ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΤΜΗΜΑ ΗΛΕΚΤΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής

Μελέτη Διάχυσης Πληροφορίας σε Γενικευμένα Δίκτυα με Χρονομεταβλητή Συμπεριφορά Χρηστών και Χρήση Μετρικών Κοινωνικής Ανάλυσης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Αδαμαντίας Χλέτσου

Επιβλέπων: Συμεών Χρ. Παπαβασιλείου

Καθηγητής, ΕΜΠ

Αθήνα, Ιούλιος 2015

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΤΜΗΜΑ ΗΛΕΚΤΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής

Μελέτη Διάχυσης Πληροφορίας σε Γενικευμένα Δίκτυα με Χρονομεταβλητή Συμπεριφορά Χρηστών και Χρήση Μετρικών Κοινωνικής Ανάλυσης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Αδαμαντίας Χλέτσου

Επιβλέπων: Συμεών Χρ. Παπαβασιλείου

Καθηγητής, ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 24^{η} Ιουλίου 2015.

....	....	....
Συμεών Παπαβασιλείου	Γιώργος Ματσόπουλος	Ιωάννα Ρουσσάκη
Καθηγητής, ΕΜΠ	Καθηγητής, ΕΜΠ	Επικ. Καθηγήτρια, ΕΜΠ

Αθήνα, Ιούλιος 2015

....

Αδαμαντία Χλέτσου

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλεκτρονικών Υπολογιστών

Περίληψη

Με τον όρο γενικευμένα δίκτυα ορίζονται τα δίκτυα στα οποία η επικοινωνία των χρηστών μπορεί να γίνει είτε μέσω φυσικού στρώματος ,με τη χρήση ασύρματων δηλαδή δικτύων, ή μέσω κοινωνικών δικτύων. Τα γενικευμένα δίκτυα διαδραματίζουν σημαντικό ρόλο στον τρόπο επικοινωνίας σήμερα αφού οι συσκευές που έχουν αναπτυχθεί επιτρέπουν και τους δύο τρόπους επικοινωνίας μεταξύ των χρηστών.

Στην παρούσα εργασία γίνεται μελέτη της δυναμικής της διάχυσης της πληροφορίας σε γενικευμένα δίκτυα, λαμβάνοντας υπόψη τη συμπεριφορα των χρηστών. Στόχος της είναι να διαμορφωθεί ένα μοντέλο που να αποτυπώνει τη διάχυση της πληροφορίας στο γενικευμένο δίκτυο με βάση τα χρονομεταβλητά ενδιαφέροντα των χρηστών ως προς την πληροφορία αυτή. Με τη βοήθεια του προγράμματος Matlab γίνεται προσομοίωση της διάδοσης σε ένα δίκτυο που αποτελείται από χρήστες με διαφορετικά ενδιαφέροντα και με χρήση μετρικών κοινωνικής ανάλυσης. Παράλληλα, αναπτύσσεται και αναλύεται θεωρητικά ένα επιδημιολογικό μοντέλο διάχυσης της πληροφορίας, ώστε να υπολογιστεί ο αριθμός των χρηστών που ενημερώνονται για συγκεκριμένα θέματα. Τα αποτελέσματα από την προσομοίωση συγκρίνονται με τα αποτελέσματα από το θεωρητικό μοντέλο. Παρατηρείται ότι μέσω του ασύρματου δικτύου οι πληροφορίες διαδίδονται με πιο ομοιόμορφο τρόπο σε αντίθεση με τη διάδοση μέσω κοινωνικού επιπέδου. Τέλος, συνοψίζονται τα αποτελέσματα της διπλωματικής και δίνονται κατευθύνσεις για μελλοντική μελέτη.

Λέξεις Κλειδιά: γενικευμένα δίκτυα, σύνθετα δίκτυα, διάχυση πληροφορίας, επιδημιολογικό μοντέλο, κόμβοι με χρονομεταβλητό ενδιαφέρον

Abstract

The term Generalized Networks is referred to the networks where the users can interact either through wireless networks or through social networks. Today, the most communication machines are developed so as to support user communication through both ways. For this reason, there is an increasing interest in the way in which information can be diffused in these networks.

This diploma thesis focuses on the study of the dynamics of information diffusion in generalized networks, taking under consideration the time-varying interests of the users in this information. The objective is to form an epidemic model describing the diffusion of information in a generalized network that consists of users with different interests on the information. With use of Matlab we develop a simulation of information diffusion in such a network. In addition we develop a mathematical model regarding the information diffusion process providing the number of users of a generalized network that get informed about topics of information and we obtain knowledge on how diverse topics of information can be spread throughout generalized networks by comparing the results of the simulation and the mathematical model. Finally, we sum up the results of the thesis and give directions for future study.

Key words: generalized networks, complex networks, information diffusion, epidemic model, time-varying interest of users

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω θερμά τον Καθηγητή κ. Συμεών Παπαβασιλείου που μου έδωσε τη δυνατότητα να πραγματοποιήσω την παρούσα διπλωματική εργασία υπό την επίβλεψη και καθοδήγησή του.

Στη συνέχεια, θα ήθελα να ευχαριστήσω τη Δρ. Έλενα Στάη για την αμέριστη βοήθεια της και τη συνεχή καθοδήγησή της καθόλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας.

Θα ήθελα ακόμα να ευχαριστήσω τον Δρ. Βασίλειο Καρυώτη για τις συμβουλές και το χρήσιμο υλικό που μου έδωσε.

Και τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου για την υποστήριξη και τη συμπαράσταση τους αυτά τα χρόνια φοίτησης μου στο ΕΜΠ.

Περιεχόμενα

Κατάλογος Σχημάτων

1 Κεφάλαιο

Εισαγωγή

1.1 Συμβολή

Στην παρούσα διπλωματική εργασία μελετάται η διάχυση της πληροφορίας σε γενικευμένα δίκτυα. Ο όρος γενικευμένα δίκτυα, ο οποίος αναλύεται και στο κεφάλαιο 3, αναφέρεται στα δίκτυα όπου η διάδοσή της πληροφορίας μπορεί να γίνει είτε μέσω φυσικού στρωμάτος (Wi-Fi/Bluetooth) ή μέσω κοινωνικού στρώματος (MMS). Με βάση τα αντίστοιχα μοντέλα στον τομέα της επιδημιολογίας δημιουργείται ένα μοντέλο για τη διάδοση της πληροφορίας από «μολυσμένους» σε «υγιείς» χρήστες λαμβάνοντας υπόψιν τα χαρακτηριστικά του δικτύου και των χρηστών που το απαρτίζουν. «Μολυσμένοι» χρήστες θεωρούνται εκείνοι που έχουν ενημερωθεί για κάποια πληροφορία, κάποιο θέμα, το οποίο αποτελεί και την «ασθένεια». Αντίστοιχα οι υγιείς χρήστες αποτελούνται από το κομμάτι του πληθυσμού που παραμένει ανενημέρωτο τη στιγμή που μελετάται το φαινόμενο. Για την ανάπτυξη του μοντέλου θεωρείται ότι οι χρήστες έχουν χρονομεταβλητά ενδιαφέροντα στη διάρκεια ενός έτους και διαφορετικά για κάθε πληροφορία. Επίσης υπολογίζονται μετρικές κοινωνικής δικτύωσης των χρηστών.

Τέλος το μοντέλο αυτό χρησιμοποιείται για την ανάλυση της διάχυση της πληροφορίας μέσα στο δίκτυο με τρεις διαφορετικούς τρόπους. Αρχικά μέσω μόνο φυσικού στρώματος, έπειτα μόνο μέσω κοινωνικού και τέλος και με τους δύο τρόπους, με ίση πιθανότητα.

1.2 Διάρθρωση

Η παρούσα διπλωματική εργασία ακολουθεί την εξής δομή :

Το Κεφάλαιο 2 αποτελεί μια εισαγωγή στη θεωρία γράφων (Graph Theory), όπου παρατίθενται τα βασικά χαρακτηριστικά αυτών. Γίνεται εισαγωγή στις έννοιες των γράφων και τα στοιχεία που τους απαρτίζουν. Περιγράφεται η διαδικασία με την οποία μελετώνται καθώς επίσης γίνεται αναφορά σε συγκεκριμένο είδος γραφημάτων, τα απλά, οι τυχαίοι γράφοι, τα μη κατευθυνόμενα και τα χωρίς βάρη. Τέλος αναφέρεται ο τρόπος αναπαράστασης των γραφημάτων, δηλαδή οι πίνακες γειτνίασης καθώς και άλλα χαρακτηριστικά τους όπως τα μονοπάτια, οι κλίκες, οι γειτονιές και οι κύκλοι.

Στο Κεφάλαιο 3 γίνεται μια εισαγωγή στα Σύνθετα Δίκτυα (Complex Networks). Περιγράφονται εκτενώς τα Ασύρματα Δίκτυα και τα Κοινωνικά και καταλήγει στα Γενικευμένα Δίκτυα. Επίσης γίνεται περιγραφή των δικτύων Μικρού Κόσμου (Small-World Networks), καθώς και των δικτύων Χωρίς Κλίμακα (Scale-free Networks). Γίνεται αναφορά επίσης σε εργασία σχετική με τη διάχυση της πληροφορίας σε γενικευμένα δίκτυα. Εφόσον η παρούσα εργασία επικεντρώνεται στη μελέτη της διάχυσης στα γενικευμένα δίκτυα είναι απαραίτητη η εισαγωγή σε αυτά τα δίκτυα και η κατανόηση ορισμένων στοιχείων τους.

Στο Κεφάλαιο 4 περιγράφεται η έννοια της διάχυσης της πληροφορίας στα δίκτυα και το επιδημιολογικό μοντέλο για την μελέτη της. Επιπρόσθετα περιγράφονται εκτενώς διάφορα είδη του επιδημιολογικού μοντέλου και δίνεται περισσότερο έμφαση στο επιδημιολογικό μοντέλο SIS το οποίο χρησιμοποιείται τελικά στο μοντέλο που αναπτύσσεται στη συνέχεια.

Ακολουθεί το Κεφάλαιο 5, που γίνεται η παρουσίαση του προβλήμτος στο οποίο επικεντρώνεται η παρούσα εργασία. Παρουσιάζεται το μαθηματικό μοντέλο που αναπτύχθηκε για τη μελέτη της διάδοσης της πληροφορίας στο δίκτυο με τους χρήστες με τα διαφορετικά και χρονομεταβλητά ενδιαφέροντα για κάθε πληροφορία και η επίλυσή του, ενώ πραγματοποιείται η αξιολόγηση του με βάση τα αποτελέσματα που συλλέχθηκαν μετά την προσομοίωση του μοντέλου αυτού.

Τέλος, στο Κεφάλαιο 6 γίνεται μια σύνοψη αυτών που παρουσιάστηκαν στα παραπάνω κεφάλαια, των αποτελεσμάτων που προέκυψαν καθώς και μελλοντικές επεκτάσεις της παρούσας εργασίας.

2 Κεφάλαιο

Γράφοι

2.1 Ορισμός Γράφου

Για τη μελέτη και ανάλυση της δυναμικής συμπεριφοράς των δικτύων (dynamic network behavior) θα χρησιμοποιηθούν στοιχεία και σχέσεις από τη Θεωρία Γραφημάτων. Η Θεωρία Γράφων (graph theory) είναι ένα γνωστικό πεδίο των διακριτών μαθηματικών, με εφαρμογές στην πληροφορική, στις επιστήμες μηχανικών, στη χημεία, στην κοινωνιολογία, στη μηχανική, στις οικονομικές και ανθρωπιστικές επιστήμες. Η θεωρία γράφων, εν γένει, μελετά τις σχέσεις αντικείμενων τα οποία έχουν κάποια συνάφεια μεταξύ τους ή συνδέονται με κάποιο ορισμένο τρόπο [RD10].

Σύμφωνα με τη Θεωρία Γράφων, ένας γράφος ή γράφημα G = (V, E) αποτελείται από ένα σύνολο σημείων ή κορυφών (vertices) ή κόμβων(nodes) που συμβολίζονται με V και ένα σύνολο E συνδέσμων μεταξύ των στοιχείων του V. Οι συνδέσεις που υφίστανται ανάμεσα στα ζεύγη κόμβους ονομάζονται ακμές (edges) του γράφου και συμβολίζοναι με E. Επομένως, ένα γράφημα είναι ένα διατεταγμένο ζεύγος G = (V, E) αποτελούμενο από ένα σύνολο V των κορυφών ή κόμβων μαζί με ένα E σύνολο από ακμές ή γραμμές, οι οποίες είναι υποσύνολα δύο στοιχείων V , δηλαδή, μια ακμή σχετίζεται με δύο κορυφές και η σχέση απεικονίζεται ως διατεταγμένο ζεύγος των κορυφών σε σχέση με τη συγκεκριμένη ακμή. Οι κορυφές που ανήκουν σε μια ακμή ονομάζονται άκρα της ακμής. Μια κορυφή μπορεί να υπάρχει σε ένα γράφημα και να μην ανήκει σε ακμή [D10].

Η τάξη ενός γραφήματος είναι |V| (ο αριθμός των κορυφών). Το μέγεθος ενός γραφήματος είναι |E|, δηλαδή ο αριθμός των ακμών. Κάθε κόμβος του γράφου χαρακτηρίζεται από το βαθμό του, ο οποίος ορίζεται ίσος με τον αριθμό των ακμών που συνδέονται με αυτόν, όπου μια ακμή η οποία συνδέεται με την κορυφή και στα δύο άκρα (ένας βρόχος) συνυπολογίζεται δύο φορές (κεφάλαιο 2 της αναφοράς [SKP13]).

Οι γράφοι βρίσκουν εφαρμογή στη μοντελοποίηση πολλών προβλημάτων όπως στα Τηλεπικοινωνιακά και Οδικά Δίκτυα, Ηλεκτρονικά Κυκλώματα, Αεροπορικές Πτήσεις κλπ.

2.2 Μη κατευθυνόμενοι και χωρίς βάρη γράφοι

Ένας γράφος είναι κατευθυνόμενος ή μη κατευθυνόμενος, αν οι ακμές του είναι ή δεν είναι προσανατολισμένες προς μία κατεύθυνση, αντίστοιχα (κεφάλαιο 2 της αναφοράς [SKP13]). Στον κατευθυνόμενο γράφο ο προσανατολισμός της ακμής συμβολίζεται με ένα βέλος. Στον μη κατευθυνόμενο γράφο τα ζεύγη των κορυφών που ορίζουν τις ακμές του στερούνται διάταξης. Για παράδειγμα, μεταξύ δύο κορυφών u και v τα βέλη από τη u στη v και από τη v στη u, θεωρούνται μία ακμή και όχι δύο. Στους κατευθυνόμενους γράφους θεωρούνται δύο ξεχωριστές ακμές ανάλογα με την κατεύθυνση τους. Για κατεύθυνση από τη u στη v, το u είναι η ουρά (tail) και το v η κεφαλή (head) και αντίστροφα για κατεύθυνση από τη v στη u. Από τα παραπάνω προκύπτει ότι ένας μη κατευθυνόμενος γράφος μπορεί να θεωρηθεί και ως συμμετρικός κατευθυνόμενος γράφος (δηλαδή για κάθε ακμή να υπάρχει και η αντίρροπή της) (κεφάλαιο 2 της αναφοράς [SKP13]).

Ο μέγιστος αριθμός ακμών για ένα μη κατευθυνόμενο γράφο με n κορυφές είναι n(n-1)/2, ενώ ο αντίστοιχος αριθμός σε κατευθυνόμενο γράφο είναι n(n-1). Ένας γράφος (κατευθυνόμενος ή μη), με τον μέγιστο αριθμό ακμών λέγεται πλήρης.

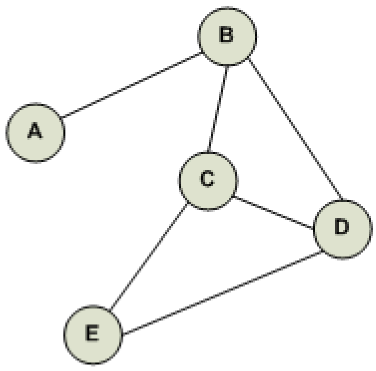
Ο μη κατευθυνόμενος γράφος είναι μια ειδική περίπτωση του κυρίου ορισμού του γράφου.

Ως μαθηματική έκφραση, ο ορισμός του μη κατευθυνόμενου γράφου έχει ως εξής:

Ο γράφος *G* είναι ένα διατεταγμένο ζεύγος *G=<V(G), E(G)>* όπου:

- $V(G)=\{v1,v2...v_n\}=n$, το πεπερασμένο και μη κενό σύνολο των κορυφών,
- $E(G)=\{e1,e2...e_m\}=m$, το μη διατεταγμένο σύνολο των ακμών.

Στην προκειμένη περίπτωση κάθε ακμή είναι ένα διμελές σύνολο αποτελούμενο από δύο κορυφές, οι οποίες αποκαλούνται τερματικές κορυφές (κόμβοι) και δεν είναι απαραίτητα διαφορετικές μεταξύ τους. Μία ακμή x=(i,j) μπορεί να αναπαρασταθεί και ως x=ij και λέμε ότι οι κόμβοι i και j είναι γειτονικοί και ο κόμβος i και η ακμή x προσπίπτουν. Αν δύο διαφορετικές ακμές x και y προσπίπτουν σε κοινό κόμβο ονομάζονται γειτονικές ακμές. Με τον όρο γράφημα θα αναφερόμαστε σε ένα απλό, μη κατευθυνόμενο γράφο [CLR09].



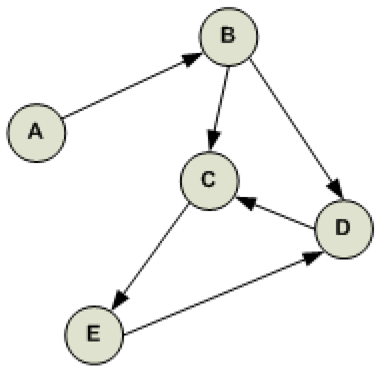


Fig 1. Undirected Graph **Fig 2. Directed Graph**

Σχήμα 1: Fig1 μη κατευθυνόμενος γράφος και Fig2 κατευθυνόμενος γράφος

Ακόμα μια ειδική περίπτωση στο γενικό ορισμό του γράφου είναι ο μη σταθμισμένος γράφος (κεφάλαιο 2 της αναφοράς [SKP13]). Συγκεκριμένα ένας γράφος G λέγεται μη σταθμισμένος αν στις ακμές του δεν έχει οριστεί μια συνάρτηση κόστους ή βάρους w: E → R. Η συνάρτηση βάρους w μπορεί να μοντελοποιεί κόστος, απόσταση, χρόνο ή το ενδιαφέρον στα κοινωνικά δίκτυα και το πεδίο τιμών της είναι το R^+ ή το N.
Δίκτυο ή σταθμισμένος γράφος λέγεται το γράφημα όπου η κάθε ακμή του χαρακτηρίζεται από κάποιο αριθμό (έχει κάποια τιμή) που ονομάζεται βάρος ή βαρύτητα της ακμής. Υποθέτουμε ότι ο G(V,E) είναι γράφος. Τότε κάθε γράφος με τη συνάρτηση βάρους του w:E → N καλείται γράφος με βάρη.

Ακόμα ένας ορισμός για τους σταθμισμένους γράφους είναι η δύναμη του κόμβου. Η δύναμη του κόμβου είναι ο συνολικός αριθμός του βάρους των ακμών που εισέρχονται ή εξέρχονται από έναν κόμβο. Έστω ότι με i συμβολίζεται ένας κόμβος τότε με s_i^{in} συμβολίζεται η δύναμη που εισέρχεται και s_i^{out} η δύναμη που εξέρχεται από τον κόμβο i (κεφάλαιο 4 της αναφοράς [SKP13]).

$$s_i^{out} = \sum_{j=1}^{N} w_{ij}. \quad (1)$$
$$s_i^{in} = \sum_{j=1}^{N} w_{ji}. \quad (2)$$

Στην περίπτωση μη σταθμισμένου γράφου ισχύει ότι η δύναμη που εισέρχεται στο γράφο θα είναι ίση με αυτή που εξέρχεται , δηλαδή:

$$s_i^{out} = s_i^{in} = s_i = \sum_{j=1}^{N} w_{ij}. \quad (3)$$

2.3 Τυχαίοι Γράφοι (Random Graphs)

Το μοντέλο του τυχαίου γράφου εισάγεται το 1959 από τους Erdős και Rényi. Η Θεωρία των Τυχαίων Γράφων διαφέρει από τη Θεωρία Γράφων στο ότι η πρώτη προκύπτει ως συνδυασμός εννοιών από την παραδοσιακή Θεωρία

Γράφων και της Θεωρίας Πιθανοτήτων. Η Θεωρία Τυχαίων Γράφων είναι το βασικό μαθηματικό εργαλείο για τη μελέτη των δικτύων επικοινωνιών τα οποία είναι και αυτά που παρουσιάζουν το μεγαλύτερο ενδιαφέρον στην Επιστήμη των Δικτύων (κεφάλαιο 2 της [SKP13]). Επίσης η Θεωρία Τυχαίων Γράφων προσφέρει ένα εναλλακτικό μοντέλο το οποίο είναι χρήσιμο για τη μελέτη δικτύων τα οποία ξεχωρίζουν για τη στοχαστική συμπεριφορά τους. Το μοντέλο αυτό επιτρέπει τη χρήση πολλών βαθμών τυχαιότητας στο σχηματισμό και εξέλιξη των δικτύων ,το οποίο δεν επιτρεπόταν με τη χρήση της απλής Θεωρίας Γράφων.
Ο τυχαίος γράφος έχει δύο παραμέτρους [SKP13]: μια παράμετρο n που ελέγχει το πλήθος των κόμβων στο γράφο και μια δεύτερη m που ελέγχει την πυκνότητα του γράφου δηλαδή το πλήθος των ακμών. Η κατανομή βαθμού των κόμβων είναι αυθαίρετη. Στην πιο κοινή εκδοχή ο τυχαίος γράφος G(n, m) αντιστοιχεί στο σύνολο των γράφων με n κόμβους όπου επιλέγονται m ακμές από το σύνολο των n(n-1)/2 και τοποθετούνται στο γράφο. Στην περίπτωση αυτή η κατανομή βαθμών μπορεί να προκύψει από ομοιόμορφη δειγματοληψία. Εναλλακτικά οι ακμές επιλέγονται με πιθανότητα p οπότε προκύπτει το μοντέλο G(n, p). Στην περίπτωση αυτή η κατανομή των βαθμών ακολουθεί την δυωνυμική κατανομή.

Το μοντέλο του τυχαίου γράφου είναι ιδιαίτερα δημοφιλές, χάρη της απλότητάς του και των ιδιοτήτων του. Χρησιμοποιείται ευρέως στη μοντελοποίηση συστημάτων όπως p2p δικτύων για να μελετηθεί η συμπεριφορά τους. Όμως το ερώτημα είναι κατά πόσο το παραγόμενο μοντέλο ταιριάζει στα δίκτυα του πραγματικού κόσμου; Αυτό εξαρτάται από τον τρόπο σχηματισμού ενός δικτύου, δηλαδή από τον τρόπο που ο κάθε κόμβος επιλέγει τους γείτονές του. Αν η επιλογή είναι τυχαία τότε τυχαίο θα είναι και το δίκτυο που θα προκύψει [SHS01].

Μερικά ακόμα βασικά μοντέλα τυχαίων γράφων τα οποία αναφέρονται και στο κεφάλαιο 2 της αναφοράς [SKP13] είναι τα παρακάτω:

- Το $G_{n,p}$ που αποτελείται από όλους τους γράφους με n κορυφές, όπου κάθε μια από τις $\binom{n}{2}$ δυνατές ακμές υπάρχει με πιθανότητα p και είναι στατιστικά ανεξάρτητες. Η πιθανότητα p=p(n) είναι μια συνάρτηση του αριθμού n των κορυφών . Το μοντέλο αυτό αναφέρεται συχνά ως πιθανοτικό μοντέλο (probabilistic model)
- Το $G_{n,M}$ που ονομάζεται και στατικό (static) , αποτελείται από όλους τους γράφους n κορυφών που έχουν M ακμές. Υπάρχουν $N = \binom{n}{2}$ δυνατές ακμές και επομένως $\binom{N}{M}$ δυνατοί γράφοι με M ακμές.

Θεωρώντας τους γράφους ισοπίθανους καθένας έχει πιθανότητα ίση με $\binom{N}{M}^{-1}$.

- Η διεργασία τυχαίων γράφων (random graph process) G^* , που είναι μια Μαρκοβιανή αλυσίδα (Markov chain) $G^* = \{Gt\}_0^\infty$ στο V={1,2,..,n} της οποίας οι καταστάσεις είναι γράφοι με σύνολο κορυφών V. Η διεργασία ξεκινάει από τον κενό γράφο και για κάθε $1 \leq t \leq \binom{n}{2}$ ο γράφος G_t προκύπτει από τον G_{t-1} προσθέτοντας μία ακμή , όπου όλες οι ακμές θεωρούνται ισοπίθανες. Το μοντέλο αυτό ονομάζεται επίσης δυναμικό (dynamic model).

  Σημειώνουμε ότι κανένα από τρία μοντέλα δεν θεωρείται περισσότερο αποδεκτό από τα υπόλοιπα. Απλά κάθε ένα έχει ορισμένα χαρακτηριστικά τα οποία ανάλογα με τη συγκεκριμένη περίπτωση το καθιστούν περισσότερο βολικό.

  Για παράδειγμα το μοντέλο $G_{n,p}$ θεωρείται περισσότερο φιλικό από τεχνική άποψη σε σχέση με το $G_{n,M}$. Αυτό οφείλεται στην πλήρη στατιστική ανεξαρτησία που χαρακτηρίζει το $G_{n,p}$. Αντίθετα το $G_{n,M}$ εμφανίζει μη επιθυμητά(από τεχνική άποψη) προβλήματα στατιστικών εξαρτήσεων λόγω της ύπαρξης συγκεκριμένου αριθμού ακμών. Πάντως συχνά είναι δυνατή η μετάβαση από το ένα μοντέλο στο άλλο. Για παράδειγμα τα μοντέλα $G_{n,p}$ και $G_{n,M}$ είναι πρακτικά ταυτόσημα όταν $M = p\binom{n}{2}$.

2.4 Πίνακας γειτνίασης (adjacency matrix), Συνδεσιμότητα(connectivity), Περίπατος και κύκλος(paths and cycles), Γειτονιά (neighborhood), Κλίκα (clique)

Στα μαθηματικά και στην επιστήμη υπολογιστών, ο πίνακας γειτνίασης χρησιμοποιείται για να αναπαραστήσει τις ακμές ενός γράφου. Ο πίνακας γειτνίασης A ενός γραφήματος G(V,E) είναι ένας τετραγωνικός πίνακας διάστασης |V|x|V| οι γραμμές και οι στήλες του οποίου αριθμούνται με βάση τις κορυφές του (κεφάλαιο 2 της αναφοράς [SKP13]). Τα στοιχεία του πίνακα γειτνίασης ορίζονται με βάση τις ακμές του γραφήματος από τη σχέση:

$$A[i,j] = \begin{Bmatrix} 1 \; αν \; \{u_i, v_j\} \in E \\ 0 \; διαφορετικά \end{Bmatrix}. \qquad (4)$$

όπου με i συμβολίζουμε τις γραμμές του πίνακα και με j τις στήλες του. Αν υπάρχει ακμή που ενώνει την κορυφή u_i με την κορυφή v_j τότε το στοιχείο a_{ij} του πίνακα γειτνίασης θα ισούται με 1, διαφορετικά με 0.

Είναι εύκολο να δούμε ότι για το ίδιο γράφημα μπορεί να προκύψουν διαφορετικοί πίνακες γειτνίασης αν χρησιμοποιηθεί διαφορετική αρίθμηση κορυφών.

Παρακάτω αναλύονται οι ιδιότητες του πίνακα γειτνίασης ενός απλού γράφου. Με τον όρο απλό γράφημα χαρακτηρίζεται ένα μη κατευθυνόμενο γράφημα που δεν έχει βρόχους και έχει το πολύ μια ακμή ανάμεσα σε δύο διαφορετικές κορυφές [CLR09]. Σε ένα απλό γράφημα οι ακμές του γραφήματος αποτελούν ένα σύνολο και κάθε ακμή είναι ένα ξεχωριστό ζευγάρι κορυφών. Σε ένα απλό γράφημα με n κορυφές κάθε κορυφή έχει ένα βαθμό που είναι μικρότερο από n (το αντίστροφο, όμως, δεν είναι αλήθεια - υπάρχουν και μη-απλά γραφήματα με n κορυφές στα οποία κάθε κορυφή έχει βαθμό μικρότερο από το n).

Οι βασικές ιδιότητες του πίνακα γειτνίασης του είναι [SKP13]:

1. Τα διαγώνια στοιχεία του είναι 0 γιατί δεν υπάρχουν ανακυκλώσεις και ο πίνακας είναι συμμετρικός ως προς τη διαγώνιο (οι ακμές δεν έχουν κατεύθυνση).
2. Το άθροισμα των στοιχείων της γραμμής ή της στήλης που αντιστοιχεί σε κάθε κορυφή v_i είναι ίσο με το βαθμό της κορυφής, δηλαδή

$$\sum_{v_j \in V} A[v_i, v_j] = \sum_{v_j \in V} A\left[v_j, v_i\right] = \deg(v_i) \; . \tag{5}$$

3. Το συνολικό άθροισμα των στοιχείων του πίνακα γειτνίασης είναι ίσο με το διπλάσιο του αριθμού των ακμών του γραφήματος , δηλαδή

$$\sum_{vi \in V} \sum_{vj \in V} A\left[v_i, v_j\right] = 2|E| \; . \tag{6}$$

Το σύνολο των ιδιοτιμών ενός γράφου είναι το φάσμα του γράφου.

Μονοπάτι από έναν κόμβο σε έναν άλλο ενός γραφήματος ονομάζεται μια ακολουθία συνεχόμενων κόμβων, όπου κάθε κόμβος της ακολουθίας συνδέεται με τον επόμενό του μέσω ακμής (κεφάλαιο 2 της αναφοράς [SKP13]). Δηλαδή είναι μια ακολουθία ακμών ($e_1,\ldots,e_k$) όπου για κάθε i , $1 \leq i \leq k-1$, το ένα άκρο (το τέλος για κατευθυνόμενο γράφημα) της ακμής e_i συμπίπτει με το άλλο άκρο (την αρχή) της ακμής e_{i+1}. Το μήκος ενός μονοπατιού, είναι το πλήθος των ακμών της ακολουθίας. Ένα μονοπάτι που αποτελείται από έναν μη επαναλαμβανόμενο κόμβο και δεν περιέχει καμία ακμή, είναι ένα τετριμμένο μονοπάτι μηδενικού μήκους. Ένα μονοπάτι είναι κατευθυνόμενο αν μπορούμε να πάμε από το ένα άκρο του στο άλλο και όχι το ανάποδο. Ενώ είναι μη κατευθυνόμενο όταν και οι δύο κατευθύνσεις του είναι δυνατές. Δύο μονοπάτια λέγονται ξένα ως προς τις ακμές, αν δεν έχουν καμία κοινή ακμή (παρότι μπορεί να τέμνονται).

