορεί να χρησιμοποιηθεί κάθε προκαθορισμένη κατανομή του OMNeT++.

Για το συγκεκριμένο σύνολο πειραμάτων, επιλέχτηκαν δύο από τις πιο συχνά

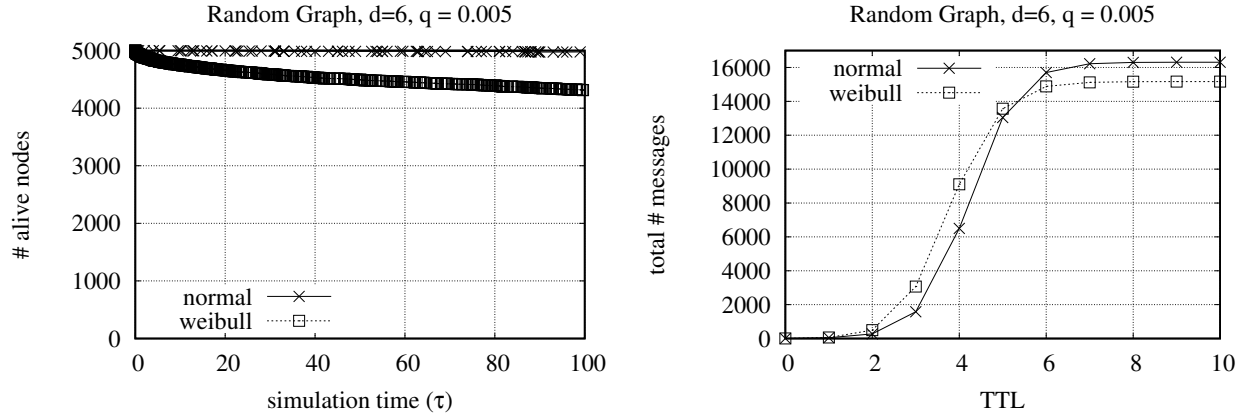
Πίνακας 6.2: Παράμετροι προσομοίωσης για δυναμικά δίκτυα

churn distr.	ER ($d = 6, r_i = 25$)	BA ($m = 2, r_i = 15$)
ND	$\mu = 4h, \sigma = \mu/3$	$\mu = 4h, \sigma = \mu/3$
WD	$a = 4h, b = 0.5$	$a = 4h, b = 0.5$

χρησιμοποιούμενες κατανομές διάρκειας ζωής στη μελέτη σύνθετων δικτύων: η Normal Distribution (ND), με παραμέτρους τη μέση τιμή (μ) και την τυπική απόκλιση (σ) και η Weibull Distribution (WD) με παραμέτρους a και b . Σε όλες τις περιπτώσεις θέσαμε την ίδια μέση τιμή ($\mu = 4$ hours). Επίσης, θεωρούμε ότι οι κόμβοι εισέρχονται στο δίκτυο σύμφωνα με μια κατανομή Poisson. Για δίκτυο ER με 5.000 αρχικούς κόμβους, χρησιμοποιείται ρυθμός άφιξης των κόμβων $1/\lambda = 4$ sec, ενώ το δίκτυο BA αρχικοποιείται με 10.000 κόμβους, και χρησιμοποιείται ρυθμός άφιξης $1/\lambda = 2$ sec. Όπως και στην περίπτωση του στατικού δικτύου, εφαρμόζουμε πολιτική ομοιόμορφης εκχώρησης και κατανομής πόρων και των αντιγραφών τους. Η προσομοίωση διεξάγεται έως ότου υποβληθούν 1000 ερωτήσεις. Για την περίπτωση των δυναμικών δικτύων εφαρμόζουμε μόνο την αναζήτηση με πλημμύρα καθώς εδώ εστιάζουμε στην αξιολόγηση του Armonia. Στο επόμενο κεφάλαιο θα μελετήσουμε διεξοδικά διάφορες τεχνικές αναζήτησης σε δυναμικά δίκτυα. Οι παράμετροι προσομοίωσης για το δυναμικό δίκτυο συνοψίζονται στον Πίνακα 6.2.

Τα αποτελέσματα των μετρήσεων απεικονίζονται στα Σχ. 6.7–6.8. Αριστερά φαίνεται η εξέλιξη του δικτύου υπό συνθήκες churn για τα πρώτα 100 sec του χρόνου προσομοίωσης. Όπως διαπιστώνεται, οι δύο κατανομές διάρκειας ζωής οδηγούν σε διαφορετική εξέλιξη του δικτύου. Βέβαια, η εξέλιξη εξαρτάται και από τις παραμέτρους της κάθε κατανομής. Επίσης, από τα αποτελέσματα παρατηρούμε ότι, οι καμπύλες των μηνυμάτων έχουν σχεδόν το ίδιο σχήμα και στα δύο δίκτυα. Για τα πρώτα λίγα βήματα, η κατανομή Weibull καταλήγει σε μεγαλύτερο αριθμό συνολικών μηνυμάτων από την κατανομή Normal· αμέσως μετά παρατηρείται το αντίθετο, που μπορεί πιθανώς να εξηγηθεί από το γεγονός ότι η κατανομή Weibull με τις συγκεκριμένες παραμέτρους οδηγεί μεγαλύτερο αριθμό κόμβων σε αποχώρηση.

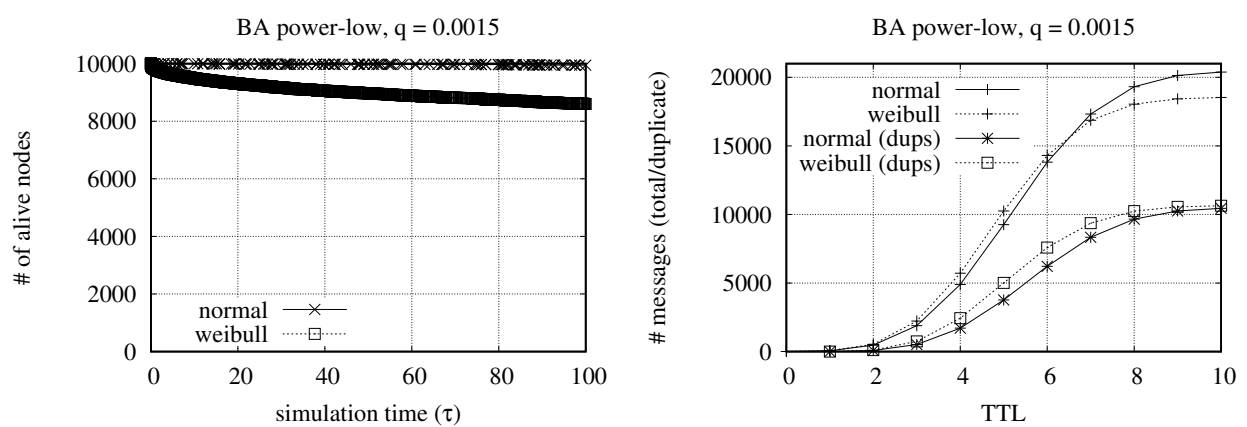
Για αυτή τη συγκεκριμένη στρατηγική, τα αποτελέσματα δείχνουν ότι η διαδικασία churn έχει μόνο μικρή επίδραση σε δίκτυα BA και ER.



Σχήμα 6.7: Armonia: Ένα δείγμα ενεργών κόμβων για τα πρώτα 100sec της προσομοίωσης με το Armonia (αριστερά) και του αριθμού των μηνυμάτων (δεξιά) που παράγει η πλημμύρα σε ένα δυναμικό δίκτυο ER με 5.000 αρχικούς κόμβους.

6.5 Συμπεράσματα

Παρουσιάσαμε το Armonia, ένα μοναδικό και ευέλικτο πλαίσιο προσομοίωσης που μπορεί να χρησιμοποιηθεί για να μοντελοποιήσει τις δυναμικές αλληλεπιδράσεις των κόμβων σε σύνθετα δίκτυα. Παρέχει ένα δομικό στοιχείο για την τοπολογία και για συγκεκριμένες δομές δικτύων, ένα δομικό στοιχείο για τους πόρους και τα αντίγραφα τους για τη μοντελοποίηση της συνολικής κατανομής των πόρων και των αντιγράφων τους, ένα δομικό στοιχείο για ορισμένους αλγορίθμους αναζήτησης που χρησιμοποιείται στον εντοπισμό πόρων και ένα δομικό στοιχείο στατιστικών για τη συλλογή χρήσιμων μετρήσεων. Το Armonia είναι το πρώτο γενικού σκοπού πλαίσιο προσομοίωσης. Στόχος μας είναι να μελετήσουμε τη συμπεριφορά διάφορων στρατηγικών αναζήτησης στα σύνθετα δίκτυα με churn, όπου οι κόμβοι μπορεί να εισέρχονται και να εξέρχονται από το δίκτυο δυναμικά. Εστιάζουμε σε πιθανοτικές στρατηγικές αναζήτησης, αν και μπορούν να υλοποιηθούν αποτελεσματικά και άλλες στρατηγικές αναζήτησης. Προς το παρόν βελτιστοποιούμε το πλαίσιο προσομοίωσης, ενώ επίσης προστίθενται νέες τοπολογίες και στρατηγικές αναζήτησης. Το Armonia θα είναι διαθέσιμο ως λογισμικό ανοικτού κώδικα, και ελπίζουμε ότι θα αποτελέσει ένα χρήσιμο εργαλείο για την ερευνητική κοινότητα.



Σχήμα 6.8: Armonia: Ένα δείγμα ενεργών κόμβων για τα πρώτα 100sec της προσομοίωσης με το Armonia (αριστερά) και του αριθμού των μηνυμάτων (δεξιά) που παράγει η πλημμύρα σε ένα δυναμικό δίκτυο BA με 10.000 αρχικούς κόμβους· στη γραφική παράσταση απεικονίζονται ο συνολικός αριθμός των μηνυμάτων και των διπλότυπων μηνυμάτων.

ΚΕΦΑΛΑΙΟ 7

ΑΝΑΖΗΤΗΣΗ ΣΕ ΔΥΝΑΜΙΚΑ ΔΙΚΤΥΑ ΜΕΓΑΛΗΣ ΚΛΙΜΑΚΑΣ

-
- 7.1 Εισαγωγή
 - 7.2 Επισκόπηση βιβλιογραφίας
 - 7.3 Μεθοδολογία και μοντέλο συστήματος
 - 7.4 Προσαρμοζόμενη Advanced Probabilistic Search (A2PF)
 - 7.5 Αξιολόγηση
 - 7.6 Σύνοψη
-

7.1 Εισαγωγή

Τα σύγχρονα σύνθετα συστήματα, όπως ασύρματα δίκτυα, δίκτυα αισθητήρων, κοινωνικά δίκτυα, δίκτυα p2p κ.ά, είναι από την φύση τους δυναμικά, καθώς εξελίσσονται και μεταβάλλονται συνεχώς και οι κόμβοι τους εισέρχονται/εξέρχονται από το σύστημα ή/και αλληλεπιδρούν με άλλους κόμβους εμφανίζοντας και εξαφανίζοντας συνδέσεις. Η μελέτη της συμπεριφοράς των συστημάτων κατά τη διάρκεια του *churn*, της συνεχόμενης δηλαδή διαδικασίας συνδέσεων/αποσυνδέσεων κόμβων από το δίκτυο [162] αποτελεί αναγκαιότητα αλλά και πρόκληση και κίνητρο για αυτή την εργασία.

Τα τελευταία χρόνια υπήρξε αυξανόμενο ενδιαφέρον και καταβλήθηκε ιδιαίτερη προσπάθεια για να αποκαλυφθεί, να κατανοηθεί και να προβλεφθεί η επίδραση της

δυναμικότητας σε πολλούς τύπους δικτύων: τηλεπικοινωνιακά δίκτυα [42], [170] διαδίκτυο [87], δίκτυα p2p [162], [69], [180], [13], [202], [67], κοινωνικά δίκτυα [53], [14], βιολογικά δίκτυα [64], κ.ά. Στα δυναμικά δίκτυα οι κόμβοι και οι μεταξύ τους συνδέσεις δεν είναι μόνιμα ενεργές με αποτέλεσμα να επηρεάζονται οι εφαρμογές του πραγματικού κόσμου που αναπτύσσονται στα αυτά.

Όπως αναφέρθηκε στην Ενότητα 2.4, πλήθος δυναμικών διεργασιών εκτυλίσσονται στα σύνθετα δίκτυα. Το ερώτημα είναι αν η δυναμικότητα του δικτύου έχει επιπτώσεις στην απόδοση και την εφαρμοσιμότητα των διαδικασιών, όπως η αναζήτηση, και κατά πόσο οι διαδικασίες προσαρμόζουν την δραστηριότητά τους και ανταπεξέρχονται σε ένα περιβάλλον συνεχώς μεταβαλλόμενο.

Σε αυτό το κεφάλαιο, εξετάσουμε το πρόβλημα της αναζήτησης σε δυναμικά δίκτυα, η οποία τώρα βρίσκεται αντιμέτωπη με νέους περιορισμούς και προκλήσεις καθώς πλέον έχει να διαχειριστεί και την προσθαφαίρεση κόμβων και συνδέσεων. Προτείνουμε μια πιθανοτική στρατηγική αναζήτησης βασιζόμενη στην πλημμύρα. Η στρατηγική αυτή είναι μια τροποποιημένη εκδοχή της APF ώστε να είναι δυνατή η εφαρμογή της σε δυναμικά δίκτυα. Έχοντας διαπιστώσει (Ενότητα 5.5.2) ότι ο χώρος αναζήτησης εκτείνεται ως συνάρτηση του μέσου βαθμού των πλησιέστερων γειτόνων, αρχικά διερευνούμε τον μέσο βαθμό των πλησιέστερων γειτόνων ενός κόμβου u , θεωρώντας ότι οι πλησιέστεροι γειτονικοί κόμβοι του u βρίσκονται τοποθετημένοι σε ομόκεντρους δακτυλίους, ανάλογα με την απόστασή τους από τον κόμβο u . Καθώς η αναλυτική εκτίμηση του μέσου βαθμού των πλησιέστερων κόμβων δεν είναι εφικτή, εκμεταλλευόμαστε, όπως εξηγούμε και παρακάτω, την τοπική πληροφορία που μπορεί να συλλέξει ο κάθε κόμβος u του δικτύου. Η πληροφορία αυτή αξιοποιείται από τη στρατηγική APF και οδηγεί σε ιδιαίτερα ικανοποιητικά αποτελέσματα. Στόχος μας είναι να μελετήσουμε τη συμπεριφορά μιας δυναμικής διαδικασίας που εκτυλίσσεται σε ένα δυναμικό δίκτυο, ένα δίκτυο δηλαδή που λαμβάνει υπόψη τη μεταβολή του χρόνου [84]. Συγκεκριμένα η συνεισφορά μας περιλαμβάνει:

- την εισαγωγή μιας νέας διάστασης στο μήνυμα-ερώτηση με την προσθήκη ενός πεδίου που πληροφορεί για τον βαθμό του κόμβου που θέτει το ερώτημα,
- την υλοποίηση ενός τοπικού μηχανισμού που “μαθαίνει” για την τοπολογία του δικτύου από τα μηνύματα που λαμβάνει,
- την τροποποίηση της στρατηγικής APF ώστε να προσαρμόζεται σε δυναμικά

Πίνακας 7.1: Συνοπτικός πίνακας συμβόλων που χρησιμοποιούνται σε αυτή την ενότητα.

Symbol List	
Σύμβολο	Περιγραφή
G_τ	Ο γράφος του δικτύου την χρονική στιγμή τ
V_τ	Το σύνολο των κόμβων που περιλαμβάνει ο γράφος G_τ
E_τ	Το σύνολο των ακμών που περιλαμβάνει ο γράφος G_τ
N	Ο αριθμός των κόμβων του δικτύου
N_t	Ο αριθμός των κόμβων που έλαβαν το μήνυμα-ερώτηση μέχρι το βήμα t
t	Ο αριθμός των βημάτων
τ	Χρόνος
n_t	Ο αριθμός των κόμβων σε απόσταση t από τον u
$\mathcal{N}(u)$	Το σύνολο των γειτονικών κόμβων του u
$N(k)$	Ο αριθμός των κόμβων που υπάρχουν στο σύστημα μετά το πέρασμα k χρονικών διαστημάτων
c	Ο ρυθμός churn
$\bar{d}$	Μέσος βαθμός του δικτύου
p_d	Η κατανομή των βαθμών
$\bar{d}_t$	ο μέσος βαθμός των κόμβων σε απόσταση t
γ	Εκθέτης της κατανομής power-law

ετερογενή δίκτυα,

- τη συγκριτική μελέτη της αναζήτησης σε δυναμικά δίκτυα χρησιμοποιώντας τον προσομοιωτή Armonia.

Το υπόλοιπο αυτού του κεφαλαίου οργανώνεται ως ακολούθως: Η Ενότητα 7.2 συζητά τη σχετική βιβλιογραφία. Στην Ενότητα 7.3 παρουσιάζεται το μοντέλο και η μέθοδος μελέτης της αναζήτησης για δυναμικά δίκτυα. Στη συνέχεια, στην Ενότητα 7.4 παρουσιάζεται ένα προσαρμοσμένο σε δυναμικά δίκτυα πλαίσιο αναζήτησης, η A2PF. Στην Ενότητα 7.5 αναφέρονται τα κύρια ευρήματα και τα αποτελέσματα προσομοιώσεων. Τέλος, στην Ενότητα 7.6 συνοψίζεται το κεφάλαιο.

7.2 Επισκόπηση βιβλιογραφίας

Η ανάλυση, μοντελοποίηση και κατανόηση των αλλαγών σε ένα δυναμικό δίκτυο είναι πολύπλοκη και απαιτητική εργασία, χωρίς να ληφθούν υπόψη οι σύνθετες διαδικασίες που εκτυλίσσονται παράλληλα όπως η αναζήτηση. Αρκετοί ερευνητές εξετάζουν τη συμπεριφορά των χρηστών σε μια προσπάθεια να αποκαλύψουν ιδιότητες του churn (π.χ ρυθμός churn) [162], [180], και τις επιπτώσεις του στο δίκτυο (π.χ. ανθεκτικότητα) [202].

Θεωρητικές προσεγγίσεις προσπαθούν να κατανοήσουν τις δυνατότητες και τους περιορισμούς κατανεμημένων υπολογισμών (απαρίθμηση, aggregation, αθροίσματα, μέσοι όροι, κ.ά.) σε δυναμικά δίκτυα, εφαρμόζοντας τυχαίους περιπατητές ή άλλες πιθανοτικές τεχνικές προώθησης των μηνυμάτων από κόμβο σε κόμβο [106], [168]. Ο Jelasity κ.ά. [90] εργαζόμενοι τόσο σε θεωρητικό όσο και σε πειραματικό πλαίσιο προτείνουν τεχνικές συνάθροισης τιμών βασιζόμενες σε πρωτόκολλα gossip.

Πολύ λίγα ζητήματα έχουν διερευνηθεί εξονυχιστικά στα δυναμικά δίκτυα όπως δρομολόγηση, ευρεία μετάδοση, σχηματισμός κοινότητων σε κοινωνικά δίκτυα, ανθεκτικότητα και αυτό-σταθεροποίηση [34], χωρίς όμως η αναζήτηση να περιλαμβάνεται σε αυτά. Σε αντίθεση με τα στατικά δίκτυα όπου η αναζήτηση απασχόλησε εντατικά την ερευνητική κοινότητα, στα δυναμικά δίκτυα η μελέτη της βρίσκεται ακόμη σε εμβρυικό στάδιο.

Οι Furness και Kolberg [69] εξέτασαν μέσω προσομοίωσης την επίδραση του churn στο ποσοστό επιτυχίας “τυφλών” τεχνικών αναζήτησης σε δομημένα συστήματα p2p. Ο Augustine κ.ά. [13] υποθέτουν περιορισμένο churn και προσαρμοσμένη αποθήκευση για να διατηρηθεί η λειτουργικότητα ενός δικτύου p2p. Στη συνέχεια εφαρμόζουν έναν τυχαιοποιημένο κατανεμημένο αλγόριθμο αναζήτησης που εγγυάται επιτυχία με μεγάλη πιθανότητα. Μια πρόσφατη εργασία [67] εξετάζει το churn χωρίς όρια και θεωρεί ότι μια διαδικασία, όπως η αναζήτηση, μπορεί να εισέρχεται και να εξέρχεται στο δίκτυο. Όταν εγκαταλείπει το δίκτυο είτε αποχωρεί άμεσα (ανταγωνιστικό churn) είτε αποχωρεί αφού πρώτα ολοκληρωθεί (συνεργατικό churn). Η δική μας προσέγγιση μελετά την επίδραση χαμηλού-υψηλού churn στο ποσοστό επιτυχίας και στο πλήθος των μηνυμάτων σε αδόμητα δίκτυα, για στρατηγικές βασιζόμενες στην πλημμύρα.

Σε κάποιες άλλες εργασίες μελετώνται οι τεχνικές της πλημμύρας και ο χρόνος ολοκλήρωσης τους, εφαρμόζοντας ένα πιθανοτικό δικτυακό μοντέλο όπου οι ακμές

προστίθενται ή αφαιρούνται σύμφωνα με μια απλή μαρκοβιανή διαδικασία [41], [42], [23], [24], [40]. Σε αντίθεση με αυτές τις εργασίες, εμείς εξετάζουμε την εφαρμογή τεχνικών βασιζόμενες σε πλημμύρα σε δυναμικά δίκτυα όπου οι κόμβοι ή/και οι ακμές εισέρχονται και εξέρχονται σύμφωνα με κάποια κατανομή.

7.3 Μεθοδολογία και μοντέλο συστήματος

Στην ενότητα αυτή θα αναπτύξουμε το θεωρητικό μοντέλο που αποτυπώνει τη δυναμικότητα του δικτύου ώστε να κατανοήσουμε τη συμπεριφορά του σε συνθήκες churn και να διερευνήσουμε αν και πως επηρεάζονται οι επιδόσεις της αναζήτησης. Για διευκόλυνση, στον Πίνακα 7.1 έχουμε συνοψίσει τα μεγέθη και τους αντίστοιχους συμβολισμούς που θα χρησιμοποιήσουμε.

Θεωρούμε ένα δυναμικό δίκτυο που περιγράφεται από το γράφο $G_\tau = (V_\tau, E_\tau)$, όπου V_τ το σύνολο των οντοτήτων και E_τ , το σύνολο των συσχετίσεων μεταξύ των οντοτήτων τη χρονική στιγμή τ . Ο αριθμός των συνδέσεων των κόμβων περιγράφεται από την κατανομή p_d . Οι οντότητες και οι συνδέσεις μπορεί να εμφανίζονται ή να εξαφανίζονται από το δίκτυο. Μια οντότητα έχει μια συγκεκριμένη διάρκεια ζωής (lifetime) που καθορίζεται με την εμφάνισή της στο σύστημα. Έτσι, μια συγκεκριμένη χρονική στιγμή τ , ένας κόμβος u , είτε είναι ενεργός, είτε έχει αποχωρήσει από το σύστημα. Αντίστοιχα, για κάθε ενεργό κόμβο u , ένας γειτονικός του κόμβος i μπορεί να είναι ή να μην ενεργός σύμφωνα με την διαδικασία $Y_i(\tau)$ [111]:

$$Y_i(\tau) = \begin{cases} 1 & \text{αν } i \text{ ενεργός τη χρονική στιγμή } \tau, \\ 0 & \text{αλλιώς} \end{cases} \quad (7.1)$$

Ο βαθμός του κόμβου u , τη χρονική στιγμή τ , υπολογίζεται ως

$$d_u(\tau) = \sum_{i \in \mathcal{N}(u)} Y_i(\tau)$$

όπου $\mathcal{N}(u)$ το σύνολο των γειτονικών κόμβων του u .

Αρχικά στο σύστημα εμφανίζονται N_0 οντότητες που η διάρκεια ζωής τους μπορεί να καθοριστεί από μια πιθανοτική κατανομή όπως εκθετική [162], Pareto [111], ή Weibull [180]. Για χάρη της απλότητας, εμείς θα υποθέσουμε ότι στη διάρκεια της ζωής της, μια οντότητα είναι μόνιμα ενεργή. Αν αποχωρήσει και επανέλθει την χειρίζομαστε ως μια νέα άφιξη. Δηλαδή δεν μπορεί να είναι στο σύστημα ως μη ενεργή.

Νέες οντότητες μπορεί να εισέρχονται στο σύστημα σύμφωνα με μια διαδικασία Poisson [160] και ρυθμό άφιξης λ .

Για να μελετήσουμε τη συμπεριφορά της πιθανοτικής πλημμύρας σε δυναμικά δίκτυα, αξιοποιούμε το μοντέλο που αναπτύχθηκε στην Ενότητα 4.5 αφού το προσαρμόσουμε ώστε να συμπεριλάβει το χαρακτηριστικό της δυναμικότητας του δικτύου. Το μοντέλο αυτό προσεγγίζει τον αριθμό των κόμβων, έστω n_t , που έλαβαν το μήνυμα-ερώτηση στο βήμα t της αναζήτησης (Εξ. 4.2). Στην Ενότητα 5.5.2 διαπιστώθηκε ότι το μοντέλο δεν μπορεί να εφαρμοστεί άμεσα σε δίκτυα άνευ κλίμακας καθώς δίνεται ως συνάρτηση του μέσου βαθμού του δικτύου. Στην ίδια Ενότητα (5.5.2) διαπιστώνεται επίσης η εφαρμοσιμότητα του μοντέλου, αν σε κάθε βήμα της αναζήτησης ήταν γνωστός ο μέσος βαθμός των κόμβων σε απόσταση t , συμβολιζόμενος ως $\bar{d}_t$. Επομένως, το πλήθος των κόμβων που έχουν λάβει το μήνυμα ερώτηση στο βήμα $t + 1$ υπολογίζεται ως:

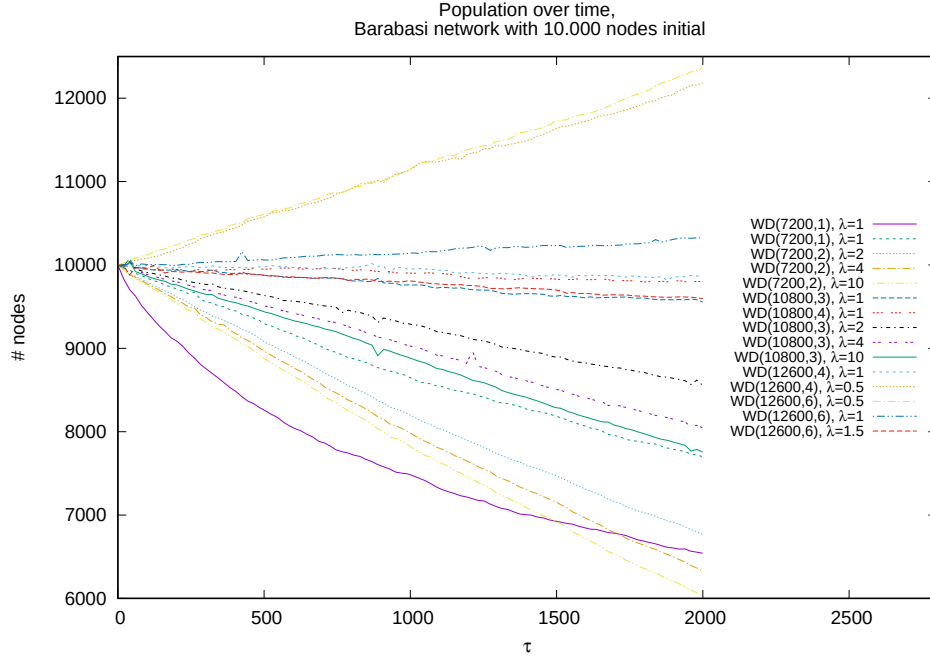
$$n_{t+1} = (\bar{d}_t - 1)n_t \left(1 - \frac{N_t}{N}\right)$$

Στα δυναμικά δίκτυα, το πλήθος των κόμβων μεταβάλλεται, και αυξομειώνεται, ανάλογα με το ρυθμό churn. Έστω $N(\tau)$, ο αριθμός των κόμβων στο δίκτυο, τη χρονική στιγμή τ . Σύμφωνα με τα παραπάνω, ο αριθμός των κόμβων που θα λάβουν την ερώτηση στο βήμα $t + 1$, θα είναι:

$$n_{t+1} = (\bar{d}_t(\tau) - 1)n_t \left(1 - \frac{N_t}{N(\tau)}\right). \quad (7.2)$$

Για να εξετάσουμε πειραματικά την επίδραση του churn στο πλήθος των κόμβων του δικτύου χρησιμοποιούμε τον προσομοιωτή Armonia. Οι συνθήκες churn καθορίζονται από δύο κατανομές: η μια αφορά τη διάρκεια ζωής του κάθε κόμβου, και η άλλη τον ρυθμό άφιξης νέων κόμβων. Στο Σχ. 7.1 παρουσιάζονται οι μεταβολές στον πληθυσμό για ένα συγκεκριμένο δίκτυο τύπου Barabasi με αρχικό πλήθος 10.000 κόμβων. Για το δίκτυο αυτό η διάρκεια ζωής καθορίζεται από μια κατανομή Weibull ($WD(\alpha, b)$) όπου a η παράμετρος κλίμακας που οδηγεί σε διαφορετικές τιμές για τη μέση διάρκεια ζωής του κάθε κόμβου, και b η παράμετρος μορφής της κατανομής. Οι αφίξεις καθορίζονται από μια κατανομή Poisson με ρυθμό αφίξεων λ .

Η εκτίμηση του μεγέθους του δικτύου ή η απαρίθμηση των κόμβων του απασχόλησε ιδιαίτερα την ερευνητική κοινότητα, καθώς η γνώση αυτής της διάστασης μπορεί να φανεί χρήσιμη στην βελτιστοποίηση της απόδοσης των πρωτοκόλλων και



Σχήμα 7.1: Μεταβολή πλήθους κόμβων σε δίκτυο Barabasi με churn, με διάφορους συνδυασμούς διάρκειας ζωής και αφίξεων για αρχικό πλήθος κόμβων 10.000.

των υπηρεσιών που παρέχονται από αυτά [90]. Για μεγάλα δυναμικά δίκτυα, οι προτεινόμενες λύσεις διακρίνονται σε δυο βασικές κατηγορίες. Στην πρώτη κατηγορία ανήκουν οι τεχνικές που βασίζονται σε πρωτόκολλα gossip [90], [186], και η εκτίμηση προκύπτει ως συνάθροιση τιμών από τα μηνύματα που ανταλλάσσουν οι κόμβοι. Στη δεύτερη κατηγορία εφαρμόζονται τυχαιοποιημένες τεχνικές [132], [106]. Μηνύματα ανιχνευτές ή τυχαίοι περιπατητές προωθούνται τυχαία στο δίκτυο από έναν αυθαίρετα τυχαίο κόμβο. Από τις αποκρίσεις που πιθανοτικά θα λάβει ο αρχικός κόμβος ή από τη λήψη του αρχικού μηνύματος θα προχωρήσει στην εκτίμηση του μεγέθους.

Για τη συνέχεια, εμείς, υποθέτοντας ότι στη διάρκεια του χρόνου δεν αλλάζουν σημαντικά τα βασικά δομικά χαρακτηριστικά του δικτύου, θεωρούμε ότι $N_t(\tau) \simeq$ σταθερό N_t . Πεποίθησή μας είναι ότι η Εξίσωση 7.2 προσεγγίζει την κάλυψη ενός δυναμικού δικτύου σε κάθε βήμα μιας διαδικασίας αναζήτησης και στην Ενότητα 7.5 επιχειρείται η πειραματική της επιβεβαίωση.

7.3.1 Μέσος βαθμός κόμβων σε απόσταση t

Ένας κόμβος u και βαθμού d_u γνωρίζει ότι σε απόσταση 1 από αυτόν υπάρχουν ($d_u(\tau) = \sum_{i \in \mathcal{N}(u)} Y_i(\tau)$) κόμβοι. Όπως είναι γνωστό η κατανομή των βαθμών του δικτύου περιγράφει την τοπολογία του δικτύου και πληροφορεί για τα καθολικά χαρακτηριστικά του [59]. Αν και ο κόμβος u μπορεί να γνωρίζει την κατανομή των βαθμών του δικτύου, δεν γνωρίζει τον βαθμό του κάθε γειτονικού του κόμβου ή τον μέσο βαθμό των d_u κόμβων. Αν υποθέσουμε ότι με κάποιο τρόπο αποκτούσε αυτή την πληροφορία, μάθαινε δηλαδή τον μέσο βαθμό των d_u κόμβων, ουσιαστικά θα αποκάλυπτε πως εκτείνεται το δίκτυο ένα βήμα πέρα από τον ορίζοντά του. Γενικότερα το ερώτημα είναι ένας κόμβος u και βαθμού d_u , θα μπορούσε να μάθει τους βαθμούς των κόμβων σε απόσταση¹ $l = 1, 2, \dots, t$ -βήματα και να αποκαλύψει ουσιαστικά την τοπολογία του δικτύου γύρω από αυτόν

Ένας τρόπος για να απαντηθεί το ερώτημα αυτό και να αποκτήσει εικόνα, ο κόμβος u , για τη δόμηση του δικτύου που εκτείνεται πίσω από τους γειτονικούς του κόμβους, είναι να υπολογίσει το μέσο βαθμό των κόμβων σε απόσταση l . Σε απόσταση ενός βήματος, η ποσότητα αυτή αποτυπώνεται από τον βαθμό του κόμβου. Η πληροφορία αυτή μπορεί να εξαχθεί και από την πρώτη παράγωγο της γεννήτριας συνάρτησης της κατανομής των βαθμών, ως:

$$d^{(1)} = \sum_{d=1}^{\infty} dp_d$$

και είναι ο μέσος βαθμός του δικτύου $\bar{d}$. Σε ομογενή δίκτυα, τακτικά και τυχαία δίκτυα (ER), οι κόμβοι έχουν τον ίδιο περίπου βαθμό. Αντίθετα, σε δίκτυα άνευ κλίμακας, μια τέτοια προσέγγιση είναι τελείως ανακριβής. Ο μέσος βαθμός των κόμβων σε απόσταση 1 από έναν κόμβο u υπολογίζεται ως:

$$\bar{d}_1^u = \frac{1}{d_u} \sum_j d_j$$

Με την βοήθεια της υπό συνθήκη πιθανότητας, $p_1(d'|d)$, η οποία ορίζεται ως η πιθανότητα ένας κόμβος με d συνδέσεις να είναι γειτονικός με έναν άλλον κόμβο με d' συνδέσεις [28], μπορεί να υπολογιστεί ο μέσος βαθμός των κόμβων που βρίσκονται σε απόσταση 1 από όλους τους κόμβους βαθμού d ως:

$$\bar{d}_1 = \sum_{d'} d' p_1(d'|d)$$

¹απόσταση είναι το μήκος της συντομότερης διαδρομής, όπως ορίζεται στην Ενότητα 2.1

Ο Mayo κ.ά. [135] χρησιμοποιώντας παρόμοια επιχειρηματολογία της υπό συνθήκη πιθανότητας $p_l(d'|d)$, προσδιορίζουν τον μέσο βαθμό των πλησιέστερων κόμβων σε απόσταση l από έναν κόμβο βαθμού d ως:

$$\bar{d}_l = \sum_{d'} d' p_l(d'|d)$$

Τυχαία δίκτυα

Όπως αναφέρθηκε στην Ενότητα 2.3.1, τα τυχαία δίκτυα τύπου Erdős - Rényi, έχουν κατανομή βαθμών Poisson (p_d). Η p_d εκφράζει την πιθανότητα, ένας τυχαία επιλεγμένος κόμβος u του δικτύου να έχει βαθμό d . Όμως εμείς ενδιαφερόμαστε όχι για τον βαθμό ενός τυχαίου κόμβου u , αλλά για τον υπολειπόμενο βαθμό (excess degree) ενός κόμβου u , στον οποίο καταλήγουμε ακολουθώντας μια τυχαία ακμή, δηλαδή για το πλήθος των υπόλοιπων συνδέσεων του κόμβου, εκτός από αυτή που μας οδήγησε στον u . Η κανονικοποιημένη κατανομή του υπολειπόμενου βαθμού δίνεται από [146]:

$$q_d = \frac{(d+1)p_{d+1}}{\sum_j j p_j}. \quad (7.3)$$

Τότε, ο μέσος όρος όλων των υπολειπόμενων βαθμών είναι:

$$\sum_{d=0}^{\infty} q_d = \frac{\sum_{d=0}^{\infty} d(d+1)p_{d+1}}{\sum_j j p_j} = \frac{\sum_{d=0}^{\infty} (d-1)d p_d}{\sum_j j p_j} = \frac{d^{(2)} - d^{(1)}}{d^{(1)}} = \frac{d^{(2)} - \bar{d}}{\bar{d}}. \quad (7.4)$$

και αποτελεί τον μέσο βαθμό των κόμβων που βρίσκονται σε απόσταση 1 από τον κόμβο u . Χρησιμοποιώντας πολυώνυμα Touchard, μπορεί να υπολογιστεί ο όρος $d^{(2)}$ που είναι ουσιαστικά η 2η παράγωγος της γεννήτριας συνάρτησης της κατανομής Poisson. Η l -οστή παράγωγος υπολογίζεται ως $\sum_{i=1}^l d^i \{d_i\}$, όπου $\{d_i\}$ υποδηλώνει τους αριθμούς Stirling δευτέρου είδους. Οπότε για $l = 1$, η Εξ. 7.4 δίνει το μέσο βαθμό κόμβων σε απόσταση 1 που είναι:

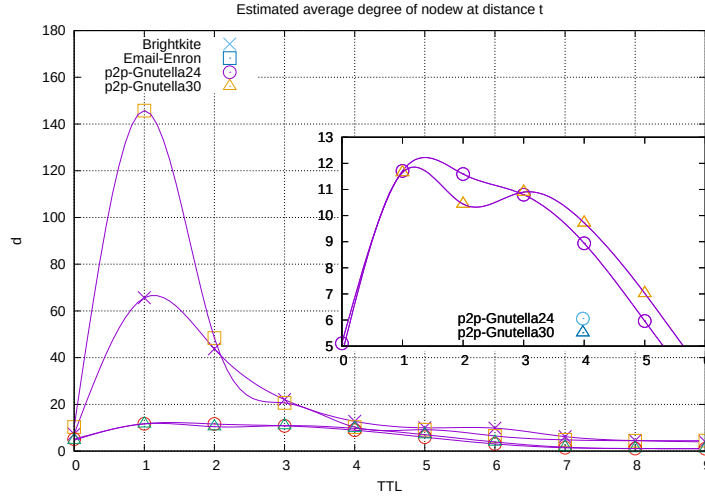
$$\bar{d}_2 = \frac{\bar{d}^2 + \bar{d} - \bar{d}}{\bar{d}} = \bar{d}$$

Για τον υπολογισμό του μέσου βαθμού των κόμβων σε απόσταση μεγαλύτερη του 1, εργαζόμαστε με παρόμοιο τρόπο. Ο μέσος αριθμός κόμβων σε απόσταση l ισούται με [146]:

$$n_l = \frac{d^{(2)} - \bar{d}}{\bar{d}} n_{l-1}$$

οπότε ο μέσος βαθμός κόμβων σε απόσταση $l - 1$ προσεγγίζεται από:

$$\bar{d}_{l-1} = \frac{n_l}{n_{l-1}} = \frac{d^{(2)} - \bar{d}}{\bar{d}} = \bar{d} \quad (7.5)$$



Σχήμα 7.2: Ο μέσος βαθμός των κόμβων σε απόσταση ακριβώς t βημάτων, όπως υπολογίζεται σε συλλογές δεδομένων από πραγματικά δίκτυα. Αντίστοιχες μετρήσεις για συνθετικά δίκτυα βρίσκονται στο Σχ. 5.7.