Η απόσταση d(u,v) μεταξύ δύο κόμβων (u,v) ενός γράφου G είναι το μικρότερο μονοπάτι που ενώνει αυτούς τους δύο κόμβους (κεφάλαιο 2 της αναφοράς [SKP13]). Η απόσταση για κάθε κόμβο u,v που ενώνονται μεταξύ τους με κάποιο μονοπάτι έχει τις παρακάτω ιδιότητες:

- d(u,v)≥0 , με d(u,v)=0 αν και μόνο αν u=v
- d(u,v)=d(v,u)

Ένα μονοπάτι λέγεται κύκλος, αν καταλήγει στον ίδιο κόμβο από τον οποίο ξεκινά (κεφάλαιο 2 της αναφοράς [SKP13]). Όταν δηλαδή η αρχική και η τελική κορυφή συμπίπτουν. Ένας κύκλος λέγεται απλός αν κανένας κόμβος δεν επαναλαμβάνεται, ενώ λέγεται σύνθετος αν υπάρχει τουλάχιστον ένας κόμβος που επαναλαμβάνεται (δηλαδή, ο σύνθετος κύκλος αποτελείται από πολλούς απλούς). Επίσης, ένας κύκλος λέγεται κατευθυνόμενος αν η διάτρεξη του γίνεται μόνο κατά μία κατεύθυνση. Αλλιώς, λέγεται μη κατευθυνόμενος. Ένας κύκλος μήκους k λέγεται άρτιος ή περιττός αν το k είναι άρτιο ή περιττό αντίστοιχα.

Ένα γράφημα ονομάζεται κλίκα ή πλήρες γράφημα αν κάθε ζευγάρι κορυφών του συνδέεται με ακμή (κεφάλαιο 2 της αναφοράς [SKP13]). Η κλίκα n κορυφών συμβολίζεται με K_n και έχει ακριβώς $\frac{n(n-1)}{2}$ ακμές. Ένα σύνολο κορυφών χωρίς καμία ακμή μεταξύ τους ονομάζεται σύνολο ανεξαρτησίας. Συνεπώς το συμπληρωματικό γράφημα μιας κλίκας είναι ένα σύνολο ανεξαρτησίας (στο ίδιο σύνολο κορυφών).

Η συνδεσιμότητα είναι μια βασική έννοια στη Θεωρία Γράφων η οποία χρησιμοποιείται για να μπορούμε να ποσοτικοποιούμε τις σχέσεις στα δίκτυα (κεφάλαιο 4 της αναφοράς [SKP13]). Η συνδεσιμότητα κ=κ(G) ενός γράφου G είναι ο ελάχιστος αριθμός των σημείων τα οποία αν τα αφαιρέσουμε θα έχουμε ένα μη συνδεδεμένο ή trivial γράφο. Η συνδεσιμότητα-ακμών (line-connectivity) λ=λ(G) ενός γράφου G είναι ο ελάχιστος αριθμός ακμών τις οποίες αν αφαιρέσουμε θα καταλήξουμε σε ένα μη συνδεδεμένο ή trivial γράφο επίσης. Ο trivial γράφος είναι ένας γράφος που αποτελείται από έναν απομονωμένο κόμβο. Ένα γράφημα ονομάζεται συνδεδεμένο αν κάθε ζεύγος διακριτών κορυφών στο γράφημα είναι συνδεδεμένο. Αλλιώς, το γράφημα ονομάζεται μη συνδεδεμένο. Για ένα μη συνδεδεμένο γράφημα ισχύει κ(G)=0 και για ένα γράφο τάξης p ισχύει $κ(K_p)=p-1$ αφού θα πρέπει να αφαιρέσουμε p-1 κόμβους για να καταλήξουμε σε trivial γράφο τάξης 1.

Έστω G γράφημα και έστω S ⊆ V (G). Σύμφωνα με τον ορισμό στο κεφάλαιο 4 της αναφοράς [SKP13] γειτονιά του S στο G καλείται το σύνολο NG(S) ={u ∈ V (G)\S | ∃ v ∈ S : {v, u} ∈ E(G)}, δηλ. το σύνολο όλων των κορυφών του G που είναι συνδεδεμένες με την v και δεν ανήκουν στο S. Αν v ∈ V (G), ορίζουμε NG(v) = NG({v}). Αν για μια κορυφή x ∈ V (G) ισχύει ότι NG(x) = ∅

τότε λέμε ότι η x είναι απομονωμένη κορυφή.

3 Κεφάλαιο

Δίκτυα

3.1 Ασύρματα Δίκτυα

Ως ασύρματο δίκτυο χαρακτηρίζεται το τηλεπικοινωνιακό δίκτυο, συνήθως τηλεφωνικό ή δίκτυο υπολογιστών, το οποίο χρησιμοποιεί ραδιοκύματα ως φορείς πληροφορίας [RMC06].
Τα δεδομένα μεταφέρονται μέσω ηλεκτρομαγνητικών κυμάτων, με συχνότητα φέροντος η οποία εξαρτάται κάθε φορά από τον ρυθμό μετάδοσης δεδομένων που απαιτείται να υποστηρίζει το δίκτυο. Η ασύρματη επικοινωνία, σε αντίθεση με την ενσύρματη, δεν χρησιμοποιεί ως μέσο μετάδοσης κάποιο τύπο καλωδίου. Σε παλαιότερες εποχές τα τηλεφωνικά δίκτυα ήταν αναλογικά, αλλά σήμερα όλα τα ασύρματα δίκτυα βασίζονται σε ψηφιακή τεχνολογία και, επομένως, κατά μία έννοια, είναι ουσιαστικώς δίκτυα υπολογιστών.

Στα ασύρματα δίκτυα εντάσσονται τα δίκτυα κινητής τηλεφωνίας, οι δορυφορικές επικοινωνίες, τα ασύρματα δίκτυα ευρείας περιοχής (WWAN), τα ασύρματα μητροπολιτικά δίκτυα (WMAN), τα ασύρματα τοπικά δίκτυα (WLAN) και τα ασύρματα προσωπικά δίκτυα (WPAN) [RMC06].

Ο βασικός διαχωρισμός των ασύρματων δικτύων γίνεται με κριτήριο το αν έχουν ή όχι κάποιο κεντρικό σημείο πρόσβασης. Υπάρχουν τα δίκτυα που έχουν ως σημείο πρόσβασης ένα σταθμό βάσης όπως για παράδειγμα τα κυψελωτά δίκτυα που χρησιμοποιούνται στην κινητή τηλεφωνία. Στα δίκτυα αυτά η απευθείας επικοινωνία μεταξύ τερματικών δεν είναι δυνατή και εξαρτάται από την ύπαρξη υποδομής (infrastructure). Στα δίκτυα με υποδομή όπως τα κυψελωτά υπάρχουν κάποια σταθερά σημεία πρόσβασης (access points-AP) μέσω των οποίων μπορούν οι διάφορες συσκευές να επικοινωνούν. Οι συσκευές έχουν ενσωματωμένους μηχανισμούς πρόσβασης στο ασύρματο μέσο μετάδοσης και επικοινωνούν με τα AP μέσω ραδιοκυμάτων. Τα AP μαζί με τις συσκευές που βρίσκονται στην δική τους κάλυψη, σχηματίζουν ένα βασικό σύνολο υπηρεσιών (Basic Service Set - BSS). Η σύνθεση των διαφόρων BSS, γίνεται μέσω των AP με ένα κατανεμημένο σύστημα και έτσι σχηματίζεται ένα δίκτυο. Οι συσκευές μμπορούν να επιλέξουν ένα AP και να συσχετιστούν μμαζί του. Τα AP παρέχουν συγχρονισμό μέσα στα BSS, υποστηρίζουν διαχείριση ενέργειας και μμπορούν να ελέγχουν το μέσο πρόσβασης για υποστήριξη υπηρεσιών με χρονικούς περιορισμούς [RMC06].

Από την άλλη πλευρά υπάρχουν τα ασύρματα Ad Hoc δίκτυα στα οποία οι κόμβοι επικοινωνούν μεταξύ τους, πιθανότατα μέσω πολλαπλών αναμεταδόσεων, χωρίς να υπάρχει συγκεκριμένη σταθερή υποδομή. Τα Ad Hoc δίκτυα είναι αυτόνομα συστήματα των οποίων η οργάνωση αποτελεί εσωτερική διαδικασία που επιτυγχάνεται χωρίς την ύπαρξη ενός αρχηγού, κεντρικού διαχειριστή ή συντονιστή του δικτύου. Κάθε κόμβος μπορεί να στείλει σε οποιονδήποτε άλλο κόμβο άμεσα, είτε στην περίπτωση που ο παραλήπτης του μηνύματος βρίσκεται εντός της εμβέλειάς του, είτε στην περίπτωση που βρίσκεται εκτός αυτής, όπου η επικοινωνία θα πραγματοποιηθεί πολυβηματικά μέσω άλλων κόμβων. Αυτοί οι κόμβοι σχηματίζουν ένα δίκτυο το οποίο δεν παρουσιάζει κεντρική οργάνωση. Οι όροι Ad Hoc δίκτυα και αυτοοργανούμενα δίκτυα χρησιμοποιούνται για να περιγράψουν ακριβώς τον ίδιο τύπο δικτύων [RMC06].

Η χρήση ασύρματου μέσου μετάδοσης έχει μία σειρά από πλεονεκτήματα:

i) Κινητικότητα χρήστη : Οι χρήστες μμπορούν να μμετακινούνται εντός της εμβέλειας του ασύρματου δικτύου, δηλαδή σε χώρο που θα έχουν επαρκές σήμα, διατηρώντας την συνδεσιμότητα τους με αυτό. Η κίνηση των κόμβων μπορεί να έχει πολλά χαρακτηριστικά , αναλόγως του περιβάλλοντος που το δίκτυο καλείται να λειτουργήσει. Έτσι είναι δυνατό να υπάρχει ατομική τυχαία κίνηση, ομαδική κίνηση, κίνηση μέσω προδιαγεγραμμένων οδών, κ.α. Το μοντέλο κινητικότητας των κόμβων μπορεί να έχει σημαντικό αντίκτυπο στην επιλογή των προδιαγραφών του δικτύου.

ii) Ευκολία, ευελιξία και απλότητα εγκατάστασης: Δεν χρειάζεται να εγκαταστήσουμε καλωδιώσεις μέσα από τοίχους και ταβάνια. Μπορεί να γίνει η δικτύωση σε μέρη όπου η καλωδίωση θα ήταν αδύνατη, ή μη επιθυμητή, όπως η δικτύωση γραφείων τα οποία βρίσκονται σε απόσταση μεταξύ τους. Η εγκατάσταση στις περισσότερες περιπτώσεις μπορεί να γίνει εύκολα αν ακολουθηθούν κάποιοι βασικοί κανόνες εγκατάστασης.

iii) Κλιμάκωση, δυνατότητα επέκτασης: Τα ασύρματα δίκτυα μπορούν να διαρθρωθούν σε ένα πλήθος από τοπολογίες, ώστε να ταιριάξουν στις απαιτήσεις των εφαρμογών. Οι τοπολογίες αλλάζουν εύκολα και επεκτείνονται από απλά δίκτυα με μικρό αριθμό χρηστών, ως μεγάλες δομές δικτύων με εκατοντάδες χρήστες και δυνατότητα περιαγωγής (roaming).

iv) Αυτοδιάρθρωση: Ένα ad hoc δίκτυο πρέπει να έχει τη δυνατότητα να καθορίζει αυτόνομα τα χαρακτηριστικά της διάρθρωσης του όπως την διευθυνσιοδότηση, τη δρομολόγηση, τη δημιουργία ομάδων, τον έλεγχο ισχύος, κ.α. Σε πολλές περιπτώσεις είναι δυνατό να υπάρχουν ειδικοί κόμβοι στους οποίους να ανατίθενται κάποιες αρμοδιότητες ελέγχου. Η επιλογή των κόμβων αυτών θα πρέπει να γίνεται αυτόνομα από το δίκτυο όπως αυτόνομα θα πρέπει να λαμβάνονται και τα μέτρα που απαιτούνται όταν αυτοί αποσυνδέονται από το δίκτυο.

v) Ταχύτητες μετάδοσης: Όσο αναπτύσσεται η τεχνολογία γίνεται δυνατή η μετάδοση μεγαλύτερων ρυθμών δεδομένων. Ήδη ο μέγιστος ρυθμός μετάδοσης δεδομένων, από τα 2Mbps που μπορούσαν να επιτευχθούν αρχικά, έφτασε σήμερα σε ταχύτητες πάνω από 100Mbps ενώ ήδη έχουν εξαγγελθεί ακόμα μεγαλύτερες ταχύτητες.

vi) Αξιοπιστία – ανεξαρτησία: Ένα ασύρματο δίκτυο κατάλληλα διαμορφωμένο μπορεί να έχει μεγάλη αξιοπιστία. Έτσι μπορεί να σχεδιαστεί έτσι ώστε να μπορεί να εργάζεται όταν συμβαίνουν διακοπές ρεύματος και να περιλαμβάνει πολλές εναλλακτικές διαδρομές.

vii) Εμβέλεια: Η εμβέλεια ενός ασύρματου δικτύου σε περιβάλλον γραφείου μπορεί να είναι μερικές δεκάδες μέτρα. Τα ραδιοκύματα σε εσωτερικό χώρο έχουν να διαπεράσουν τοίχους και οροφές οπότε υφίστανται σημαντική απόσβεση. Σε ανοικτό χώρο όπου υπάρχει οπτική επαφή ανάμεσα στις ασύρματες συσκευές, οι αποστάσεις που μπορεί να καλυφθούν είναι μεγαλύτερες.

viii) Συμβατότητα με το υπάρχον δίκτυο: Τα περισσότερα ασύρματα δίκτυα έχουν προτυποποιημένο τρόπο σύνδεσης με τα υπάρχοντα ενσύρματα δίκτυα. Έτσι, η προσθήκη ασύρματης δικτύωσης σε υπάρχουσες δομές δικτύων μπορεί να γίνει με τον ευκολότερο τρόπο. Πολλές φορές δε, αποτελούν επέκταση ενός ενσύρματου δικτύου

Έχει αναπτυχθεί ένας αριθμός από ασύρματες τεχνολογίες. Οι πιο διαδεδομένες είναι: Bluetooth HomeRF Openair IEEE 802.11 IEEE 802.16 HyperLan I & II . Κάθε μία έχει διαφορετική εφαρμογή, άρα μπορεί να ειπωθεί ότι είναι συμπληρωματικές μεταξύ τους παρά ανταγωνιστικές. Το Bluetooth και το HomeRF για παράδειγμα είναι σχεδιασμένα για ζεύξεις μμικρών αποστάσεων για σύνδεση μεταξύ συσκευών και των περιφερειακών τους, τα IEEE 802.11 για την υλοποίηση ασύρματων τοπικών δικτύων, ενώ το IEEE 802.16 για την υλοποίηση ευρύτερων ασύρματων μητροπολιτικών δικτύων.

3.2 Κοινωνικά / Σύνθετα Δίκτυα

Στην εργασία αυτή μελετάται η διάχυση πληροφορίας σε συνδυασμό κοινωνικών και ασύρματων δικτύων. Για το λόγο αυτό στη συνέχεια παρατίθενται οι ορισμοί για τα κοινωνικά και σύνθετα δίκτυα και αναλύονται κάποιες από τις χαρακτηριστικές δομές τους.

Ένα κοινωνικό δίκτυο είναι μία κοινωνική δομή που συμπεριλαμβάνει κόμβους (που γενικά είναι φυσικά πρόσωπα ή οργανισμοί) οι οποίοι συνδέονται με έναν ή περισσότερους παρόμοιου τύπου δεσμούς ή σχέσεις, όπως αξίες, οράματα, στόχοι, ιδέες, οικονομικές συναλλαγές, εμπορικές συναλλαγές, φιλία [KNW11]. Τα κοινωνικά δίκτυα εκφράζουν την ανάγκη των ανθρώπων να δημιουγούν δεσμούς με άλλα άτομα με τα οποία μοιράζονται κάποιο κοινό χαρακτηριστικό. Ο όρος «social network» αναφέρεται στη δημιουργία και την αξιοποίηση κοινοτήτων για τη διασύνδεση ανθρώπων με κοινά ενδιαφέροντα. Σήμερα, υπάρχουν πολλοί ιστότοποι κοινωνικής δικτύωσης. Οι ιστότοποι διακρίνονται σε κατηγορίες ανάλογα με:

- το αντικείμενό τους, δηλαδή το στόχο της δικτύωσης,
- τον τρόπο εγγραφής και συμμετοχής μελών (ελεύθερη ή περιορισμένη),
- τον τρόπο επικοινωνίας μεταξύ των μελών τους και
- το είδος του περιεχομένου που ανταλλάσσουν οι χρήστες μεταξύ τους.

Τα περισσότερα κοινωνικά δίκτυα εμφανίζουν «δομές κοινοτήτων» , δηλαδή ομάδες κόμβων οι οποίες έχουν υψηλή πυκνότητα πλευρών μεταξύ των μελών τους και σχετικά χαμηλή πυκνότητα μεταξύ ομάδων . Οι κοινότητες αυτές δημιουργούνται βάσει κάποιου κοινού χαρακτηριστικού και έχουν την τάση να παρουσιάζουν κοινή συμπεριφορά σε εξωτερικά ερεθίσματα.

Ο κλασσικός τρόπος για την εύρεση κοινοτήτων είναι η ανάλυση ομάδων (cluster analysis) [SHS01]. Η μέθοδος αυτή προϋποθέτει ότι οι ακμές έχουν κάποιο βάρος το οποίο υποδηλώνει τη σημαντικότητα της σύνδεσης. Ο αλγόριθμος ξεκινά με ένα σύνολο n κόμβων χωρίς ακμές και διαδοχικά προστίθενται οι ακμές κατά σειρά μεγαλύτερου βάρους.

Μια άλλη μέθοδος η οποία έχει προταθεί από τον Girvan and Newman βασίζεται στην «ενδιαμεσική κεντρικότητα» (betweenness centrality) των ακμών, η οποία ορίζεται ως ο αριθμός των συντομότερων μονοπατιών που περνούν από μια ακμή (κεφάλαιο 4 της αναφοράς [SKP13]). Η ιδέα είναι ότι σε ένα δίκτυο με ισχυρό το χαρακτηριστικό των κοινοτήτων θα υπάρχουν κάποιες ακμές οι οποίες θα συνδέουν τις διάφορες ομάδες. Προφανώς οι ακμές αυτές θα έχουν υψηλή τιμή ενδιαμεσικής κεντρικότητας. Βρίσκοντας και αφαιρώντας αυτές τις ακμές το δίκτυο μπορεί να χωριστεί σε κοινότητες. Η ποσότητα αυτή μπορεί να υπολογιστεί σε O(nm) χρόνο για ένα δίκτυο με n κόμβους και m ακμές. Ο αλγόριθμος υπολογίζει την ενδιαμεσική κενρικότητα κάθε ακμής σε κάθε βήμα και αφαιρεί από το δίκτυο την ακμή με την υψηλότερη τιμή , έως ότου αφαιρεθούν όλες οι ακμές. Ο αλγόριθμος μπορεί, όπως και προηγούμενα, να σταματήσει σε οποιόδηποτε σημείο, στο οποίο και μπορεί κανείς να διακρίνει τις διαφορετικές κοινότητες. Είναι λοιπόν σημαντικό να βρεθεί κάποιος τρόπος ο οποίος να δείχνει αν ο χωρισμός που έχει επιτευχθεί είναι ικανοποιητικός ώστε να σταματάει ο αλγόριθμος. Για το λόγο αυτό οι Girvan και Newman πρότειναν το μέτρο της τμηματικότητας (modularity). Για μια διαίρεση του δικτύου που αποτελείται από g ομάδες ορίζεται ένας πίνακας e μεγέθους gxg όπου κάθε στοιχείο e_{ij} είναι το ποσοστό των ακμών του αρχικού δικτύου που συνδέουν τους κόμβους της ομάδας i με αυτούς της ομάδας j. Σε αυτή την περίπτωση η τμηματικότητα ορίζεται ως :

$$Q = \sum_i e_{ii} - \sum_{ijk} e_{ij}\, e_{ki} = Tre - ||e^2|| \,. \qquad (7)$$

Όπου ||χ|| είναι το άθροισμα όλων των στοιχείων του χ. Πρακτικά το Q είναι το ποσοστό των ακμών που βρίσκονται ανάμεσα σε κοινότητες μείον την αναμενόμενη τιμή της ίδιας ποσότητας για ένα δίκτυο με τα ίδια χαρακτηριστικά με τη διαφορά ότι οι ακμές είναι τυχαίες. Μία τιμή Q=0 δείχνει ότι η εκτιμώμενη δομή κοινοτήτων δεν είναι καλύτερη από ένα τυχαίο γράφο. Τοπικά μέγιστα του μέτρου δείχνουν μία καλή εκτίμηση για τις ομάδες. Έτσι λοιπόν με χρήση του μέτρου της τμηματικότητας είναι δυνατό να γίνεται μια εκτιμήση για την ποιότητα των ομάδων σε κάθε βήμα και ο αλγόριθμος να σταματάει όταν βρεθεί μια ικανοποιητική διαίρεση του δικτύου. Σοβαρό μειονέκτημα του αλγορίθμου είναι ότι απαιτεί στη χειρότερη περίπτωση $O(m^2n)$, αν και έχουν προταθεί παραλλαγές που έχουν πολυπλοκότητα $O(n^2)$. Η εύρεση κοινοτήτων σε δίκτυα δίνει τη δυνατότητα δημιουργίας ιεραρχικών δομών οι οποίες παρέχουν μια μορφή αφαίρεσης σε πολλά προβλήματα μειώνοντας πρακτικά το μέγεθος του δικτύου. Για παράδειγμα, στα ad hoc δίκτυα , όπου η κινητικότητα των κόμβων δημιουργεί αρκετά προβλήματα , είναι δυνατό μέσω της ανακάλυψης ισχυρών κοινοτήτων , πολλές λειτουργίες να γίνονται σε επίπεδο ομάδων, οι οποίες είναι συνήθως αρκετά πιο στατικές από ότι οι μεμονωμένοι κόμβοι.

Ως σύνθετο ή πολύπλοκο δίκτυο (complex networks) ορίζεται ένα δίκτυο ή γράφος το οποίο έχει μη τετριμμένα τοπολογικά χαρακτηριστικά. Τα περισσότερα δίκτυα τα οποία εμφανίζονται στη φύση μπορούν να χαρακτηριστούν πολύπλοκα καθώς παρουσιάζουν χαρακτηριστικά που δεν εμφανίζονται στα απλά δίκτυα [SHS01] . Παραδείγματα τέτοιων δικτύων είναι ο παγκόσμιος ιστός, τα κοινωνικά δίκτυα που δημιουργούνται μεταξύ ανθρώπων ή ακόμα και τα βιολογικά δίκτυα που δημιουργούνται μέσα σε οργανισμούς. Γίνεται προσπάθεια κατανόησης των στατιστικών ιδιοτήτων των πραγματικών δικτύων για να δοθούν απαντήσεις για τον τρόπο δημιουργίας των δικτύων καθώς και για τη γενικότερη συμπεριφορά τους. Γι αυτό το λόγο έχουν προταθεί αρκετά μοντέλα που περιγράφουν τις γενικότερες μακροσκοπικές ιδιότητες πραγματικών δικτύων. Οι δύο βασικές κατηγορίες είναι τα δίκτυα απούσης κλίμακας (scale free networks) και τα δίκτυα μικρού κόσμου (small world networks) τα οποία παρουσιάζουν κάποια επιπλέον χαρακτηριστικά σε σχέση με τα απλά δίκτυα. Αυτά τα χαρακτηριστικά, τα οποία θα αναλυθούν παρακάτω, είναι η κατανομή βαθμού που ακολουθεί κατανομή power law, ο υψηλός βαθμός του συντελεστή συσσώρευσης (clustering coefficient) και η ύπαρξη κοινοτήτων και ιεραρχικής δομής.

Για να συνεχιστεί η ανάλυση χαρακτηριστικών δομών ενός σύνθετου δικτύου χρειάζονται οι παρακάτω ορισμοί.

Ως μέσο μήκος μονοπατιού ορίζεται ο μέσος όρος των συντομότερων σε μήκος μονοπατιών μεταξύ όλων των ζευγών των κόμβων σε ένα δίκτυο (κεφάλαιο 4 της αναφοράς [SKP13]). Η μέτρηση αυτή υποδεικνύει την απόσταση που περιμένουμε να έχουν δύο τυχαία επιλεγμένοι κόμβοι ενός ζευγαριού στο δίκτυο. Η απόσταση αυτή τις περισσότερες φορές μετριέται σε βήματα (hops), δηλαδή στον ελάχιστο αριθμό συνδέσεων που χωρίζουν αυτούς τους κόμβους. Εναλλακτικά η απόσταση αυτή μπορεί να είναι άθροισμα ή γινόμενο του βάρους των συνδέσεων σε ένα μονοπάτι, αν μελετάμε δίκτυα με βάρη. Εξ ορισμού λοιπόν ο υπολογισμός του μέσου μήκους μονοπατιού απαιτεί τη γνώση όλων των συντομότερων μονοπατιών σε ένα δίκτυο και άρα θα πρέπει να έχουν αρχικά προσδιοριστεί όλα τα ζεύγη κόμβων ενός δικτύου (το πολύ δηλαδή $\binom{|V|}{2}$) .Το μέσο μήκος μονοπατιού χαρακτηρίζει την έκταση του δικτύου. Είναι μια μέτρηση που χρησιμοποιείται για την αξιολόγηση της επίδοσης ενός παραδοσιακού επικοινωνιακού δικτύου και είναι μια παράμετρος που ορίζεται για όλο το δίκτυο.

Ο συντελεστής ομαδοποίησης – C(i) ενός κόμβου χρησιμοποιείται στα σύνθετα δίκτυα για να χαρακτηρίσει τη δομή ενός δικτύου τοπικά, δηλαδή σε επίπεδο κόμβου και παγκόσμια, δηλαδή σε επίπεδο δικτύου (κεφάλαιο 4 της αναφοράς [SKP13]). Ο συντελεστής αυτός εκφράζει την τριαδική διαδικασία σε ένα δίκτυο. Με τον όρο τριαδική διαδικασία εννοούμε τη διαδικασία κατά την οποία δύο γείτονες ενός κόμβου γίνονται και μεταξύ τους γείτονες. Ο συντελεστής ομαδοποίησης εκφράζει την πιθανότητα οι κόμβοι με κοινό γείτονα, να έχουν επίσης γείτονα έναν ακόμη κόμβο. Δηλαδή υπολογίζει την απευθείας σύνδεση μεταξύ των γειτόνων του κόμβου i. Αυτό συμβαίνει αρκετά συχνά στα κοινωνικά δίκτυα. Ο συντελεστής ομαδοποίησης ορίζεται ως:

$$C_i = \frac{\text{αριθμός ακμών μεταξύ των γειτόνων του κόμβου } i}{\text{αριθμός όλων των δυνατών ακμών μεταξύ των γειτονων του κόμβου}} \quad (8)$$

Μια άλλη εξίσωση για το συντελεστή ομαδοποίησης είναι η εξής:

$$C(i) = \frac{2\mathrm{E_i}}{d_i(d_i-1)} . \quad (9)$$

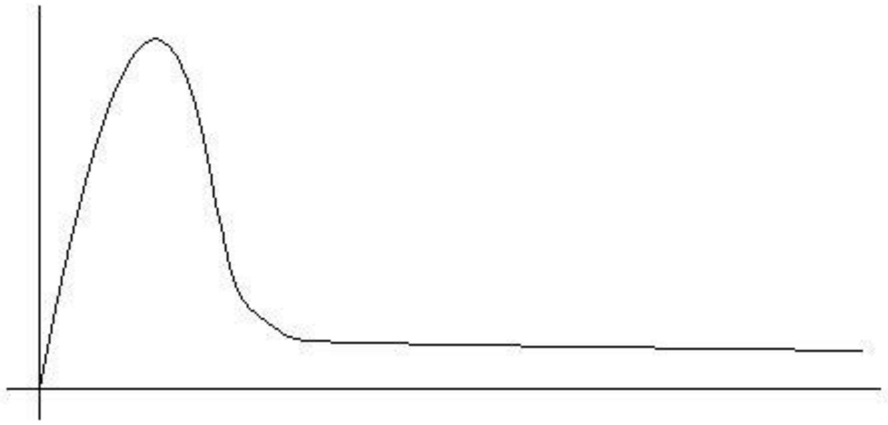
όπου d_i ο βαθμός του κόμβου i και E_i ο αριθμός των συνδέσεων μεταξύ των γειτόνων του κόμβου i. Αν ο βαθμός του κόμβου είναι ένα, τότε ο συντελεστής ομαδοποίησης δεν ορίζεται από την εξίσωση αυτή.

Ενώ πιο λεπτομερής εξίσωση είναι η παρακάτω. Θεωρώντας ότι ο κόμβος i έχει k_i γείτονες και A=[a_{ij}] είναι ο NxN πίνακας γειτνίασης του δικτύου τότε ο συντελεστής ομαδοποίησης ενός τυχαίου γράφου ισούται με:

$$C_i = \frac{\sum_{j,k} a_{ij} a_{jk} a_{ki}}{(\sum_j a_{ij})\text{^}2 - \sum_j a_{ij}} \ . \qquad (10)$$

Ο συντελεστής ομαδοποίησης ενός γράφου με βάρη υπολογίζεται με βάση τα βάρη κάθε ακμής ενώ για τον υπολογισμό του συντελεστή σε ένα κατευθυνόμενο, ο οποίος χρησιμοποιείται για την καλύτερη μοντελοποίηση των σύνθετων δικτύων λαμβάνονται υπόψη η κατανομή της εισερχόμενης ροής καθώς και της εξερχόμενης κάθε φορά [SKP13].

Κατανομή βαθμών (degree distribution): Κατά την άφιξη ενός νέου κόμβου στο κατανεμημένο σύστημα και για να συνδεθεί (join) σε αυτό ανταλλάσσει μηνύματα με κόμβους που ήδη μετέχουν στο δίκτυο [SKP13]. Στη συνέχεια, επιλέγει κάποιους από αυτούς για να συνδεθεί άμεσα (σε ένα hop) και στο εξής τους θεωρεί γείτονες. Το πλήθος των γειτόνων ενός κόμβου είναι ο βαθμός του. Η ακολουθία βαθμών του δικτύου που προκύπτει εξαρτάται από τον τρόπο που ο κόμβος επιλέγει τους γείτονές του. Η κατανομή των βαθμών των κόμβων διαμορφώνει την τοπολογία του δικτύου. Η κατανομή βαθμών μπορεί να έχει ντετερμινιστική ή πιθανολογική μορφή ανάλογα με το πλαίσιο ανάλυσης του δικτύου και το πεδίο εφαρμογής του. Σε περιπτώσεις όπου οι σχέσεις γειτνίασης δεν ποικίλουν, όπως σε ένα χάρτη , η κατανομή βαθμών περιλαμβάνει όλο το φάσμα των τιμών των βαθμών κόμβων . Σε περιπτώσεις όμως που οι συνδέσεις του δικτύου αλλάζουν με το χρόνο ή εξαρτώνται από άλλες παραμέτρους τότε η κατανομή βαθμών γίνεται στοχαστική. Σε αυτές τις περιπτώσεις η κατανομή βαθμών είναι η πιθανότητα P(k) δηλαδή η πιθανότητα ότι ένας κόμβος έχει k γείτονες. Αυτός ο ορισμός όμως καλύπτει μόνο τα μη κατευθυνόμενα δίκτυα. Στα κατευθυνόμενα δίκτυα θα πρέπει να είναι γνωστά τόσο η (in-degree) μέσα κατανομή βαθμών για την εισερχόμενη ροή όσο και η (out-degree) έξω κατανομή βαθμών για την εξερχόμενη. Όπως και στα μη κατευθυνόμενα δίκτυα έτσι και στα κατευθυνόμενα και οι δύο αυτές κατανομές μπορεί να είναι ντετερμινιστικές ή πιθανολογικές. Η κατανομή βαθμών είναι χαρακτηριστικό μιας τοπολογίας δικτύου. Στην ανάλυση σύνθετων δικτύων η κατανομή βαθμών χρησιμοποιείται για ταξινόμηση των δικτύων. Δηλαδή προκύπτουν διάφοροι τύποι δικτύων σύμφωνα με τις σχέσεις διασύνδεσης τους, το οποίο φαίνεται από την κατανομή βαθμού των κόμβων, όπως αναφέρεται στο κεφάλαιο 4 της αναφοράς [SKP13]. Στα κοινωνικά δίκτυα, η κατανομή βαθμών είναι συνήθως μια heavy tail κατανομή , όπως φαίνεται και στο σχήμα 2 παρακάτω.