Δίκτυα scale-free

Στα δίκτυα άνευ κλίμακας η κατανομή βαθμών είναι $p_x = Cx^{-\gamma}$ και βρίσκει εφαρμογή για τιμές μεγαλύτερες μιας ελάχιστης τιμής (x_{min}). Σε αυτή την περίπτωση, ο παράγοντας κανονικοποίησης C , ισούται με $C = (\gamma - 1)x_{min}^{\gamma-1}$. Οπότε η p_x μπορεί να εκφραστεί σε συνάρτηση της x_{min} ως:

$$p_x = \frac{\gamma - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\gamma}$$

Ο μέσος βαθμός του δικτύου υπολογίζεται από την 1η παράγωγο της p_x ως:

$$\bar{d} = \frac{\gamma - 1}{\gamma - 2} x_{min}$$

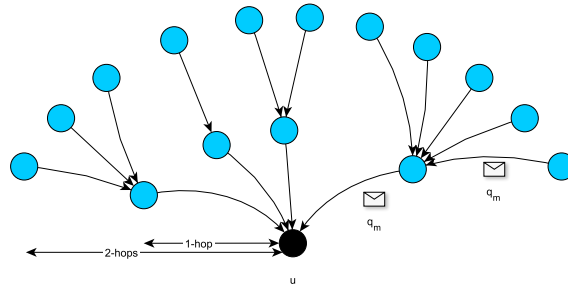
που υπάρχει και είναι καλώς ορισμένη μόνο για $\gamma > 2$. Για την 2η παράγωγο διακρίνουμε δύο περιπτώσεις:

α) $2 < \gamma \leq 3$: η παράγωγος αποκλίνει. Αυτό σημαίνει ότι στην περιοχή $2 \leq \gamma \leq 3$ υπάρχει μεγάλη διακύμανση των βαθμών που αποκλίνουν από τη μέση τιμή. Για παράδειγμα, ο βαθμός ενός τυχαία επιλεγμένου κόμβου μπορεί να είναι μικρός ή αυθαίρετα μεγάλος.

β) $\gamma > 3$: η παράγωγος υπάρχει και είναι καλώς ορισμένη ως:

$$x^{(2)} = \frac{\gamma - 1}{\gamma - 3} x_{min}$$

Η παραπάνω ανάλυση δεν μπορεί να χρησιμοποιηθεί ευρύτερα για τον υπολογισμό του μέσου βαθμού σε απόσταση t για δυο λόγους: ο πρώτος είναι γιατί



Σχήμα 7.3: Η δομή του δικτύου όπως την αντιλαμβάνεται ο κόμβος u από την πληροφορία που συλλέγει από τα εισερχόμενα μηνύματα-ερωτήσεις q_m .

επιτρέπει τον υπολογισμό του μέσου βαθμού των κόμβων μόνο σε απόσταση 1, και ο δεύτερος ότι δεν μπορεί να εφαρμοσθεί σε δίκτυα με $\gamma = 3$ όπως το δίκτυο που παράγει το μοντέλο των Barabási-Albert.

Χρησιμοποιώντας σύνολα δεδομένων από πραγματικά δίκτυα ελεύθερης κλίμακας [116] και με τη βοήθεια του πλαισίου προσομοίωσης Armonia, υπολογίζουμε πειραματικά τον μέσο βαθμό σε απόσταση t από κάθε κόμβο. Στο Σχ. 7.2 απεικονίζεται ο μέσος βαθμός για τέσσερα πραγματικά δίκτυα: το δίκτυο επικοινωνίας email-Enron που περιλαμβάνει 36.692 κόμβους, το κοινωνικό δίκτυο Brightkite με 58.228 κόμβους και δύο στιγμιότυπα από το δίκτυο Gnutella p2p που ελήφθησαν το 2002, το p2p-Gnutella24 με 26.518 κόμβους και το p2p-Gnutella230 με 36.682 κόμβους. Τα αποτελέσματα καταδεικνύουν ότι ο μέσος βαθμός των πλησιέστερων γειτονικών κόμβων σε απόσταση 1 είναι αρκετά μεγαλύτερος από το μέσο βαθμό του δικτύου. Στη συνέχεια, και καθώς η απόσταση αυξάνεται, ο μέσος βαθμός μειώνεται. Παρόμοια αποτελέσματα διαπιστώνονται από τον Wang κ.ά. [194] σε δίκτυα ISP.

7.4 Προσαρμοζόμενη Advanced Probabilistic Search (A2PF)

Στην ενότητα αυτή παρουσιάζουμε μια μέθοδο αναζήτησης που προσαρμόζεται αποτελεσματικά στις αλλαγές του δικτύου και αποτελεί μια τροποποιημένη εκδοχή της APF που παρουσιάστηκε στο Κεφ. 5. Στη στρατηγική APF ο κάθε κόμβος αποφασίζει την προώθηση ενός μηνύματος-ερώτησης σε συνάρτηση με τη δημοτικότητα του αναζητούμενου πόρου και του εκτιμώμενου ποσοστού κάλυψης του δικτύου. Σε ένα πραγματικό δίκτυο, ο κάθε κόμβος δεν γνωρίζει την πλήρη δομή του

δικτύου, παρά μόνο για τους κόμβους με τους οποίους έχει απευθείας συνδέσεις. Μπορεί όμως να επωφεληθεί και να συλλέξει γνώση από τα μηνύματα-ερωτήσεις, που ανταλλάσσουν οι κόμβοι μεταξύ τους, και διέρχονται από αυτόν για να ανακαλύψει “κρυφές” ιδιότητες του δικτύου, όπως ο μέσος αριθμός των συνδέσεων σε ορισμένη απόσταση από αυτόν. Υποθέτουμε ότι η γνώση αυτή συμπεριλαμβάνεται ή μπορεί να εξαχθεί από τα μηνύματα αναζήτησης που υποβάλλονται στο δίκτυο. Συγκεκριμένα, στο πρόβλημα που μας απασχολεί, το οποίο είναι η εκτίμηση του μέσου βαθμού των κόμβων που βρίσκονται σε συγκεκριμένη απόσταση από έναν κόμβο u , θα χρησιμοποιήσουμε την εξής τεχνική: σε κάθε μήνυμα θα επισυνάπτεται πληροφορία για το πλήθος των συνδέσεων του κόμβου που εισάγει το ερώτημα.

Για χάρη της απλότητας, εστιάζουμε σε έναν κόμβο u , που μετέχει στο δίκτυο για κάποιο χρονικό διάστημα (έχει μια διάρκεια ζωής), και δέχεται/υποβάλλει ερωτήσεις αναζήτησης στο δίκτυο. Καθώς η αναζήτηση βασίζεται σε τεχνικές πλημμύρας και θεωρώντας ότι εφαρμόζεται μηχανισμός DDM, για τις μεταδόσεις χρησιμοποιούνται ελάχιστα μονοπάτια, ο κόμβος u που λαμβάνει το ερώτημα θα γνωρίζει και σε ποια απόσταση βρίσκεται ο κόμβος που το έστειλε. Ο κόμβος u αποτελεί το κέντρο ομόκεντρων κύκλων ακτίνας $l = 1, 2, \dots$, πάνω στους οποίους τοποθετούνται οι κόμβοι που βρίσκονται στην αντίστοιχη απόσταση l από τον κόμβο u , όπως φαίνεται και στο Σχ. 7.3. Από τις εισερχόμενες ερωτήσεις, ο u γνωρίζει τον βαθμό των άλλων κόμβων (και την απόσταση) και συναθροίζοντας, κάνει μία εκτίμηση για τον μέσο βαθμό ανάλογα με την απόσταση από αυτόν. Έτσι, ο κόμβος u εκτιμά τη δόμηση του δικτύου γύρω από αυτόν. Η γνώση αυτή τον διευκολύνει στην απόφαση της προώθησης ή όχι μιας ερώτησης. Η διαδικασία αυτή, εφαρμόζεται από όλους τους κόμβους και περιγράφεται συνοπτικά παρακάτω:

1. Όταν ένας κόμβος v θέτει ένα ερώτημα (q_m), συμπεριλαμβάνει σε αυτό το βαθμό του d_v .
2. Κάθε κόμβος u που λαμβάνει για πρώτη φορά το μήνυμα q_m στο βήμα t της διαδικασίας αναζήτησης, συλλέγει από αυτό την πληροφορία του βαθμού (d_v), για να υπολογίσει το μέσο βαθμό των κόμβων σε απόσταση t .

Υπολογισμός μέσου βαθμού. Για να υπολογίσουμε τον μέσο βαθμό των κόμβων σε απόσταση t από έναν κόμβο u αξιοποιούμε τις παρατηρήσεις που συλλέγει ο κόμβος u από τις εισερχόμενες ερωτήσεις. Έτσι μπορούμε να εκτιμήσουμε το μέσο

βαθμό για μια συγκεκριμένη απόσταση ως τον αριθμητικό μέσο όρο των k πιο πρόσφατων παρατηρήσεων (βαθμών). Αυτό πρακτικά σημαίνει ότι δίνουμε βαρύτητα στις k πιο πρόσφατες τιμές των παρατηρήσεων. Όμως, λόγω της δυναμικότητας του δικτύου, οι πραγματικοί βαθμοί των κόμβων μεταβάλλονται και οι τιμές τους διαφοροποιούνται από αυτές που συμπεριλήφθηκαν και στάλθηκαν με το μήνυμα-ερώτηση. Η παραπάνω τεχνική θεωρεί ότι όλες οι k παρατηρήσεις έχουν την ίδια βαρύτητα, χωρίς αυτό να ισχύει πάντα καθώς, για παράδειγμα, ένας κόμβος (που η τιμή του βαθμού του συμπεριλαμβάνεται στις k τιμές) μπορεί να αποχωρήσει. Σε αυτή την περίπτωση, πιθανόν οι πιο πρόσφατες παρατηρήσεις να είναι και πιο ακριβείς. Για τον λόγο αυτό χρησιμοποιούμε έναν συντελεστή w ως μέτρο βαρύτητας για τις νέες τιμές. Έτσι, όταν ληφθεί μήνυμα με νέο βαθμό, η νέα εκτίμηση διαμορφώνεται ως γραμμικός συνδυασμός της τρέχουσας εκτίμησης και του νέου βαθμού, ως εξής:

$$\hat{d}_t(new) = (1 - w) \times \hat{d}_t(current) + w \times value$$

όπου $value$ είναι η ληφθείσα τιμή βαθμού ενός κόμβου σε απόσταση t και $0 \leq w \leq 1$. Όσο μικρότερη είναι η τιμή του w τόσο μεγαλύτερη βαρύτητα δίνουμε στην τελευταία εκτίμηση. Στην αρχή της διαδικασίας υπολογισμού, όταν ο κόμβος u λάβει το πρώτο μήνυμα-ερώτηση με το βαθμό ενός κόμβου σε απόσταση t ($value$) θεωρούμε ότι $\hat{d}_t(current) = value$.

Για να εξετάσουμε την αποτελεσματικότητα του μοντέλου, χρησιμοποιούμε τον προσομοιωτή Armonia που παρουσιάστηκε στην Ενότητα 6.3 αφού κάνουμε τις απαραίτητες προσθήκες στα μηνύματα-ερωτήσεις και στα τοπικά στατιστικά. Στη συνέχεια διεξάγουμε πλήθος πειραματικών μετρήσεων σε τυχαία δίκτυα και δίκτυα άνευ κλίμακας.

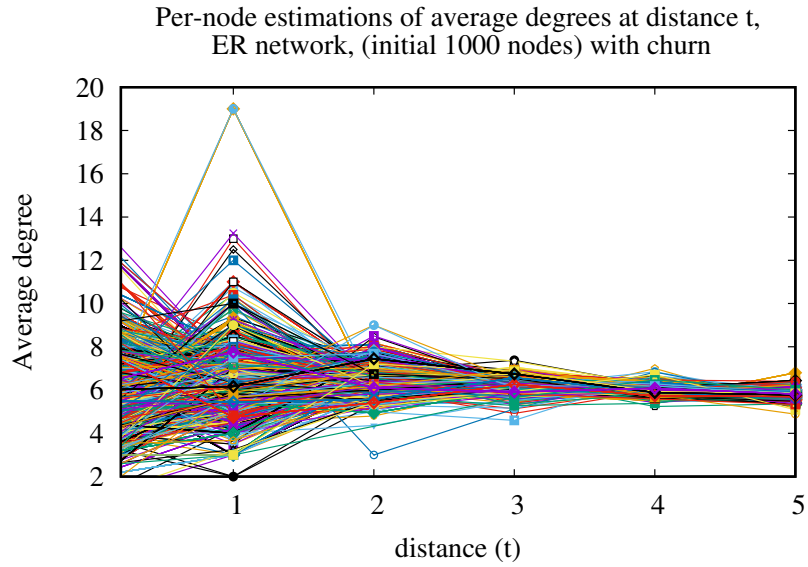
Αρχικά, εξετάζουμε ένα δυναμικό δίκτυο ER με 1.000 κόμβους και μέσο βαθμό 5. Η διάρκεια ζωής του κάθε κόμβου καθορίζεται τη στιγμή της δημιουργίας του από την κατανομή Weibull με παραμέτρους $a = 7200sec$ και $b = 4$. Νέοι κόμβοι εισέρχονται στο δίκτυο ακολουθώντας κατανομή Poisson, με $\lambda = 2$. Οι κόμβοι παρέχουν πόρους χωρίς όμως να διατηρούν αντίγραφα τους. Στο δίκτυο υποβάλλονται συνολικά 1.000 ερωτήσεις εφαρμόζοντας τον αλγόριθμο A2PF, ενώ ο αναμενόμενος χρόνος ολοκλήρωσης της προσομοίωσης ορίζεται σε 0.5 ώρες. Στο Σχ. 7.4 παρουσιάζεται ο μέσος βαθμός όπως παρατηρείται από κάθε κόμβο. Να επισημάνουμε ότι για λόγους ευκρίνειας, στο Σχ. 7.4 απεικονίζονται οι εκτιμήσεις ενός υποσυνόλου

των κόμβων και όχι όλων των κόμβων. Όπως παρατηρούμε, ο μέσος βαθμός των κόμβων σε απόσταση t από έναν κόμβο u είναι περίπου ίσος με το μέσο βαθμό του δικτύου. Μάλιστα, όσο η απόσταση αυξάνεται τόσο ο μέσος βαθμός των κόμβων σε αυτή την απόσταση συγκλίνει στο μέσο βαθμό του δικτύου όπως αναμέναμε από τα αναλυτικά αποτελέσματα σύμφωνα με την (7.5). Ο αριθμητικός μέσος όρος των αποτελεσμάτων που απεικονίζονται στο Σχ. 7.4 παρουσιάζεται στο Σχ. 7.5 μαζί με τον μέσο βαθμό κόμβων σε συνάρτηση με την απόσταση t όλων των κόμβων του δικτύου, όπως υπολογίζεται πειραματικά για το ίδιο δίκτυο και τις ίδιες ρυθμίσεις. Όπως αποδεικνύεται από τα αποτελέσματα η διαφορά μεταξύ των δύο καμπυλών είναι πάρα πολύ μικρή.

Στη συνέχεια, χρησιμοποιώντας το μοντέλο των Barabási-Albert [19] παράγουμε δίκτυα άνευ κλίμακας και εξετάζουμε τον παρατηρούμενο μέσο βαθμό των κόμβων σε σχέση με την απόσταση. Θεωρούμε δύο σενάρια για δίκτυο BA των 1.000 κόμβων: στο πρώτο, ο αριθμός των κόμβων είναι σταθερός ενώ στο δεύτερο ο αριθμός των κόμβων μεταβάλλεται λόγω churn. Διατηρούμε τις υπόλοιπες ρυθμίσεις όπως ορίστηκαν παραπάνω για το δίκτυο ER. Συγκεκριμένα, για το δυναμικό δίκτυο οι παράμετροι της κατανομής Weibull είναι $a = 7200_{sec}$ και $b = 4$. Ο ρυθμός άφιξης νέων κόμβων είναι $\lambda = 2$. Στο δίκτυο υποβάλλονται συνολικά 2000 ερωτήσεις για την εύρεση πόρων εφαρμόζοντας τον αλγόριθμο A2PF, ενώ ο αναμενόμενος χρόνος ολοκλήρωσης της προσομοίωσης ορίζεται σε 0.5 ώρες.

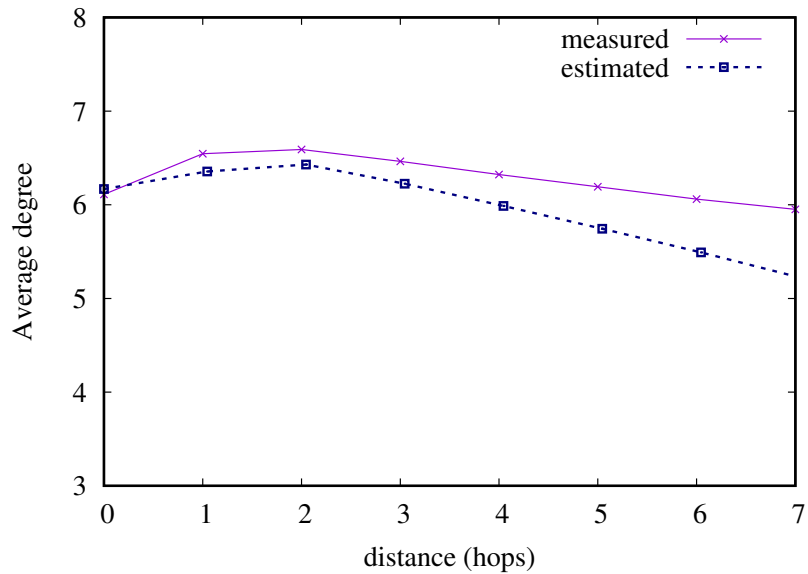
Στα Σχ. 7.6 και 7.8 απεικονίζεται ο μέσος βαθμός κόμβων σε απόσταση t από έναν κόμβο του δικτύου όπως παρατηρείται από κάθε κόμβο του δικτύου. Όπως διαπιστώνουμε οι μορφές των καμπυλών είναι παρόμοιες με αυτές στο Σχ. 7.2. Για την πλειονότητα των κόμβων, ο μέσος βαθμός σε απόσταση 1 είναι αρκετά μεγαλύτερος από τον μέσο βαθμό του δικτύου. Στη συνέχεια αυτή η ποσότητα μειώνεται σχετικά απότομα και συγκλίνει προς μια ελάχιστη τιμή. Επισημαίνεται ότι για λόγους ευκρίνειας στα Σχ. 7.6–7.8 παρουσιάζονται οι παρατηρήσεις από ένα τυχαίο υποσύνολο των κόμβων του δικτύου, ικανό όμως να αναδείξει την παρατηρούμενη συμπεριφορά.

Όπως στην περίπτωση του δικτύου ER, έτσι και εδώ συγκρίνουμε τον αριθμητικό μέσο των εκτιμήσεων του κάθε κόμβου με τον μέσο βαθμό σε σχέση με την απόσταση που υπολογίζουμε πειραματικά, για όλους τους κόμβους δικτύου. Τα αποτελέσματα απεικονίζονται στα Σχ. 7.7 και 7.9. Και εδώ, και για τα δύο δίκτυα, διαπιστώνουμε την εγγύτητα των αποτελεσμάτων που επιβεβαιώνουν την αποτελεσματικότητα της

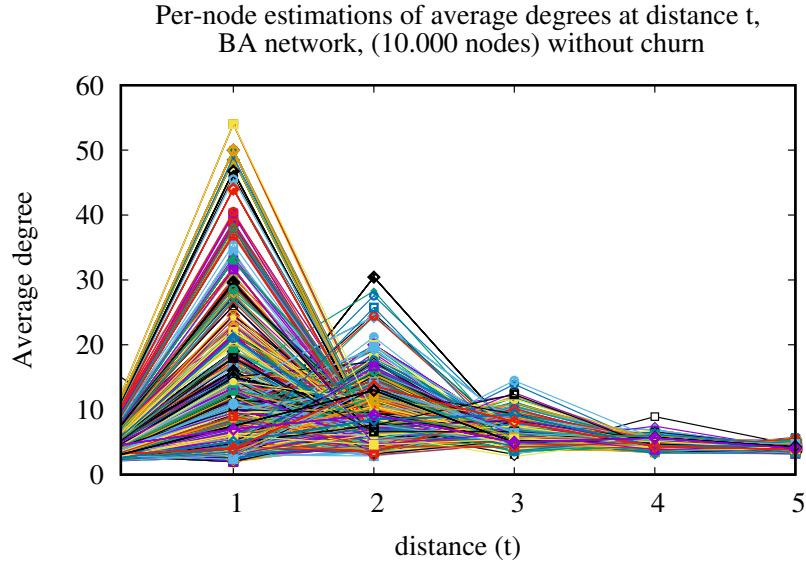


Σχήμα 7.4: Ο μέσος βαθμός των κόμβων σε απόσταση t , όπως υπολογίζεται από διαφορετικούς κόμβους ενός δυναμικού δικτύου ER με αρχικό πλήθος 1.000 κόμβων.

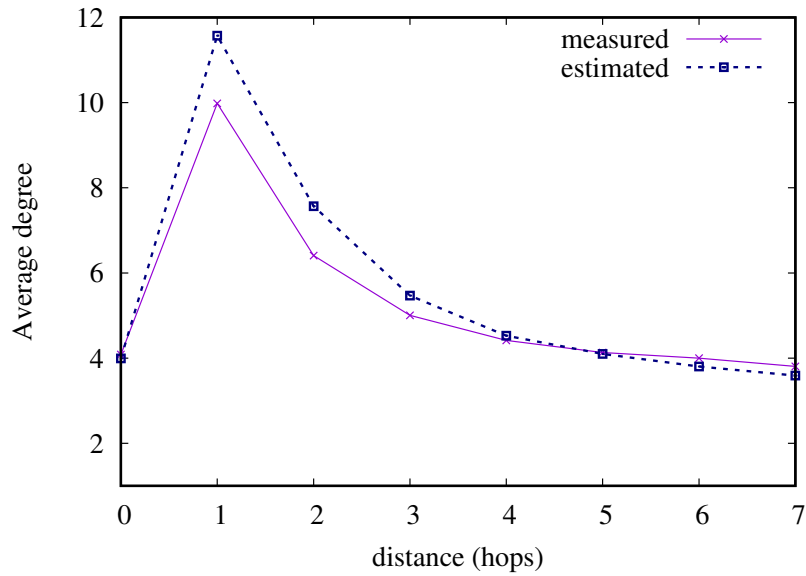
τεχνικής μας.



Σχήμα 7.5: Αριθμητικός μέσος όρος των εκτιμήσεων των κόμβων από το Σχ. 7.4 (καμπύλη estimated) και πραγματικός μέσος βαθμός όπως μετρήθηκε για το ίδιο δυναμικό δίκτυο ER (καμπύλη measured).



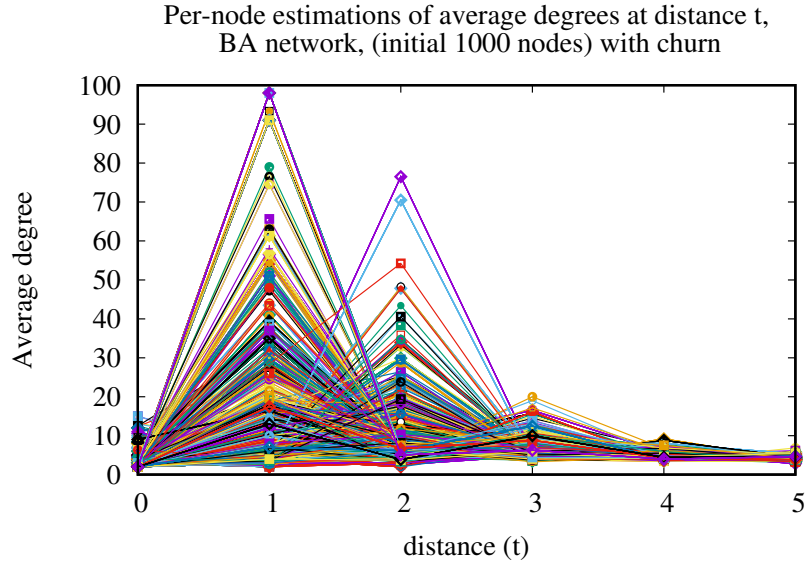
Σχήμα 7.6: Ο μέσος βαθμός των κόμβων σε απόσταση t σε δίκτυο Barabasi 10.000 κόμβων χωρίς churn.



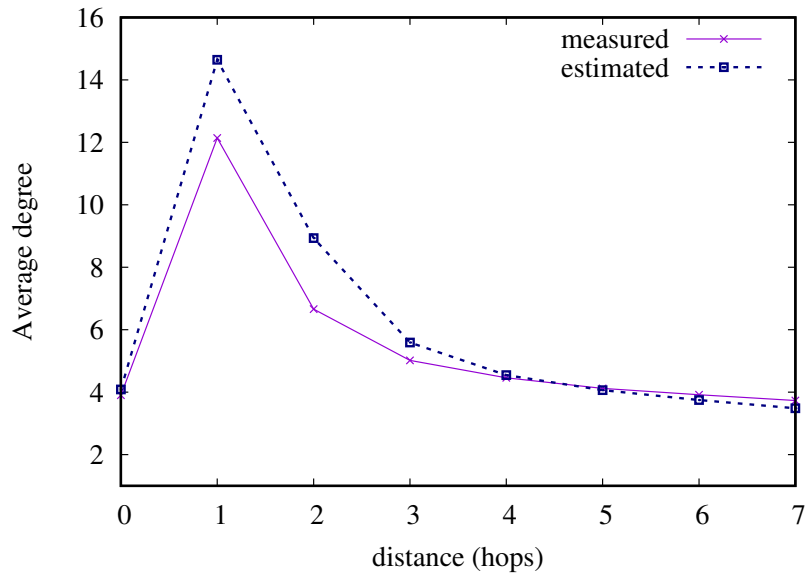
Σχήμα 7.7: Αριθμητικός μέσος όρος των εκτιμήσεων των κόμβων από το Σχ. 7.6 (καμπύλη estimated) και πραγματικός μέσος βαθμός όπως μετρήθηκε για το ίδιο δυναμικό δίκτυο BA (καμπύλη measured).

7.5 Αξιολόγηση

Μια πρώτη συζήτηση για την αναζήτηση σε δυναμικά δίκτυα έγινε στην Ενότητα 6.4 στα πλαίσια αξιολόγησης του προσομοιωτή Armonia. Εδώ, χρησιμοποιώντας το ίδιο πλαίσιο εργασίας, αναλύεται η επίδραση του churn στην αναζήτηση και παράλληλα



Σχήμα 7.8: Ο μέσος βαθμός των κόμβων σε απόσταση t σε δίκτυο Barabasi με αρχικό πλήθος 1.000 κόμβων.



Σχήμα 7.9: Αριθμητικός μέσος όρος των εκτιμήσεων των κόμβων από το Σχ. 7.8 (καμπύλη estimated) και πραγματικός μέσος βαθμός όπως μετρήθηκε για το ίδιο δυναμικό δίκτυο BA (καμπύλη measured).

εξετάζουμε την εφαρμογή της προτεινόμενης μεθόδου σε διάφορα δίκτυα εστιάζοντας κυρίως στα άνευ κλίμακας δίκτυα. Συγκεκριμένα εξετάζουμε τη συμπεριφορά τεσσάρων στρατηγικών αναζήτησης σε συνθήκες υψηλού και χαμηλού churn, των: απλή πλημμύρα (flood), modified BFS (mBFS), Adaptive Resource-based Probabilis-

tic Search (ARPS) και adaptive APF (A2PF). Επιπλέον αξιολογούμε την απόδοσή της A2PF συγκριτικά με τις υπόλοιπες τεχνικές ως προς την πιθανότητα επιτυχίας και τα διπλότυπα μηνύματα.

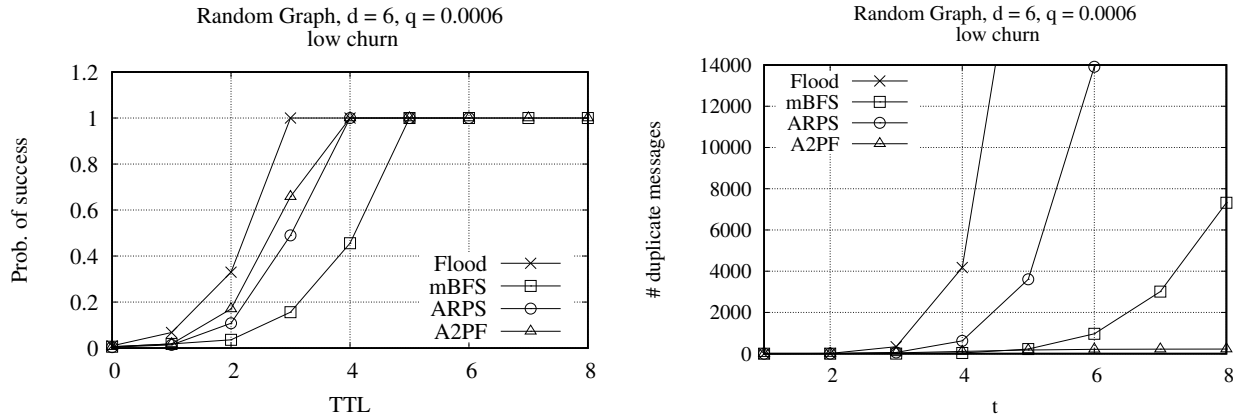
Χρησιμοποιώντας το πλαίσιο προσομοίωσης Armonia, διεξάγουμε πλήθος διαφορετικών σεναρίων σε τυχαία δυναμικά δίκτυα. Για να προσομοιώσουμε την διαδικασία churn, θεωρούμε ότι η διάρκεια ζωής των κόμβων δίδεται από την κατανομή Weibull, όπως προτείνουν οι Stutzbach κ.ά. [180]. Επίσης θεωρούμε ότι οι κόμβοι εισέρχονται στο δίκτυο ακολουθώντας κατανομή Poisson. Οι πόροι και τα αντίγραφά τους κατανέμονται ομοιόμορφα τυχαία στους κόμβους του δικτύου. Το δίκτυο αρχικά περιλαμβάνει έναν συγκεκριμένο αριθμό κόμβων, ο οποίος μεταβάλλεται στη διάρκεια της εκτέλεσης.

Δυναμικά Τυχαία Δίκτυα. Σε αυτή την περίπτωση προσομοίωσης εξετάζουμε δύο διαφορετικά σενάρια churn: ένα με υψηλό και ένα με χαμηλό ρυθμό churn. Για να πετύχουμε διαφορετικό ρυθμό churn μεταβάλλουμε τις παραμέτρους a και b της κατανομής Weibull. Ο συνολικός ρυθμός churn διαμορφώνεται στο 45% και στο 14,5% για υψηλό και χαμηλό ρυθμό churn αντίστοιχα.

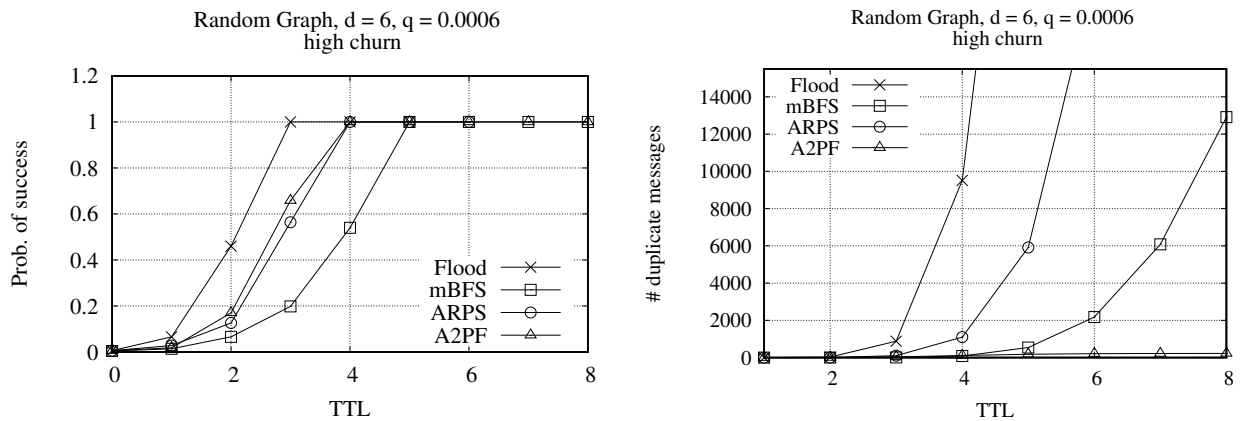
Οι υπόλοιπες ρυθμίσεις του δικτύου διατηρούνται ίδιες και για τα δύο σενάρια. Αρχικά εμφανίζονται 30.000 κόμβοι στο δίκτυο, στους οποίους κατανέμονται ομοιόμορφα τυχαία οι πόροι και τα αντίγραφά τους. Για κάθε πόρο διατίθενται $r = 20$ αντίγραφα. Κάθε κόμβος μπορεί να θέσει ερωτήσεις αναζήτησης στο δίκτυο, που προωθούνται ανάλογα με την στρατηγική αναζήτησης που εφαρμόζεται κάθε φορά. Τα συγκριτικά αποτελέσματα αναζήτησης όλων των στρατηγικών που αφορούν την πιθανότητα επιτυχίας και τα διπλότυπα μηνύματα απεικονίζονται στα Σχ. 7.10 και 7.11. Συγκριτικά, η συμπεριφορά των στρατηγικών παραμένει ίδια με αυτή στα στατικά δίκτυα, με την A2PF να διακρίνεται χάρη στο χαμηλό αριθμό διπλότυπων μηνυμάτων.

Εξετάζοντας τώρα τη συμπεριφορά της κάθε στρατηγικής ξεχωριστά, όπως διαπιστώνουμε από τα απεικονιζόμενα αποτελέσματα στο Σχ. 7.12, η πιθανότητα επιτυχίας παραμένει σχεδόν αμετάβλητη. Όμως παρατηρούμε μια μικρή αύξηση στα διπλότυπα μηνύματα όταν στο δίκτυο εμφανίζεται υψηλό churn. Από τα αποτελέσματα φαίνεται ότι αυτό ισχύει σε όλες τις στρατηγικές εκτός της A2PF. Αυτό πιθανότητα οφείλεται στο ότι αυξάνεται ο αριθμός των ακμών στο δίκτυο καθώς εισέρχονται νέοι κόμβοι.