Σχήμα 2: Κατανομή heavy tail , ο άξονας y αναπαριστά τον αριθμό των κόμβων και ο x το βαθμό κατανομής

Η κατανομή αυτή μοιάζει με την κατανομή power law σε γραμμική κλίμακα όπου η καμπύλη παρουσιάζει μια μακριά ουρά κατά μήκος του άξονα του βαθμού κόμβων, αποδεικνύοντας με αυτό τον τρόπο ότι ένας μεγάλος αριθμός κόμβων έχει μικρό βαθμό [SKP13].

Τέλος η κατανομή power law συνοπτικά μπορεί να οριστεί ως εξής: Μια τυχαία μεταβλητή X λέγεται ότι έχει power law κατανομή αν:

$$\mathrm{P_r}(X \geq x) = cx^{-a} \ . \qquad (11)$$

όπου c θετική σταθερά (c>0) και α>0. Η Εξίσωση (11) εκφράζει την πιθανότητα ο κόμβος r να έχει τουλάχιστον x γείτονες όπου η τυχαία μεταβλητή X αντιστοιχεί στο βαθμό του κόμβου. Η πιθανότητα αυτή ελέγχεται από τις δύο παραμέτρους (α, c). Η παράμετρος c είναι παράγοντας κανονικοποίησης που ελέγχει το μέγιστο βαθμό k_{max} και η παράμετρος α καλείται εκθέτης της power law (power law exponent).

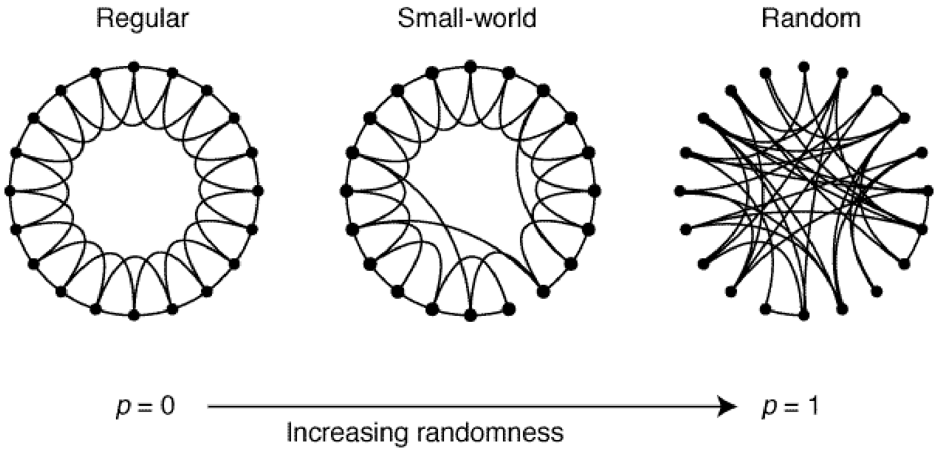
Μετά τους παραπάνω ορισμούς θα περιγραφούν ορισμένες χαρακτηριστικές δομές. Οι διάφορες χαρακτηριστικές δομές δικτύων αναπτύχθηκαν για την κατανόηση των σύνθετων δικτύων, ώστε να μοντελοποιηθούν με αυτά φαινόμενα της καθημερινότητας μας. Από τις χαρακτηριστικές δομές , για την παρούσα εργασία , θα χρειαστούν δύο, οι οποίες και αναλύονται παρακάτω.

Μια σημαντική χαρακτηριστική δομή δικτύου είναι τα δίκτυα μικρού κόσμου [SHS98]. Ως «δίκτυα τύπου μικρού κόσμου» ή «(Small world networks)» ορίζονται τα δίκτυα τα οποία έχουν υψηλό βαθμό ομαδοποίησης αλλά το μήκος μονοπατιού είναι μικρό ενώ η κλιμάκωση του ως προς τον αριθμό των κόμβων του δικτύου είναι όμοια με την αντίστοιχη των τυχαίων γράφων. Στα δίκτυα αυτά τα περισσότερα ζεύγη κόμβων συνδέονται μεταξύ τους με ένα μονοπάτι λίγων ακμών (έχουν δηλαδή μικρό μέσο μήκος διαδρομής), σε σχέση με το μέγεθος n του δικτύου. Τα δίκτυα αυτά τα όρισαν αρχικά οι Watts και Strogatz για να περιγράψουν δίκτυα που μπορεί να είναι σε μεγάλο βαθμό ομαδοποιημένα και έχουν μικρό μέσο μήκος μονοπατιού ενώ δεν είναι ούτε κανονικά ούτε τυχαία. Το μέσο μήκος μονοπατιού ενός τέτοιου δικτύου είναι:

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \ . \qquad (12)$$

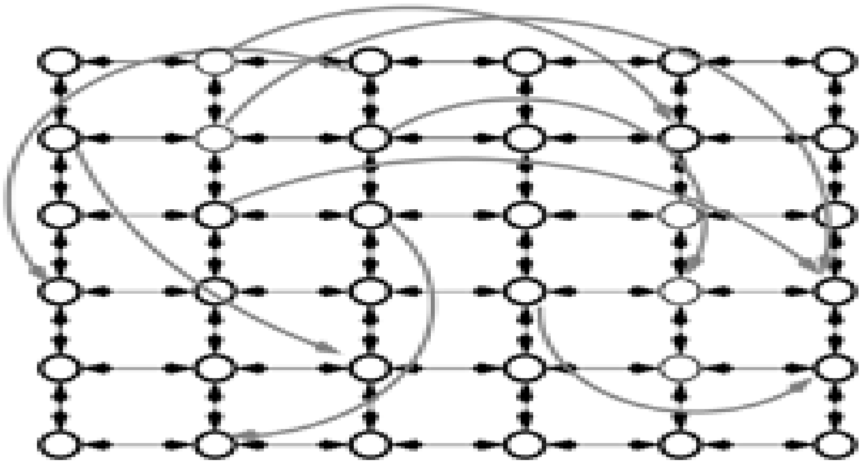
Όπου d_{ij} είναι η ελάχιστη απόσταση από τον κόμβο i στον κόμβο j και n ο αριθμός των κόμβων του δικτύου. Στο άθροισμα προστίθεται και η απόσταση από τον κάθε κόμβο στον εαυτό του, το οποίο θεωρείται μηδενικό. Από μαθηματική άποψη τα δίκτυα θεωρείται ότι εμφανίζουν το φαινόμενο του μικρού κόσμου αν το l αυξάνει λογαριθμικά (ή και πιο αργά) συγκριτικά με την αύξηση του μεγέθους του δικτύου, δηλαδή με την προσθήκη νέων κόμβων. Το μέσο μήκος του μονοπατιού αυτού είναι περίπου έξι βήματα (hops) γι αυτό στα δίκτυα αυτά αποδίδεται συχνά το φαινόμενο των έξι βαθμών διαχωρισμού.

Το μοντέλο των Watts και Strogatz περιγράφει τη διαδικασία με την οποία γίνεται η μετάβαση σε δίκτυο τύπου μικρού κόσμου ξεκινώντας από ένα απλό πλέγμα, δηλαδή η μετάβαση από απλό δίκτυο σε τυχαίο γράφο (κεφάλαιο 5 της αναφοράς [SKP13]). Σύμφωνα με το μοντέλο αυτό αλλάζοντας τις τιμές βασικών παραμέτρων οι οποίες εκφράζουν την πιθανότητα να προστεθούν επιπλέον συντομεύσεις, ελέγχεται ο βαθμός της τυχαιότητας και αλλάζει τελικά το δίκτυο. Συγκεκριμένα, ξεκινώντας από ένα απλό κυκλικό πλέγμα και προσθέτοντας τυχαία ακμές που ενώνουν κόμβους οι οποίοι είναι μακριά σε απόσταση βημάτων (hop distance) θα επιτευχθεί υψηλός συντελεστής ομαδοποίησης ενώ οι επιπλέον συντομεύσεις (shortcuts) ,όπως ονομάζονται οι τυχαία αυτές προστιθέμενες ακμές, μπορούν να μειώσουν το μέσο μήκος μονοπατιού τόσο ώστε το δίκτυο αυτό να θεωρείται πλέον δίκτυο μικρού κόσμου. Γενικά σε ένα δίκτυο μικρού κόσμου με n κορυφές και k ακμές σε κάθε κορυφή ισχύει $L \approx L_{random}(n,k)$ και $\gamma \geq \gamma_{random}(n,k)$ όπου L_{random} είναι το μέσο μήκος μονοπατιού ενός τυχαίου γράφου και γ_{random} ο συντελεστής ομαδοποίησης ενός τυχαίου γράφου. Πιο συγκεκριμένα για p=0 ο γράφος παραμένει ένα κυκλικό πλέγμα ενώ για p=1 το κυκλικό πλέγμα μετατρέπεται σε τυχαίο γράφο. Για τιμές 0<p<1 ,όπως φαίνεται και παρακάτω, το δίκτυο βρίσκεται σε μια ενδιάμεση κατάσταση.



Σχήμα 3 : Μετατροπή κυκλικού πλέγματος σε τυχαίο γράφο

Εκτός των Watts και Strogatz τα δίκτυα μικρού κόσμου μελετήθηκαν και από τον Kleinberg (κεφάλαιο 5 της αναφοράς [SKP13]). Σύμφωνα με το μοντέλο του Kleinberg δίκτυα μικρού κόσμου θεωρούνται τα δίκτυα στα οποία υπάρχουν μικρού μήκους μονοπάτια που χωρίζουν τους κόμβους και οι κόμβοι έχουν την ικανότητα να βρουν αυτά τα μονοπάτια χρησιμοποιώντας μόνο τοπικές σε αυτούς πληροφορίες. Ο όρος τοπικές πληροφορίες αναφέρεται στο γεγονός ότι σε κάθε βήμα του αλγορίθμου ο κάθε κόμβος γνωρίζει μόνο τους γείτονες του, τη θέση του κόμβου για τον οποίο προορίζεται το μήνυμα και τη θέση των κόμβων από τους οποίους έχει περάσει το μήνυμα μέχρι να φτάσει σε αυτόν. Ο Kleinberg στο μοντέλο του θεωρεί ότι το δίκτυο μοντελοποιείται από ένα τετράγωνο πλέγμα δύο διαστάσεων με n^2 κόμβους στο οποίο προστίθενται επιπλέον συνδέσεις μεγάλης εμβέλειας, σε αντίθεση με το μοντέλο των Watts και Strogatz το οποίο ξεκινά με ένα κυκλικό πλέγμα όπου μερικές ακμές «επαναδιατάσσονται». Το μοντέλο του Kleinberg μπορεί να αποτυπωθεί στην παρακάτω εικόνα:



Σχήμα 4: Το μοντέλο του Kleinberg για Small Word δίκτυα

Τα δίκτυα μικρού κόσμου θεωρούνται σημαντικά από την άποψη ότι η διάδοση της πληροφορίας σε αυτά γίνεται πιο γρήγορα σε σχέση με τα άλλα, αφού το ελάχιστο μονοπάτι είναι σχετικά μικρό και άρα η απόσταση που πρέπει να διανυθεί είναι μικρότερη από τη διάμετρο του δικτύου [SKP13].

Η επόμενη χαρακτηριστική δομή η οποία θα αναλυθεί είναι τα scale free δίκτυα ή αλλιώς δίκτυα χωρίς κλίμακα τα οποία χρησιμοποιούνται, επίσης, για τη μοντελοποίηση των κοινωνικών δικτύων. Σύμφωνα με την αναφορά [AB99] τα δίκτυα χωρίς κλίμακα μοιάζουν με τα δίκτυα μικρού κόσμου στο γεγονός ότι είναι πολύ πιθανό το μέσο μήκος διαδρομής να είναι μικρό. Ο όρος χωρίς κλίμακα χρησιμοποιείται εξ αιτίας της έλλειψης κλιμάκωσης του βαθμού του κόμβου, δηλαδή την έλλειψη ομοιομορφίας στην κατανομή των βαθμών. Παρατηρούνται λοιπόν σχηματικές διαφορές στη συνδεσιμότητα των γειτονικών κόμβων εξ αιτίας της υψηλής δημοτικότητας ορισμένων κόμβων και γενικότερα της κατανομής βαθμού των κόμβων. Συγκεκριμένα, στα δίκτυα χωρίς κλίμακα ,οι διαφορετικές ομάδες κόμβων παρουσιάζουν διαφορετικούς βαθμούς κόμβου το οποίο μεταφράζεται σε διαφορετική κλίμακα συνδεσιμότητας και αριθμού γειτόνων.

Αρχικά όλα τα σύνθετα δίκτυα θεωρούνταν τυχαία. Στα τυχαία δίκτυα η κατανομή του βαθμού διανομής ακολουθεί κατανομή Poisson με σχήμα καμπάνας [BRS01]. Σύμφωνα με την κατανομή Poisson , η πιθανότητα ένας κόμβος να έχει k συνδέσεις μειώνεται εκθετικά για μεγάλο αριθμό k. Επίσης στα τυχαία δίκτυα δεν συναντάς συχνά κόμβους με συνδέσεις παραπάνω ή λιγότερες από το μέσο όρο συνδέσεων του δικτύου. Σε πολλά όμως σύνθετα δίκτυα η δομή και η αρχιτεκτονική είναι πιο πολύπλοκη. Πρώτον, τα δίκτυα αναπτύσσονται με τη συνεχή προσθήκη κόμβων σε αυτά και δεύτερον, όταν οι κόμβοι δημιουργούν νέες συνδέσεις τείνουν να συνδέονται με άλλους κόμβους με πιθανότητα που εξαρτάται από τη δημοτικότητα των άλλων κόμβων, δηλαδή από τον αριθμό των ακμών που έχει ήδη ο κόμβος αυτός. Αυτό είναι γνωστό ως προσάρτηση κατά προτίμηση (preferential attachment).

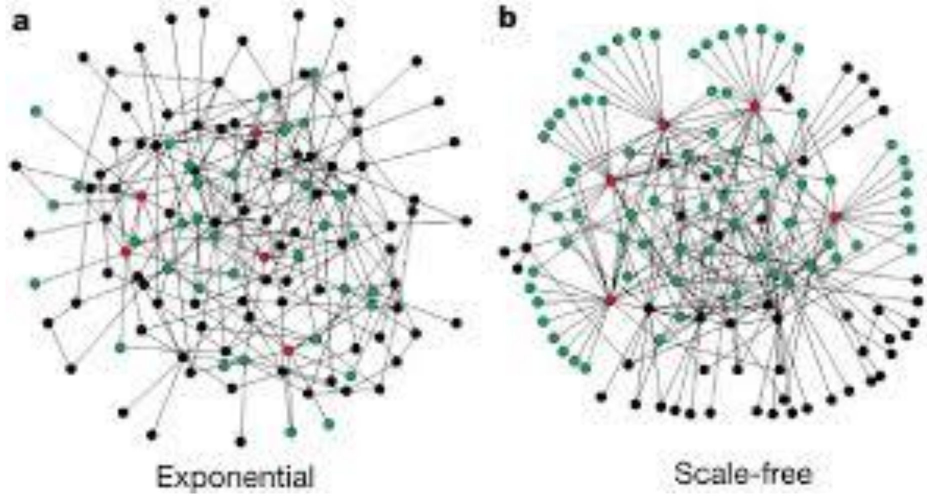
Στα σύνθετα δίκτυα δεν είναι όλοι οι κόμβοι το ίδιο δημοφιλείς. Ένας νέος κόμβος έχει μεγαλύτερη πιθανότητα να συνδεθεί σε έναν υπάρχοντα με μεγαλύτερο βαθμό σε σχέση με κάποιον άλλο με λιγότερες συνδέσεις. Οι κόμβοι που έχουν εξαιρετικά μεγάλο αριθμό συνδέσεων ονομάζονται hubs. Στα δίκτυα αυτά δεν υπάρχει καμία μορφή κλίμακας και επικρατούν οι λίγοι κόμβοι με μεγάλο αριθμό συνδέσεων (κεφάλαιο 5 της αναφοράς [SKP13]).

Οι Barabasi και Albert πρότειναν ένα μοντέλο το οποίο οδηγεί σε δίκτυα χωρίς κλίμακα [AB99]. Το μοντέλο τους προσπαθεί να αναπαράγει την προσάρτηση κατά προτίμηση των νέων κόμβων. Το δίκτυο που παράγεται μέσω των μοντέλων αυτών παρουσιάζει κατανομή βαθμού κόμβου power-law. Η κατανομή power-law είναι μια φθίνουσα συναρτηση χωρίς κορυφή, σε αντίθεση με την κατανομή Poisson που παρατηρείται στα τυχαία δίκτυα. Η κατανομή των συνδέσεων δηλαδή δεν είναι δημοκρατική, αφού υπάρχουν λίγοι κόμβοι οι οποίοι κυριαρχούν στο δίκτυο και πολλοί άλλοι με μικρό αριθμό συνδέσεων. Ο αλγόριθμος των Barabasi και Albert ξεκινάει με ένα μικρό αριθμό κόμβων , m_0 ,και σε κάθε βήμα προστίθεται ένας κόμβος με m ακμές που τον συνδέουν με ήδη υπάρχοντες κόμβους. Η πιθανότητα P με την οποία ο καινούριος κόμβος θα συνδεθεί με ένα κόμβο i εξαρτάται από τις ήδη υπάρχουσες συνδέσεις του i. Αν k_i είναι ο αριθμός των ακμών του i τότε $P(k_i) = \frac{k_i}{\sum_j k_j}$. Μετά από t βήματα ο αλγόριθμος οδηγεί σε ένα τυχαίο δίκτυο αποτελουμενο από $t+m_0$ κόμβους και mt ακμές. Η μελέτη του μοντέλου οδήγησε στο συμπέρασμα ότι το παραγόμενο δίκτυο παρουσιάζει κατανομή power-law με εκθέτη $\gamma=2.9\pm0.1$ που είναι ανεξάρτητος του m , της μοναδικής παραμέτρου του μοντέλου.

Αν και το μοντέλο των Barabasi και Albert αναπαράγει δίκτυα χωρίς κλίμακα , δεν παρέχει κάποια εξήγηση για τους εκθέτες των κατανομών των πραγματικών δικτύων, οι οποίοι κυμαίνονται μεταξύ 1 και 3 , και επιπλέον, δεν ενσωματώνει μηχανισμούς δημιουργίας ακμών που έχουν παρατηρηθεί σε διάφορα δίκτυα (κεφάλαιο 5 της αναφοράς [SKP13]). Για αυτό το λόγο πολλοί ερευνητές προσπάθησαν να προσδιορίσουν την εξάρτηση του εκθέτη από διάφορες διαδικασίες που παρουσιάζουν πραγματικά δίκτυα. Οι Krapivsky, Render και Leyvraz μελέτησαν την επιρροή μη γραμμικής προσάρτησης κατά προτίμησης στη δυναμική του συστήματος αντικαθιστώντας την πιθανότητα με την οποία ένας κόμβος συνδέεται σε ένα κόμβο i με βαθμό k_i με k_i^{a}. Βρήκαν ότι για μη γραμμική προσάρτηση κατά προτίμηση ($\alpha\neq1$) δεν προκύπτει δίκτυο χωρίς κλίμακα και ότι αυτό συμβαίνει μόνο για αυμπτωτικά γραμμική προσάρτηση κατά προτίμηση. Οι Drogovtsev, Mendes και Samukhin μελέτησαν την επιρροή της αρχικής ελκυστικότητας (initial attractiveness) στην προσάρτηση κατά προτίμηση αντικαθιστώντας την πιθανότητα με την οποία ένας κόμβος αποκτάει μια καινούρια εισερχομενη ακμή με $P(k_i)\approx A+k_i$ όπου A είναι η αρχική ελκυστικότητα και περιγράφει την πιθανότητα ενός απομονωμένου κόμβου να αποκτήσει μια εισερχόμενη ακμή.

Στο μοντέλο των Barabasi και Albert η πιθανότητα ενός κόμβου χωρίς ακμές να αποκτήσει μια καινούρια είναι μηδενική, πράγμα το οποίο δεν ισχύει σε πραγματικά δίκτυα. Οι Dorogovtsev, Mendes και Samukhin υπολόγισαν ότι στο μοντέλο που προτείνουν η κατανομή του βαθμού των κόμβων ακολουθεί νόμο δύναμης με εκθέτη $\gamma=2+A/m$ και άρα δημιουργείται όντως ένα δίκτυο χωρίς κλίμακας στο οποίο η αρχική ελκυστικότητα πηρεάζει μόνο την τιμή του εκθέτη. Επιπλέον το μοντέλο των Barabasi και Albert αδυνατεί να προσομοιώσει το χαρακτηριστικό ότι ο ρυθμός με τον οποίο οι κόμβοι αυξάνουν τις ακμές τους δεν ίδιος. Στο μοντέλο τους όλοι οι κόμβοι αυξάνουν τις ακμές τους με τον ίδιο ρυθμό και επομένως οι μεγαλύτεροι σε ηλικία κόμβοι έχουν το μεγαλύτερο αριθμό ακμών, το οποίο δεν ισχύει στα σύνθετα δίκτυα (κεφάλαιο 5 της αναφοράς [SKP13]).

Τα δίκτυα χωρίς κλίμακα παρουσιάζουν δύο σημαντικά χακτηριστικά, αρχικά, είναι αρκετά ανθεκτικά σε τυχαίες βλάβες αλλά από την άλλη είναι πολύ ευάλωτα σε οργανωμένες επιθέσεις. Το πρώτο χαρακτηριστικό οφείλεται στο γεγονός ότι η πλειοψηφία των κόμβων σε ένα δίκτυο χωρίς κλίμακας έχει μικρό αριθμό συνδέσεων για αυτό μια βλάβη σε ένα τυχαίο δίκτυο θα ήταν πιο πιθανό να καταστρέψει έναν κόμβο με μικρό βαθμό ο οποίος δεν είναι και πολύ σημαντικός στο δίκτυο. Με αυτό τον τρόπο εγγυάται η συνεκτικότητα του δικτύου. Είναι όμως κατανοητό ότι η κατάργηση μεγάλου αριθμού κεντρικών κόμβων θα οδηγήσει σε σημαντικές διαταράξεις του δικτύου. Για αυτό το λόγο όταν κάποιος οργανώσει επίθεση σε ένα δίκτυο χωρίς κλίμακα τότε θα χτυπήσει έναν hub ο οποίος παίζει πολύ σημαντικό ρόλο στο δίκτυο ώστε να απειλήσει την ισορροπία του δικτύου. Όταν ένας κεντρικός hub μολυνθεί τότε θα μολυνθούν με μεγάλη πιθανότητα και οι συνδέσεις του και ο ιός-πληροφορία τελικά θα μεταδοθεί σε μεγάλο αριθμό κόμβων τείνοντας να μολυνθεί όλο το δίκτυο.Ένας αποτελεσματικός τρόπος για να αντιμετωπιστεί αυτό είναι ο προσδιορισμός και εντοπισμός των κόμβων hubs ενός δικτύου το οποίο όμως είναι δύσκολο για τα σύνθετα δίκτυα (κεφάλαιο 5 τηςαναφοράς [SKP13]).



Σχήμα 5: a)Δίκτυα που ακολουθούν την Εκθετική Κατανομή b) Δίκτυα χωρίς κλίμακα

3.3 Γενικευμένα Δίκτυα

Στα δίκτυα πραγματικού κόσμου, όπως στα κοινωνικά δίκτυα ή στον παγκόσμιο ιστό (World Wide Web) μπορεί να συμβεί κάποια πρόσθεση, διαγραφή ή ένας επαναπροσδιορισμός ακμής ή κάποια πρόσθεση ή διαγραφή κόμβου (κεφάλαιο 5 της αναφοράς [SKP13]) . Επίσης η όλο και πιο διαδεδομένη χρήση των έξυπνων κινητών συσκευών οδηγεί στην ανάπτυξη διάφορων τρόπων για την κοινωνική αλληλεπίδραση των χρηστών. Αρχικά, η επικοινωνία μεταξύ ενός ατόμου και των φίλων του μπορούσε να γίνει μέσω προσωπικού κοινωνικού δικτύου, δηλαδή μέσω τηλεφωνικών κλήσεων οι οποίες εξαρτώνται από τη φορητότητα της συσκευής. Επειδή όμως τα έξυπνα τηλέφωνα διαθέτουν και τεχνολογία μικρής κλίμακας ασύρματης επικοινωνίας, δηλαδή το WiFi και το Bluetooth, υπάρχει και η επικοινωνία μεταξύ ομότιμων η οποία βασίζεται στην γεωγραφική εγγύτητα τους [KNW11]. Η γεωγραφική αυτή αλληλεξάρτηση των ατόμων επιβάλει τη δημιουργία ενός δικτύου το οποίο θα ενώνει τα προσωπικά κοινωνικά δίκτυα με τα χωρικά κοινωνικά δίκτυα. Άρα για τη μελέτη των πραγματικών δικτύων χρειάζεται ένα πιο γενικό πλαίσιο το οποίο θα μιμείται την εξέλιξη των πραγματικών δικτύων. Αυτά τα δίκτυα είναι τα γενικευμένα δίκτυα. Οι παραπάνω αναφερόμενες διεργασίες που μπορεί να συμβούν κατά τη διάρκεια της εξέλιξης ενός πραγματικού δικτύου διαχωρίζονται σε δύο κατηγορίες, γνωστές ως «Μεταβολή Ακμών» (“Edge Churn”) και «Μεταβολή Κόμβων» (“Node Churn”). Οι Μεταβολές Ακμών περιλαμβάνουν όλες τις πιθανές διεργασίες που μπορεί να γίνουν στις ακμές ενός δικτύου ενώ οι Μεταβολές Κόμβων όλες αυτές που μπορεί να γίνουν στους κόμβους ενός δικτύου. Στα γραφήματα που μελετώνται στην παρούσα εργασία όλα τα ζευγάρια κόμβων μπορεί να είναι γείτονες. Συμβολίζονται με N(i) οι απευθείας γείτονες του κόμβου i και με N_t ο αριθμός των κόμβων τη χρονική στιγμή t. Πιο αναλυτικά, στις διεργασίες Μεταβολής Ακμών περιλαμβάνονται οι [SKP13]:

- Προσθήκη ακμής: όπου με πιθανότητα p_1 προστίθενται m_1 νέες ακμές μεταξύ υπαρχόντων κόμβων και για κάθε νέα ακμή από τον κόμβο i στον κόμβο j ο i διαλέγεται με πιθανότητα $P_1^a(k_i)$ και ο j με πιθανότητα $P_1^b(k_j)$.
- Επαναπροσδιορισμός ακμής: όπου με πιθανότητα p_2 επαναπροσδιορίζονται m_2 υπάρχουσες ακμές. Ο κόμβος i από τον οποίο διαγράφεται η ακμή και ο κόμβος j στον οποίο προστίθεται διαλέγονται με πιθανότητα $P_2^a(k_i)$ και $P_2^b(k_j)$ αντίστοιχα.
- Διαγραφή ακμής: με πιθανότητα p_3 διαγράφονται m_3 υπάρχουσες ακμές. Ένας κόμβος i επιλέγεται με πιθανότητα $P_3(k_i)$ και τυχαία διαγράφει κάποια από τις συνδέσεις του.

Στις διαδικασίες Μεταβολής Κορυφών περιλαμβάνονται [SKP13] :

- Πρόσθεση κόμβου: όπου με πιθανότητα q_1 προστίθενται n_1 νέοι κόμβοι από τους οποίους ο καθένας δημιουργεί Μ νέες συνδέσεις με παλιότερους κόμβους που διαλέγονται με πιθανότητα $P_a(k_i)$
- Διαγραφή κόμβου: όπου με πιθανότητα q_2 διαγράφονται n_2 κόμβοι μαζί με όλες τις ακμές τους. Οι κόμβοι που θα διαγραφούν επιλέγονται με πιθανότητα $P_b(k_i)$

Έχει υπολογιστεί λοιπόν ότι ο ρυθμός της αλλαγής του μέσου βαθμού κόμβου k_i ∀ i ισούται με :

$$\frac{dk_i}{dt} = p_1 m_1 [P_1^a(k_i) + P_1^b(k_i)] + p_2 m_2 [P_2^b(k_i) - P_2^a(k_i)] - p_3 m_3 \left[P_3(k_i) + \sum_{j\in N(i)} \frac{P_3(k_j)}{k_j}\right] + q_1 n_1 M P_a(k_i) - q_2 n_2 \sum_{j\in N(i)} P_b(k_j)\ . \quad (13)$$

Αν θεωρηθεί ότι συμβαίνει μόνο πρόσθεση κάποιου κόμβου τότε η παραπάνω εξίσωση θα καταλήξει στο μοντέλο των Barabasi και Albert όπου M=m, q_1=p, $P_a(k_i)=\Pi(k_i)$, n_1=n και $P_1=P_2=P_3=q_2=0$.

Στα κατευθυνόμενα και με βάρη γραφήματα δικτύων το παραπάνω μοντέλο θα πρέπει να λαμβάνει υπόψη σύμφωνα με τα κεφάλαιο 5 της αναφοράς [SKP13] και άλλες παραμέτρους όπως:

- Τα βάρη στην προσάρτηση κατά προτίμηση: ότι δηλαδή σε κάθε χρονική στιγμή t ένας καινούριος κόμβος προστίθεται και συνδέεται με άλλους m υπάρχοντες από τους οποίους ο καθένας επιλέγεται με βάση την προνομιακή προσάρτηση, δηλαδή με πιθανότητα

$$\prod(s_i) = \frac{s_i(t)}{\sum_{\forall j} s_j(t)}\ . \quad (14)$$

όπου $s_i(t) = \sum_j w_{ij}(t) = \sum_j w_{ji}(t)$ και με w_{ij} συμβολίζεται το βάρος της ακμής που συνδέει τον κόμβο i με τον κόμβο j ενώ με s συμβολίζεται η δύναμη του. Το βάρος κάθε νέας ακμής θεωρείται ίσο με w_0.

- Την εξέλιξη του βάρους: η προσθήκη μιας νέας ακμής στον κόμβο i ,τη χρονική στιγμή t, η οποία θα έχει βάρος w_0 θα δημιουργήσει διαταραχές στα βάρη των άλλων ακμών του κόμβου i. Η διαταραχή αυτή για την ακμή που ενώνει τους κόμβους i και j μπορεί να θεωρηθεί ίση με $\Delta w_{ij}(t)$ και περιγράφεται από τον κανόνα: $w_{ij}(t) \leftarrow w_{ij}(t)+\Delta w_{ij}(t)$. Στο προτεινόμενο μοντέλο θεωρείται ότι για κάθε νέα ζεύξη που προστίθεται στον κόμβο i επάγεται μια αύξηση ίση με δ_i στο s_i η οποία κατανέμεται σε όλα τα βάρη $w_{ij}(t)$ αναλογικά με τις τιμές τους, δηλαδή $\Delta w_{ij}(t) = \delta_i \frac{w_{ij}(t)}{s_i(t)}$.