Δίκτυα άνευ κλίμακας. Για να εξετάσουμε την απόδοση της A2PF αρχικά, θεω-



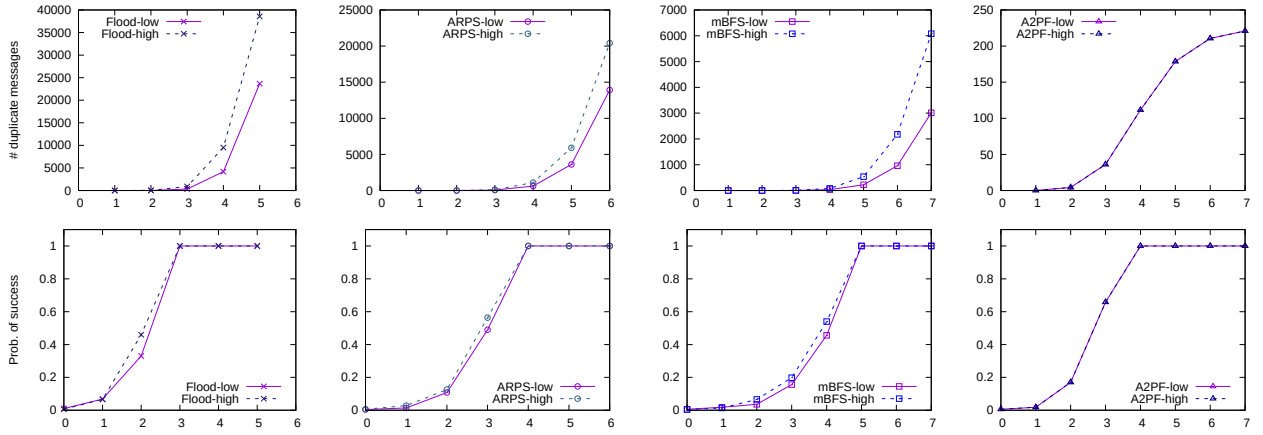
Σχήμα 7.10: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο ER 30.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0006$ και με ρυθμό churn 14,5%.



Σχήμα 7.11: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο ER 30.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0006$ και με ρυθμό churn 45%.

ρούμε ένα δίκτυο άνευ κλίμακας σύμφωνα με το μοντέλο των Barabási-Albert [19] και σταθερό πλήθος κόμβων ίσο με 30.000. Για κάθε διαθέσιμο πόρο, δημιουργούνται $r = 50$ αντίγραφα που κατανέμονται ομοιόμορφα τυχαία στο δίκτυο. Στο δίκτυο αυτό εφαρμόζουμε την στρατηγική μας καθώς και τις άλλες τρεις στρατηγικές αναζήτησης και παρατηρούμε τη συμπεριφορά τους ως προς την πιθανότητα επιτυχίας και τα διπλότυπα μηνύματα.

Τα αποτελέσματα, που απεικονίζονται στο Σχ. 7.13 είναι ικανοποιητικά για την στρατηγική μας. Όπως αναμέναμε διατηρεί την υψηλή πιθανότητα επιτυχίας, ενώ παράλληλα εξαλείφει σχεδόν τα διπλότυπα μηνύματα. Αυτό επιτυγχάνεται με τις

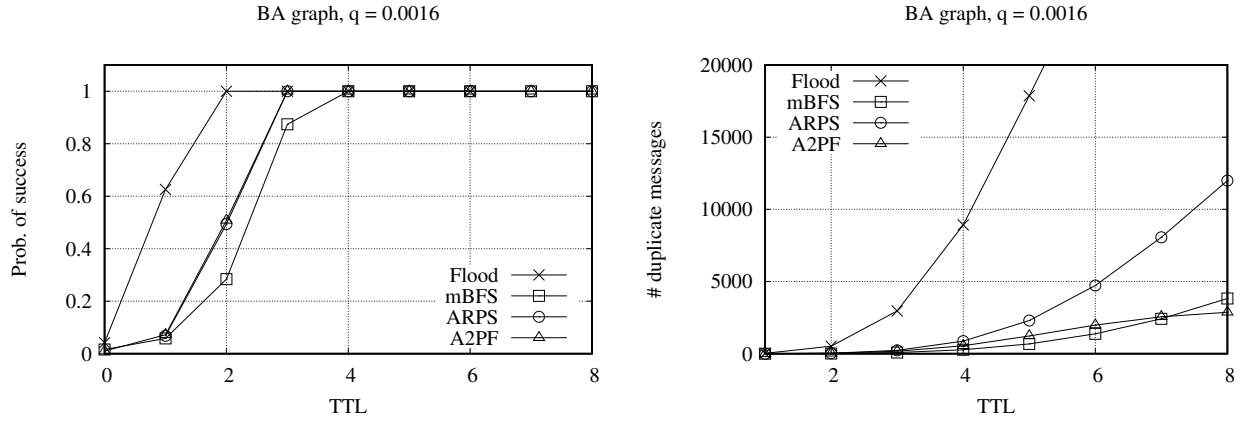


Σχήμα 7.12: Συγκριτικά αποτελέσματα αναζήτησης με υψηλό και χαμηλό ρυθμό churn (14, 5% και 45% αντίστοιχα) σε δυναμικό δίκτυο ER με αρχικό πλήθος 30.000 κόμβων, με $\bar{d} = 6$ και $q = 0.0006$.

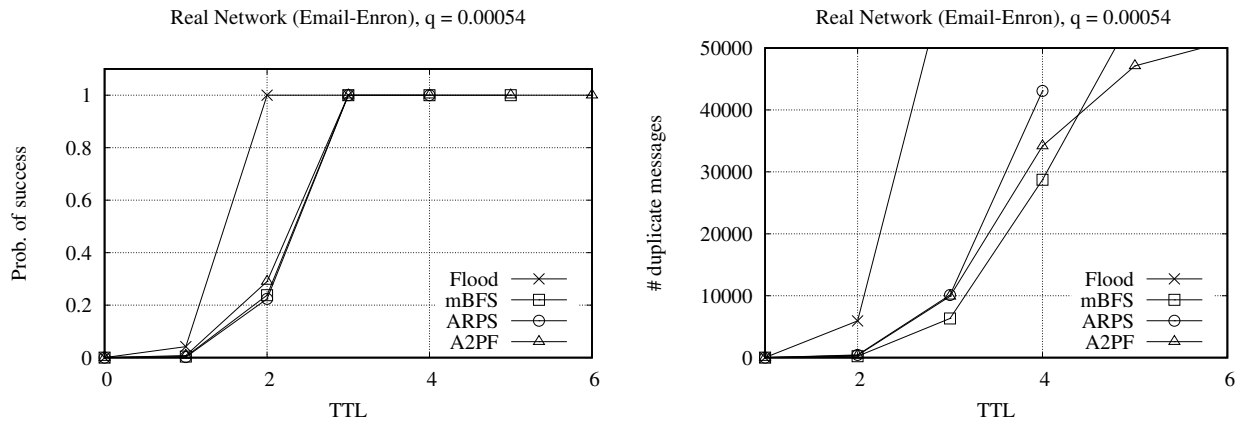
πληροφορίες που συλλέγει ο κάθε κόμβος από τα εισερχόμενα μηνύματα και που αποτυπώνουν τη δόμηση της ευρύτερης γειτονιάς του. Έτσι, η προώθηση της κάθε ερώτησης καθοδηγείται από τη γνώση αυτή, συμβάλλοντας ουσιαστικά στη μείωση των διπλότυπων μηνυμάτων.

Όπως φαίνεται και στη γραφική αναπαράσταση των αποτελεσμάτων στο Σχ. 7.13(α), η πιθανότητα επιτυχίας για την A2PF γίνεται 1 μετά τα τρία πρώτα βήματα της διαδικασίας όπως επίσης συμβαίνει και στη στρατηγική ARPS, ενώ για την πλημμύρα αυτό συμβαίνει από το 2ο βήμα και για την mBFS από το 4ο βήμα και μετά. Ταυτόχρονα, όπως παρατηρείται στο Σχ. 7.13(β), το πλήθος των διπλότυπων μηνυμάτων αυξάνεται σχεδόν εκθετικά από ένα βήμα και μετά για τις περισσότερες στρατηγικές, σε αντίθεση με την A2PF, όπου η ποσότητά τους σταθεροποιείται σε μια σχετικά μικρή τιμή από το 7ο βήμα της αναζήτησης και μετά.

Στη συνέχεια διεξάγουμε την ίδια σειρά προσομοιώσεων σε πραγματικά δίκτυα. Εδώ παρουσιάζουμε τα αποτελέσματα για ένα από αυτά, το δίκτυο email-Enron [116] που αποτελείται από 36.692 κόμβους και 183.831 ακμές θεωρώντας ότι για κάθε πόρο τοποθετούνται στο δίκτυο 20 αντίγραφα. Τα αποτελέσματα των μετρήσεων παρουσιάζονται στο Σχ. 7.14. Όπως παρατηρούμε όλες οι στρατηγικές επιτυγχάνουν στα 2 πρώτα βήματα, όμως όλες εκτός της τροποποιημένης A2PF οδηγούν και σε εκθετική αύξηση των διπλότυπων μηνυμάτων. Αντίθετα η τροποποιημένη A2PF κατορθώνει να ελέγξει την παραγωγή των διπλότυπων μηνυμάτων και να τα περιορίσει αισθητά.



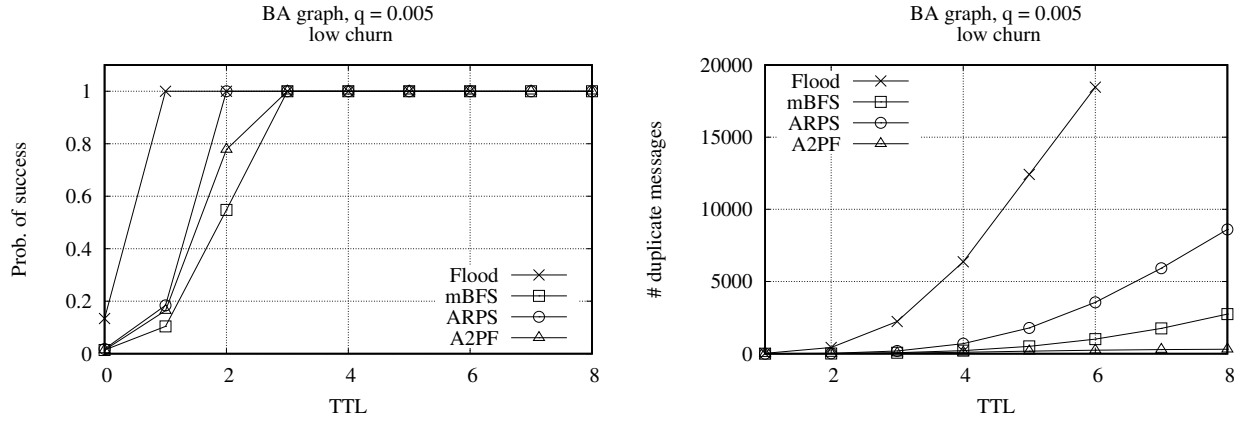
Σχήμα 7.13: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) σε ένα δίκτυο BA power-law με σταθερό πλήθος 30.000 κόμβων με $m = 2$, $q = 0.0016$.



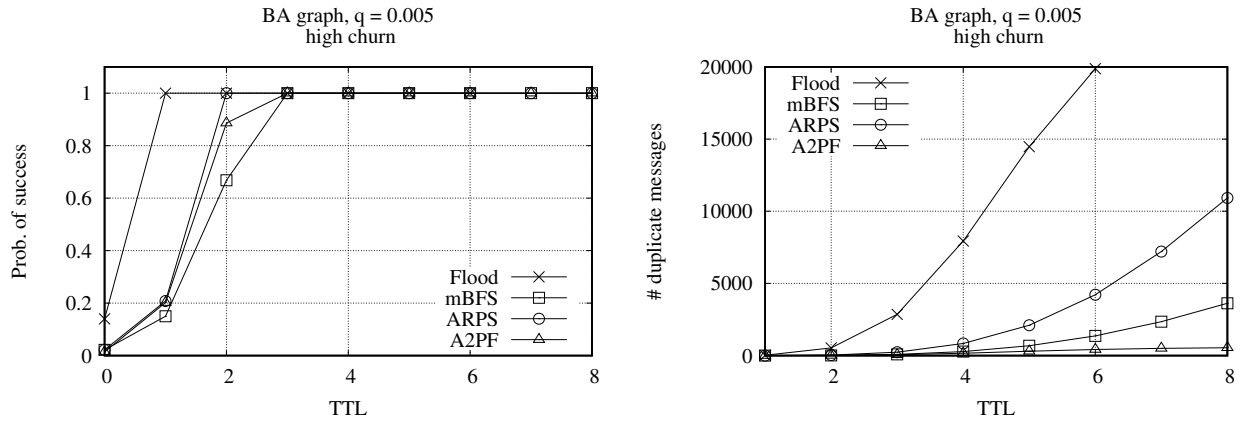
Σχήμα 7.14: Πιθανότητα επιτυχίας (α) και διπλότυπα μηνύματα (β) στο πραγματικό δίκτυο email-Enron 36.692 κόμβων και 183.831 ακμών, $q = 0.0054$ αντίγραφα.

Επίσης, εξετάζουμε τις ίδιες στρατηγικές αναζήτησης σε δίκτυα που παράγουμε σύμφωνα με το μοντέλο των Barabási-Albert σε συνθήκες υψηλού και χαμηλού ρυθμού churn και αξιολογούμε τόσο την αποτελεσματικότητα της δικιάς μας στρατηγικής όσο και την συμπεριφορά όλων των άλλων στρατηγικών σε συνθήκες churn. Το δίκτυο αποτελείται από 20.000 κόμβους αρχικά, και για κάθε διατιθέμενο πόρο τοποθετούνται ομοιόμορφα τυχαία 100 αντίγραφα του. Όπως και στα δίκτυα *ER*, η διάρκεια ζωής του κάθε κόμβου καθορίζεται από την κατανομή Weibull. Η μέση διάρκεια ζωής κυμαίνεται από 2 έως 4 ώρες που οδηγεί σε συνολικό ρυθμό churn 14% έως 45%.

Στα Σχ. 7.15, 7.16 και 7.17 παρουσιάζονται τα αποτελέσματα από τις πειραματικές μετρήσεις των τεσσάρων στρατηγικών αναζήτησης σε αυτό το δίκτυο σε

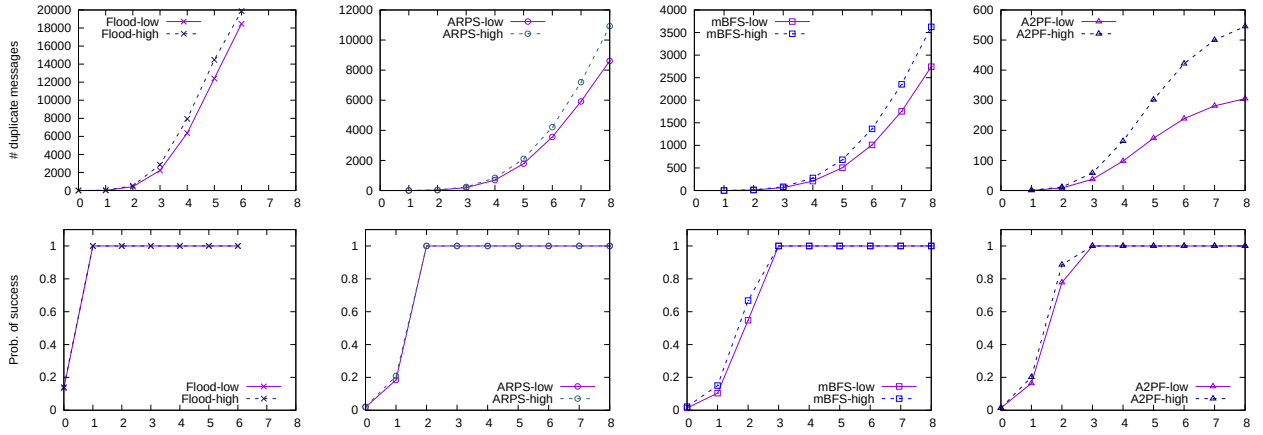


Σχήμα 7.15: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο BA power-law 20.000 κόμβων, με $m = 2$ και $q = 0.005$ και με ρυθμό churn 14%.



Σχήμα 7.16: Πιθανότητα επιτυχίας (αριστερά) και διπλότυπα μηνύματα (δεξιά) σε ένα δυναμικό δίκτυο BA power-law 20.000 κόμβων, με $m = 2$ και $q = 0.005$ και με ρυθμό churn 45%.

συνθήκες υψηλού (45%) και χαμηλού (14%) ρυθμού churn. Όπως παρατηρούμε, η εικόνα για την απόδοση των στρατηγικών αναζήτησης παραμένει ίδια: η πλημμύρα επιτυγχάνει από το 1ο βήμα, αλλά ταυτόχρονα παράγει υπολογίσιμο αριθμό διπλότυπων μηνυμάτων, τα οποία αυξάνονται αισθητά μετά το 3ο βήμα της διαδικασίας. Η ARPS και η A2PF εμφανίζουν σχεδόν όμοια πιθανότητα επιτυχίας, ενώ η πιθανότητα επιτυχίας στην mBFS γίνεται 1 μετά το 3ο βήμα της διαδικασίας. Όμως όλες οι στρατηγικές εκτός της A2PF, εισάγουν υπολογίσιμο αριθμό διπλότυπων μηνυμάτων που αυξάνεται όσο αυξάνονται τα βήματα της διαδικασίας. Αντίθετα, ο αριθμός των διπλότυπων μηνυμάτων για την A2PF, είναι σχεδόν αμελητέος. Αυτό είναι αποτέλε-



Σχήμα 7.17: Συγκριτικά αποτελέσματα αναζήτησης με υψηλό και χαμηλό ρυθμό churn (14,5% και 45% αντίστοιχα) σε δυναμικό δίκτυο BA power-law με αρχικό πλήθος 20.000 κόμβων, με $m = 2$ και $q = 0.005$.

σμα της στρατηγικής προώθησης που εφαρμόζει η A2PF και που σχετίζεται με το πλήθος των αντιγράφων που υπάρχουν στο δίκτυο. Η A2PF προσαρμόζει την πιθανότητα προώθησης ανάλογα με το πλήθος των αντιγράφων και τη δομή του δικτύου με αποτέλεσμα ο αριθμός των παραγόμενων διπλότυπων μηνυμάτων να είναι σχεδόν αμελητέος.

Από τα ίδια αποτελέσματα που παρουσιάζονται στο Σχ. 7.17 για κάθε στρατηγική ξεχωριστά, διαπιστώνουμε ότι η επίδραση του churn στην αναζήτηση είναι σχεδόν αμελητέα. Η συμπεριφορά της κάθε στρατηγικής είναι παρόμοια σε χαμηλό ή υψηλό churn. Αυτό οφείλεται στο γεγονός πως σε αυτά τα δίκτυα παρά την επίδραση του churn δεν εμφανίζονται απομονωμένες συνιστώσες. Η μεγάλη απόκλιση στο μέσο βαθμό των γειτονικών κόμβων [29] εξασφαλίζει την ύπαρξη ικανού αριθμού συνδέσεων ώστε να μεταδοθεί το μήνυμα-ερώτηση. Όμως, οι υπάρχουσες συνδέσεις, αλλά και οι νέες από τους εισερχόμενους κόμβους οδηγούν σε μικρή αύξηση των διπλότυπων μηνυμάτων.