Οπότε τελικά μεταβάλλεται η αρχική εξίσωση του ρυθμού μεταβολής του δικτύου.

4 Κεφάλαιο

Θεωρητικό Υπόβαθρο και Σχετικές Εργασίες στο Επιδημιολογικό Μοντέλο

Διάχυση Πληροφορίας

Διάχυση της πληροφορίας είναι η διαδικασία κατά την οποία μια πληροφορία ή μια ασθένεια διαχέεται σε ένα δίκτυο. Θεωρούμε ότι μια συμπεριφορά (πληροφορία ή ασθένεια) διαδίδεται στο περιβάλλον το οποίο μοντελοποιείται μέσω απλών δομών γραφημάτων. Δυστυχώς στα δίκτυα πραγματικού κόσμου η διάχυση πληροφορίας είναι δύσκολο να προσεγγισθεί με μεγάλη ακρίβεια με τα μοντέλα που διαθέτουμε. Για τη διάχυση συμπεριφορών σε ένα πληθυσμό χρησιμοποιούνται μοντέλα από την επιδημιολογία.

4.1 Ομογενές Μοντέλο Συνεχούς Χρόνου

Για τη χρήση του επιδημιολογικού μοντέλου σε δίκτυα υπολογιστών και τηλεπικοινωνιών γίνεται η θεώρηση ότι αναφέρεται σε ένα κλειστό ομογενή πληθυσμό ατόμων που έχουν κοινά χαρακτηριστικά και εμφανίζουν παρόμοιες συμπεριφορές [SPZ05]. Το ομογενές μοντέλο βασίζεται στο ότι κάθε «μολυσμένος» κόμβος - άτομο έχει ίδια πιθανότητα να διαδώσει τη συμπεριφορά σε έναν «ευάλωτο». Ως μολυσμένος χαρακτηρίζεται ο κόμβος-άτομο του δικτύου ο οποίος έχει μια πληροφορία ή ασθένεια ενώ ως ευάλωτος, ο κόμβος που δεν διαθέτει κάποια συμπεριφορά. Οι σύνδεσμοι μεταξύ των ατόμων παρουσιάζονται ως ένας μη κατευθυνόμενος γράφος G ο οποίος θεωρείται σταθερός κατά τη διάρκεια του φαινομένου της διάδοσης της συμπεριφοράς. Το ομογενές μοντέλο χωρίζεται επίσης σε συνεχούς χρόνου μοντέλο και διακριτού χρόνου. Για την παρούσα εργασία μας ενδιαφέρει το συνεχούς χρόνου μοντέλο. Ένα μοντέλο συνεχούς χρόνου αναλύεται με χρήση ενός συνόλου διαφορικών εξισώσεων. Τη στιγμή $t=0$ θεωρείται ότι υπάρχει ένας συγκεκριμένος αριθμός «μολυσμένων» χρηστών στο δίκτυο που συμβολίζεται με I_0 και βάση των οποίων γίνεται και η εξέλιξη του φαινομένου.

4.2 Τα επιδημιολογικά μοντέλα

Για τη μελέτη της διάδοσης της πληροφορίας μέσα σε ένα δίκτυο γίνεται χρήση επιδημιολογικών μοντέλων. Το επιδημιολογικό μοντέλο χωρίζεται σε τρεις κατηγορίες, τις οποίες μελετάμε παρακάτω.

Υγιής – Μολυσμένος (Susceptible – Infected) μοντέλο

Αυτό είναι το πιο κλασσικό επιδημιολογικό μοντέλο [D00]. Σε αυτό το μοντέλο υπάρχουν δύο κατηγορίες χρηστών , οι υγιείς (susceptible) το σύνολο των οποίων συμβολίζεται με S και οι μολυσμένοι που συμβολίζονται με I. Η δυνατή μετάβαση οποιουδήποτε κόμβου είναι μόνο μία , από S σε I και θεωρείται ότι όταν ένας κόμβος μεταβεί στην κατάσταση I τότε θα παραμείνει για πάντα σε αυτή. Για αυτό το μοντέλο ονομάζεται S-I. Συμβολίζουμε με S(t) το πλήθος των υγιών χρηστών τη χρονική στιγμή t, I(t) το σύνολο των μολυσμένων χρηστών τη χρονική στιγμή t και N(t) το σύνολο των χρηστών τη συγκεκριμένη χρονική στιγμή στο δίκτυο [TP00]. Οπότε κάθε χρονική στιγμή θα ισχύει:

$$N(t) = S(t) + I(t) \ . \qquad (15)$$

Θεωρώντας τη σταθερά γ ως την πιθανότητα να μολύνει ένας μολυσμένος κόμβος, έναν ευάλωτο-υγιή η διαφορική εξίσωση που περιγράφει την εξέλιξη της διάχυσης της συμπεριφοράς είναι η εξής:

$$\frac{dS}{dt} = -\gamma * S(t) * I(t) \ . \qquad (16)$$

Η παραπάνω εξίσωση αποδεικνύει ότι ο ρυθμός με τον οποίο μειώνονται οι υγιείς χρήστες ενός δικτύου στο χρόνο εξαρτάται από την πιθανότητα ένας ευάλωτος-υγιής να μολυνθεί από έναν μολυσμένο καθώς και από τον αριθμό των μολυσμένων και υγιών χρηστών του δικτύου. Εξ αιτίας του ότι δεν υπάρχει μηχανισμός ανάρρωσης στο επιδημιολογικό αυτό μοντέλο μακροπρόθεσμα όλοι οι χρήστες του πληθυσμού θα μολυνθούν.

Υγιής – Μολυσμένος - Απομακρυσμένος (Susceptible – Infected - Removed) μοντέλο

Το μοντέλο αυτό αναλύθηκε για πρώτη φορά από τους Kermack και McKendrick [KM01]. Στο μοντέλο αυτό, συγκριτικά με το προηγούμενο, υπάρχει μία επιπλέον κατάσταση, οι απομακρυσμένοι χρήστες (Recovered) στην οποία μπορεί να οδηγηθεί κάποιος μολυσμένος χρήστης είτε λόγω θανάτου ή λόγω ανάρρωσης. Άρα υπάρχουν τρεις δυνατές καταστάσεις στις οποίες μπορεί να βρίσκεται κάποιος χρήστης κάθε χρονική στιγμή. Συγκεκριμένα στο μοντέλο αυτό, το S(t) αντιπροσωπεύει τον αριθμό των υγιών χρηστών τη χρονική στιγμή t που είναι ευάλωτοι και μπορεί δηλαδή να μολυνθούν από κάποιο μολυσμένο. Με I(t) συμβολίζονται οι μολυσμένοι χρήστες του δικτύου τη χρονική στιγμή t οι οποίοι μπορεί να διαδώσουν την πληροφορία ή ασθένεια και σε άλλους κόμβους του δικτύου. Και τέλος με R(t) συμβολίζονται οι χρήστες οι οποίοι μολύνθηκαν και στη συνέχεια απομακρύνθηκαν από το σύνολο των μολυσμένων είτε γιατί ανάρρωσαν ή λόγω θανάτου , τη χρονική στιγμή t. Οι χρήστες αυτοί δεν είναι δυνατό να μολυνθούν ξανά ή να διαδώσουν την ασθένεια σε άλλους χρήστες, θεωρείται πως έχουν ανοσία πλέον σε αυτή την ασθένεια ή πληροφορία.

Για αυτό το μοντέλο σε κάθε χρονική στιγμή ο συνολικός αριθμός κόμβων ισούται με N(t) όπου:

$$N(t) = S(t) + I(t) + R(t)\ . \tag{17}$$

Αν επιπλέον θεωρήσουμε ότι με πιθανότητα β ένας μολυσμένος κόμβος μπορεί να μολύνει έναν υγιή-ευάλωτο τότε ο αριθμός των μολύνσεων θα είναι:

$$\beta * S(t) * I(t)\ . \tag{18}$$

Όπου β*S(t) είναι ο αριθμός ο αριθμός των νέων μολυσμένων κόμβων. Σε αυτό το μοντέλο, όπως και στο κλασσικό επιδημιολογικό S-I μοντέλο ο ρυθμός με τον οποίο ένας ευάλωτος-υγιής κόμβος μολύνεται θα ισούται με:

$$\frac{dS(t)}{dt} = -\beta * I(t) * S(t)\ . \tag{19}$$

Αλλά σε αυτό το μοντέλο υπάρχει επίσης η πιθανότητα ένας μολυσμένος κόμβος να πεθάνει ή να αναρρώσει. Έστω ότι ένας μολυσμένος κόμβος μπορεί να αναρρώσει με ρυθμό γ τότε ο ρυθμός με τον οποίο ένας μολυσμένος κόμβος αναρρώνει θα ισούται με:

$$\frac{dR(t)}{dt} = \gamma * I(t)\ . \tag{20}$$

Άρα ο αριθμός των μολυσμένων κόμβων θα αλλάζει με ρυθμό:

$$\frac{dI(t)}{dt} = \beta * I(t) * S(t) - \gamma * I(t)\ . \tag{21}$$

Εξαρτάται δηλαδή από το ρυθμό μολύνσεων των ευάλωτων κόμβων καθώς και από το ρυθμό ανάρρωσης των μολυσμένων.

Υγιής-Μολυσμένος-Υγιής (Susceptible-Infected-Susceptible) μοντέλο

Στο SIS μοντέλο οι υγιείς κόμβοι περιμένουν την άφιξη κάποιας πληροφορίας για να μεταβούν στην κατάσταση του «μολυσμένου» [KCS13]. Εφόσον μεταβούν σε αυτή την κατάσταση είναι δυνατό να αναρρώσουν, πετώντας την πληροφορία-ασθένεια από αυτούς και να μεταβούν ξανά στην κατάσταση του υγιούς-ευάλωτου χρήστη. Εφόσον λοιπόν αναρρώσουν μπορεί να ξαναμολυνθούν από την ίδια ασθένεια ή πληροφορία, δεν παθαίνουν ανοσία, σε αντίθεση με το SIR μοντέλο. Οπότε σε αυτό το μοντέλο οι κόμβοι μπορούν να διαχωριστούν κάθε χρονική στιγμή σε δύο ειδών, τους μολυσμένους και τους υγιείς-ευάλωτους. Άρα κάθε χρονική στιγμή t ο συνολικός αριθμός των κόμβων του δικτύου θα ισούται με N(t):

$$N(t) = S(t) + I(t) \qquad (22)$$

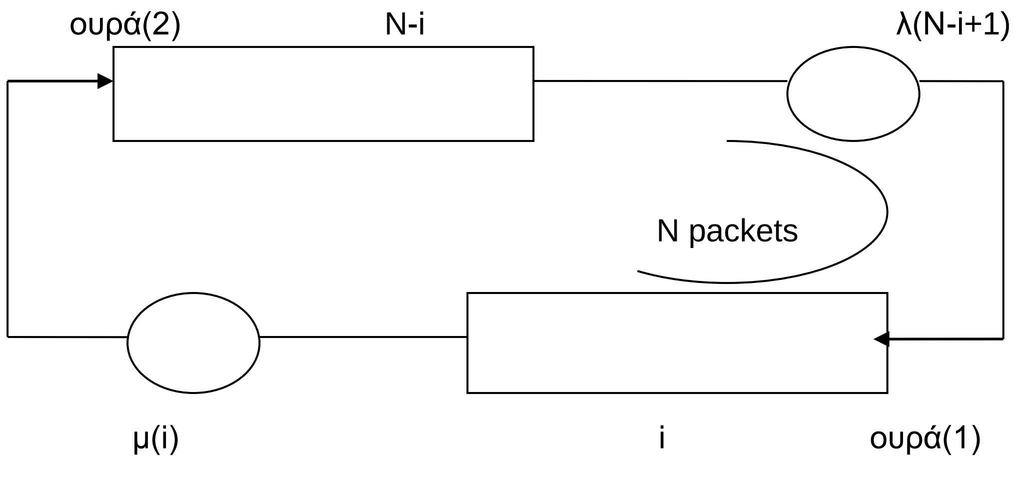
Όπου τώρα στο σύνολο των υγιών κόμβων ανήκουν και αυτοί που μπορεί να είχαν μολυνθεί αλλά πλέον έχουν αναρρώσει.Αν επιπλέον θεωρήσουμε β την πιθανότητα ένας μολυσμένος να μολύνει έναν υγιή τότε ο ρυθμός αύξησης των μολυσμένων θα είναι β*S(t)*I(t) ενώ αν θεωρήσουμε ίσο με α το μέρος των μολυσμένων χρηστών οι οποίοι ανάρρωσαν και επανεντάσσονται στο σύνολο των υγιών τότε ο αριθμός των υγιών χρηστών θα αλλάζει με ρυθμό:

$$\frac{dS(t)}{dt} = -\beta * S(t) * I(t) + \alpha * I(t)\ . \qquad (23)$$

Ενώ ο αριθμός των μολυσμένων θα μεταβάλλεται με ρυθμό:

$$\frac{dI(t)}{dt} = \beta * S(t) * I(t) - \alpha * I(t)\ . \qquad (24)$$

Επειδή κάθε φορά οι χρήστες μπορεί να είναι σε μια από τις δύο καταστάσεις το δίκτυο αυτό μπορεί να μοντελοποιηθεί και σαν ένα κλειστό σύστημα αποτελούμενο από δύο ουρές όπου κυκλοφορούν τα N συνολικά πακέτα. Αυτό το κλειστό σύστημα ανάδρασης μοντελοποιεί τη μετάβαση ενός χρήστη από τη μια κατάσταση στην άλλη δηλαδή την διαδικασία μόλυνσης και ανάρρωσης. Κάθε χρονική στιγμή αν θεωρήσουμε ότι I χρήστες είναι μολυσμένοι τότε N-i θα είναι υγιείς. Θεωρούμε ότι η ουρά(1) αντιπροσωπεύει το σύνολο των μολυσμένων κόμβων ενώ η ουρά(2) το σύνολο των υγιών κόμβων. Ο ρυθμός εξυπηρέτησης της ουράς των υγιών-ευάλωτων κόμβων θα είναι λ(N-i+1) και μ(i) ο ρυθμός εξυπηρέτησης των μολυσμένων [KCC13]. Αν λ είναι ο ρυθμός μόλυνσης στη μονάδα του χρόνου των υγιών κόμβων και μ ο ρυθμός γεννήσεων, δηλαδή ο ρυθμός με τον οποίο γίνονται και πάλι υγιείς. Οπότε το μοντέλο θα μοιάζει με το παρακάτω σύστημα:



Σχήμα 6: Προσομοίωση ρυθμού εξυπηρέτησης υγιών και μολυσμένων κόμβων

Για τη μελέτη της διάδοσης της πληροφορίας στη συγκεκριμένη εργασία θα χρησιμοποιηθεί το μοντέλο του Υγιούς-Μολυσμένου-Υγιούς. Θεωρούμε αρχικά ότι υπάρχει ένας συγκεκριμένος αριθμός μολυσμένων κόμβων και στη συνέχεια μολύνονται και άλλοι από τους οποίους μερικοί επανεντάσσονται στους υγιείς ενώ άλλοι παραμένουν μολυσμένοι.

4.3 Σχετική Εργασία

Μία εργασία σχετική με την παρούσα είναι αυτή των Shin-Ming Cheng, Weng Chon Ao, Pin-Yu Chen και Kwang-Cheng Chen με τίτλο «On Modeling Malware Propagation in Generalized Social Networks» [CAC11]. Στην εργασία αυτή μελετάται η διάδοση της κακόβουλης πληροφορίας στα «έξυπνα κινητά» όπου υπάρχουν δυο ειδών επικοινωνίες, οι προσωπικές επικοινωνίες μακράς εμβέλειας μέσω κοινωνικών δικτύων, οι οποίες στην εργασία αυτή θα συμβολίζονται με MMS, καθώς και οι χωρικές ασύρματες επικοινωνίες μικρής εμβέλειας , όπου ο τρόπος αυτός συμβολίζεται με BT. Για την ανάλυση της διάδοσης της πληροφορίας βάσει και των δυο αυτών συμπεριφορών/τρόπων επικοινωνίας αναπτύσσεται και μελετάται μαθηματικό μοντέλο και πραγματοποιείται προσομοίωση. Το μοντέλο που χρησιμοποιείται σε αυτή την εργασία βασίζεται στο Υγιές-Μολυσμένο επιδημιολογικό μοντέλο όπου ένας υγιής, εφόσον μολυνθεί δεν μπορεί να ξαναγίνει υγιής. Ο συνολικός πληθυσμός των κόμβων συμβολίζεται με N και όλοι οι κόμβοι μπορούν να μεταδώσουν και να λάβουν πληροφορίες μέσω MMS και BT υπηρεσιών. Άρα τη χρονική στιγμή t θα υπάρχουν συνολικά

$$I(t) = I_{BT}(t) + I_{MMS}(t)\ . \tag{25}$$

μολυσμένοι κόμβοι, όπου με I_{BT} συμβολίζονται οι κόμβοι (το πλήθος τους) που έχουν μολυνθεί μέσω BT και με I_{MMS} αυτοί που μολύνθηκαν μέσω MMS. Με S(t) συμβολίζονται οι υγιείς οπότε σε μια χρονική στιγμή t ο συνολικός αριθμός των κόμβων θα είναι :

$$N = I_{MMS}(t) + I_{BT}(t) + S(t)\ . \tag{26}$$

Και ο ρυθμός αύξησης των μολυσμένων ισούται με:

$$\frac{dI(t)}{dt} = \frac{dI_{BT}(t)}{dt} + \frac{dI_{MMS}(t)}{dt}\ . \tag{27}$$

Στο μοντέλο που μελετήθηκε στην εργασία αυτή, αρχικά μολυσμένος θεωρείται ένας κόμβος, ο οποίος μπορεί να έχει μολυνθεί μόνο μέσω MMS δηλαδή: $I(0)=I_{MMS}(0)=1$ και $I_{BT}(0)=0$. Επίσης θεωρείται ίση με β_{MMS} η πιθανότητα ένας μολυσμένος κόμβος να μολύνει μέσω MMS έναν υγιή και β_{BT} η πιθανότητα ένας μολυσμένος να μολύνει έναν υγιή μέσω BT. Τέλος, για την περιγραφή του αναλυτικού μοντέλου ορίζεται ως η_{MMS} ο μέσος βαθμός των κόμβων που συνδέονται μέσω MMS επικοινωνίας και ως η_{BT} ο μέσος βαθμός των κόμβων που επικοινωνούν με BT. Η εξίσωση του μαθηματικού μοντέλου βάσει της οποίας διαδίδεται η κακόβουλη πληροφορία μέσω MMS δηλαδή χωρίς να λαμβάνεται υπόψιν η χωρική θέση των κόμβων, δίνεται από την παρακάτω διαφορική εξίσωση:

$$\frac{dI_{MMS}(t)}{dt} = \beta_{MMS}\frac{S(t)(\eta_{MMS}-1)}{N}I(t)\ . \tag{28}$$

Όπου το η_{MMS}-1 σημαίνει ότι εφόσον υπάρχει ένας μολυσμένος κόμβος τότε τουλάχιστον ένας ακόμα από τους γείτονες του θα είναι μολυσμένος. Κατά τη διάδοση μέσω χωρικής επικοινωνίας , δηλαδή μέσω BT επικοινωνίας, οι κόμβοι αρχικά ψάχνουν τους κοντινούς τους κόμβους μέσα στην ακτίνα εμβέλειας τους R_C , και επικοινωνούν μέσω ασύρματης επικοινωνίας με τους γείτονές τους ώστε να μολύνουν τους υγιείς. Ο μέσος αριθμός των γειτόνων ισούται με $\rho\pi R_C^2$. Σε αυτό το μοντέλο η κινητικότητα του ανθρώπου αγνοείται και η διάδοση μοιάζει με κυματισμό του οποίου η πηγή είναι ένας μολυσμένος κόμβος και απλώνεται προς τα έξω με το χρόνο. Θεωρώντας ότι ένας κύκλος μόλυνσης δημιουργείται τη στιγμή r από έναν κόμβο ο οποίος αρχικά έχει μολυνθεί μέσω MMS και έχει επεκταθεί μέσα σε s χρονικές μονάδες , τότε ή χωρική μόλυνση τη χρονική στιγμή r+s θα ισούται με:

$$G^{'}(r,s) = \frac{dG(r,s)}{ds} = \beta_{BT}\frac{S(r+s)\frac{1}{2}\eta_{BT}}{N}c\sqrt{G(r,s)}\ . \tag{29}$$

Όπου το $\frac{1}{2}\eta_{BT}$χρησιμοποιείται γιατί ένας μολυσμένος κόμβος που ανήκει στην περιφέρεια στην περιφέρεια του κύκλου θα έχει τουλάχιστον μισούς γείτονες, εκτός του μολυσμένου κύκλου, οι οποίοι θα είναι υγιείς. Επίσης $c = 2R_C\sqrt{\rho\pi}$ για ομοιόμορφη κατανομή.

Οπότε η διαφορική εξίσωση που περιγράφει τη διάδοση μέσω BT επικοινωνίας τη χρονική στιγμή t είναι η παρακάτω:

$$\frac{dI_{BT}(t)}{dt} = \int_0^t I_{MMS}^{'}(\tau)G^{'}(\tau, t-\tau)d\tau\ . \tag{30}$$

Η παραπάνω εξίσωση προκύπτει θεωρώντας ότι τη χρονική στιγμή τ υπάρχουν $I_{MMS}^{'}$(τ)dτ πηγές μόλυνσης από τις οποίες η κάθε μία μολύνει με χωρική συνάρτηση μόλυνσης $G^{'}$(τ,t-τ).

Από την ανάλυση σε αυτή την εργασία , οι ερευνητές καταλήγουν ότι η διάδοση μέσω BT είναι πιο αργή σε σχέση με τη διάδοση μέσω MMS λόγω των χωρικών χαρακτηριστικών διάδοσης καθώς επίσης και ότι η διάδοση της κακόβουλης πληροφορίας γίνεται με μεγαλύτερη ταχύτητα στα αρχικά στάδια της διάδοσης.

5 Κεφάλαιο

Μελέτη Διάχυσης Πληροφορίας σε Κόμβους με Χρονομεταβλητά Ενδιαφέροντα με Μετρικές Κοινωνικής Δικτύωσης

5.1 Ορισμός του Δικτύου και Βασικές Έννοιες

Στην εργασία αυτή μελετάται η διάχυση της μη κακόβουλης πληροφορίας σε γενικευμένα δίκτυα λαμβάνοντας υπόψη και το ενδιαφέρον των χρηστών.

Υπάρχουν πολλά παραδείγματα τα οποία δείχνουν ότι η διάχυση της πληροφορίας εξαρτάται από το ενδιαφέρον των χρηστών για συγκεκριμένες πληροφορίες. Η διάχυση της πληροφορίας είναι φανερό και μέσω των μέσων κοινωνικής δικτύωσης ότι εξαρτάται από τις χωρικές και χρονικές μεταβολές στα ενδιαφέροντα των χρηστών. Για παράδειγμα, διαφημίσεις που αφορούν τις καλοκαιρινές διακοπές μπορεί να διαχέονται περισσότερο κατά τη διάρκεια των ανοιξιάτικων και των καλοκαιρινών μηνών, αλλά όχι κατά τη διάρκεια του φθινοπώρου ή του χειμώνα. Παρατηρείται δηλαδή μια περιοδικότητα στο ενδιαφέρον των χρηστών για αυτού του είδους τις πληροφορίες κατά τη διάρκεια του χρόνου. Το γεγονός αυτό δεν είναι εμφανές από τα τωρινά μοντέλα διάχυσης της πληροφορίας τα οποία θεωρούν ότι υπάρχει ίδιος βαθμός διάχυσης καθόλη τη διάρκεια του έτους, ανεξάρτητα από την εποχή του χρόνου. Ένα δεύτερο παράδειγμα είναι τα νέα για έναν ποδοσφαιρικό αγώνα τα οποία δεν θα ενδιαφέρουν κάποια ομάδα χορού οπότε δεν θα υπάρξει διάχυση για αυτή την πληροφορία μέσα σε αυτή την ομάδα. Είναι φανερό λοιπόν πως για τη διάχυση της πληροφορίας σε ένα πληθυσμό θα πρέπει να λαμβάνονται υπόψη τα ενδιαφέροντα του συγκεκριμένου πληθυσμού τα οποία είναι χρονικά μεταβαλλόμενα και διαφορετικά για κάθε ομάδα πληθυσμού.

Ο όρος γενικευμένα δίκτυα χρησιμοποιείται διότι μελετάται η διάχυση της πληροφορίας μέσω Υπηρεσιών Πολυμεσικών Μηνυμάτων (MMS) στο κοινωνικό επίπεδο καθώς και μέσω Wi-Fi/Bluetooth στο φυσικό επίπεδο. Η διάχυση μέσω MMS αναφέρεται σε μεγάλης εμβέλειας διάδοση, μακριά από την πηγή, ενώ η διάδοση μέσω Wi-Fi/Bluetooth σε μικρής εμβέλειας διάδοση [CAC11].

Για τη μελέτη της διάχυσης της μη κακόβουλης πληροφορίας, όπως αναλύθηκε και παραπάνω, θεωρείται ένα δίκτυο το οποίο αποτελείται από N κόμβους ομοιόμορφα κατανεμημένους σε μια τετράγωνη περιοχή , πλευράς L. Ο κάθε κόμβος έχει μια ακτίνα μήκους R. Αγνοείται η κινητικότητα των χρηστών. Θεωρούνται Μ κλάσεις πληροφορίας κάθε μία από τις οποίες συμβολίζονται με m. Η κάθε κλάση αποτελείται από μηνύματα για ένα συγκεκριμένο θέμα. Η διάχυση της πληροφορίας μελετάται ξεχωριστά για κάθε κανάλι, παρόλα αυτά λαμβάνονται υπόψη στον ορισμό του μοντέλου του συστήματος οι αλληλεπιδράσεις μεταξύ διαφορετικών κλάσεων. Κάθε κόμβος χαρακτηρίζεται από το ενδιαφέρον του για κάθε κλάση m , το οποίο συμβολίζεται ως $R_i^m(t)$. Συγκεκριμένα το ενδιαφέρον $R_i^m(t)$ του κάθε κόμβου εκφράζει το ποσοστό του ενδιαφέροντος του κόμβου i για την κλάση m τη χρονική στιγμή t, δηλαδή το ενδιαφέρον του κόμβου i για την κλάση m κανονικοποιημένο για όλες τις κλάσεις. Για αυτό ισχύει

$$\sum_{\forall m} R_i^m \leq 1 \, , \forall i \, . \tag{31}$$

Ένας κόμβος i θεωρείται μολυσμένος (ή ενημερωμένος) για μια συγκεκριμένη κλάση αν έχει τουλάχιστον ένα μήνυμα από αυτά που ανήκουν στην κλάση αυτή, διαφορετικά θεωρείται υγιής για την κλάση αυτή.

Σε όλη την παρακάτω ανάλυση χρησιμοποιούνται οι εξής συμβολισμοί:

- $I^m(t)$: ο αριθμός των μολυσμένων κόμβων για την κλάση m
- $S^m(t)$: ο αριθμός των υγιών κόμβων για την κλάση m
- $N_S(i)$: το σύνολο των φίλων του κόμβου i στο κοινωνικό επίπεδο
- $I^m(t)$: το σύνολο των μολυσμένων κόμβων για την κλάση m
- $S^m(t)$: το σύνολο των υγιών κόμβων για την κλάση m
- $N_{BT}(i)$: το σύνολο των συνδέσεων του i στο φυσικό επίπεδο

Ο μέσος βαθμός (θα είναι ο έξω-βαθμός αν οι ακμές του δικτύου είναι κατευθυνόμενες) όλων των κόμβων του δικτύου σε κοινωνικό επίπεδο θεωρείται ίσος με N_S^{avg}. Θεωρούνται οι παρακάτω συναρτήσεις:

- $f_1(x) : [0,1] \rightarrow [0,1]$: μονότονα αύξουσα ως προς το x
- $f_2(x) : [0,1] \rightarrow [0,1]$: μονότονα φθίνουσα ως προς το x
- $f_{1E}(x): [0,1] \rightarrow [0,1]$: μονότονα αύξουσα ως προς το x

Το επιδημιολογικό μοντέλο που χρησιμοποιείται για τη μελέτη της διάχυσης της πληροφορίας είναι το Υγιής-Μολυσμένος-Υγιής (SIS). Ένας κόμβος i θεωρείται μολυσμένος για μια κλάση αν έχει έστω και ένα μήνυμα για την κλάση αυτή, διαφορετικά θεωρείται υγιής. Ένας κόμβος μολύνεται και μεταβαίνει από την κατάσταση του υγιούς στην κατάσταση του μολυσμένου αν λάβει κάποιο μήνυμα-πληροφορία για κάποια κλάση. Ενώ ένας μολυσμένος κόμβος μπορεί να μεταβεί στην κατάσταση του υγιούς διαγράφοντας όλα τα μηνύματα –πληροφορίες που έχει για μια συγκεκριμένη κλάση.

Ένας κόμβος i μπορεί σε κάθε χρονική στιγμή να βρίσκεται σε μια από τις παρακάτω καταστάσεις:

1) Για τις κλάσεις για τις οποίες ο i είναι μολυσμένος:
   a) Ο i μπορεί να κάνει διάχυση της πληροφορίας για την κλάση m: $i \in I^m(t)$ με πιθανότητα $f_1(R_i^m(t))$.
   b) Ο i μπορεί να διαγράψει όλα τα μηνύματα για την κλάση m: $i \in I^m(t)$ με πιθανότητα $f_2(R_i^m(t))$.

Σε κάθε χρονοσχισμή επιτρέπεται μόνο μία ενέργεια από τις παραπάνω. Δηλαδή μόνο μία κλάση m επιλέγεται και είτε θα συμβεί το (a) ή το (b). Αυτό σημαίνει ότι:

$$\sum_{m:i\in I^m(t)} f_1\big(R_i^m(t)\big) + f_2\big(R_i^m(t)\big) \le 1\ . \tag{32}$$

2) Ο κόμβος i κάνει κάτι άλλο το οποίο δεν αφορά την διάχυση της πληροφορίας, με πιθανότητα ίση με

$$1 - \sum_{m:i\in I^m(t)} (f_1\big(R_i^m(t)\big) + (f_2\big(R_i^m(t)\big)\ . \tag{33}$$

Αν επιλέγει μία κλάση m για την περίπτωση (a) με πιθανότητα $f_1\big(R_i^m(t)\big)$ τότε ο κόμβος i θα ενεργήσει με έναν από τους παρακάτω τρόπους:

- Με πιθανότητα p_1, ο κόμβος θα στείλει MMS, περιλαμβάνοντας ως παραλήπτες κάθε $j \in N_S(i)$ τον οποίο θα επιλέξει με πιθανότητα $f_{1E}\left(R_j^m(t)\right)$ όπου το f_{1E} είναι απλά μια πιθανότητα και όχι μέρος κάποιας κατανομής όπως οι f_1 και f_2
- Με πιθανότητα p_2 ο κόμβος i θα στείλει με ευρυεκπομπή σε όλους τους γείτονες του, δηλαδή στο σύνολο N_{BT} την πληροφορία.