7.6 Σύνοψη

Στο κεφάλαιο αυτό μελετήσαμε την εφαρμογή μεθόδων αναζήτησης σε δυναμικά δίκτυα με κυμαινόμενους ρυθμούς churn. Παρόλο που η επίδραση του είναι σχετικά μικρή στη διαδικασία της αναζήτησης, υπάρχει ανάγκη να περιοριστούν τα

διπλότυπα μηνύματα που με ή χωρίς churn παράγει η διαδικασία σε υπολογίσιμες ποσότητες. Εμείς παρουσιάσαμε την στρατηγική A2PF που σχεδόν εξαλείφει αυτά τα μηνύματα. Αυτό επιβεβαιώνεται και από τις πειραματικές μετρήσεις.

ΚΕΦΑΛΑΙΟ 8

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

8.1 Συμπεράσματα

8.2 Μελλοντική εργασία

Σκοπός αυτής της διατριβής ήταν η ανάπτυξη, η εφαρμογή και η αξιολόγηση μοντέλων, αλγορίθμων και τεχνικών αναζήτησης σε σύνθετα δίκτυα έχοντας ως βάση την πιθανοτική πλημμύρα. Στη συνέχεια παρουσιάζουμε συνοπτικά τα αποτελέσματα και τα εξαγόμενα συμπεράσματα. Τέλος, δίνονται κατευθύνσεις για μελλοντική εργασία.

8.1 Συμπεράσματα

Σε αυτή τη διατριβή, εστιάζουμε στο πρόβλημα της αναζήτησης σε αδόμητα συστήματα μεγάλης κλίμακας, όπως είναι τα δίκτυα $p2p$. Τα συστήματα αυτά, γνωστά ως σύνθετα δίκτυα, δεν έχουν προκαθορισμένο μέγεθος, χαρακτηρίζονται από στατιστικές ιδιότητες όπως η κατανομή των βαθμών των κόμβων τους, και διαθέτουν περιεχόμενα κατανεμημένα στους κόμβους τους. Σε ένα τέτοιο περιβάλλον, είναι αναγκαίο ο μηχανισμός αναζήτησης να είναι ευέλικτος, αποτελεσματικός και αποδοτικός ώστε να αποφευχθούν προβλήματα όπως παραγωγή τεράστιου όγκου μηνυμάτων, σπατάλη πόρων (εύρος ζώνης, αποθήκευση, κ.ά.), και συμφόρηση. Για να εξετάσουμε το πρόβλημα της αναζήτησης, θέτουμε ένα πλαίσιο που περιλαμβάνει

τρία διαφορετικά πεδία: την τοπολογία του δικτύου, τους πόρους και τα αντίγραφα τους και τον μηχανισμό αναζήτησης.

Παραδοσιακά και καθώς η διαθέσιμη πληροφορία είναι περιορισμένη, η αναζήτηση βασίζεται στην πλημμύρα. Οι στρατηγικές αναζήτησης βασιζόμενες στην πλημμύρα είναι ταχύτατες, όμως ευθύνονται για μια σειρά από προβλήματα που πηγάζουν κυρίως από την υπέρμετρη παραγωγή μηνυμάτων με συνέπεια να μετριάζουν την απόδοση και τη λειτουργικότητα όλου του συστήματος. Προκειμένου να βελτιστοποιήσουμε τη διαδικασία αναζήτησης, θα πρέπει πρώτα να κατανοήσουμε τη συμπεριφορά της. Εμείς εξηγήσαμε πως ακριβώς παράγεται ο τεράστιος όγκος μηνυμάτων, τον ποσοτικοποιούμε και υπολογίζουμε όρια για το πλήθος των μηνυμάτων σε κάθε βήμα της διαδικασίας χρησιμοποιώντας αναλυτικές εκφράσεις. Επίσης, προτείνουμε ένα προσεγγιστικό μοντέλο που υπολογίζει το πλήθος των διαφορετικών κόμβων που συναντά η διαδικασία σε κάθε της βήμα, καθώς εξαπλώνεται στο δίκτυο. Όλα τα παραπάνω έχουν δημοσιευτεί στο [3].

Η βασική ιδέα για να μειωθεί η παραγόμενη κίνηση της πλημμύρας και να εξλειφθούν τα προβλήματά της είναι η πιθανοτική πλημμύρα, ο περιορισμός δηλαδή των αναμεταδόσεων των μηνυμάτων που λαμβάνει ένας κόμβος. Εμείς, εστιάζοντας σε αυτή την προσέγγιση, προτείνουμε στρατηγικές αναζήτησης που σε αντίθεση με άλλες παρόμοιες προσεγγίσεις εφαρμόζονται σε ετερογενή και με υψηλό churn περιβάλλοντα. Πιο συγκεκριμένα, αρχικά προτείνουμε, σχεδιάζουμε και αναπτύσσουμε μια πιθανοτική στρατηγική αναζήτησης, που ονομάζεται APF και κατορθώνει να: α) αποκόψει τις περιττές διαδρομές της απλής πλημμύρας (overshooting), β) εξαλείψει σχεδόν τον αριθμό των διπλότυπων μηνυμάτων, γ) προσαρμόσει την επέκτασή της στο δίκτυο βελτιώνοντας έτσι την απόδοσή της αναζήτησης. Το σημείο κλειδί της APF είναι η πιθανότητα προώθησης που υπολογίζεται από τον κάθε κόμβο, από μια σύνθετη έκφραση που περιλαμβάνει την απόσταση από τον αρχικό κόμβο που εισάγει το μήνυμα-ερώτηση, τη δημοτικότητα του επιθυμητού πόρου, και τους βαθμούς των κόμβων του δικτύου. Η APF συμπεριφέρεται ιδιαίτερα ικανοποιητικά σε δίκτυα που οι βαθμοί των κόμβων του δεν παρουσιάζουν μεγάλη απόκλιση από το μέσο βαθμό του δικτύου. Επίσης, τα αποτελέσματα προσομοιώσεων δείχνουν την υπεροχή της σε σχέση με άλλες παρόμοιες στρατηγικές. Όμως σε δίκτυα με μεγάλες αποκλίσεις στους βαθμούς των κόμβων (π.χ. δίκτυα άνευ κλίμακας) οι επιδόσεις της APF είναι χαμηλότερες. Τα συμπεράσματα αυτά παρουσιάστηκαν και δημοσιεύτηκαν στα [4], [2].

Για να ξεπεράσουμε αυτόν τον περιορισμό της APF και να επεκτείνουμε την εφαρμογή της σε κάθε δίκτυο προτείνουμε την προσαρμοζόμενη APF (A2PF) αφού πρώτα εισάγουμε μια νέα διάσταση στο μήνυμα-ερώτηση με την προσθήκη ενός πεδίου με το βαθμό του κόμβου που ξεκινάει την αναζήτηση. Ο κάθε κόμβος του δικτύου αξιοποιεί αυτή την πληροφορία που προέρχεται από τις ερωτήσεις που λαμβάνει για να αποκαλύψει δομικά χαρακτηριστικά του δικτύου που εκτείνεται γύρω από αυτόν, όπως για παράδειγμα ο μέσος βαθμός των κόμβων σε μια συγκεκριμένη απόσταση. Σύμφωνα με την A2PF ο κάθε κόμβος εκμεταλλεύεται αυτή τη γνώση για να υπολογίσει την πιθανότητα προώθησης ενός μηνύματος που λαμβάνει και που εξαρτάται από τον μέσο βαθμό των κόμβων σε απόσταση ίση με το βήμα της διαδικασίας και όχι με βάση το μέσο βαθμό του δικτύου. Αυτή η τροποποίηση είναι απαραίτητη καθώς στα δίκτυα άνευ κλίμακας οι βαθμοί των κόμβων μπορεί να διαφέρουν πολύ από το μέσο βαθμό του δικτύου. Επιπλέον, είναι ένας τρόπος προσαρμογής στις μεταβολές που συμβαίνουν στις συνδέσεις των κόμβων λόγω δυναμικότητας του δικτύου.

Η A2PF εφαρμόζεται σε κάθε τύπο δικτύου και βελτιώνει ουσιαστικά την απόδοση της αναζήτησης. Επίσης, πειραματικές μετρήσεις σε σύνολα δεδομένων από πραγματικά δίκτυα είχαν εξίσου ικανοποιητικά αποτελέσματα. Αυτό μας οδηγεί στο συμπέρασμα πως η μέθοδός μας είναι κατάλληλη για εφαρμογή και σε πραγματικά δυναμικά εξελισσόμενα δίκτυα.

Τέλος, σχεδιάζουμε και αναπτύσσουμε έναν πρωτότυπο, γενικού σκοπού προσομοιωτή, ονομαζόμενο Armonia [1]. Το Armonia, ένα ολοκληρωμένο και ευέλικτο πλαίσιο προσομοίωσης το οποίο βασίζεται στο OMNeT++, παρέχει υλοποιημένα σύνολα για διάφορα στατικά και δυναμικά μοντέλα δικτύων, διαφορετικές τεχνικές αναζήτησης και στρατηγικές εκχώρησης αντιγράφων προερχόμενες από τη σχετική βιβλιογραφία. Ο χρήστης μπορεί να χρησιμοποιήσει οποιοδήποτε επιθυμητό συνδυασμό μοντέλου δικτύου–πολιτικής εκχώρησης αντιγράφων–στρατηγικής αναζήτησης, να καθορίσει τις ρυθμίσεις ή/και να προσθέσει δικές του υλοποιήσεις.

8.2 Μελλοντική εργασία

Στη συνέχεια παραθέτουμε ορισμένα από τα πιο σημαντικά ανοικτά ζητήματα που αναδείχτηκαν στη διάρκεια εκπόνησης αυτής της διατριβής και αφορούν συνεχιζό-

μενη και μελλοντική εργασία.

Σε αυτή τη διατριβή εστιάσαμε στην εύρεση τουλάχιστον ενός επιθυμητού αποτελέσματος. Μια ενδιαφέρουσα κατεύθυνση για μελλοντική εργασία είναι να διερευνηθεί η συμπεριφορά των πιθανοτικών στρατηγικών στην περίπτωση που το ζητούμενο είναι ένα ποσοστό των αποτελεσμάτων ή και όλα τα δυνατά αποτελέσματα. Τα μέχρι τώρα αποτελέσματα εδραιώνουν την πεποίθησή μας ότι τόσο η APF όσο και η A2PF θα λειτουργήσουν με αποτελεσματικό τρόπο και σε αυτή την περίπτωση.

Η επιτυχία της αναζήτησης, όπως αναφέρθηκε και στην Ενότητα 3.4.2 σχετίζεται σε μεγάλο βαθμό με τη διαθεσιμότητα των πόρων και των αντιγράφων τους. Η διαχείρισή τους εγείρει μια σειρά από ζητήματα, όπως ποιος θα είναι αριθμός αντιγράφων, που θα τοποθετηθούν/κατανεμηθούν τα αντίγραφα και για πόσο διάστημα θα διατηρηθούν σε ένα σύστημα. Στα πραγματικά δίκτυα, η δημοτικότητα των πόρων και η ζήτησή τους μεταβάλλεται και αυτό αντικατοπτρίζεται στη διάρκεια ζωής τους αλλά και στις αλληλεπιδράσεις που συμβαίνουν στο δίκτυο (π.χ. προσθήκη αφαίρεση συνδέσεων). Η επιστήμη των δικτύων σήμερα παρέχει εργαλεία και τεχνικές (π.χ. στατιστικούς μηχανισμούς) που μπορούν να δώσουν μια νέα διάσταση στην έρευνα της αναζήτησης σε συνδυασμό με τη διαχείριση των πόρων και των αντιγράφων τους.

Στην Ενότητα 4.4 έγινε μια πρώτη προσέγγιση βελτιστοποίησης της αναζήτησης σε σύνθετα δίκτυα με τη χρήση της θεωρίας παιγνίων. Αυτή η προσέγγιση ανοίγει νέους δρόμους μελέτης σε διάφορες κατευθύνσεις, όπως στην εκχώρηση πόρων, στην προώθηση των ερωτήσεων ή στο σχηματισμό του δικτύου. Για την πρώτη κατεύθυνση, την κατανομή των πόρων, οι Arfeen κ.ά. [12] υιοθετούν τη θεωρία παιγνίων σε περιβάλλον cloud computing. Πρόκληση αποτελεί η υλοποίηση τεχνικών κατανομής πόρων και σε άλλους τύπους δικτύων εφαρμόζοντας στρατηγικές αλληλεξάρτησης ώστε να ελαχιστοποιείται το κόστος τοποθέτησης αλλά και το κόστος της αναζήτησης. Στη μέχρι τώρα μελέτη μας, υποθέσαμε ότι όλοι οι κόμβοι είναι συνεργάσιμοι και προωθούν τις ερωτήσεις που λαμβάνουν. Στην πραγματικότητα όμως η συμπεριφορά των κόμβων είναι αυτόνομη και ενδεχομένως δεν προωθούν πάντα τις ερωτήσεις που λαμβάνουν. Ιδιαίτερο ενδιαφέρον έχει να διερευνήσουμε την συμπεριφορά τους αν υποθέσουμε ότι υπάρχει κάποια στρατηγική που αποφέρει απολαβές/ποινές στους κόμβους εφόσον την εφαρμόζουν. Ο Kleinberg κ.ά [100] θέτει τη βάση σε αυτό το ζήτημα, υποθέτοντας ανταμοιβή σε κάθε κόμβο που προ-

ωθεί μια ερώτηση σε ένα τυχαίο δίκτυο-δέντρο. Τέλος, ιδιαίτερα ελκυστική είναι η μελέτη του σχηματισμού και της εξέλιξης των σύνθετων δικτύων σε συνδυασμό με τη θεωρία παιγνίων. Όπως είναι γνωστό [60], σε πολλά δίκτυα (π.χ. κοινωνικά, οικονομικά) οι ρόλοι των κόμβων και οι μεταξύ τους συνδέσεις καθορίζονται από το ατομικό όφελος του καθενός. Περαιτέρω διερεύνηση και χρήση ευρετικών μεθόδων θα μπορούσε να οδηγήσει στην πρόβλεψη σχηματισμού ομάδων ή σε βελτιστοποίηση της δρομολόγησης των ερωτήσεων.

Για να αποφύγουμε το κόστος από την πληθώρα των παραγόμενων μηνυμάτων, στο Κεφάλαιο 5 προτείνουμε την εφαρμογή μιας πιθανοτικής στρατηγικής αναζήτησης όπου ο κάθε κόμβος εκτιμά την πιθανότητα προώθησης από μια έκφραση που περιλαμβάνει τυπικές παραμέτρους του δικτύου. Η στρατηγική μας είναι επεκτάσιμη καθώς επιτρέπει τον ορισμό νέων παραμέτρων στον υπολογισμό της πιθανότητας προώθησης. Ενδεικτικά θα μπορούσε να χρησιμοποιηθεί ένα μέτρο ομοιότητας (similarity) των κόμβων και που ανάλογα με την τιμή του θα αυξάνει ή θα μειώνει την πιθανότητα προώθησης. Επίσης, η στρατηγική μας επιλέγει τυχαία τους γειτονικούς κόμβους στους οποίους θα προωθήσει το μήνυμα ερώτηση. Ενδιαφέρον είναι να εξετάσουμε τι θα συμβεί αν προτρέψουμε την προώθηση προς συγκεκριμένους κόμβους, με βάση κάποιο κριτήριο, όπως αυτό της ομοιότητας των κόμβων.

Σημείο-κλειδί για την αποτελεσματική λειτουργία διαδικασιών στα σύνθετα δίκτυα, είναι η ικανότητα προσαρμογής τους στην εκάστοτε τοπολογία. Για να το κατορθώσουμε αυτό στην περίπτωση της αναζήτησης, προτείνουμε έναν μηχανισμό που βοηθάει τον κάθε κόμβο να αποκτήσει εικόνα για την “άγνωστη” τοπολογία του δικτύου. Συγκεκριμένα, συμπεριλάβαμε στο μήνυμα-ερώτηση τον βαθμό του κάθε κόμβου, με αποτέλεσμα όσοι κόμβοι λαμβάνουν το μήνυμα, μπορούν να μάθουν και για το μέσο πλήθος των συνδέσεων σε ορισμένη απόσταση από αυτούς. Θα ήταν ιδιαίτερα ενδιαφέρον να μελετηθεί η εφαρμογή αυτού του μηχανισμού και σε άλλες δυναμικές διαδικασίες των σύνθετων δικτύων. Για παράδειγμα θα μπορούσε αυτός ο μηχανισμός να δώσει μια νέα διάσταση στη μελέτη της διάδοσης στα σύνθετα δίκτυα, ή στη λήψη αποφάσεων για την υιοθέτηση μιας καινοτομίας, ή και ακόμη στην πρόβλεψη διαμόρφωση γνώμης. Τέλος, προσδοκούμε ο μηχανισμός αυτός να συμβάλει στην αποκάλυψη τοπολογικών χαρακτηριστικών των δικτύων, όπως για παράδειγμα η συσχέτιση βαθμού (degree correlation) που εκφράζει τη σχέση των βαθμών μεταξύ συνδεδεμένων κόμβων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] N. S. 3, “ns-3.” [Online]. Available: <https://www.nsnam.org/>
- [2] L. A. Adamic and N. Glance, “The political blogosphere and the 2004 us election: divided they blog,” in *Proceedings of the 3rd international workshop on Link discovery*. ACM, 2005, pp. 36–43.
- [3] L. A. Adamic, R. M. Lukose, and B. A. Huberman, “Local search in unstructured networks,” *Handbook of graphs and networks: from the genome to the internet*, 2006.
- [4] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, “Search in power-law networks,” *Physical review E*, vol. 64, no. 4, p. 046135, 2001.
- [5] W. Aiello, F. Chung, and L. Lu, “A random graph model for massive graphs,” in *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. Acm, 2000, pp. 171–180.
- [6] T. Akiba, Y. Iwata, and Y. Yoshida, “Fast exact shortest-path distance queries on large networks by pruned landmark labeling,” in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. ACM, 2013, pp. 349–360.
- [7] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.
- [8] R. Albert, H. Jeong, and A.-L. Barabási, “Internet: Diameter of the world-wide web,” *Nature*, vol. 401, no. 6749, pp. 130–131, 1999.
- [9] B. Alspach, “Searching and sweeping graphs: a brief survey,” *Le matematiche*, vol. 59, no. 1, 2, pp. 5–37, 2006.

- [10] A. Arenas, A. Cabrales, A. Díaz-Guilera, R. Guimerà, and F. Vega-Redondo, “Search and congestion in complex networks,” in *Statistical Mechanics of Complex Networks*. Springer, 2003, pp. 175–194.
- [11] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, “Synchronization in complex networks,” *Physics Reports*, vol. 469, no. 3, pp. 93–153, 2008.
- [12] M. A. Arfeen, K. Pawlikowski, and A. Willig, “A framework for resource allocation strategies in cloud computing environment,” in *Computer Software and Applications Conference Workshops (COMPSACW), 2011 IEEE 35th Annual*. IEEE, 2011, pp. 261–266.
- [13] J. Augustine, A. R. Molla, E. Morsy, G. Pandurangan, P. Robinson, and E. Upfal, “Storage and search in dynamic peer-to-peer networks,” in *Proceedings of the twenty-fifth annual ACM symposium on Parallelism in algorithms and architectures*. ACM, 2013, pp. 53–62.
- [14] L. Backstrom and J. Leskovec, “Supervised random walks: predicting and recommending links in social networks,” in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 635–644.
- [15] H. Balakrishnan, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica, “Looking up data in p2p systems,” *Communications of the ACM*, vol. 46, no. 2, pp. 43–48, 2003.
- [16] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, V. Colizza *et al.*, “Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility,” *BMC medicine*, vol. 7, no. 1, p. 1, 2009.
- [17] F. Banaei-Kashani and C. Shahabi, “Criticality-based analysis and design of unstructured peer-to-peer networks as” complex systems”, in *Cluster Computing and the Grid, 2003. Proceedings. CCGrid 2003. 3rd IEEE/ACM International Symposium on*. IEEE, 2003, pp. 351–358.
- [18] A.-L. Barabási, *Network science*. Cambridge university press, 2016.
- [19] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *science*, vol. 286, no. 5439, pp. 509–512, 1999.

- [20] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical processes on complex networks*. Cambridge University Press, 2008.
- [21] M. Bastian, S. Heymann, and M. Jacomy, “Gephi: An open source software for exploring and manipulating networks,” in *Proc. International AAAI Conference on Weblogs and Social Media*, 2009.
- [22] A. Basu, S. Fleming, J. Stanier, S. Naicken, I. Wakeman, and V. K. Gurbani, “The state of peer-to-peer network simulators,” *ACM Comput. Surv.*, vol. 45, no. 4, pp. 46:1–46:25, Aug. 2013.
- [23] H. Baumann, P. Crescenzi, and P. Fraigniaud, “Parsimonious flooding in dynamic graphs,” *Distributed Computing*, vol. 24, no. 1, pp. 31–44, 2011.
- [24] H. Baumann, P. Crescenzi, and P. Fraigniaud, “Flooding in dynamic graphs with arbitrary degree sequence,” *Journal of Parallel and Distributed Computing*, vol. 74, no. 5, pp. 2433–2437, 2014.
- [25] I. Baumgart, B. Heep, and S. Krause, “OverSim: A flexible overlay network simulation framework,” in *Proc. 10th IEEE Global Internet Symposium (GI '07) in conjunction with IEEE INFOCOM 2007, Anchorage, AK, USA*, May 2007, pp. 79–84.
- [26] O. Bénichou, C. Loverdo, M. Moreau, and R. Voituriez, “Intermittent search strategies,” *Reviews of Modern Physics*, vol. 83, no. 1, p. 81, 2011.
- [27] N. Bisnik and A. Abouzeid, “Modeling and analysis of random walk search algorithms in p2p networks,” in *Second International Workshop on Hot Topics in Peer-to-Peer Systems*. IEEE, 2005, pp. 95–103.
- [28] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics reports*, vol. 424, no. 4, pp. 175–308, 2006.
- [29] M. Boguná, R. Pastor-Satorras, and A. Vespignani, “Absence of epidemic threshold in scale-free networks with degree correlations,” *Physical review letters*, vol. 90, no. 2, p. 028701, 2003.
- [30] B. Bollobás, *Random Graphs*, 2nd ed. Cambridge University Press, 2001.

- [31] S. Brin and L. Page, “Reprint of: The anatomy of a large-scale hypertextual web search engine,” *Computer networks*, vol. 56, no. 18, pp. 3825–3833, 2012.
- [32] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. Newman, and S. H. Strogatz, “Are randomly grown graphs really random?” *Physical Review E*, vol. 64, no. 4, p. 041902, 2001.
- [33] S. Carmi, R. Cohen, and D. Dolev, “Searching complex networks efficiently with minimal information,” *EPL (Europhysics Letters)*, vol. 74, no. 6, p. 1102, 2006.
- [34] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro, “Time-varying graphs and dynamic networks,” *International Journal of Parallel, Emergent and Distributed Systems*, vol. 27, no. 5, pp. 387–408, 2012.
- [35] J. Chandra and N. Ganguly, “On coverage bounds of unstructured peer-to-peer networks,” *Advances in Complex Systems*, vol. 14, no. 04, pp. 611–633, 2011.
- [36] J. Chandra, S. Shaw, and N. Ganguly, “Analyzing network coverage in unstructured peer-to-peer networks: A complex network approach,” in *International Conference on Research in Networking*. Springer, 2009, pp. 690–702.
- [37] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, “Making gnutella-like p2p systems scalable,” in *Proc. 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM ’03. New York, NY, USA: ACM, 2003, pp. 407–418.
- [38] S. Chen, W. Huang, C. Cattani, and G. Altieri, “Traffic dynamics on complex networks: a survey,” *Mathematical Problems in Engineering*, vol. 2012, 2011.
- [39] A. Clauset, C. R. Shalizi, and M. E. Newman, “Power-law distributions in empirical data,” *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.
- [40] A. Clementi, R. Silvestri, and L. Trevisan, “Information spreading in dynamic graphs,” *Distributed Computing*, vol. 28, no. 1, pp. 55–73, 2015.
- [41] A. E. Clementi, C. Macci, A. Monti, F. Pasquale, and R. Silvestri, “Flooding time in edge-markovian dynamic graphs,” in *Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing*. ACM, 2008, pp. 213–222.

- [42] A. E. Clementi, A. Monti, F. Pasquale, and R. Silvestri, “Broadcasting in dynamic radio networks,” *Journal of Computer and System Sciences*, vol. 75, no. 4, pp. 213–230, 2009.
- [43] E. Cohen, A. Fiat, and H. Kaplan, “Associative search in peer to peer networks: Harnessing latent semantics,” *Computer Networks*, vol. 51, no. 8, pp. 1861–1881, 2007.
- [44] E. Cohen and S. Shenker, “Replication strategies in unstructured peer-to-peer networks,” in *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 4. ACM, 2002, pp. 177–190.
- [45] R. Cohen and S. Havlin, *Complex networks: structure, robustness and function*. Cambridge University Press, 2010.
- [46] B. F. Cooper, “An optimal overlay topology for routing peer-to-peer searches,” in *Middleware 2005*. Springer, 2005, pp. 82–101.
- [47] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, “Characterization of complex networks: A survey of measurements,” *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.
- [48] L. d. F. Costa, O. N. Oliveira Jr, G. Travieso, F. A. Rodrigues, P. R. Villas Boas, L. Antigueira, M. P. Viana, and L. E. Correa Rocha, “Analyzing and modeling real-world phenomena with complex networks: a survey of applications,” *Advances in Physics*, vol. 60, no. 3, pp. 329–412, 2011.
- [49] A. Crespo and H. Garcia-Molina, “Routing indices for peer-to-peer systems,” in *Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on*. IEEE, 2002, pp. 23–32.
- [50] A. Crespo and H. Garcia-Molina, “Semantic overlay networks for p2p systems,” in *Agents and Peer-to-Peer Computing*. Springer, 2004, pp. 1–13.
- [51] S. Crisóstomo, U. Schilcher, C. Bettstetter, and J. a. Barros, “Analysis of probabilistic flooding: how do we choose the right coin?” in *Proceedings of the 2009 IEEE international conference on Communications*, ser. ICC’09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 2080–2085.

- [52] S. Crisóstomo, U. Schilcher, C. Bettstetter, and J. Barros, “Probabilistic flooding in stochastic networks: Analysis of global information outreach,” *Computer Networks*, vol. 56, no. 1, pp. 142–156, 2012.
- [53] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, “Social ties and their relevance to churn in mobile telecom networks,” in *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*. ACM, 2008, pp. 668–677.
- [54] D. J. de Solla Price, “Networks of scientific papers,” *Science*, vol. 149, no. 3683, pp. 510–515, 1965.
- [55] V. V. Dimakopoulos, S. Margariti, M. Ntetsika, and E. Pitoura, “Data replication in p2p systems,” in *Handbook of Research on P2P and Grid Systems for Service-Oriented Computing: Models, Methodologies and Applications: Models, Methodologies and Applications*. IGI Global, 2010, pp. 589–615.
- [56] V. V. Dimakopoulos and E. Pitoura, “On the performance of flooding-based resource discovery,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 11, pp. 1242–1252, 2006.
- [57] P. S. Dodds, R. Muhamad, and D. J. Watts, “An experimental study of search in global social networks,” *science*, vol. 301, no. 5634, pp. 827–829, 2003.
- [58] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, “Structure of growing networks with preferential linking,” *Physical review letters*, vol. 85, no. 21, p. 4633, 2000.
- [59] S. N. Dorogovtsev and J. F. Mendes, “Evolution of networks,” *Advances in physics*, vol. 51, no. 4, pp. 1079–1187, 2002.
- [60] B. Dutta and M. O. Jackson, *Networks and groups: Models of strategic formation*. Springer Science & Business Media, 2013.
- [61] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
- [62] P. Erdős and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, pp. 17–61, 1960.

- [63] E. Estrada, “Introduction to complex networks: structure and dynamics,” in *Evolutionary Equations with Applications in Natural Sciences*. Springer, 2015, pp. 93–131.
- [64] F. E. Faisal and T. Milenković, “Dynamic networks reveal key players in aging,” *Bioinformatics*, vol. 30, no. 12, pp. 1721–1729, 2014.
- [65] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On power-law relationships of the internet topology,” *SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, Aug. 1999. [Online]. Available: <http://doi.acm.org/10.1145/316194.316229>
- [66] S. Ferretti, “Gossiping for resource discovering: An analysis based on complex network theory,” *Future Generation Computer Systems*, vol. 29, no. 6, pp. 1631–1644, 2013.
- [67] D. Foreback, M. Nesterenko, and S. Tixeul, “Infinite unlimited churn,” *arXiv preprint arXiv:1608.00726*, 2016.
- [68] S. Fortunato, “Community detection in graphs,” *Physics reports*, vol. 486, no. 3, pp. 75–174, 2010.
- [69] J. Furness and M. Kolberg, “Considering complex search techniques in dhts under churn,” in *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*. IEEE, 2011, pp. 559–564.
- [70] R. Gaeta, G. Balbo, S. Bruell, M. Gribaudo, and M. Sereno, “A simple analytical framework to analyze search strategies in large-scale peer-to-peer networks,” *Performance Evaluation*, vol. 62, no. 1, pp. 1–16, 2005.
- [71] R. Gaeta and M. Sereno, “Random graphs as models of hierarchical peer-to-peer networks,” *Performance Evaluation*, vol. 64, no. 9, pp. 838–855, 2007.
- [72] R. Gaeta and M. Sereno, “On the evaluation of flooding-based search strategies in peer-to-peer networks,” *Concurrency and Computation: Practice and Experience*, vol. 20, no. 6, pp. 713–734, 2008.
- [73] R. Gaeta and M. Sereno, “Generalized probabilistic flooding in unstructured peer-to-peer networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 12, pp. 2055–2062, 2011.

- [74] A. J. Ganesh, A.-M. Kermarrec, and L. Massoulié, “Peer-to-peer membership management for gossip-based protocols,” *Computers, IEEE Transactions on*, vol. 52, no. 2, pp. 139–149, 2003.
- [75] E. N. Gilbert, “Random graphs,” *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.
- [76] C. Gkantsidis, M. Mihail, and A. Saberi, “Random walks in peer-to-peer networks,” in *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 1. IEEE, 2004.
- [77] C. Gkantsidis, M. Mihail, and A. Saberi, “Hybrid search schemes for unstructured peer-to-peer networks,” in *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 3. IEEE, 2005, pp. 1526–1537.
- [78] J. Gray, P. Helland, P. O’Neil, and D. Shasha, “The dangers of replication and a solution,” in *ACM SIGMOD Record*, vol. 25, no. 2. ACM, 1996, pp. 173–182.
- [79] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, “Optimal network topologies for local search with congestion,” *Physical Review Letters*, vol. 89, no. 24, p. 248701, 2002.
- [80] H. Han, J. He, and C. Zuo, “A hybrid p2p overlay network for high efficient search,” in *Information and Financial Engineering (ICIFE), 2010 2nd IEEE International Conference on*. IEEE, 2010, pp. 241–245.
- [81] L. S. Heath and N. Parikh, “Generating random graphs with tunable clustering coefficients,” *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 23, pp. 4577–4587, 2011.
- [82] D. R. Himali and S. K. Prasad, “Spun: A p2p probabilistic search algorithm based on successful paths in unstructured networks,” *2013 IEEE International Symposium on Parallel and Distributed Processing, Workshops and Phd Forum*, vol. 0, pp. 1610–1617, 2011.
- [83] P. Holme and B. J. Kim, “Growing scale-free networks with tunable clustering,” *Physical review E*, vol. 65, no. 2, p. 026107, 2002.

- [84] P. Holme and J. Saramäki, “Temporal networks,” *Physics reports*, vol. 519, no. 3, pp. 97–125, 2012.
- [85] C. Hota, V. Nunia, M. Di Francesco, J. K. Nurminen, and A. Ylä-Jääski, “Enhanced search in unstructured peer-to-peer overlay networks,” in *Grid and Pervasive Computing*. Springer, 2013, pp. 270–279.
- [86] H.-C. Hsiao and H.-W. Su, “On optimizing overlay topologies for search in unstructured peer-to-peer networks,” *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, no. 5, pp. 924–935, 2012.
- [87] B. Q. Huang, M. T. Kechadi, and B. Buckley, “Customer churn prediction for broadband internet services,” in *International Conference on Data Warehousing and Knowledge Discovery*. Springer, 2009, pp. 229–243.
- [88] B. A. Huberman and L. A. Adamic, “Internet: growth dynamics of the world-wide web,” *Nature*, vol. 401, no. 6749, pp. 131–131, 1999.
- [89] M. O. Jackson *et al.*, *Social and economic networks*. Princeton university press Princeton, 2008, vol. 3.
- [90] M. Jelasity, A. Montresor, and O. Babaoglu, “Gossip-based aggregation in large dynamic networks,” *ACM Transactions on Computer Systems (TOCS)*, vol. 23, no. 3, pp. 219–252, 2005.
- [91] S. Jiang, L. Guo, and X. Zhang, “Lightflood: an efficient flooding scheme for file search in unstructured peer-to-peer systems,” in *Parallel Processing, 2003. Proceedings. 2003 International Conference on*. IEEE, 2003, pp. 627–635.
- [92] S. Jin and H. Jiang, “Novel approaches to efficient flooding search in peer-to-peer networks,” *Computer Networks*, vol. 51, no. 10, pp. 2818–2832, 2007.
- [93] R. Johnsonbaugh and M. Kalin, “A graph generation software package,” in *ACM SIGCSE Bulletin*, vol. 23, no. 1. ACM, 1991, pp. 151–154.
- [94] K. V. Jónsson, Y. Vigfússon, and O. Helgason, “Simulating large-scale dynamic random graphs in omnet++,” in *Proc. 5th International ICST Conference on Simulation Tools and Techniques*, ser. SIMUTOOLS ’12. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2012, pp. 311–318.

- [95] V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti, “A local search mechanism for peer-to-peer networks,” in *Proceedings of the eleventh international conference on Information and knowledge management*. ACM, 2002, pp. 300–307.
- [96] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, “Social media? get serious! understanding the functional building blocks of social media,” *Business horizons*, vol. 54, no. 3, pp. 241–251, 2011.
- [97] J. Kleinberg, “The small-world phenomenon: An algorithmic perspective,” in *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. ACM, 2000, pp. 163–170.
- [98] J. Kleinberg, “The small-world phenomenon: An algorithmic perspective,” in *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. ACM, 2000, pp. 163–170.
- [99] J. Kleinberg, “Complex networks and decentralized search algorithms,” in *Proceedings of the International Congress of Mathematicians (ICM)*, vol. 3, 2006, pp. 1019–1044.
- [100] J. Kleinberg and P. Raghavan, “Query incentive networks,” in *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*. IEEE, 2005, pp. 132–141.
- [101] J. M. Kleinberg, “Navigation in a small world,” *Nature*, vol. 406, no. 6798, pp. 845–845, 2000.
- [102] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, “The web as a graph: measurements, models, and methods,” in *Computing and combinatorics*. Springer, 1999, pp. 1–17.
- [103] A. Klemm, C. Lindemann, M. K. Vernon, and O. P. Waldhorst, “Characterizing the query behavior in peer-to-peer file sharing systems,” in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. ACM, 2004, pp. 55–67.
- [104] V. A. Korhonen and J. Kurhinen, “Logistic model for modeling mobile encounter network,” in *2007 ITI 5th International Conference on Information and Communications Technology*. IEEE, 2007, pp. 151–155.

- [105] P. L. Krapivsky and S. Redner, “Organization of growing random networks,” *Physical Review E*, vol. 63, no. 6, p. 066123, 2001.
- [106] F. Kuhn, N. Lynch, and R. Oshman, “Distributed computation in dynamic networks,” in *Proceedings of the forty-second ACM symposium on Theory of computing*. ACM, 2010, pp. 513–522.
- [107] A. Kumar, J. Xu, and E. W. Zegura, “Efficient and scalable query routing for unstructured peer-to-peer networks,” in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 2. IEEE, 2005, pp. 1162–1173.
- [108] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, “Stochastic models for the web graph,” in *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*. IEEE, 2000, pp. 57–65.
- [109] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tompkins, and E. Upfal, “The web as a graph,” in *Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2000, pp. 1–10.
- [110] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, “Extracting large-scale knowledge bases from the web,” in *VLDB*, vol. 99. Citeseer, 1999, pp. 639–650.
- [111] D. Leonard, Z. Yao, V. Rai, and D. Loguinov, “On lifetime-based node failure and stochastic resilience of decentralized peer-to-peer networks,” *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 644–656, 2007.
- [112] E. Leontiadis, V. V. Dimakopoulos, and E. Pitoura, “Creating and maintaining replicas in unstructured peer-to-peer systems,” in *Euro-Par 2006 Parallel Processing*. Springer, 2006, pp. 1015–1025.
- [113] J. Leskovec. (2011) Stanford network analysis project. [Online]. Available: <http://snap.stanford.edu/index.html>
- [114] J. Leskovec, J. Kleinberg, and C. Faloutsos, “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.* ACM, 2005, pp. 177–187.
- [115] J. Leskovec, J. Kleinberg, and C. Faloutsos, “Graph evolution: Densification and shrinking diameters,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, p. 2, 2007.
 - [116] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection,” <http://snap.stanford.edu/data>, Jun. 2014.
 - [117] J. Leskovec, K. J. Lang, and M. Mahoney, “Empirical comparison of algorithms for network community detection,” in *Proceedings of the 19th international conference on World wide web.* ACM, 2010, pp. 631–640.
 - [118] J. Leskovec, M. McGlohon, C. Faloutsos, N. S. Glance, and M. Hurst, “Patterns of cascading behavior in large blog graphs,” in *SDM*, vol. 7. SIAM, 2007, pp. 551–556.
 - [119] J. Leskovec, *Dynamics of large networks.* ProQuest, 2008.
 - [120] Y. Liu, N. Xiong, L. Zhu, J. H. Park, and J. Gao, “An effective simulation method for search strategy in unstructured p2p network,” *Simulation Modelling Practice and Theory*, vol. 18, no. 4, pp. 456–469, 2010.
 - [121] Y. Liu, L. Xiao, A.-H. Esfahanian, and L. M. Ni, “Approaching optimal peer-to-peer overlays,” in *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2005. 13th IEEE International Symposium on.* IEEE, 2005, pp. 407–414.
 - [122] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, “Search and replication in unstructured peer-to-peer networks,” in *Proceedings of the 16th international conference on Supercomputing.* ACM, 2002, pp. 84–95.
 - [123] A. S. Maiya and T. Y. Berger-Wolf, “Expansion and decentralized search in complex networks,” *Knowledge and information systems*, vol. 38, no. 2, pp. 469–490, 2014.
 - [124] N. Makino, S. Arakawa, and M. Murata, “A flooding method for exchanging routing information in power-law networks,” in *2005 Asia-Pacific Conference on Communications.* IEEE, 2005, pp. 812–816.