Ισχύει:

$$p_1 + p_2 \le 1\ . \tag{34}$$

Για την παραπάνω διαδικασία διάχυσης χρειάζεται σε κάθε στιγμή να υπάρχει τουλάχιστον ένας μολυσμένος κόμβος για κάθε κλάση έτσι ώστε να μη σταματάει η διάδοση. Επειδή όμως στα πραγματικά κοινωνικά δίκτυα επιτρέπεται και η διαγραφή , όπως και στην παρούσα εργασία, υπάρχει η πιθανότητα να σταματήσει η διάδοση της πληροφορίας για μια κλάση αν όλοι οι κόμβοι διαγράψουν την πληροφορία για αυτή.

Παρακάτω περιγράφεται μέσω διαφορικών εξισώσεων ο ρυθμός αύξησης των μολυσμένων κόμβων $I^m(t)$ για κάθε κλάση m στο χρόνο. Η διαφορική εξίσωση που περιγράφει τη δυναμική της εξέλιξης του αριθμού των μολυσμένων κόμβων είναι η εξής:

$$\frac{dI^m(t)}{dt} = p_1 f_1^{avg}(t) f_{1E}^{avg}(t) \frac{S^m(t)}{N} I^m(t) N_S^{avg} + p_2 N \frac{\pi R^2}{L^2} \frac{S^m(t)}{N} I^m(t) f_1^{avg}(t) - q I^m(t) f_2^{avg}(t)\ . \quad (35)$$

Όπου το $f_{1E}^{avg}(t)$είναι ο μέσος όρος των συναρτήσεων f_{1E}τη χρονική στιγμή t για την αντίστοιχη κλάση m. Όμοια ορίζονται και οι f_1^{avg}και f_2^{avg}. Το q είναι η πιθανότητα διαγραφής μηνυμάτων για την κλάση από τους κόμβους. Η αρχική συνθήκη είναι $I^m(0)=1$ δηλαδή μόνο ένας κόμβος είναι αρχικά μολυσμένος για κάθε κλάση. Η παραπάνω διαφορική εξίσωση έχει μοναδική λύση όταν οι $f_{1E}^{avg}, f_1^{avg}\ \kappa\alpha\iota\ f_2^{avg}$ είναι συνεχείς συνεχείς συναρτήσεις σύμφωνα με το θεώρημα Cauchy-Lipschitz [TG12].

Στην εργασία αυτή μελετώνται οι παρακάτω περιπτώσεις για τα ενδιαφέροντα των κόμβων:

- Περίπτωση 1: Περιοδικά Ενδιαφέροντα Χρηστών

Θεωρούνται 2 κλάσεις πληροφορίας, με

$$R_i^1(t) = 1 - A_i(\sin a(t+b_i))^2 + B_i, \forall i\ . \quad (36)$$

για την κλάση m=1,όπου το α προσδιορίζει την περίοδο των ενδιαφερόντων των κόμβων και i και θεωρείται σταθερά και ίση με α για όλους τους κόμβους και A_i,B_i,b_i είναι σταθερές κατάλληλα ορισμένες.

Και για την κλάση m=2

$$R_i^2(t) = A_i(\sin a(t+b_i))^2 - B_i, \forall i\ . \quad (37)$$

Έτσι ώστε να ισχύει:

$$R_i^1(t) + R_i^2(t) = 1, \forall i\ . \quad (38)$$

Σε αυτή την περίπτωση η διαφορική εξίσωση για την κλάση 1 θα πάρει τη μορφή:

$$\frac{dI^1(t)}{dt} = p_1 \frac{(1-\mathrm{A}(\sin a(t+b))^2+B)^2}{M} \frac{N-I^1(t)}{N} I^1(t) N_S^{avg} + p_2 \frac{\pi R^2}{L^2}(N - I^1(t)) I^1(t) \frac{1-\mathrm{A}(\sin a(t+b))^2+B)}{M} - I^1(t) \frac{1-(1-\mathrm{A}(\sin a(t+b))^2+B))}{M}. \quad (39)$$

Ομοίως η διαφορική για την κλάση 2 θα είναι:

$$\frac{dI^2(t)}{dt} = p_1 \frac{(\mathrm{A}(\sin a(t+b))^2-B)^2}{M} \frac{N-I^1(t)}{N} I^1(t) N_S^{avg} + p_2 \frac{\pi R^2}{L^2}(N - I^1(t)) I^1(t) \frac{\mathrm{A}(\sin a(t+b))^2-B)}{M} - I^1(t) \frac{1-(\mathrm{A}(\sin a(t+b))^2-B)}{M}. \quad (40)$$

Όπου A, B και b υπολογίζονται από το μέσο όρο των ενδιαφερόντων όλων των χρηστών τη χρονική στιγμή t.

Η λύση των παραπάνω διαφορικών εξισώσεων είναι σύνθετη για αυτό το λόγο στην παρούσα εργασία εφαρμόζεται η μέθοδος πεπερασμένων διαφορών (finite difference scheme) με το οποία επιτυγχάνεται η λύση της διαφορικής.

Θεωρείται ίσο με $M_1(t)$ το δεξί μέλος της εξίσωσης (9) και $M_2(t)$ αυτό της εξίσωσης (10) τότε με πολύ μικρό χρονικό βήμα Δt>0 θα ισχύει τελικά:

$$I^1(t+\Delta t) = I^1(t) + M_1(t)\Delta t \tag{48}$$

και

$$I^2(t+\Delta t) = I^2(t) + M_2(t)\Delta t \ .$$

Παρατηρείται ότι για

$$1 - \mathrm{A}(\sin a(t+b))^2 + B \cong 0 \ . \tag{42}$$

Ισχύει :

$$M_1(t) \cong -\frac{qI^1(t)}{M} \ . \tag{439}$$

Δηλαδή

$$I^1(t+1) < I^1(t) \ . \tag{44}$$

Για αυτές τις χρονικές περιόδους, και άρα

$$S^1(t+1) > S^1(t) \ . \tag{45}$$

Το αντίστροφο ισχύει για τις χρονικές περιόδους όπου

$$1 - \mathrm{A}(\sin a(t+b))^2 + B \cong 1.$$

- Περίπτωση 2: Στη δεύτερη περίπτωση θεωρούνται σταθερά και όχι περιοδικά μεταβαλλόμενα ενδιαφέροντα για τους κόμβους αφού εξετάζεται η διάχυση της πληροφορίας σε ομάδες με διαφορετικά ενδιαφέροντα.

Ισχύει ότι $R_i^1(t) = \alpha$ και $R_i^2(t) = 1 - \alpha$. Όπου α μία σταθερά.

Τα είδη υπάρχοντα συστήματα είναι ειδικές περιπτώσεις του προτεινόμενου μοντέλου αφού οι παράμετροι (f_1^{avg}, f_2^{avg}, f_{1E}^{avg}) τώρα είναι σταθεροί και δεν μεταβάλλονται με το χρόνο.

5.2 Αποτελέσματα Προσομοίωσης και Μαθηματικού Μοντέλου

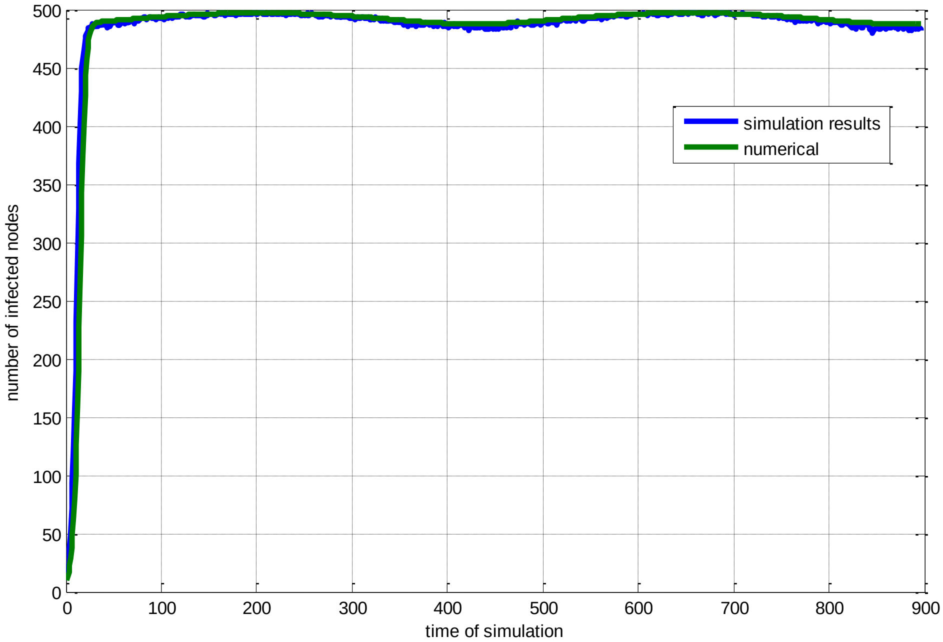
Για την ανάλυση των δύο περιπτώσεων χρησιμοποιήθηκε το MATLAB. Με τη βοήθεια του Matlab έγινε προσομοίωση της διάχυσης της πληροφορίας σε δίκτυο με N=500 κόμβους και συγκρίθηκαν τα αποτελέσματα της προσομοίωσης με μαθηματικά αποτελέσματα από τις λύσεις των διαφορικών εξισώσεων.

Δημιουργήθηκε ασύρματο δίκτυο αποτελούμενο από N=500 κόμβους συνολικά, ομοιόμορφα κατανεμημένοι σε τετράγωνη τοπολογία πλευράς L=350m. Οι κόμβοι έχουν ακτίνα μετάδοσης R=25m. Τα τελικά αποτελέσματα της προσομοίωσης λήφθηκαν ως ο μέσος όρος προσομοιώσεων πολλών τοπολογιών. Επιλέχθηκε Δt=0.4 ώστε να είναι επαρκώς μικρό και να μπορεί να προσδιοριστεί η λύση της διαφορικής μέσω της εξίσωσης (10). Το μοντέλο αυτό αναλύεται με δύο διαφορετικούς αρχικά μολυσμένους κόμβους. Οι μολυσμένοι κόμβοι από τους οποίους θα ξεκινήσει η διάχυση της πληροφορίας είναι τη μια φορά τυχαία επιλεγμένοι, ενώ στη συνέχεια είναι αυτοί με τους περισσότερους φίλους στο κοινωνικό επίπεδο.

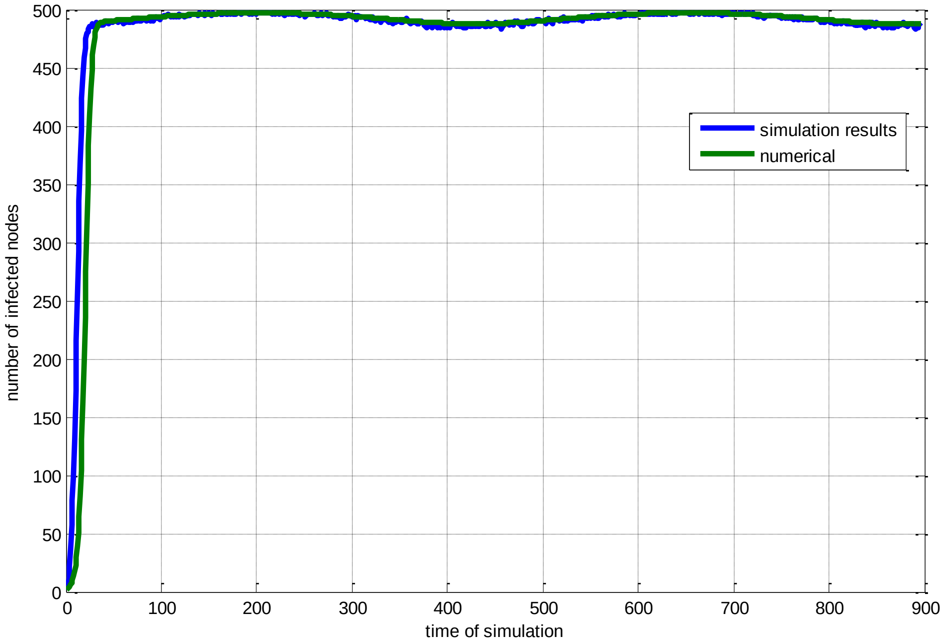
- Περίπτωση 1: Περιοδικό Ενδιαφέρον Χρηστών

Σε αυτή την περίπτωση το ενδιαφέρον για κάθε κλάση θεωρείται ομοιόμορφα και τυχαία μεταβαλλόμενο με μέση τιμή ίση με: $1-\frac{1}{4}(\sin(\frac{\pi}{180}(t+100)))^2-\frac{1}{7}$ για την κλάση m=1 και η μέση τιμή για την άλλη κλάση θα είναι ίση με $\frac{1}{4}(\sin(\frac{\pi}{180}(t+100)))^2+\frac{1}{7}$ δηλαδή $A=\frac{1}{4}$, $B=-\frac{1}{7}$, $a=\frac{\pi}{180}$ $και\, b=100$. Άρα η περίοδος των συναρτήσεων ενδιαφέροντος είναι ένας χρόνος. Οι τιμές p_1, p_2, και q μεταβάλλονται σε κάθε προσομοίωση.

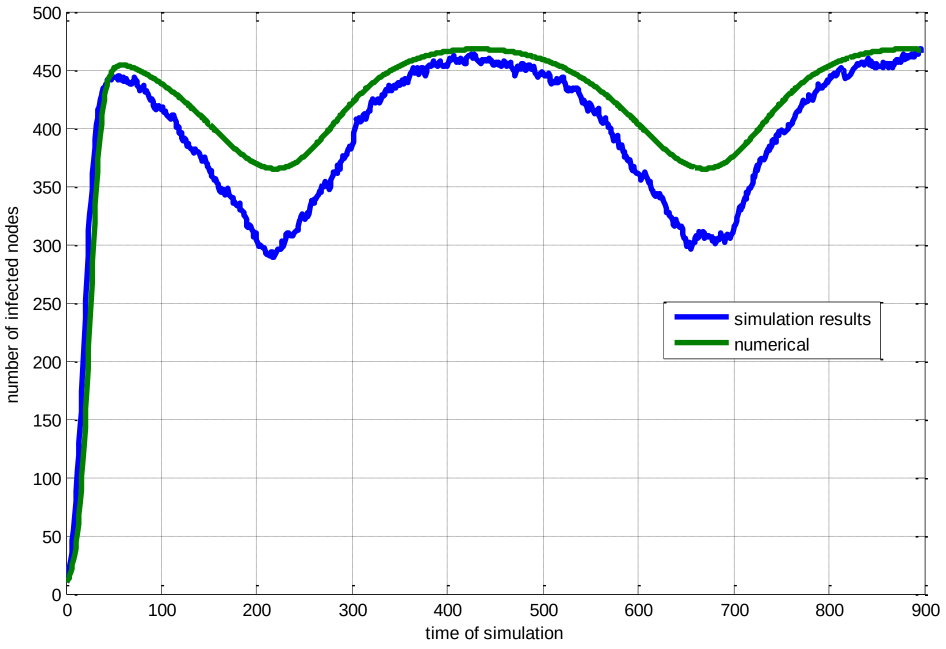
Στο Σχήμα 7.1.1 φαίνεται η δυναμική της διάχυσης της πληροφορίας για την κλάση 1 ξεκινώντας από μολυσμένους κόμβους οι οποίοι είναι τυχαία επιλεγμένοι και στη συνέχεια παρουσιάζεται η διάχυση για την ίδια κλάση ξεκινώντας τη μόλυνση κόμβοι οι οποίοι είναι οι πιο δημοφιλείς κοινωνικά (σχήμα 7.1.2). Στο Σχήμα 7.2.1 για την κλάση 2, δηλαδή ο ρυθμός μόλυνσης των κόμβων ξεκινώντας τη διάχυση από τυχαία επιλεγμένους κόμβους, ενώ στο σχήμα 7.2.2 παρουσιάζεται η διάχυση για την κλάση 2 ξεκινώντας από μολυσμένους κόμβους τους πιο δημοφιλείς στο κοινωνικό επίπεδο. Οι παράμετροι έχουν τις τιμές $p_1=p_2=0.5$ δηλαδή η πιθανότητα να ενημερωθούν οι υπόλοιποι κόμβοι για την κάθε κλάση μέσω MMS ή Wi-Fi/Bluetooth είναι η ίδια. Επίσης Τα ίδια με παραπάνω ισχύουν για τα σχήματα 8.1.1, 8.1.2, 8.2.1, 8.2.2 αντίστοιχα. Η μονη διαφορά είναι η πιθανότητα να διαγράψει ένας κόμβος την πληροφορία, η οποία είναι ίση με q=0.2 για το σχήμα 7 και ίση με q=0.6 στο σχήμα 8.



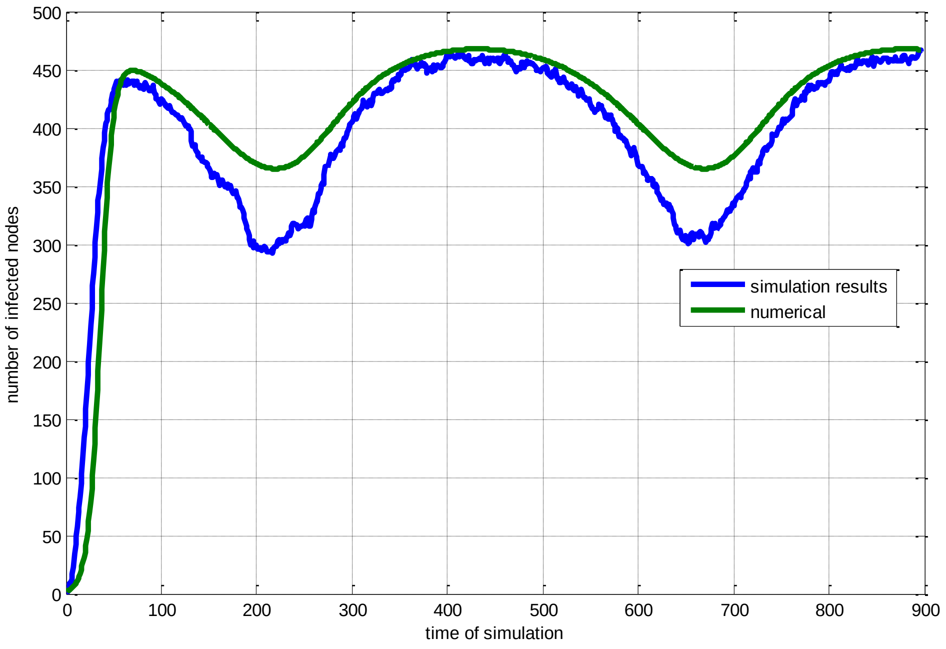
Σχήμα 7.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 7.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1

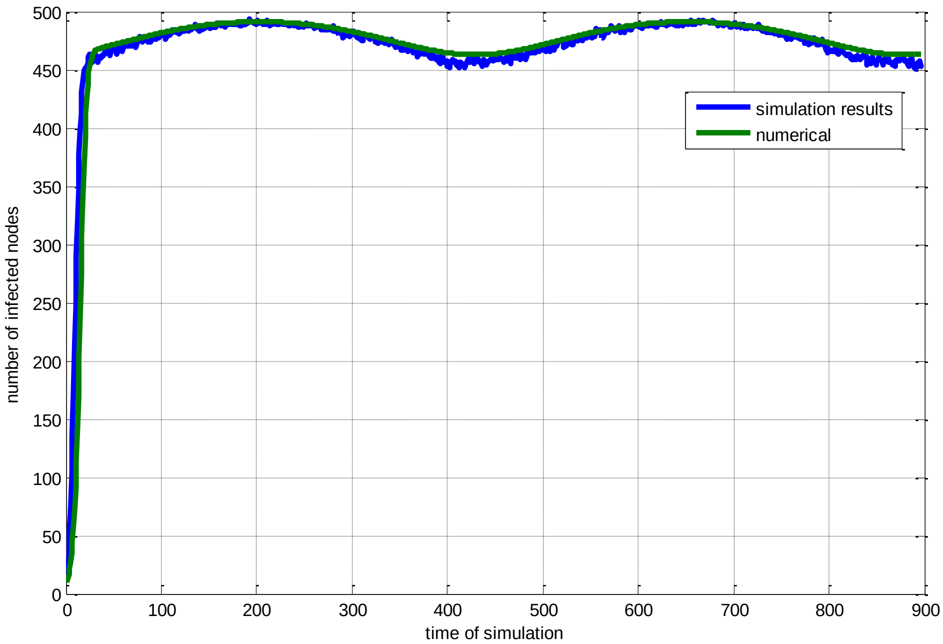


Σχήμα 7.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2

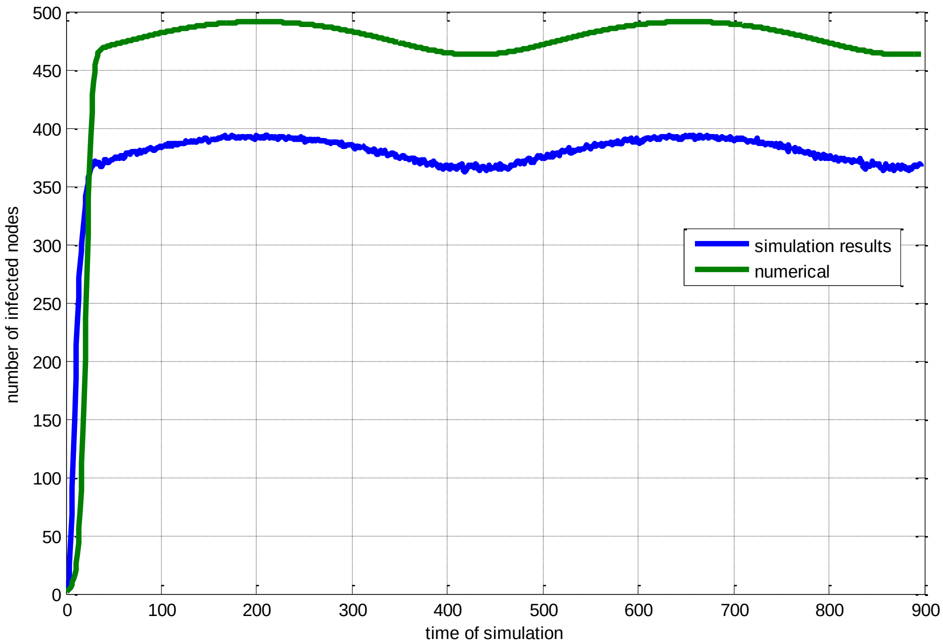


Σχήμα 7.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

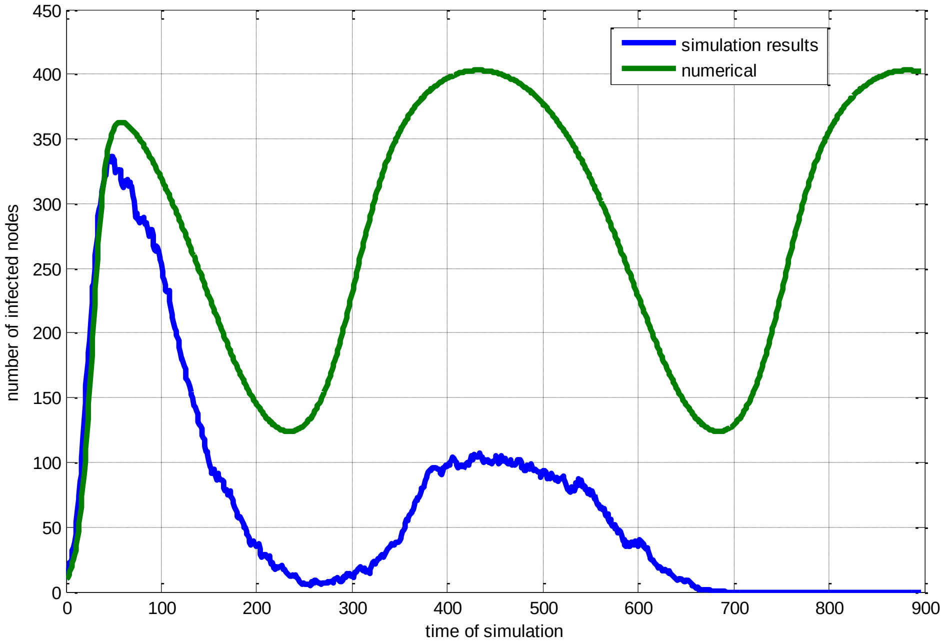
Σχήμα 7: Διάχυση πληροφορίας με ίση πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, δηλαδή p_1=p_2=0.5 και q=0.2



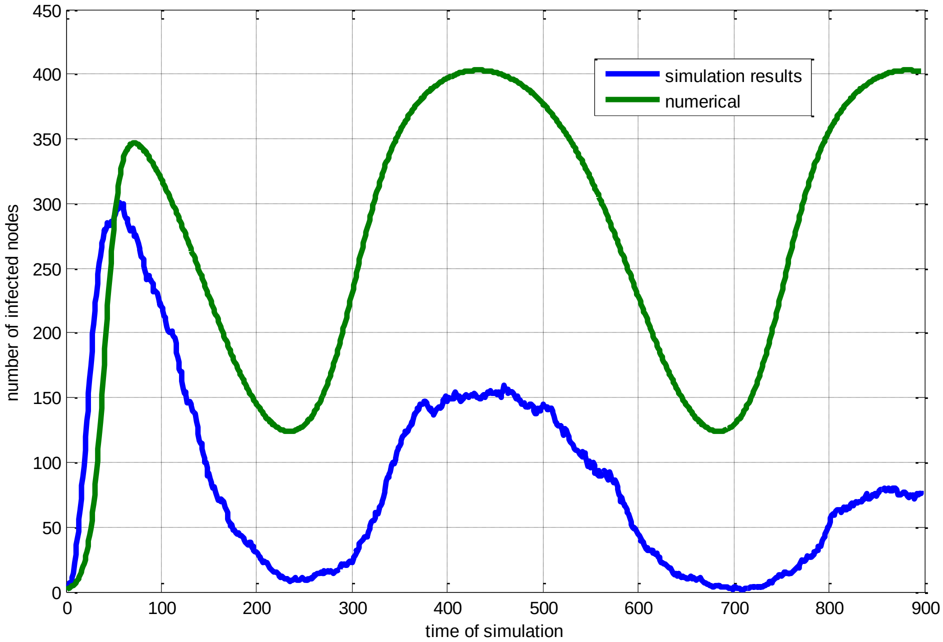
Σχήμα 8.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 8.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1



Σχήμα 8.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2



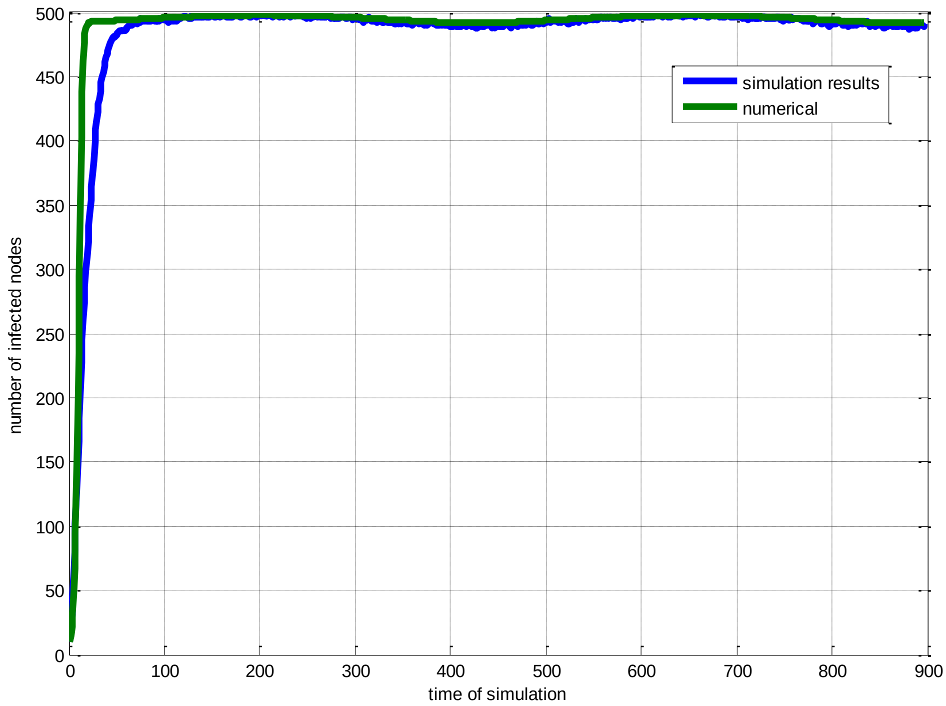
Σχήμα 8.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=22

Σχήμα 8: Διάχυση πληροφορίας με ίση πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, δηλαδή p_1=p_2=0.5 και q=0.6

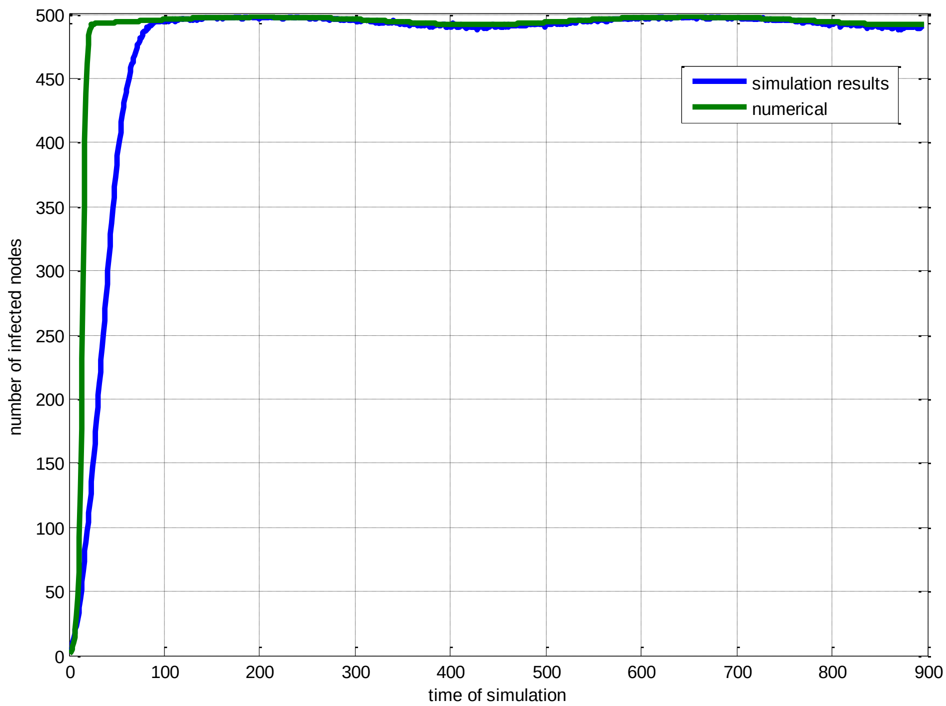
Από τις παραπάνω γραφικές αναπαραστάσεις είναι φανερό ότι η μαθηματική αναπαράσταση της διάχυσης της πληροφορίας πλησιάζει αρκετά αυτή της προσομοίωσης αν και σε συγκεκριμένες περιπτώσεις, όταν το ενδιαφέρον είναι μικρότερο, όπως στην κλάση 2 , το μαθηματικό μοντέλο υπερεκτιμά την προσομοίωση. Όταν στην προσομοίωση συμβεί κάποια διαγραφή μηνύματος από έναν κόμβο τότε λόγω του ότι παίρνουμε το μέσο όρο των προσομοιώσεων, μπορεί να μηδενιστεί το πλήθος των κόμβων με την πληροφορία αυτή, γεγονός το οποίο δεν συμβαίνει στο μαθηματικό μοντέλο Στο μαθηματικό μοντέλο δεν σταματάει η διάχυση της πληροφορίας με τη διαγραφή μηνυμάτων. Γι αυτό το λόγο με αύξηση της πιθανότητας διαγραφής της πληροφορίας (q) από τους κόμβους παρατηρείται απότομη μείωση των μολυσμένων κόμβων στην προσομοίωση, την οποία δεν προβλέπει τόσο άμεσα το μαθηματικό μοντέλο , με αποτέλεσμα να φαίνεται και εδώ υπερεκτίμηση του μαθηματικού μοντέλου για τον αριθμό των μολυσμένων κόμβων.

Στο Σχήμα 9.1.1 φαίνεται η δυναμική της διάχυσης της πληροφορίας για την κλάση 1 ξεκινώντας από μολυσμένους κόμβους οι οποίοι είναι τυχαία επιλεγμένοι και στη συνέχεια παρουσιάζεται η διάχυση για την ίδια κλάση ξεκινώντας τη μόλυνση κόμβοι οι οποίοι είναι οι πιο δημοφιλείς κοινωνικά (σχήμα 9.1.2). Στο Σχήμα 9.2.1 αναπαριστάται η διάχυση για την κλάση 2, δηλαδή ο ρυθμός μόλυνσης των κόμβων ξεκινώντας τη διάχυση από τυχαία επιλεγμένους κόμβους, ενώ στο σχήμα 9.2.2 παρουσιάζεται η διάχυση για την κλάση 2 ξεκινώντας από μολυσμένους κόμβους τους πιο δημοφιλείς στο κοινωνικό επίπεδο. Επίσης τα ίδια με παραπάνω ισχύουν για τα σχήματα 10.1.1, 10.1.2, 10.2.1, 10.2.2 αντίστοιχα. Η μόνη διαφορά είναι η πιθανότητα να διαγράψει ένας κόμβος την πληροφορία, η οποία είναι ίση με q=0.2 για το σχήμα 9 και ίση με q=0.6 στο σχήμα 10.

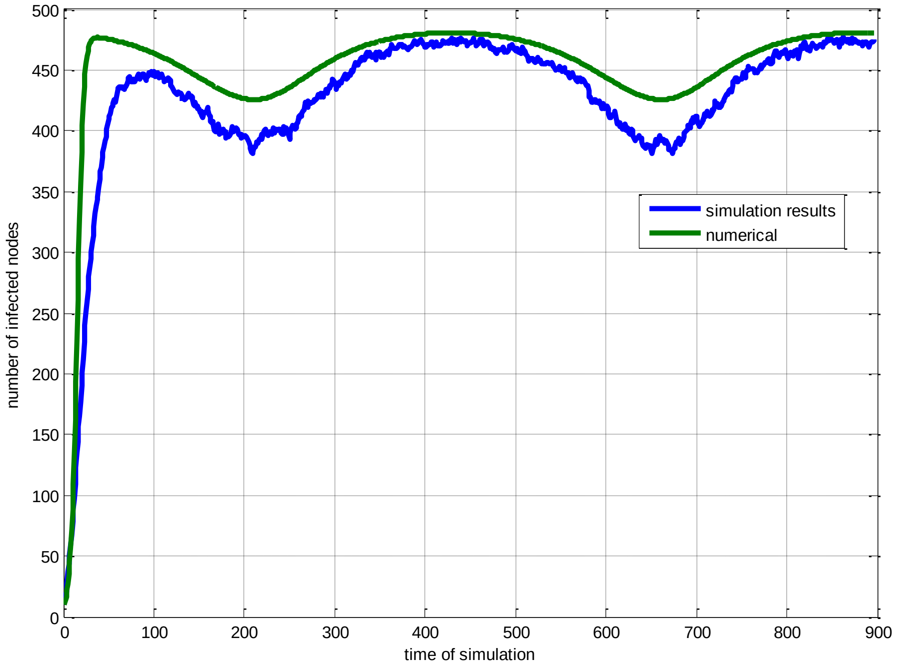
Οι παράμετροι έχουν τις τιμές $p_1=0$,$p_2=1$ δηλαδή η πιθανότητα να γίνει η διάχυση μέσω MMS είναι μηδενική. Άρα η διάχυση γίνεται μέσω Wi-Fi/Bluetooth.



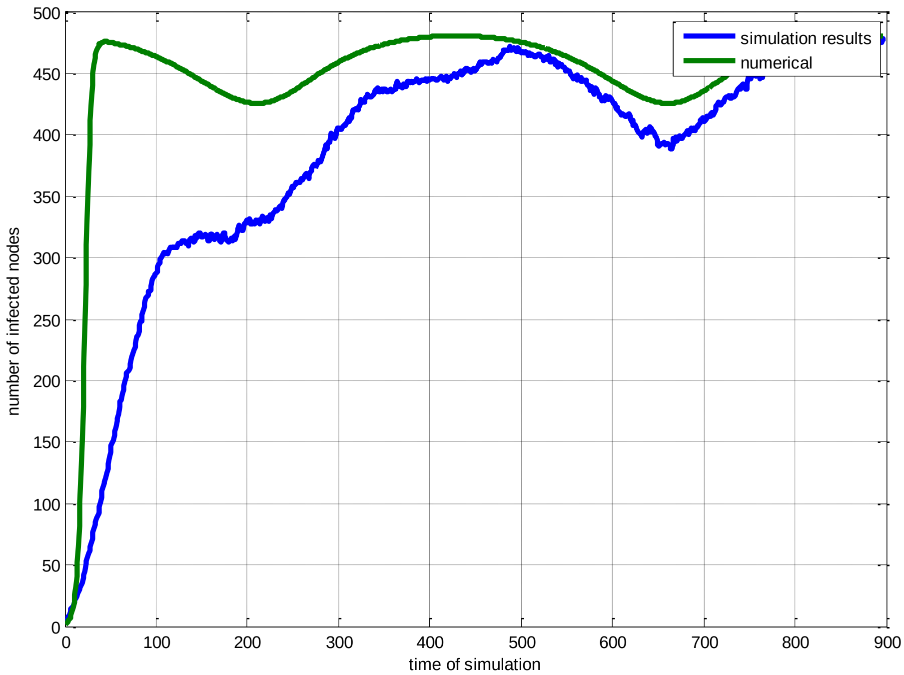
Σχήμα 9.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 9.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1

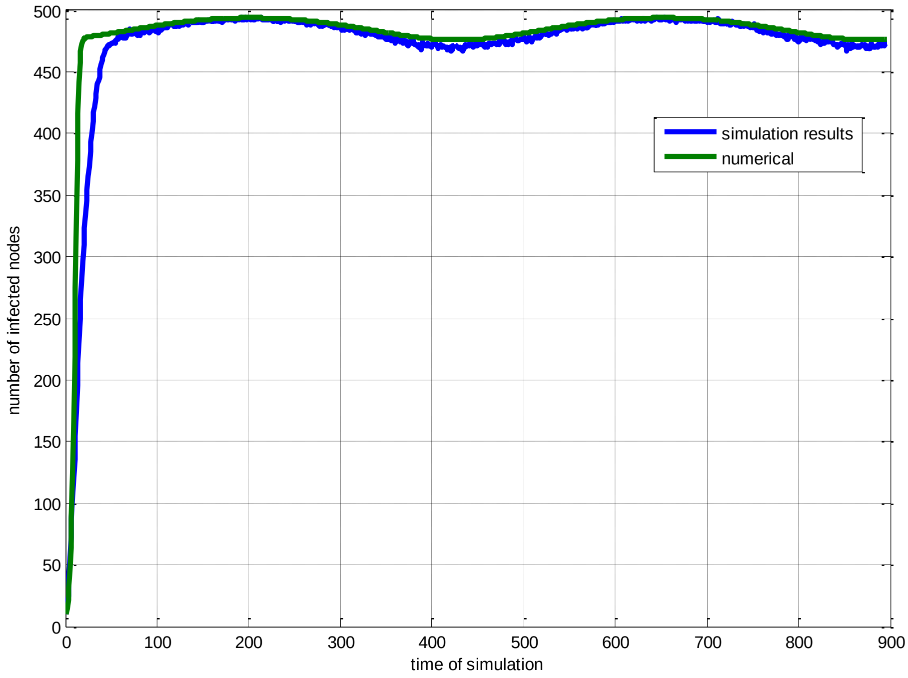


Σχήμα 9.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2

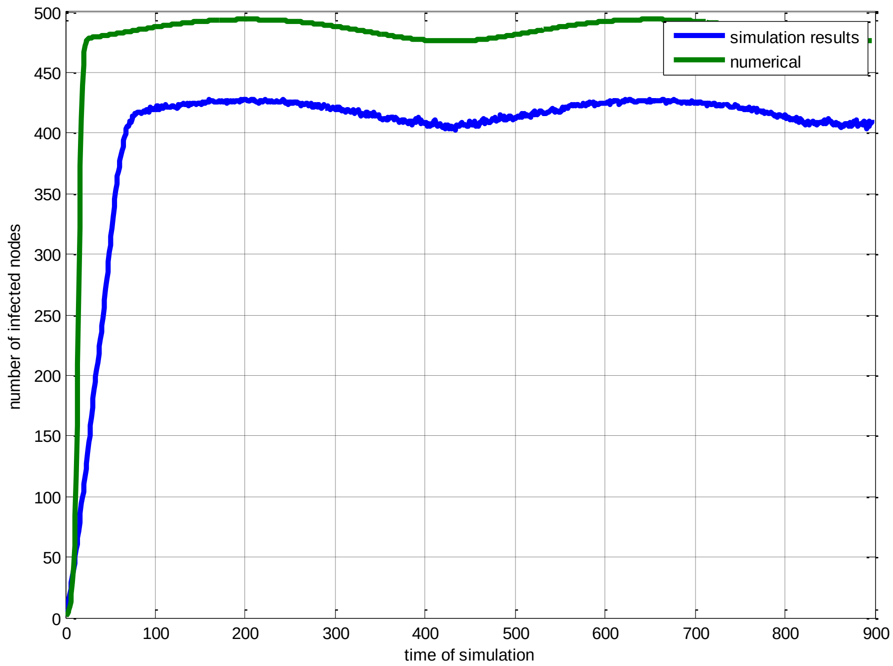


Σχήμα 9.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

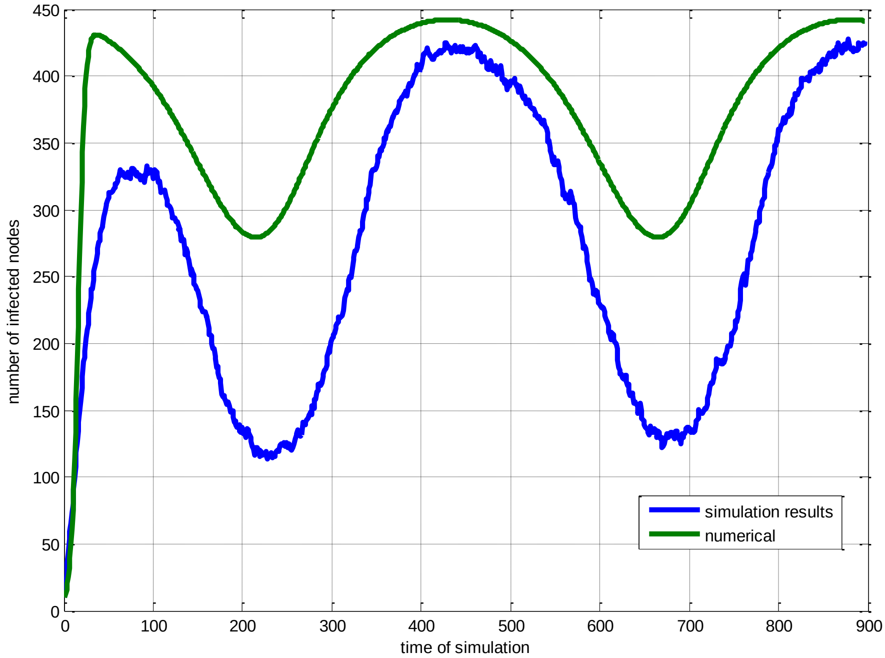
Σχήμα 9: Διάχυση πληροφορίας με πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, p_1=0 p_2=1 και q=0.2



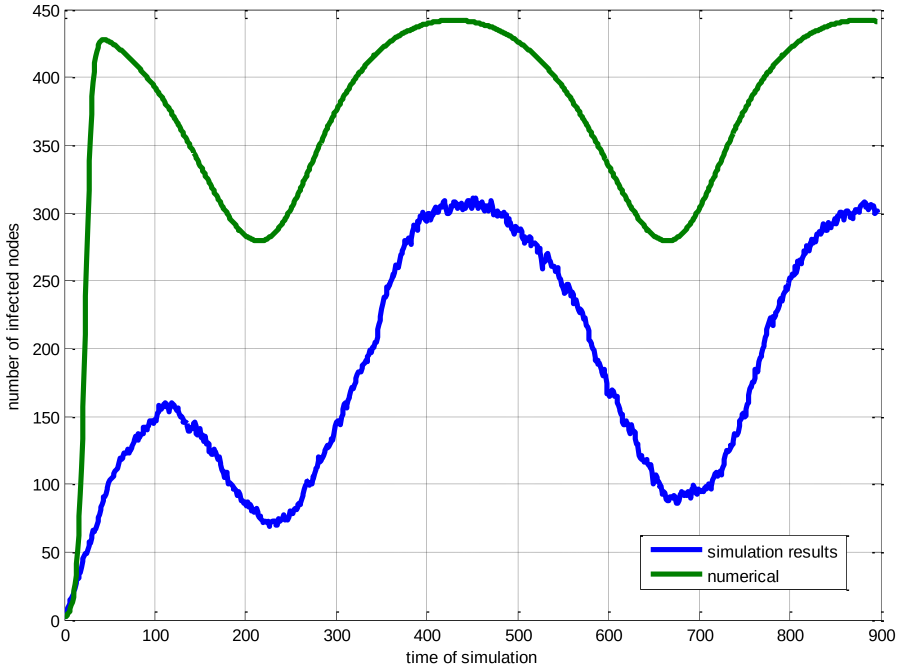
Σχήμα 10.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 10.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1



Σχήμα 10.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2



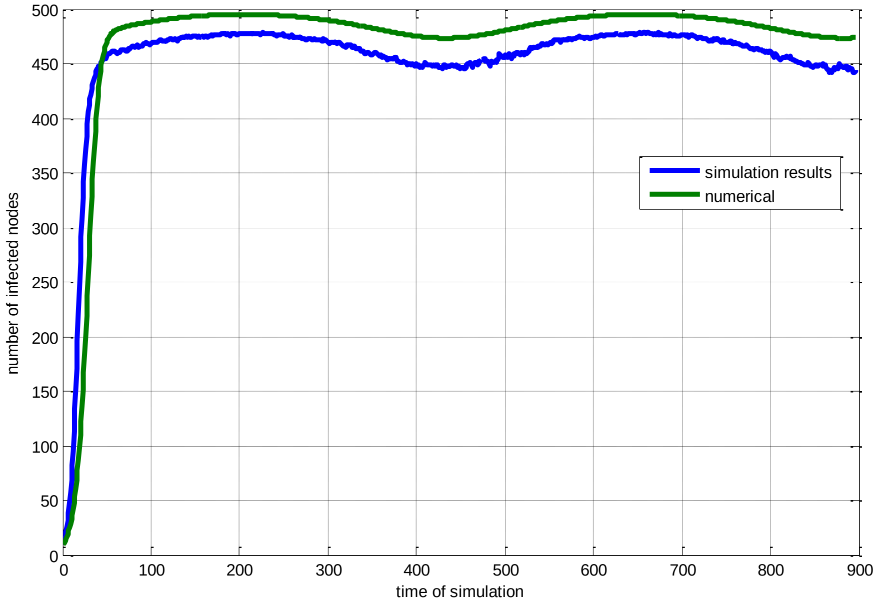
Σχήμα 10.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

Σχήμα 10: Διάχυση πληροφορίας με πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, p_1=0 p_2=1 και q=0.6

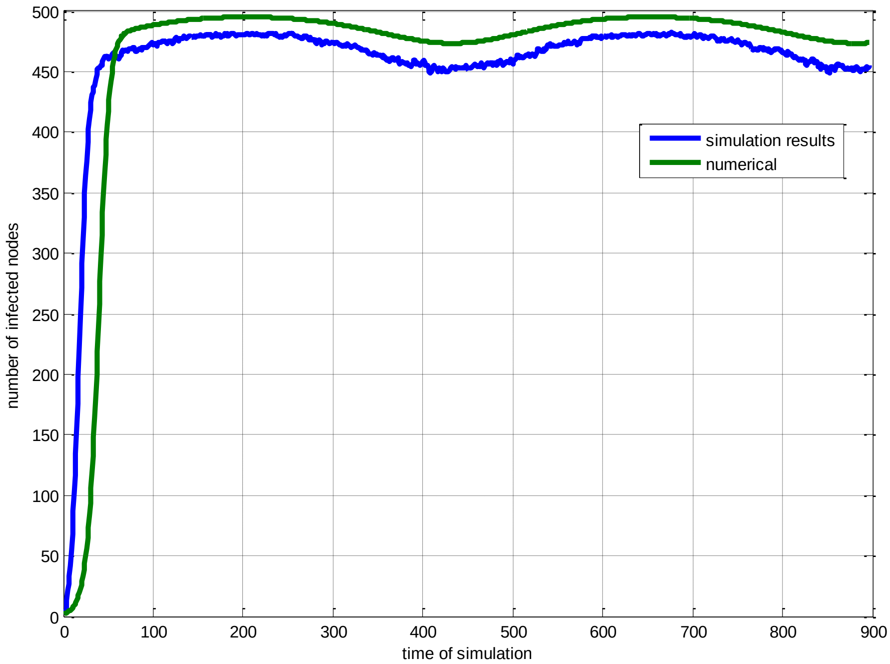
Στην περίπτωση αυτή ισχύει ότι και προηγουμένως για τη μαθηματική ανάλυση και την προσομοίωση. Επιπλέον, παρατηρείται ότι η διάχυση της πληροφορίας μέσω Wi-Fi/Bluetooth, δηλαδή μέσω του φυσικού επίπεδου μοιάζει πολύ με την παραπάνω διάχυση όπου η διάχυση γινόταν μέσω MMS και Wi-Fi/Bluetooth με την ίδια πιθανότητα.

Στο Σχήμα 11.1.1 φαίνεται η δυναμική της διάχυσης της πληροφορίας για την κλάση 1 ξεκινώντας από μολυσμένους κόμβους οι οποίοι είναι τυχαία επιλεγμένοι και στη συνέχεια παρουσιάζεται η διάχυση για την ίδια κλάση ξεκινώντας τη μόλυνση κόμβοι οι οποίοι είναι οι πιο δημοφιλείς κοινωνικά (σχήμα 11.1.2). Στο Σχήμα 11.2.1 αναπαριστάται η διάχυση για την κλάση 2, δηλαδή ο ρυθμός μόλυνσης των κόμβων ξεκινώντας τη διάχυση από τυχαία επιλεγμένους κόμβους, ενώ στο σχήμα 11.2.2 παρουσιάζεται η διάχυση για την κλάση 2 ξεκινώντας από μολυσμένους κόμβους τους πιο δημοφιλείς στο κοινωνικό επίπεδο. Επίσης τα ίδια με παραπάνω ισχύουν για τα σχήματα 12.1.1, 12.1.2, 12.2.1, 12.2.2 αντίστοιχα. Η μονη διαφορά είναι η πιθανότητα να διαγράψει ένας κόμβος την πληροφορία, η οποία είναι ίση με q=0.2 για το σχήμα 11 και ίση με q=0.6 στο σχήμα 12.

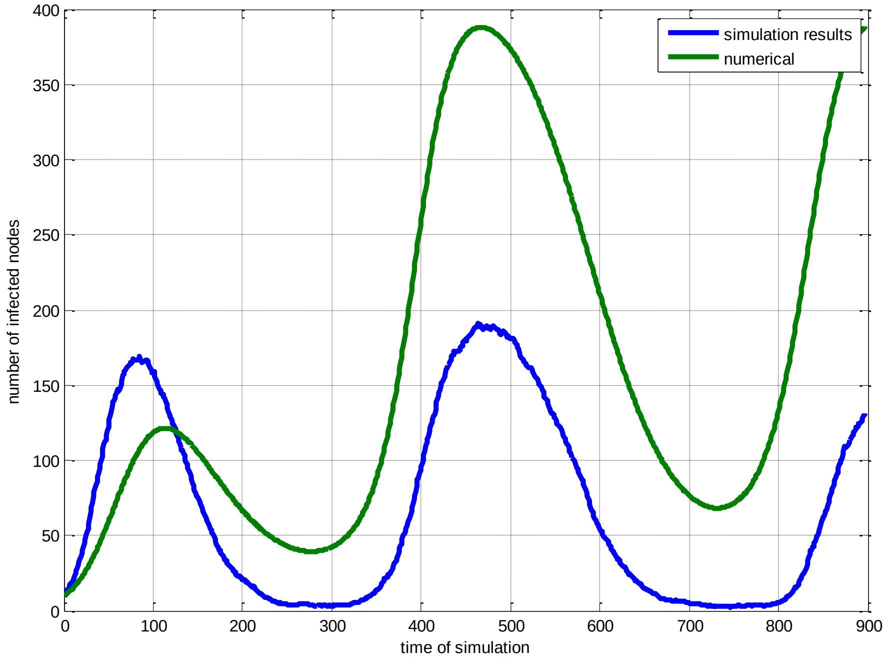
Οι παράμετροι όμως τώρα έχουν τις τιμές p_1=1 ,p_2=0 δηλαδή η πιθανότητα να γίνει η διάχυση μέσω Wi-Fi/Bluetooth είναι μηδενική. Άρα η διάχυση γίνεται μέσω MMS.



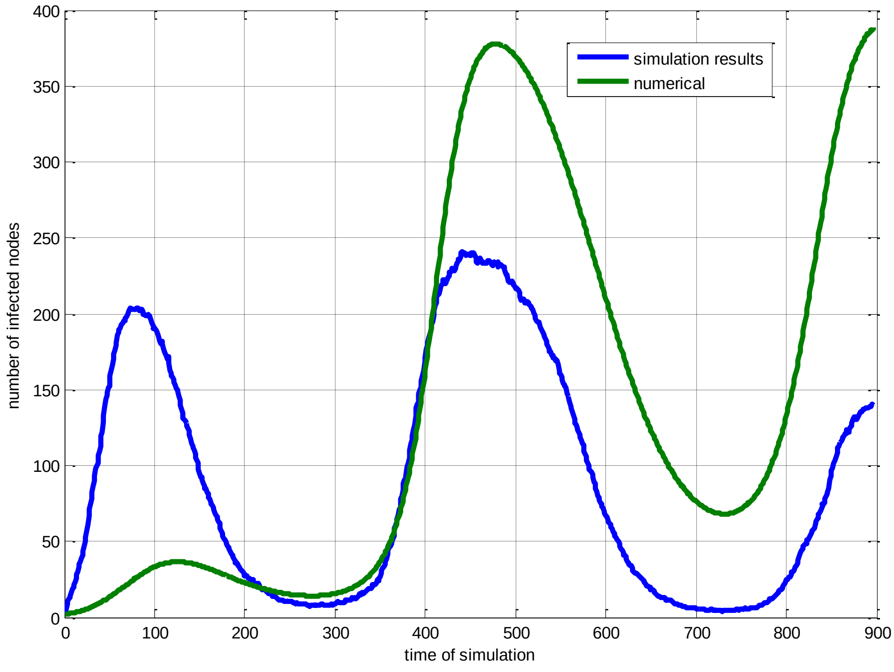
Σχήμα 11.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 11.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1

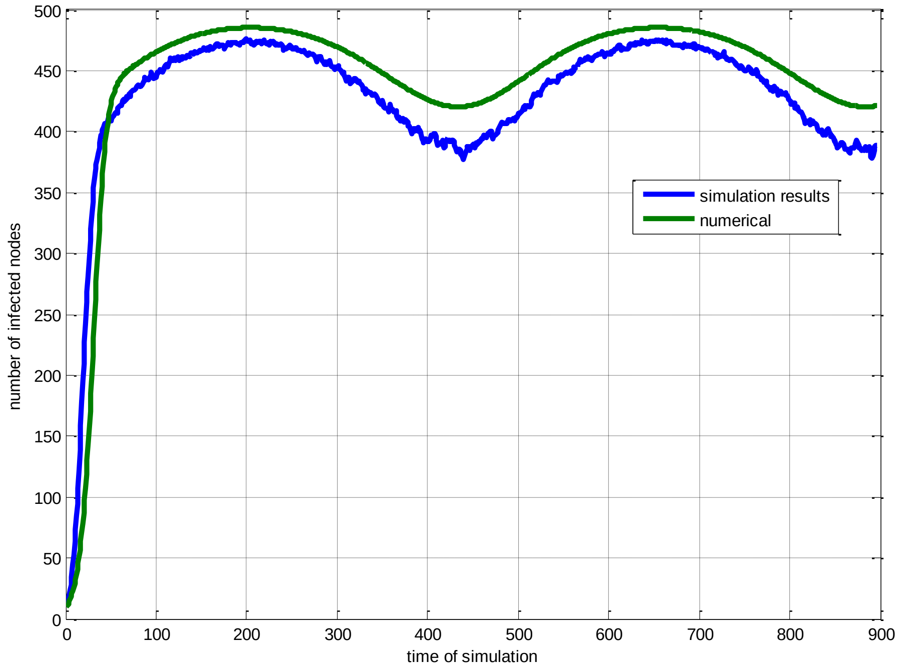


Σχήμα 11.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2

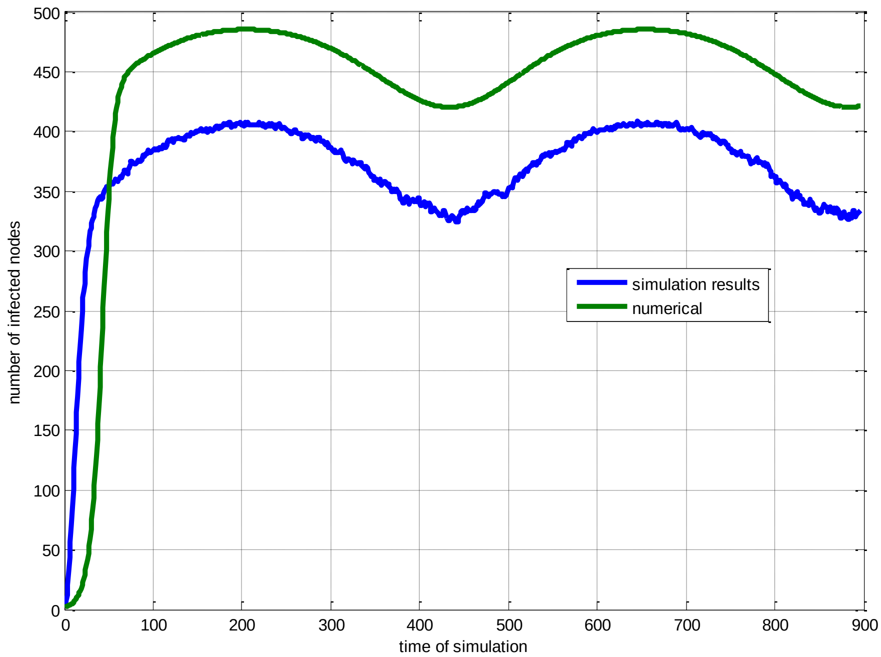


Σχήμα 11.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

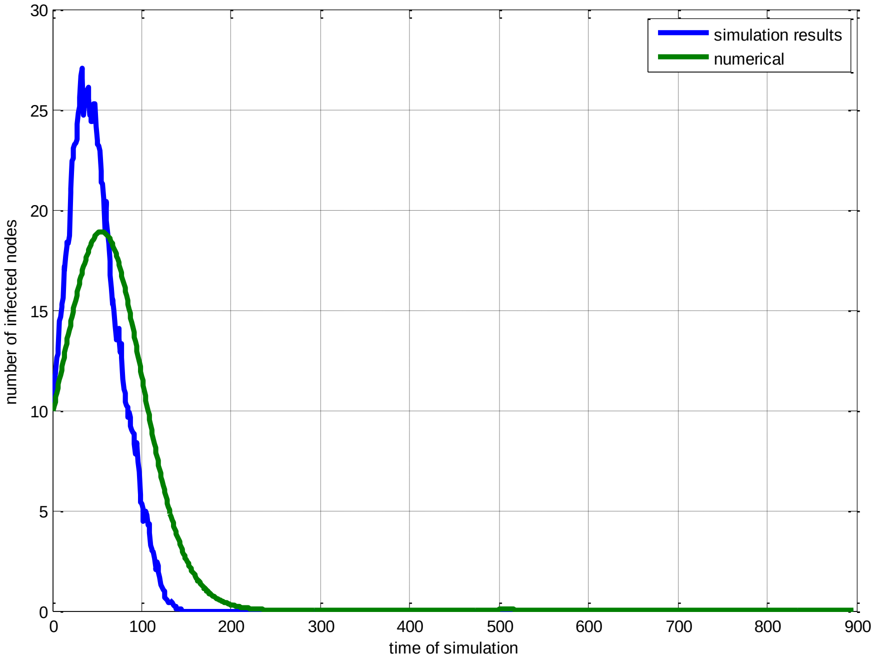
Σχήμα 11: Διάχυση πληροφορίας με ίση πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, δηλαδή p_1=1 p_2=0 και q=0.2



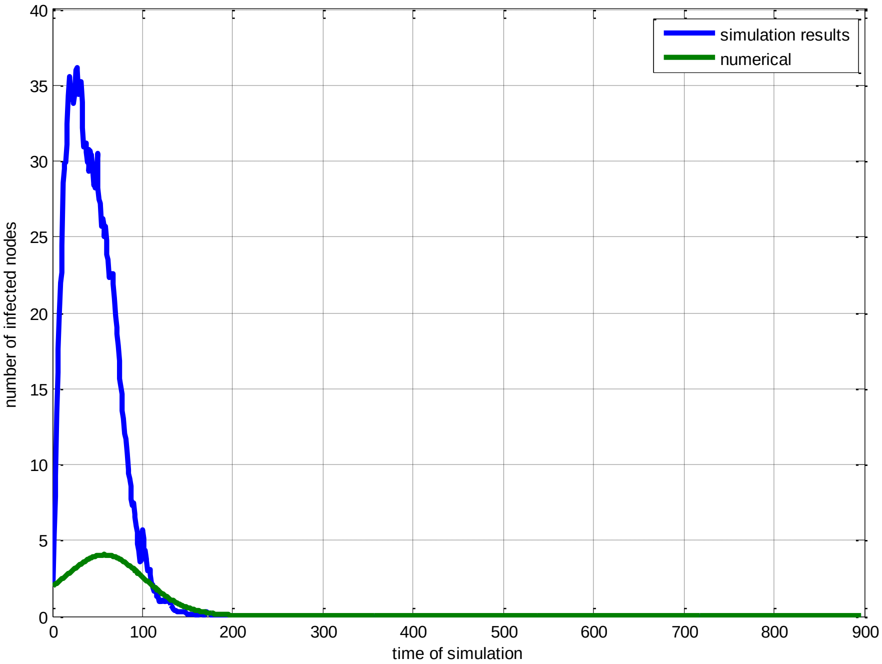
Σχήμα 12.1.1Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 12.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1



Σχήμα 12.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2



Σχήμα 12.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

Σχήμα 12: Διάχυση πληροφορίας με πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, p_1=1 p_2=0 και q=0.6

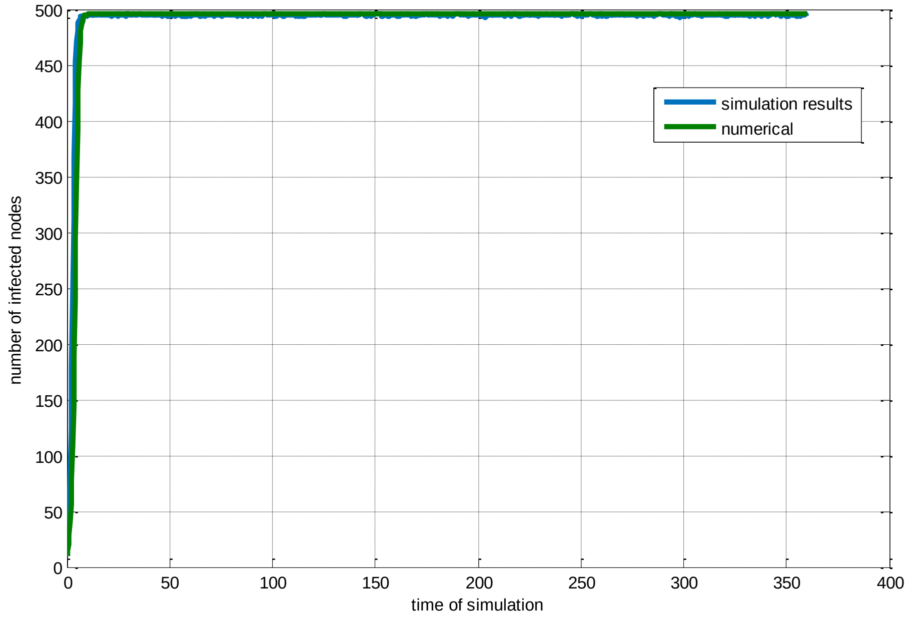
Άρα τελικά η διάχυση μέσω του κοινωνικού επιπέδου επηρεάζεται πολύ όταν αυξάνεται η πιθανότητα διαγραφής των μηνυμάτων από τους κόμβους καθώς επίσης όταν το ενδιαφέρον είναι μικρότερο, όπως συμβαίνει για την κλάση 2, με αποτέλεσμα να πέφτει απότομα. Είναι φανερό ότι όταν η διάχυση γίνεται μέσω του φυσικού στρώματος διατηρείται ζωντανή και δεν επηρεάζεται τόσο από τα παραπάνω, σε αντίθεση με τη διάχυση μέσω κοινωνικού επιπέδου.