- [125] D. Malkhi, M. Naor, and D. Ratajczak, “Viceroy: A scalable and dynamic emulation of the butterfly,” in *Proceedings of the twenty-first annual symposium on Principles of distributed computing*. ACM, 2002, pp. 183–192.
- [126] F. D. Malliaros, M.-E. G. Rossi, and M. Vazirgiannis, “Locating influential nodes in complex networks,” *Scientific reports*, vol. 6, p. 19307, 2016.
- [127] P. Marendy *et al.*, “A review of world wide web searching techniques, focusing on hits and related algorithms that utilise the link topology of the world wide web to provide the basis for a structure based search technology,” 2001.
- [128] S. V. Margariti and V. V. Dimakopoulos, “A novel probabilistic flooding strategy for unstructured peer-to-peer networks,” in *Informatics (PCI), 2011 15th Panhellenic Conference on*. IEEE, 2011, pp. 149–153.
- [129] S. V. Margariti and V. V. Dimakopoulos, “A study on the redundancy of flooding in unstructured p2p networks,” *International Journal of Parallel, Emergent and Distributed Systems*, vol. 28, no. 3, pp. 214–229, 2013.
- [130] S. V. Margariti and V. V. Dimakopoulos, “On probabilistic flooding search over unstructured peer-to-peer networks,” *Peer-to-Peer Networking and Applications*, vol. 8, no. 3, pp. 447–458, 2015.
- [131] S. V. Margariti and V. V. Dimakopoulos, “Simulating search protocols in large-scale dynamic networks,” in *Parallel, Distributed, and Network-Based Processing (PDP), 2016 24th Euromicro International Conference on*. IEEE, 2016, pp. 176–183.
- [132] L. Massoulié, E. Le Merrer, A.-M. Kermarrec, and A. Ganesh, “Peer counting and sampling in overlay networks: random walk methods,” in *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*. ACM, 2006, pp. 123–132.
- [133] D. Maxwell, L. Azzopardi, K. Järvelin, and H. Keskustalo, “Searching and stopping: An analysis of stopping rules and strategies,” in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 313–322.

- [134] P. Maymounkov and D. Mazieres, “Kademlia: A peer-to-peer information system based on the xor metric,” in *Peer-to-Peer Systems*. Springer, 2002, pp. 53–65.
- [135] M. Mayo, A. Abdelzaher, and P. Ghosh, “Long-range degree correlations in complex networks,” *Computational Social Networks*, vol. 2, no. 1, p. 4, 2015.
- [136] S. Mccanne, S. Floyd, and K. Fall, “NS2 (Network Simulator 2).” [Online]. Available: <http://www-nrg.ee.lbl.gov/ns>
- [137] E. Meshkova, J. Riihijärvi, M. Petrova, and P. Mähönen, “A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks,” *Computer networks*, vol. 52, no. 11, pp. 2097–2128, 2008.
- [138] S. Milgram, “The small world problem,” *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.
- [139] M. Mitzenmacher, “A brief history of generative models for power law and lognormal distributions,” *Internet mathematics*, vol. 1, no. 2, pp. 226–251, 2004.
- [140] M. Molloy and B. Reed, “A critical point for random graphs with a given degree sequence,” *Random structures & algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.
- [141] A. Montresor and M. Jelasity, “PeerSim: A scalable P2P simulator,” in *Proc. of the 9th Int. Conference on Peer-to-Peer (P2P’09)*, Seattle, WA, Sep. 2009, pp. 99–100.
- [142] S. Naicken, A. Basu, B. Livingston, S. Rodhetbhai, and I. Wakeman, “Towards yet another peer-to-peer simulator,” in *Proc. 4th International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs)*, Ilkley, UK, August 2006.
- [143] S. Naicken, B. Livingston, A. Basu, S. Rodhetbhai, I. Wakeman, and D. Chalmers, “The state of peer-to-peer simulators and simulations,” *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 2, pp. 95–98, Mar. 2007.
- [144] M. E. J. Newman, “Random Graphs with Clustering,” *Physical Review Letters*, vol. 103, no. 5, p. 058701, Jul. 2009.

- [145] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.
- [146] M. E. Newman, “Random graphs as models of networks,” *arXiv preprint cond-mat/0202208*, 2002.
- [147] M. E. Newman, “The structure and function of complex networks,” *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.
- [148] M. E. Newman, “Power laws, pareto distributions and zipf’s law,” *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.
- [149] M. E. Newman, S. H. Strogatz, and D. J. Watts, “Random graphs with arbitrary degree distributions and their applications,” *Physical review E*, vol. 64, no. 2, p. 026118, 2001.
- [150] M. E. Newman and D. J. Watts, “Renormalization group analysis of the small-world network model,” *Physics Letters A*, vol. 263, no. 4, pp. 341–346, 1999.
- [151] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu, “The broadcast storm problem in a mobile ad hoc network,” in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, ser. MobiCom ’99. New York, NY, USA: ACM, 1999, pp. 151–162. [Online]. Available: <http://doi.acm.org/10.1145/313451.313525>
- [152] J. D. Noh and H. Rieger, “Random walks on complex networks,” *Physical review letters*, vol. 92, no. 11, p. 118701, 2004.
- [153] K. Oikonomou, D. Kogias, and I. Stavrakakis, “Probabilistic flooding for efficient information dissemination in random graph topologies,” *Computer Networks*, vol. 54, no. 10, pp. 1615–1629, 2010.
- [154] Y. Oyama, H. Sasaki, H. Iwasaki, and Y. Nishiura, “Data dissemination dynamics in mobile p2p network,” *Modelling, Simulation, and Optimization–2005*, pp. 471–802, 2005.
- [155] C. Papadakis, P. Fragopoulou, E. P. Markatos, E. Athanasopoulos, M. Dikaiakos, and A. Labrinidis, “A feedback-based approach to reduce duplicate messages in unstructured peer-to-peer networks,” in *Integrated Research in GRID Computing*. Springer, 2007, pp. 103–118.

- [156] H. Park, R. RatzinM, and M. van der Schaar, “Peer-to-peer networks-protocols, cooperation and competition,” *Streaming Media Architectures, Techniques, and Applications: Recent Advances*, pp. 262–294, 2010.
- [157] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of modern physics*, vol. 87, no. 3, p. 925, 2015.
- [158] R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, 2007.
- [159] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, “Random walks and search in time-varying networks,” *Physical review letters*, vol. 109, no. 23, p. 238701, 2012.
- [160] K. Pussep, C. Leng, and S. Kaune, “Modeling user behavior in p2p systems,” in *Modeling and Tools for Network Simulation*. Springer, 2010, pp. 447–461.
- [161] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, “A scalable content-addressable network,” *SIGCOMM Comput. Commun. Rev.*, vol. 31, no. 4, pp. 161–172, Aug. 2001.
- [162] S. Rhea, D. Geels, T. Roscoe, and J. Kubiawicz, “Handling churn in a dht,” in *Proc. Annual Conference on USENIX Annual Technical Conference*, ser. ATEC ’04. Berkeley, CA, USA: USENIX Association, 2004, pp. 10–10.
- [163] S. C. Rhea and J. Kubiawicz, “Probabilistic location and routing,” in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2002, pp. 1248–1257.
- [164] M. Ripeanu and I. Foster, “Mapping the gnutella network: Macroscopic properties of large-scale peer-to-peer systems,” in *International Workshop on Peer-to-Peer Systems*. Springer, 2002, pp. 85–93.
- [165] C. Rohrs, “Query routing for the gnutella network,” *lime Wire LLC*, <http://rfc-gnutella.sourceforge.net/src/qrp.html>, 2001.

- [166] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, “Networks and cities: An information perspective,” *Physical Review Letters*, vol. 94, no. 2, p. 028701, 2005.
- [167] A. Rowstron and P. Druschel, “Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems,” in *Middleware 2001*. Springer, 2001, pp. 329–350.
- [168] A. D. Sarma, A. R. Molla, and G. Pandurangan, “Distributed computation in dynamic networks via random walks,” *Theoretical Computer Science*, vol. 581, pp. 45–66, 2015.
- [169] M. A. Serrano and M. Boguná, “Tuning clustering in random networks with arbitrary degree distributions,” *Physical Review E*, vol. 72, no. 3, p. 036133, 2005.
- [170] K. Shanmugam, K. Subburathinam, and A. Velayuthampalayam Palanisamy, “A dynamic probabilistic based broadcasting scheme for manets,” *The Scientific World Journal*, vol. 2016, 2016.
- [171] C. Shirky, “Power laws, weblogs, and inequality,” *Clay Shirky’s writings about the Internet*, vol. 8, 2003.
- [172] M. F. Shlesinger, “Mathematical physics: Search research,” *Nature*, vol. 443, no. 7109, pp. 281–282, 2006.
- [173] H. A. Simon, “On a class of skew distribution functions,” *Biometrika*, vol. 42, no. 3/4, pp. 425–440, 1955.
- [174] A. Singla and C. Rohrs, “Ultrapeers: Another step towards gnutella scalability,” 2002.
- [175] E. Spaho, A. Barolli, F. Xhafa, and L. Barolli, “P2p data replication: Techniques and applications,” in *Modeling and Processing for Next-Generation Big-Data Technologies*. Springer, 2015, pp. 145–166.
- [176] K. Sripanidkulchai, B. Maggs, and H. Zhang, “Efficient content location using interest-based locality in peer-to-peer systems,” in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3. IEEE, 2003, pp. 2166–2176.

- [177] A. O. Stauffer and V. C. Barbosa, “Probabilistic heuristics for disseminating information in networks,” *IEEE/ACM Transactions on Networking (TON)*, vol. 15, no. 2, pp. 425–435, 2007.
- [178] M. v. Steen, *Graph theory and complex networks : an introduction*. Lexington: Maarten van Steen, 2010. [Online]. Available: <http://opac.inria.fr/record=b1130915>
- [179] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A scalable peer-to-peer lookup service for internet applications,” *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 4, pp. 149–160, 2001.
- [180] D. Stutzbach and R. Rejaie, “Understanding churn in peer-to-peer networks,” in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*. ACM, 2006, pp. 189–202.
- [181] S. Tewari and L. Kleinrock, “Analysis of search and replication in unstructured peer-to-peer networks,” in *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1. ACM, 2005, pp. 404–405.
- [182] H. Thadakamalla, R. Albert, and S. Kumara, “Search in spatial scale-free networks,” *New Journal of Physics*, vol. 9, no. 6, p. 190, 2007.
- [183] S. M. Thampi *et al.*, “Survey of search and replication schemes in unstructured p2p networks,” *arXiv preprint arXiv:1008.1629*, 2010.
- [184] D. Tsoumakos and N. Roussopoulos, “Adaptive probabilistic search for peer-to-peer networks,” in *Peer-to-Peer Computing, 2003.(P2P 2003). Proceedings. Third International Conference on*. IEEE, 2003, pp. 102–109.
- [185] D. Tsoumakos and N. Roussopoulos, “A comparison of peer-to-peer search methods,” in *WebDB*. Citeseer, 2003, pp. 61–66.
- [186] R. Van De Bovenkamp, F. Kuipers, and P. Van Mieghem, “Gossip-based counting in dynamic networks,” in *International Conference on Research in Networking*. Springer, 2012, pp. 404–417.
- [187] K. Vanthournout, G. Deconinck, and R. Belmans, “A taxonomy for resource discovery,” *Personal and Ubiquitous Computing*, vol. 9, no. 2, pp. 81–89, 2005.

- [188] A. Varga and R. Hornig, “An overview of the omnet++ simulation environment,” in *Proc. 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops*, ser. Simutools ’08. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 60:1–60:10.
- [189] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, “Large-scale topological and dynamical properties of the internet,” *Physical Review E*, vol. 65, no. 6, p. 066130, 2002.
- [190] A. Vespignani, “Modelling dynamical processes in complex socio-technical systems,” *Nature Physics*, vol. 8, no. 1, pp. 32–39, 2012.
- [191] A. Vlachou, C. Doulkeridis, and Y. Kotidis, “Metric-based similarity search in unstructured peer-to-peer systems,” in *Transactions on Large-Scale Data-and Knowledge-Centered Systems V*. Springer, 2012, pp. 28–48.
- [192] S. Vuong and J. Li, “Efa: an efficient content routing algorithm in large peer-to-peer overlay networks,” in *Peer-to-Peer Computing, 2003.(P2P 2003). Proceedings. Third International Conference on*. IEEE, 2003, pp. 216–217.
- [193] J. Wang, Y. Li, F. Gong, and W. Chen, “A new strategy of resource searching in unstructured p2p network,” in *Communication Software and Networks, 2010. ICCSN’10. Second International Conference on*. IEEE, 2010, pp. 32–36.
- [194] L. Wang, S. Bayhan, J. Ott, J. Kangasharju, A. Sathiaselalan, and J. Crowcroft, “Pro-diluvian: Understanding scoped-flooding for content discovery in information-centric networking,” in *Proceedings of the 2nd International Conference on Information-Centric Networking*. ACM, 2015, pp. 9–18.
- [195] D. J. Watts, “Networks, dynamics, and the small-world phenomenon 1,” *American Journal of sociology*, vol. 105, no. 2, pp. 493–527, 1999.
- [196] D. J. Watts, P. S. Dodds, and M. E. Newman, “Identity and search in social networks,” *science*, vol. 296, no. 5571, pp. 1302–1305, 2002.
- [197] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

- [198] T. White, *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., 2012.
- [199] B. Yang and H. Garcia-Molina, "Improving search in peer-to-peer networks," in *Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on*. IEEE, 2002, pp. 5–14.
- [200] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *Data mining (ICDM), 2013 IEEE 13th international conference on*. IEEE, 2013, pp. 1151–1156.
- [201] S.-J. Yang, "Exploring complex networks by walking on them," *Physical Review E*, vol. 71, no. 1, p. 016107, 2005.
- [202] Z. Yao, D. B. Cline, X. Wang, and D. Loguinov, "Unifying models of churn and resilience for unstructured p2p graphs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 9, pp. 2475–2485, 2014.
- [203] H. Zhang, L. Zhang, X. Shan, and V. O. Li, "Probabilistic search in p2p networks with high node degree variation." in *ICC*, vol. 7, 2007, pp. 1710–1715.
- [204] R. Zhang and Y. C. Hu, "Assisted peer-to-peer search with partial indexing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 8, pp. 1146–1158, 2007.
- [205] Z. Zhang, T. Shan, and G. Chen, "Random walks on weighted networks," *Physical Review E*, vol. 87, no. 1, p. 012112, 2013.
- [206] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. D. Kubiatowicz, "Tapestry: A resilient global-scale overlay for service deployment," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 1, pp. 41–53, 2004.
- [207] K. Zhao, Z. Niu, Y. Zhao, and J. Yang, "Group-based search in unstructured peer-to-peer networks," in *Global Telecommunications Conference, 2009. GLOBE-COM 2009. IEEE*. IEEE, 2009, pp. 1–6.
- [208] Z. Zhu, P. Kalnis, and S. Bakiras, "Dcmp: A distributed cycle minimization protocol for peer-to-peer networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 3, pp. 363–377, 2008.

- [209] Z. Zhuang, Y. Liu, L. Xiao, and L. M. Ni, “Hybrid periodical flooding in unstructured peer-to-peer networks,” in *Parallel Processing, 2003. Proceedings. 2003 International Conference on*. IEEE, 2003, pp. 171–178.

ΔΗΜΟΣΙΕΥΣΕΙΣ ΣΥΓΓΡΑΦΕΑ

1. S. V. Margariti and V. V. Dimakopoulos, “Simulating search protocols in large-scale dynamic networks,” in *Parallel, Distributed, and Network-Based Processing (PDP), 2016 24th Euromicro International Conference on*. IEEE, 2016, pp. 176–183.
2. S. V. Margariti and V. V. Dimakopoulos, “On probabilistic flooding search over unstructured peer-to-peer networks,” *Peer-to-Peer Networking and Applications*, vol. 8, no. 3, pp. 447–458, 2015.
3. S. V. Margariti and V. V. Dimakopoulos, “A study on the redundancy of flooding in unstructured p2p networks,” *International Journal of Parallel, Emergent and Distributed Systems*, vol. 28, no. 3, pp. 214–229, 2013.
4. S. V. Margariti and V. V. Dimakopoulos, “A novel probabilistic flooding strategy for unstructured peer-to-peer networks,” in *Informatics (PCI), 2011 15th Panhellenic Conference on*. IEEE, 2011, pp. 149–153.
5. V. V. Dimakopoulos, S. Margariti, M. Ntetsika, and E. Pitoura, “Data replication in p2p systems,” in *Handbook of Research on P2P and Grid Systems for Service-Oriented Computing: Models, Methodologies and Applications: Models, Methodologies and Applications*. IGI Global, 2010, pp. 589–615.

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Μαργαρίτη Σπυριδούλα γεννήθηκε το 1970 στην Άρτα. Σπούδασε Ηλεκτρονική στο Τμήμα Ηλεκτρονικής του Τ.Ε.Ι. Λαμίας (1992) και έκανε Παιδαγωγικές σπουδές (2000) στο Τμήμα ΠΑΤΕΣ/ΣΕΛΕΤΕ Ιωαννίνων. Απέκτησε πτυχίο και μεταπτυχιακό δίπλωμα ειδίκευσης από το Τμήμα Πληροφορικής του Πανεπιστημίου Ιωαννίνων, το 2002 και 2007, αντίστοιχα. Σήμερα είναι υποψήφια διδάκτορας στο Τμήμα Μηχανικών Η/Υ και Πληροφορικής του ίδιου πανεπιστημίου. Από το 1994 εργάστηκε σε διάφορες θέσεις και τμήματα του Τ.Ε.Ι. Ηπείρου. Σήμερα είναι μέλος Εργαστηριακού Διδακτικού Προσωπικού (Ε.ΔΙ.Π.) στο Τμήμα Μηχανικών Πληροφορικής του Τ.Ε.Ι. Ηπείρου.