- Περίπτωση 2: Σταθερά ενδιαφέροντα

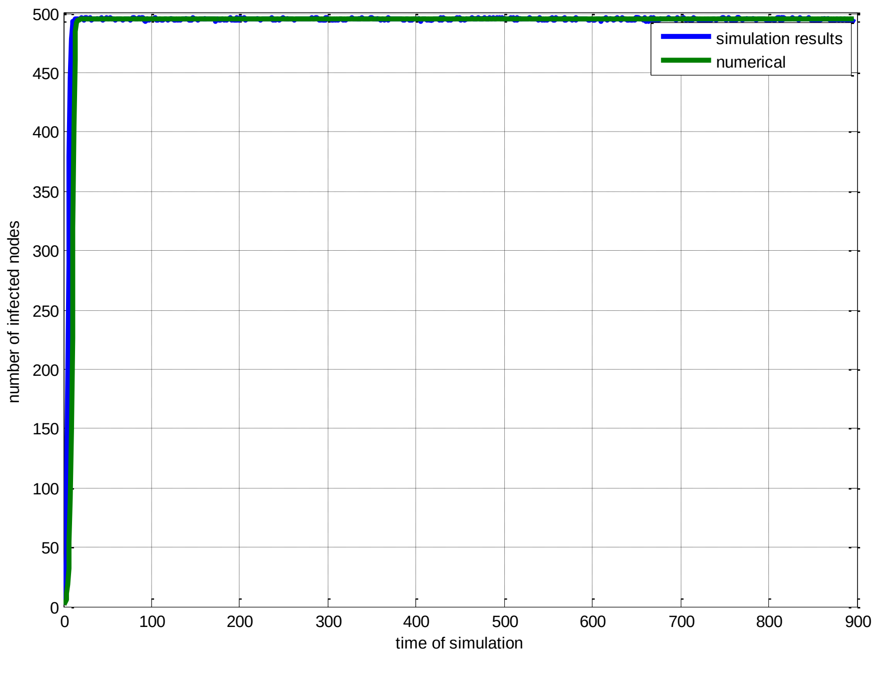
Σε αυτή την περίπτωση θεωρούμε δύο διαφορετικές ομάδες με διαφορετικά, σταθερά ενδιαφέροντα. Η ομάδα 1 έχει ενδιαφέρον ίσο με 0.8 και η ομάδα 2 ίσο με 0.2. Τα αποτελέσματα από το Matlab για την περίπτωση αυτή, αναλύοντας τη με μαθηματική ανάλυση και με προσομοίωση, είναι τα παρακάτω:

Στο Σχήμα 13.1.1 φαίνεται η δυναμική της διάχυσης της πληροφορίας για την κλάση 1 ξεκινώντας από μολυσμένους κόμβους οι οποίοι είναι τυχαία επιλεγμένοι και στη συνέχεια παρουσιάζεται η διάχυση για την ίδια κλάση ξεκινώντας τη μόλυνση κόμβοι οι οποίοι είναι οι πιο δημοφιλείς κοινωνικά (σχήμα 13.1.2). Στο Σχήμα 13.2.1 αναπαριστάται η διάχυση για την κλάση 2, δηλαδή ο ρυθμός μόλυνσης των κόμβων ξεκινώντας τη διάχυση από τυχαία επιλεγμένους κόμβους, ενώ στο σχήμα 13.2.2 παρουσιάζεται η διάχυση για την κλάση 2 ξεκινώντας από μολυσμένους κόμβους τους πιο δημοφιλείς στο κοινωνικό επίπεδο. Επίσης τα ίδια με παραπάνω ισχύουν για τα σχήματα 14.1.1, 14.1.2, 14.2.1, 14.2.2 αντίστοιχα. Η μόνη διαφορά είναι η πιθανότητα να διαγράψει ένας κόμβος την πληροφορία, η οποία είναι ίση με q=0.2 για το σχήμα 13 και ίση με q=0.6 στο σχήμα 14.

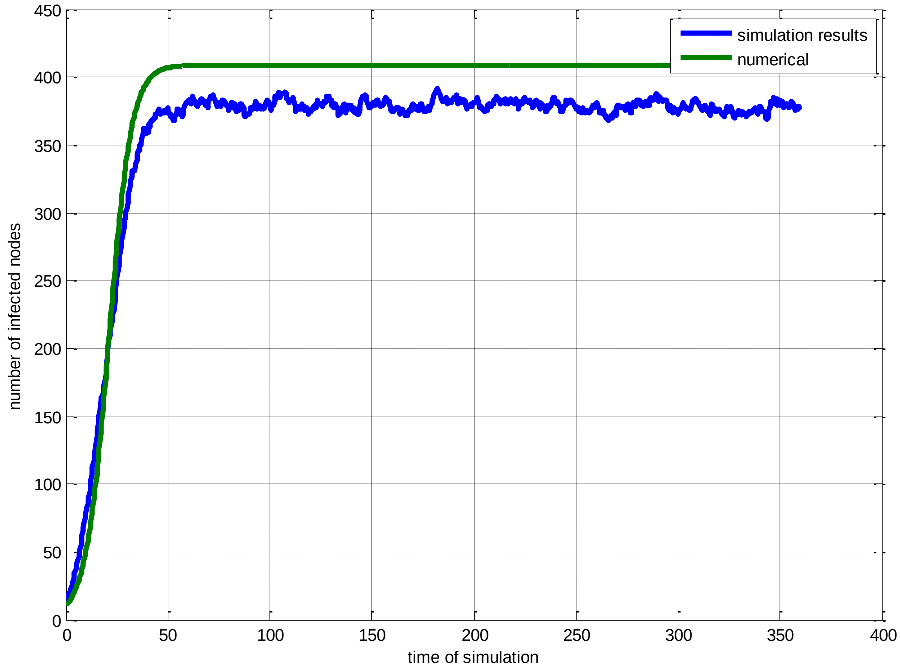
Οι παράμετροι έχουν τις τιμές p_1=0.5 ,p_2=0.5 δηλαδή η πιθανότητα να γίνει η διάχυση μέσω MMS είναι ίδια με την πιθανότητα να γίνει διάχυση μέσω Wi-Fi/Bluetooth.



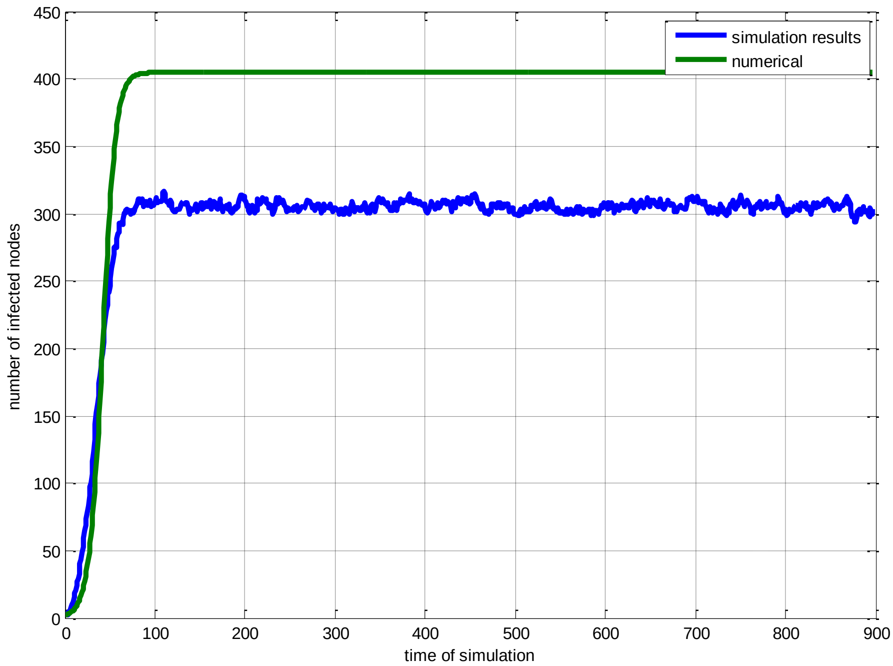
Σχήμα 13.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 13.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1

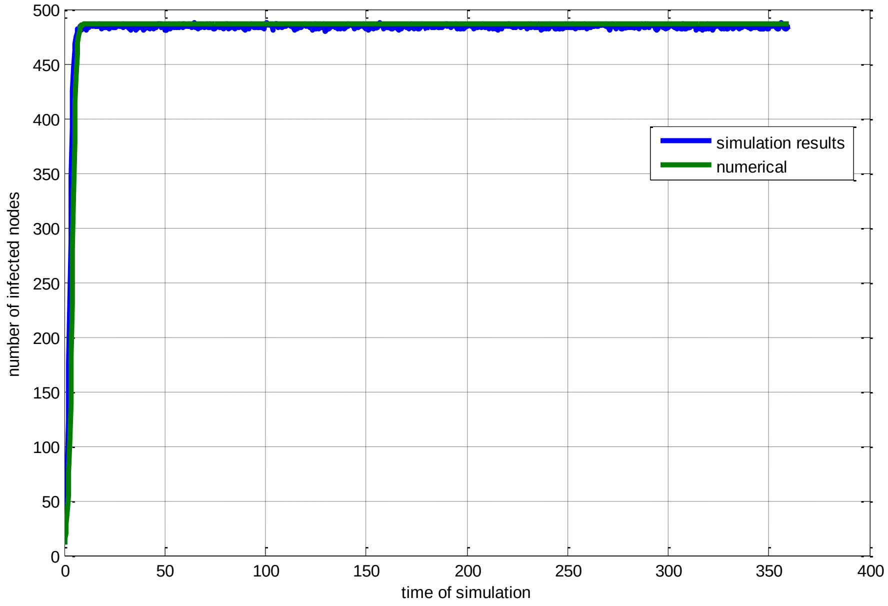


Σχήμα 13.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2

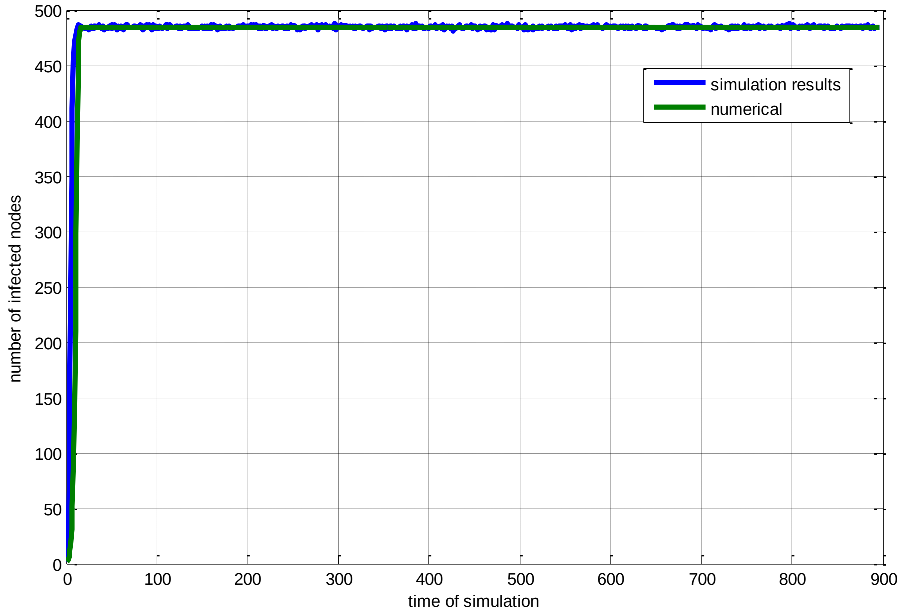


Σχήμα 13.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

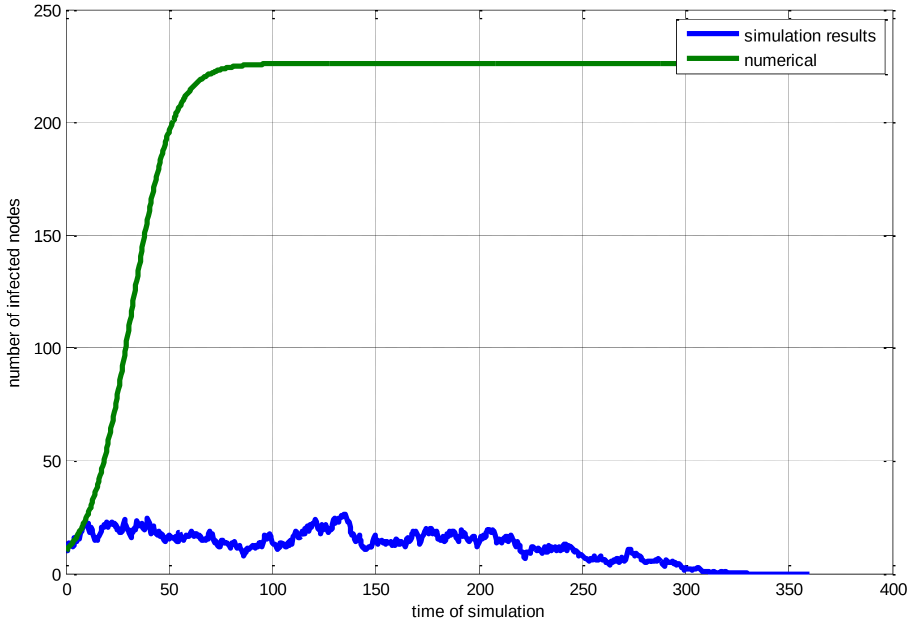
Σχήμα 13: Διάχυση πληροφορίας με ίση πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, δηλαδή p_1=p_2=0.5 και q=0.2



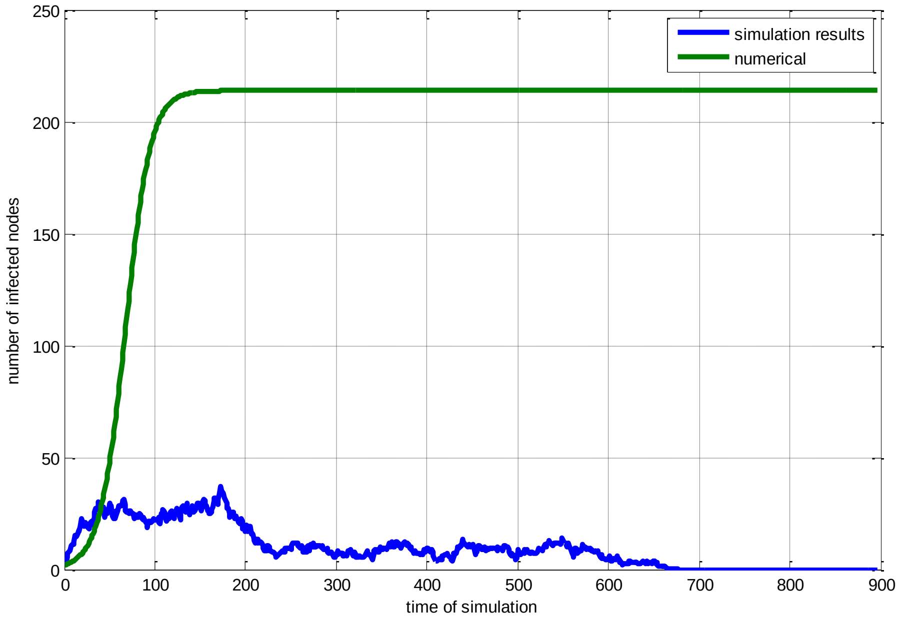
Σχήμα 14.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 14.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1



Σχήμα 14.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2



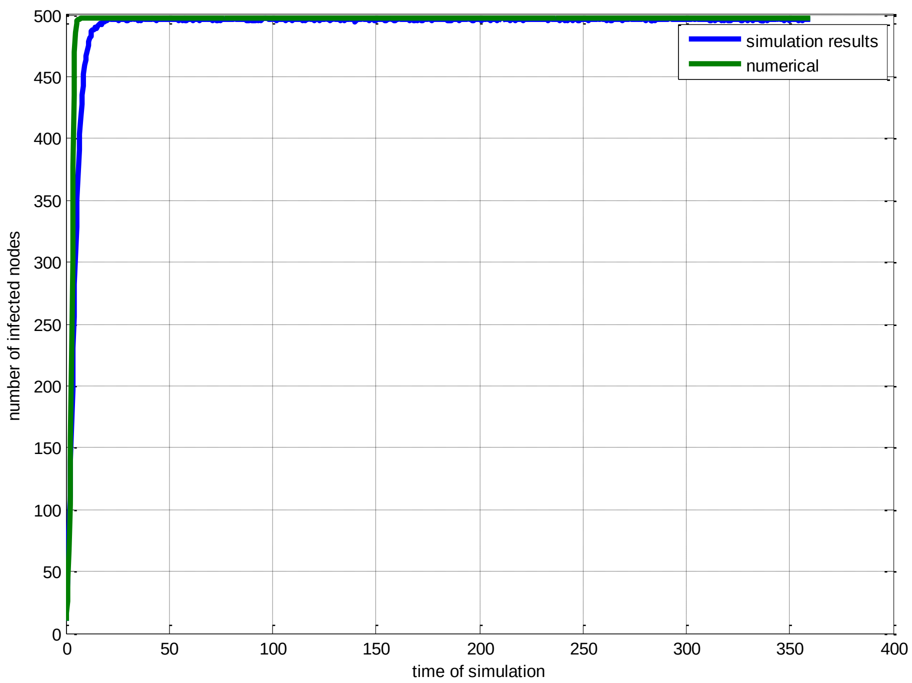
Σχήμα 14.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

Σχήμα 14: Διάχυση πληροφορίας με ίση πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, δηλαδή p_1=p_2=0.5 και q=0.6

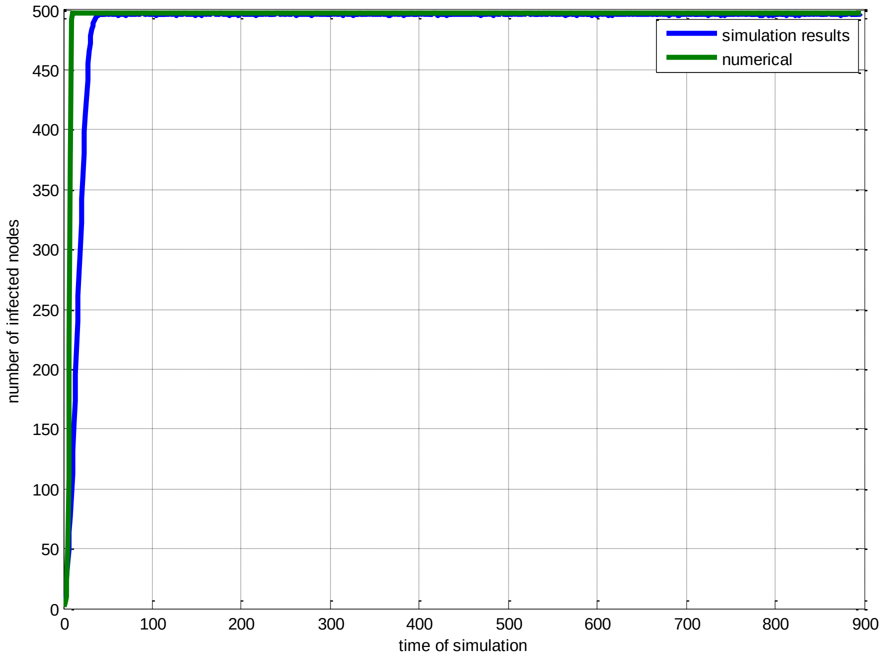
Παρατηρούμε ότι και πάλι το μαθηματικό μοντέλο είναι πολύ κοντά στην προσομοίωση. Στην περίπτωση όμως που η πιθανότητα διαγραφής μηνύματος από τους κόμβους είναι μεγαλύτερη καθώς επίσης όταν το ενδιαφέρον της ομάδας είναι μικρότερο, το μαθηματικό μοντέλο υπερεκτιμά την προσομοίωση. Η διάχυση είτε από αρχικά τυχαία μολυσμένους κόμβους ή από αρχικά μολυσμένους τους πιο δημοφιλείς στο κοινωνικό επίπεδο είναι η ίδια.

Στο Σχήμα 15.1.1 φαίνεται η δυναμική της διάχυσης της πληροφορίας για την κλάση 1 ξεκινώντας από μολυσμένους κόμβους οι οποίοι είναι τυχαία επιλεγμένοι και στη συνέχεια παρουσιάζεται η διάχυση για την ίδια κλάση ξεκινώντας τη μόλυνση κόμβοι οι οποίοι είναι οι πιο δημοφιλείς κοινωνικά (σχήμα 15.1.2). Στο Σχήμα 15.2.1 αναπαριστάται η διάχυση για την κλάση 2, δηλαδή ο ρυθμός μόλυνσης των κόμβων ξεκινώντας τη διάχυση από τυχαία επιλεγμένους κόμβους, ενώ στο σχήμα 15.2.2 παρουσιάζεται η διάχυση για την κλάση 2 ξεκινώντας από μολυσμένους κόμβους τους πιο δημοφιλείς στο κοινωνικό επίπεδο. Επίσης τα ίδια με παραπάνω ισχύουν για τα σχήματα 16.1.1, 16.1.2, 16.2.1, 16.2.2 αντίστοιχα. Η μόνη διαφορά είναι η πιθανότητα να διαγράψει ένας κόμβος την πληροφορία, η οποία είναι ίση με q=0.2 για το σχήμα 15 και ίση με q=0.6 στο σχήμα 16.

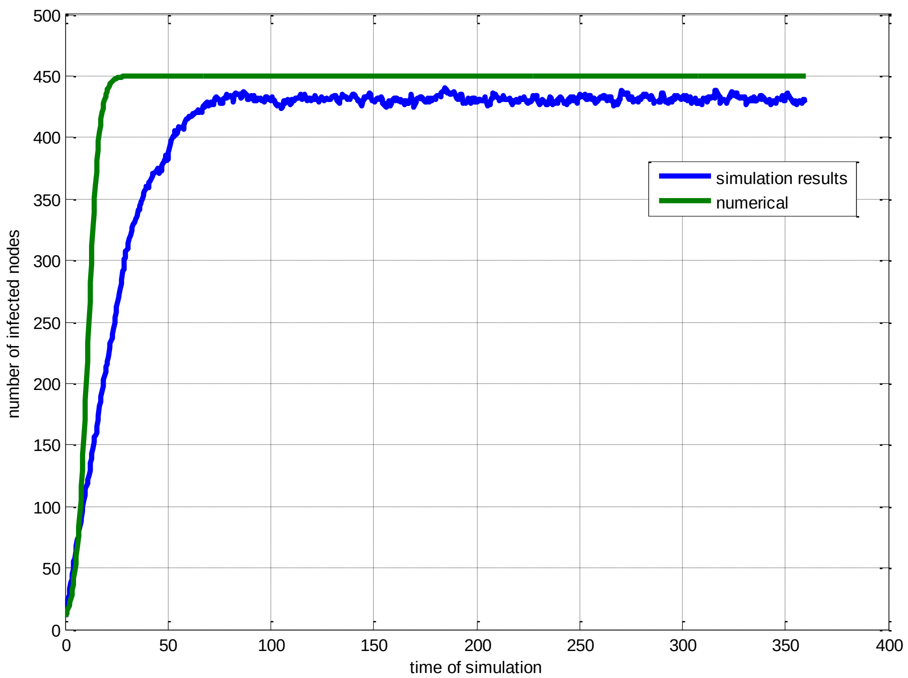
Οι παράμετροι όμως τώρα έχουν τις τιμές p_1=0 ,p_2=1 δηλαδή η πιθανότητα να γίνει η διάχυση μέσω MMS είναι μηδενική. Άρα η διάχυση γίνεται μέσω Wi-Fi/Bluetooth.



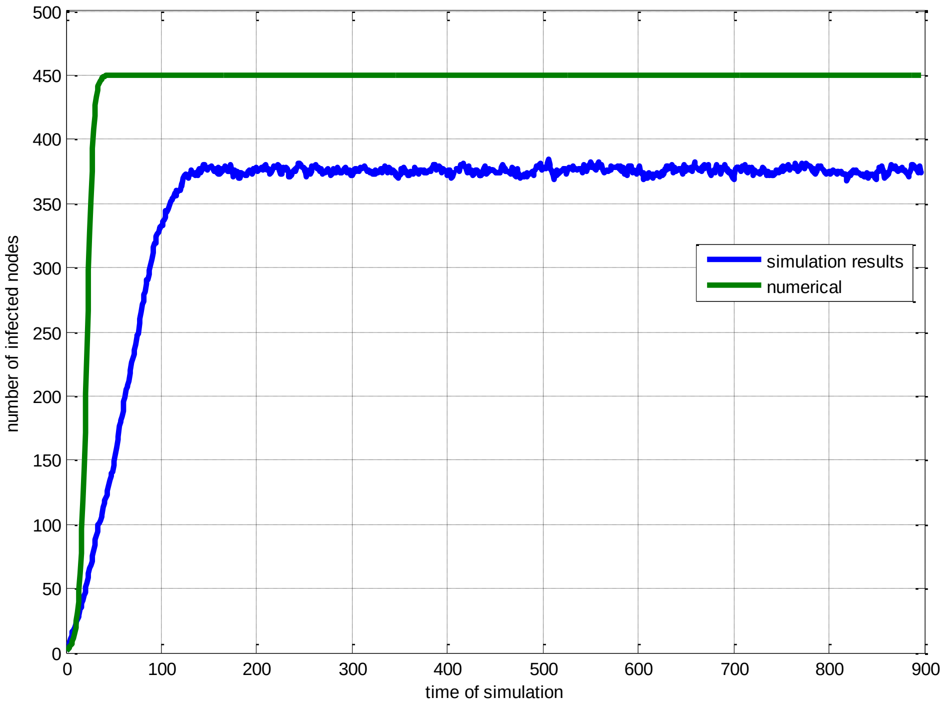
Σχήμα 15.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 15.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1

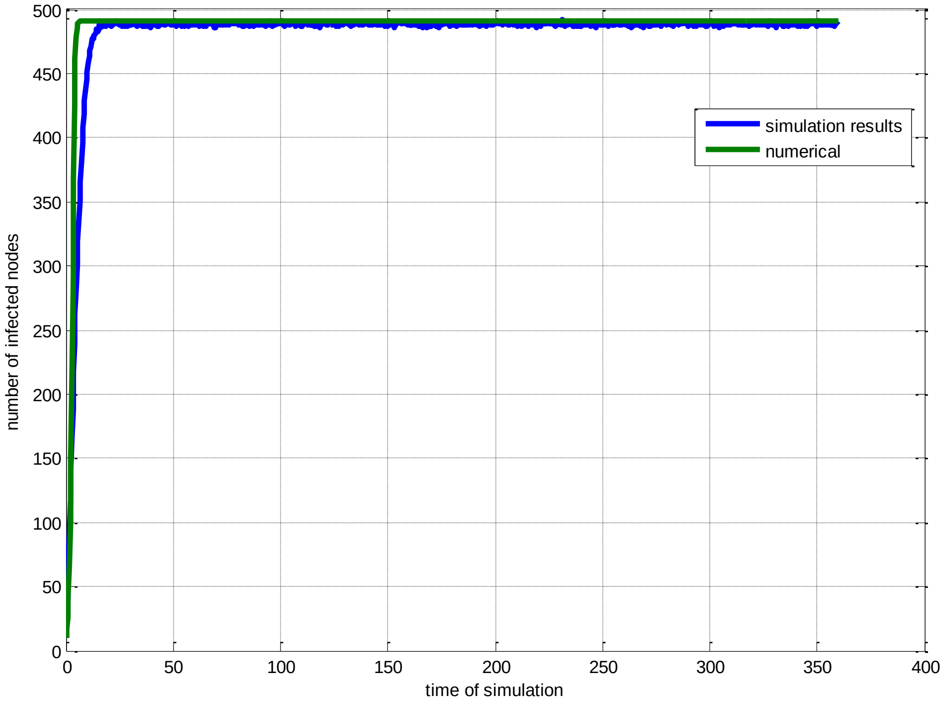


Σχήμα 15.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2

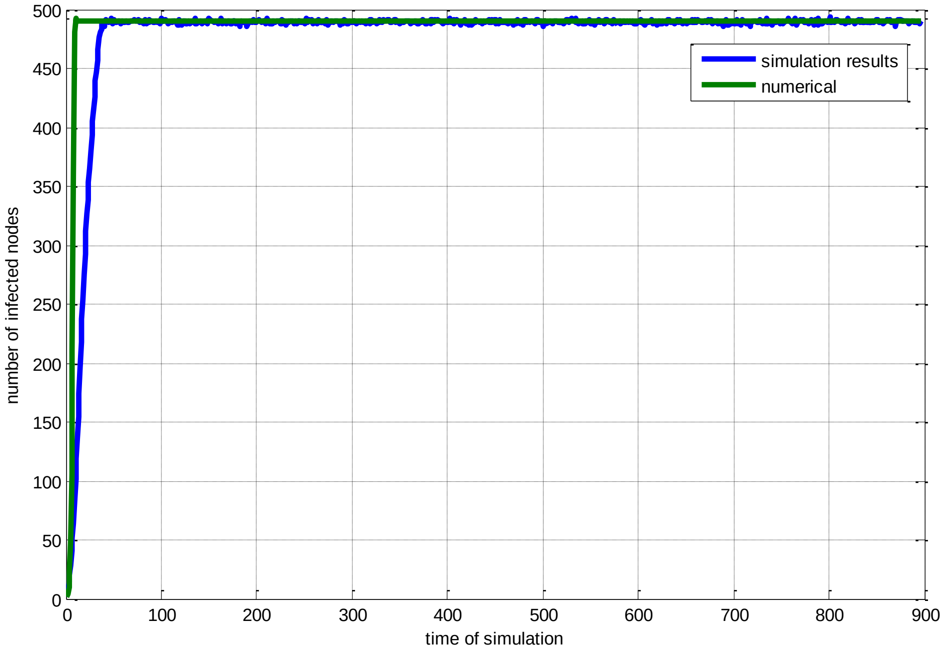


Σχήμα 15.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

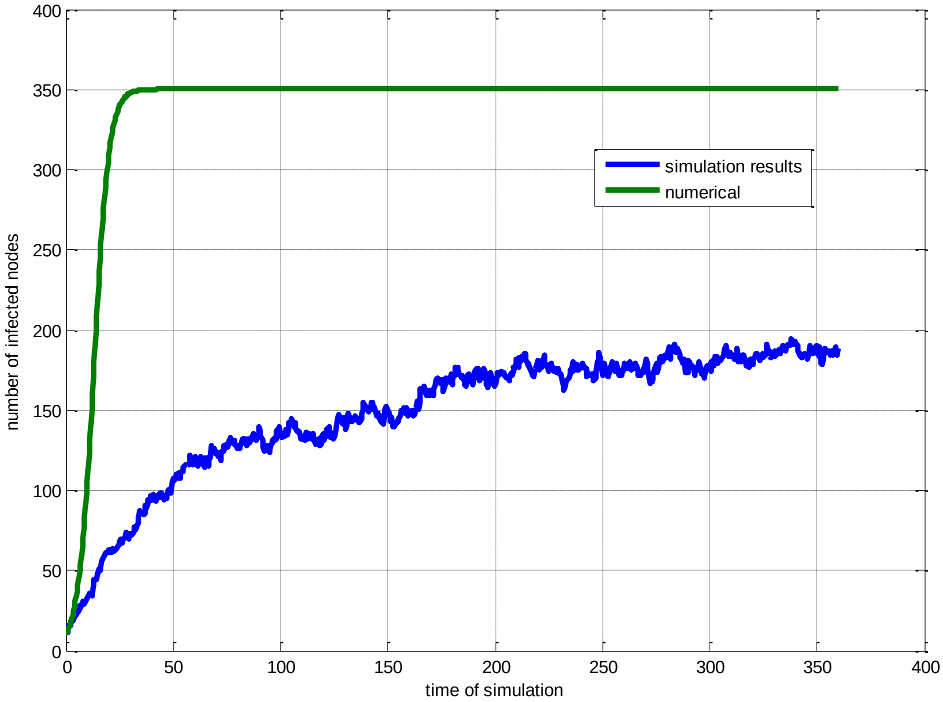
Σχήμα 15 : Διάχυση πληροφορίας με πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, p_1=0 p_2=1 αντίστοιχα και q=0.2



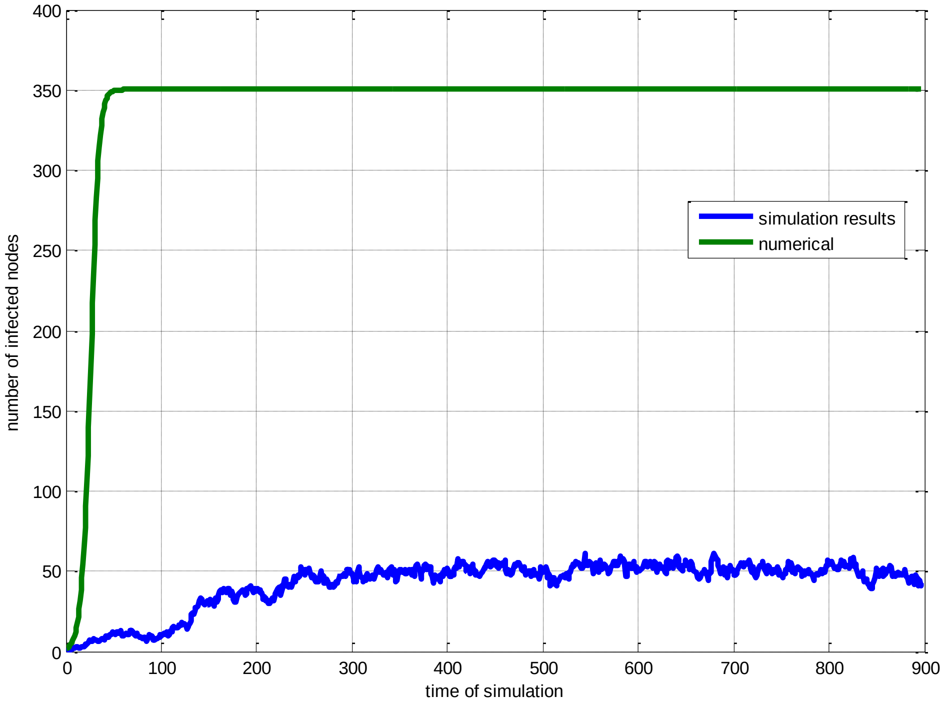
Σχήμα 16.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 16.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1



Σχήμα 16.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2



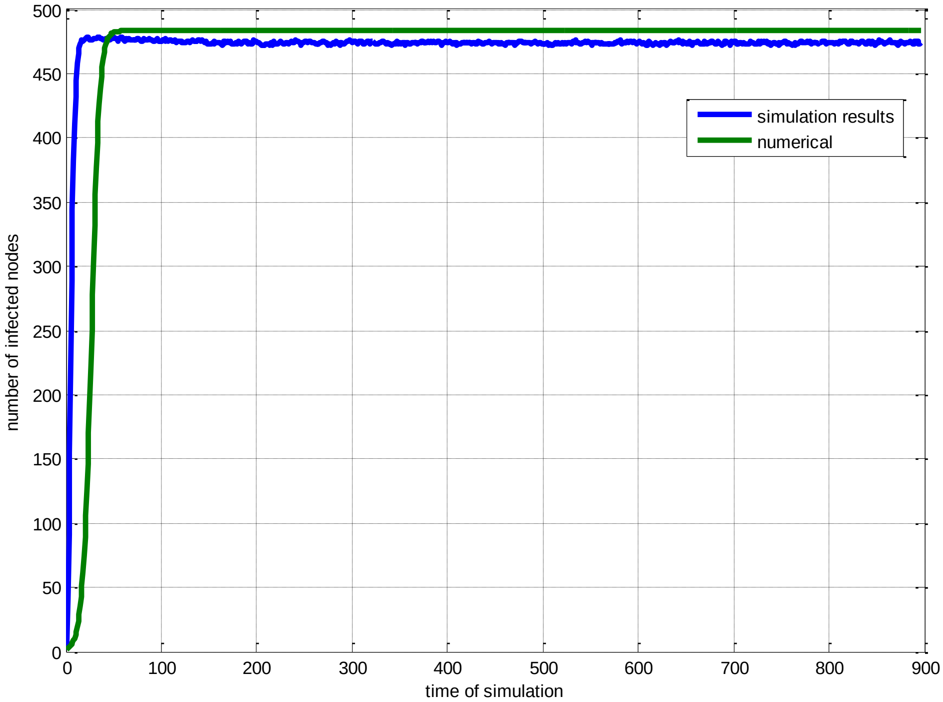
Σχήμα 16.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

Σχήμα 16: Διάχυση πληροφορίας με πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, p_1=0 p_2=1 και q=0.6

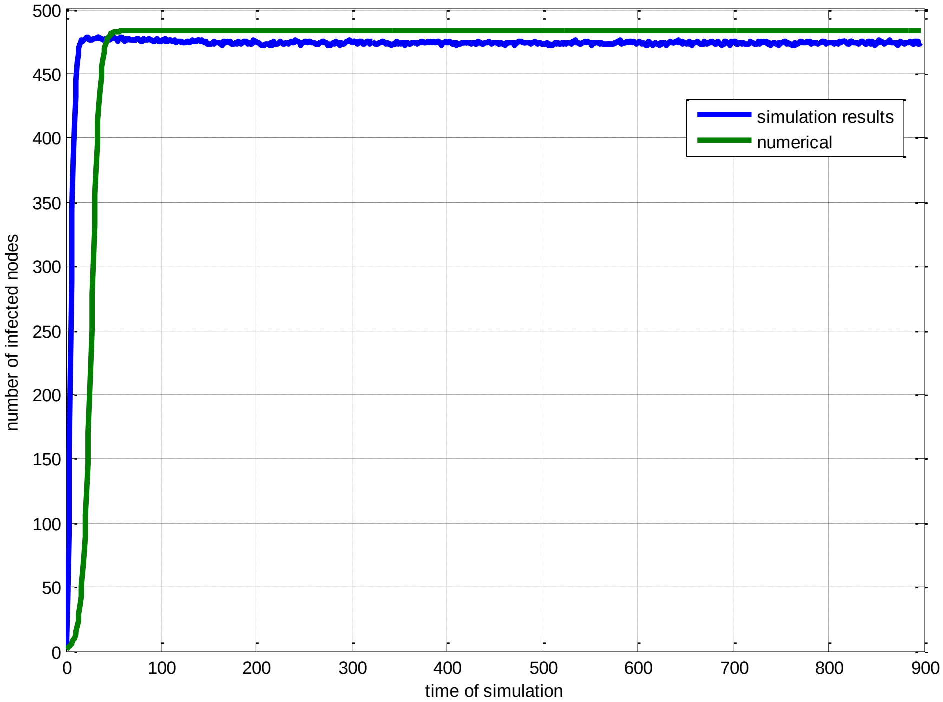
Στην περίπτωση αυτή ισχύει ότι και προηγουμένως για τη μαθηματική ανάλυση και την προσομοίωση. Τώρα όμως παρατηρείται επιπλέον ότι όταν η διάχυση γίνει μέσω Wi-Fi/Bluetooth, δηλαδή μέσω του φυσικού επίπεδου μοιάζει πολύ με την παραπάνω διάχυση όπου η διάχυση γινόταν μέσω MMS ή Wi-Fi/Bluetooth με την ίδια πιθανότητα.

Στο Σχήμα 17.1.1 φαίνεται η δυναμική της διάχυσης της πληροφορίας για την κλάση 1 ξεκινώντας από μολυσμένους κόμβους οι οποίοι είναι τυχαία επιλεγμένοι και στη συνέχεια παρουσιάζεται η διάχυση για την ίδια κλάση ξεκινώντας τη μόλυνση κόμβοι οι οποίοι είναι οι πιο δημοφιλείς κοινωνικά (σχήμα 17.1.2). Στο Σχήμα 17.2.1 αναπαριστάται η διάχυση για την κλάση 2, δηλαδή ο ρυθμός μόλυνσης των κόμβων ξεκινώντας τη διάχυση από τυχαία επιλεγμένους κόμβους, ενώ στο σχήμα 17.2.2 παρουσιάζεται η διάχυση για την κλάση 2 ξεκινώντας από μολυσμένους κόμβους τους πιο δημοφιλείς στο κοινωνικό επίπεδο. Επίσης τα ίδια με παραπάνω ισχύουν για τα σχήματα 18.1.1, 18.1.2, 18.2.1, 18.2.2 αντίστοιχα. Η μόνη διαφορά είναι η πιθανότητα να διαγράψει ένας κόμβος την πληροφορία, η οποία είναι ίση με q=0.2 για το σχήμα 17 και ίση με q=0.6 στο σχήμα 8.

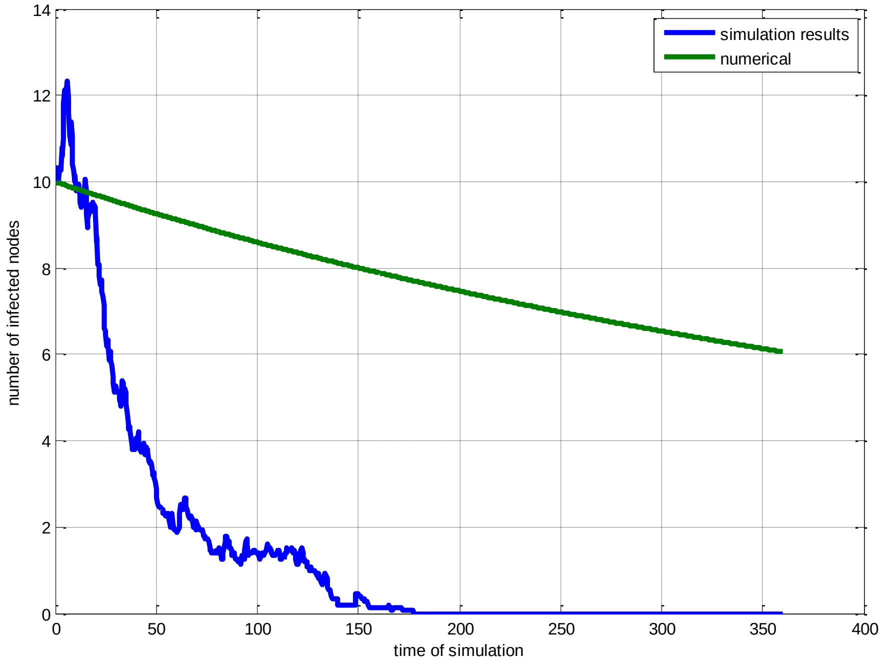
Οι παράμετροι όμως τώρα έχουν τις τιμές p_1=1 ,p_2=0 δηλαδή η πιθανότητα να γίνει η διάχυση μέσω Wi-Fi/Bluetooth είναι μηδενική. Άρα η διάχυση γίνεται μέσω MMS.



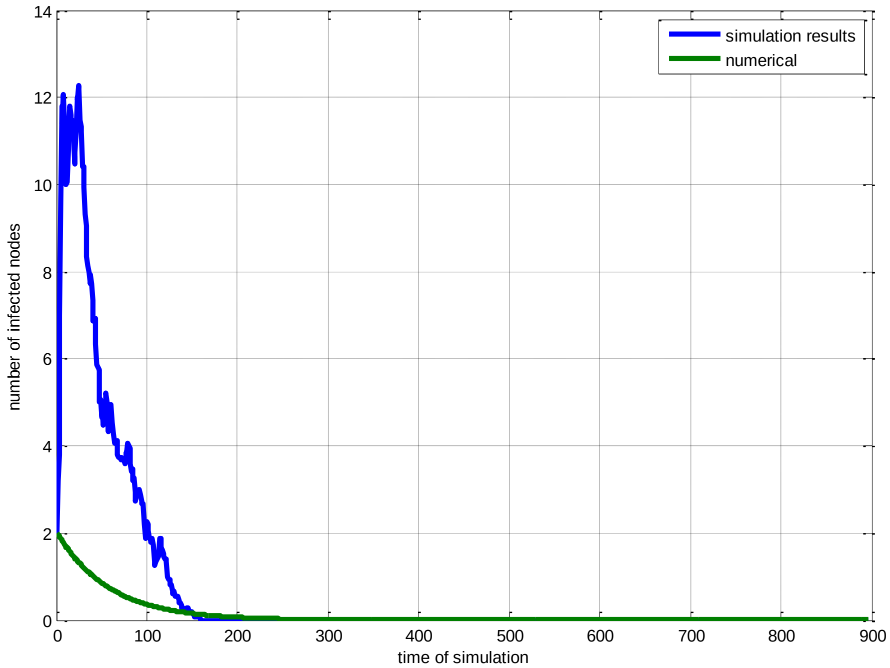
Σχήμα 17.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 17.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1

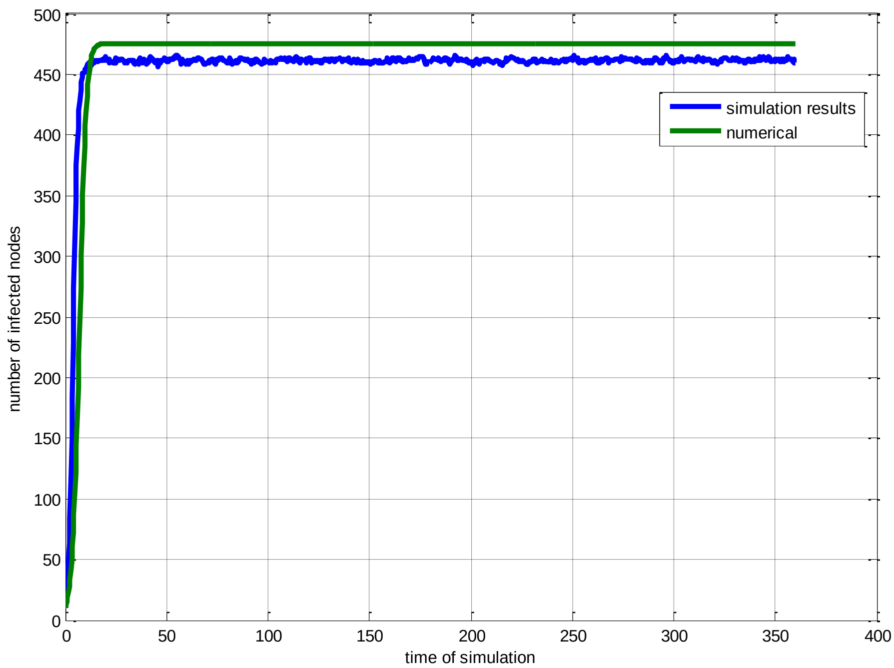


Σχήμα 17.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2

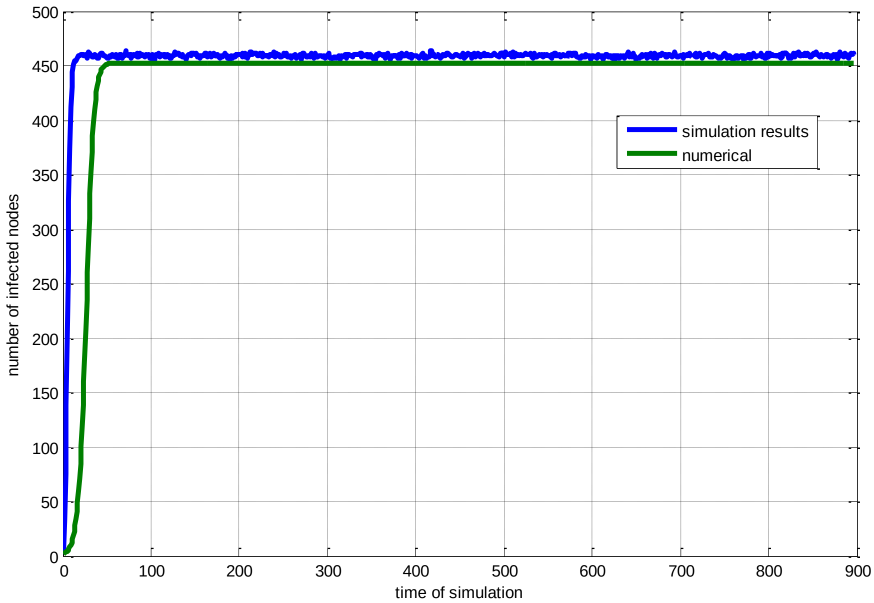


Σχήμα 17.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

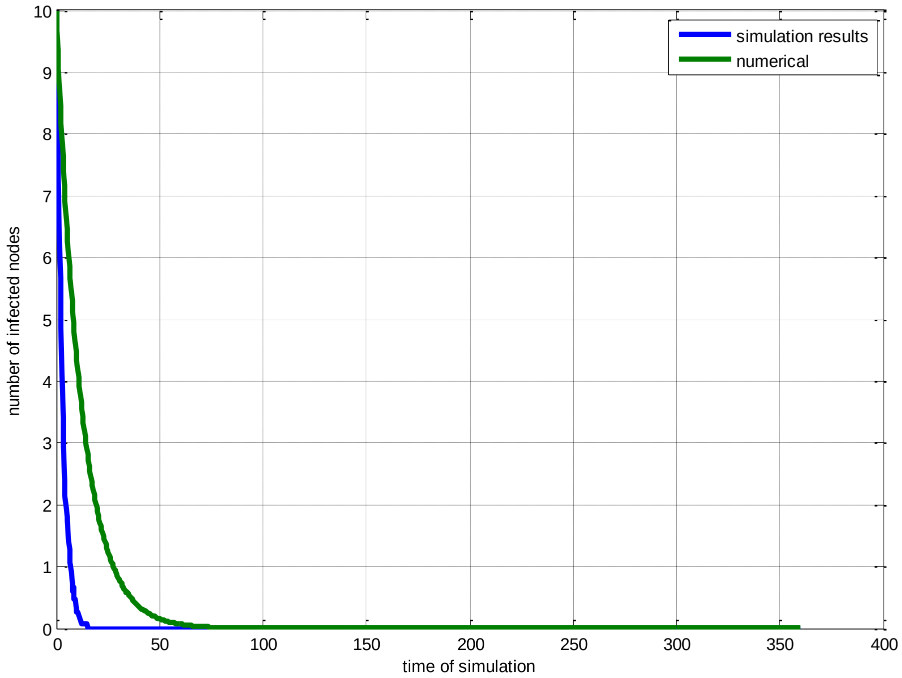
Σχήμα 17: Διάχυση πληροφορίας με πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, p_1=1 p_2=0 και q=0.2



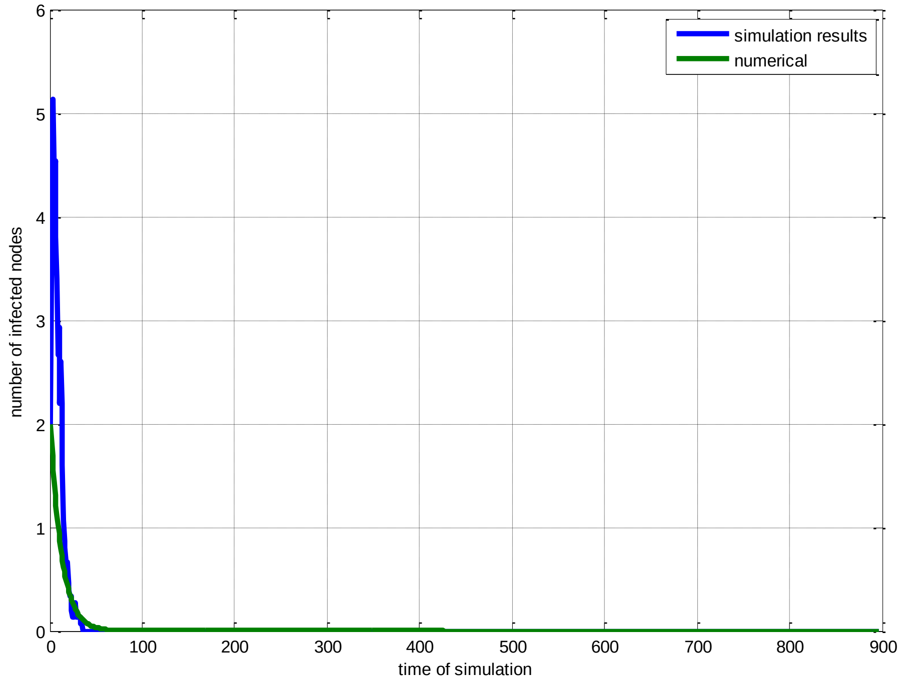
Σχήμα 18.1.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=1



Σχήμα 18.1.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=1



Σχήμα 18.2.1 Διάχυση πληροφορίας από αρχικά μολυσμένους κόμβους που επιλέγονται τυχαία για την κλάση m=2



Σχήμα 18.2.2 Διάχυση πληροφορίας από αρχικά μολυσμένους επιλεγμένους ως οι πιο δημοφιλείς κόμβοι στο κοινωνικό επίπεδο για την κλάση m=2

Σχήμα 18: Διάχυση πληροφορίας με πιθανότητα για MMS και Wi-Fi/Bluetooth διάχυση, p_1=1 p_2=0 και q=0.6

Από τις παραπάνω γραφικές παραστάσεις είναι φανερό ότι η διάχυση μέσω του κοινωνικού επιπέδου πέφτει απότομα όταν το ενδιαφέρον είναι μικρότερο, όπως συμβαίνει για την κλάση 2 και ιδιαίτερα όταν αυξάνεται η πιθανότητα διαγραφής των μηνυμάτων από τους κόμβους. Είναι φανερό ότι όταν η διάχυση γίνεται μέσω του φυσικού στρώματος η πληροφορία διατηρείται ζωντανή και οι μολυσμένοι κόμβοι είναι περισσότεροι σε σχέση με τη διάχυση μέσω κοινωνικού επιπέδου όπου οι κόμβοι που μολύνονται είναι λιγότεροι και η διάχυση σε πολλές περιπτώσεις σταματάει.

6 Κεφάλαιο

Επίλογος

6.1 Σύνοψη

Η παρούσα διπλωματική εργασία αφορά τη μελέτη της διάδοσης της πληροφορίας σε γενικευμένα δίκτυα, δηλαδή σε δίκτυα όπου οι χρήστες μπορούν και επικοινωνούν με διαφορετικούς τρόπους. Στα δίκτυα αυτά μελετήθηκε η διάχυση της πληροφορίας λαμβάνοντας υπόψη τα χρονομεταβλητά ενδιαφέροντα των χρηστών και τις μετρικές κοινωνικής δικτύωσης τους. Αναπτύχθηκε ένα μαθηματικό μοντέλο για την περιγραφή της διάδοσης και τον υπολογισμό της έκτασης στην οποία μπορεί να διαδοθεί ένα θέμα, το οποίο βασίζεται σε διαφορικές εξισώσεις. Για την μελέτη της διάχυσης της πληροφορίας μέσα στο δίκτυο και την ανάπτυξη του μαθηματικού μοντέλου χρησιμοποιήθηκε η θεωρία της επιδημιολογίας και συγκεκριμένα το επιδημιολογικό μοντέλο SIS. Το ενδιαφέρον των χρηστών για κάθε πληροφορία θεωρήθηκε τη μια φορά περιοδικό, με περίοδο ενός χρόνου και στη συνέχεια σταθερό και διαφορετικό για κάθε είδος πληροφορίας. Για τη μελέτη της διάχυση της πληροφορίας στην παρούσα εργασία υπολογίστηκε και ο τρόπος που επικοινωνούν οι χρήστες, δηλαδή αν επικοινωνούν μέσω του φυσικού ή κοινωνικού στρώματος. Στο τέλος της εργασίας τα αποτελέσματα από το μαθηματικό μοντέλο συγκρίθηκαν με τα αποτελέσματα της προσομοίωση της διάχυσης. Για την προσομοίωση της διάχυσης της πληροφορίας στο δίκτυο και τη σύγκριση των αποτελεσμάτων χρησιμοποιήθηκε το πρόγραμμα Matlab.

6.2 Μελλοντική Εργασία

Τα αποτελέσματα της διπλωματικής εργασίας αποτελούν μια διαφορετική προσέγγιση σε σχέση με υπάρχουσες εργασίες καθώς μελετάται η διάδοση με βάση τα χρονομεταβλητά ενδιαφέροντα και μετρικές κοινωνικής δικτύωσης των χρηστών. Μελλοντικά θα μπορούσε να μελετηθεί η ανάπτυξη τεχνικών βέλτιστου ελέγχου της διάδοσης πληροφορίας σχετικά με το ενδιαφέρον των χρηστών λαμβάνοντας υπόψη τις μετρικές ανάλυσης κοινωνικής δικτύωσης τους. Ταυτόχρονα τα αποτελέσματα για τη δυναμική της διάδοσης της πληροφορίας στο δίκτυο μπορούν να χρησιμοποιηθούν πρακτικά από διαφημιστικές και εμπορικές εταιρίες που θέλουν να χρησιμοποιήσουν το δίκτυο ως μέσω προώθησης της επιχείρησης και διαφήμισης του έργου τους. Μια επέκταση που θα μπορούσε να βελτιώσει τα αποτελέσματα του μοντέλου είναι να γίνει η παραπάνω διαδικασία για πολύ μεγαλύτερο χρονικό διάστημα και σε πιο μεγάλη κλίμακα χρηστών, ώστε να μπορεί να μελετηθεί πιο σφαιρικά χωρίς έντονες διακυμάνσεις και καλύπτοντας την πλήρη έκταση του φαινομένου που θα μελετηθεί.

Αναφορές

[SKP13] E. Stai, V. Karyotis and S. Papavasileiou, *"Evoluionary Dynamics of Complex Communications Networks", CRC Press,* 2013.

[SHS01] Steven H. Strogatz , *"Exploring Complex Network"*, Nature, Vol. 410, pp. 268-276, 2001.

[CLR09] T.H. Cormen, C.E. Leiserson, R.L. Rivest and C. Stein, *"Introduction to Algorithms"*, The MIT Press, 2009.

[RD10] R. Diestel, *"Graph Theory" ,* Springer, 4th edition, 2010.

[KNW11] N. Kayastha, D. Niyato, P. Wang and Hossain E., *"Applications, Architectures, and Protocol Design Issues for Mobile Social Networks: A Survey",* Vol. 99, Issue 12, pp. 2130 - 2158 , 2011

[RMC06] M. G. Rubinstein and I. M. Moraes, M. E. M. Campista, L. H. M. K. Costa, O. C. M. B. Duarte, *"A Survey on Wireless Ad Hoc Networks"* , Mobile and Wireless Communications Networks, Springer Verlag, pp. 1-34, 2006.

[SHS98] Steven H. Strogatz , *"Collective dynamics of 'small-world' networks"* , Nature, Vol. 393, pp. 440-442, 1998.

[BRS01] B. Bollobas, O. Riordan, J. Spencer and G. Tusnady, *"The degree sequence of a scale-free random graph process" ,* Random Structures and Algorithms, Vol. 18, pp. 279-290, 2001.

[AB99] R. Albert and A.-L. Barabasi, *"Emergence of Scaling in Random Networks"*, Science, Vol.286 ,No. 5439, pp. 509-512, 1999.

[SPZ05] S. Papavassileiou and J. Zhu, *"A Continuum Theory-Based Approach to the Modeling of Dynamic Wireless Sensor Networks*" , IEEE Communication Letters, Vol.9 , No.4 , pp.337-339, 2005.

[CAC11] S.-M. Cheng, W.-C. Ao, P.-Y. Chen and K.-C. Chen, *"On Modeling Malware Propagation in Generalized Social Networks",* IEEE Communication Letters, Vol. 15, No 1, pp. 25-27, 2011

[D00] M. Deijfen, "*Epidemics on social network graphs"*, 2000.

[KM01] K. Matt, *"The Mathematics of Diseases"* , Plus Magazine: Living Mathematics, 2001.

[TP00] H. Trottiera and P. Philippe, "*Deterministic Modeling Of Infectious Diseases: Theory And Methods"* ,The Internet Journal of Infectious Diseases, Vol. 1, 2000.

[KCS13] V. Karyotis, S.-M. Cheng, P.-Y. Chen, K. -C. Chen, and S. Papavasileiou, *"Diffusion Models for Information Dissemination Dynamics in Wireless Complex Communication Networks",* Journal of Complex Systems, Vol. 2013, pp. 1-13, 2013.

[TG12] Teschl, Gerald, *"Ordinary Differential Equations and Dynamical Systems",* American Mathematical Society, 2012.