



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΕΚΤΙΜΗΣΗ ΚΟΙΝΩΝΙΚΩΝ ΣΧΕΣΕΩΝ ΧΡΗΣΤΩΝ
ΜΕ ΑΞΙΟΠΟΙΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ
ΑΠΟ ΤΟ ΦΥΣΙΚΟ ΚΑΙ ΤΟΝ ΕΙΚΟΝΙΚΟ ΚΟΣΜΟ**

ΠΟΛΥΜΕΝΗ ΣΟΦΙΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΔΡ. ΛΟΥΤΑ ΜΑΛΑΜΑΤΗ

ΙΟΥΛΙΟΣ 2018

ΚΟΖΑΝΗ



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΕΚΤΙΜΗΣΗ ΚΟΙΝΩΝΙΚΩΝ ΣΧΕΣΕΩΝ ΧΡΗΣΤΩΝ
ΜΕ ΑΞΙΟΠΟΙΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ
ΑΠΟ ΤΟ ΦΥΣΙΚΟ ΚΑΙ ΤΟΝ ΕΙΚΟΝΙΚΟ ΚΟΣΜΟ**

ΠΟΛΥΜΕΝΗ ΣΟΦΙΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΔΡ. ΛΟΥΤΑ ΜΑΛΑΜΑΤΗ

ΙΟΥΛΙΟΣ 2018

ΚΟΖΑΝΗ



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

DIPLOMA THESIS

ASSESSMENT OF USERS' SOCIAL RELATIONS USING INFORMATION FROM THE PHYSICAL AND VIRTUAL WORLDS

POLYMENI SOFIA

SUPERVISOR: DR. LOUTA MALAMATI

JULY 2018

KOZANI

Περίληψη

Οι κοινωνικές σχέσεις που αναπτύσσονται μεταξύ των χρηστών ενός κοινωνικού δικτύου, αξιοποιούν πληροφορίες και από το φυσικό και από τον εικονικό κόσμο. Οι σχέσεις αυτές, σε κάθε περίπτωση, λογίζονται ως ένα βασικό συστατικό του προφίλ ενός χρήστη, οπότε και ο κατάλληλος τρόπος αξιοποίησής τους προσφέρει τόσο εξατομικευμένες υπηρεσίες στους τελικούς χρήστες, όσο και προτάσεις βελτίωσης των δικτυακών λειτουργιών από πλευράς παρόχων. Η παρούσα Διπλωματική Εργασία εστιάζει αφενός, στον υπολογισμό του μέσου χρόνου που μεσολαβεί μεταξύ διαδοχικών συναντήσεων των χρηστών ενός τηλεπικοινωνιακού δικτύου φυσικού κόσμου, και αφετέρου, στην υλοποίηση μιας απλοποιημένης έκδοσης του προτεινόμενου αλγορίθμου ενός συστήματος εμπιστοσύνης. Στόχος είναι η υλοποίηση ενός συνδυαστικού μοντέλου που να μπορεί να συνδυάζει τους δύο τύπους πληροφοριών και να αποτελέσει ένα πρώτο βήμα εκτίμησης της κινητικότητας των χρηστών.

Προκειμένου, λοιπόν, να υπολογιστεί ο μέσος αυτός χρόνος, θα πρέπει πρώτα να εντοπίσουμε τις ανάγκες του κάθε χρήστη, σύμφωνα με δεδομένα που παρέχονται από τις κινήσεις του στο παρελθόν, την ομάδα των φίλων του, τις τοποθεσίες τις οποίες έχει επισκεφθεί, καθώς και πολλά άλλα παρόμοια δεδομένα. Όλες αυτές οι πληροφορίες θεωρούμε ότι αποτελούν το τελικό προφίλ του χρήστη αυτού μέσα στο δίκτυό του, το οποίο προσφέρει τη δυνατότητα ένταξής του σε κάποια ομάδα. Μετά τη συγκέντρωση αυτών των πληροφοριών που αναφέρθηκαν, ακολουθεί και η υλοποίηση του αλγορίθμου.

Το δεύτερο κομμάτι αφορά, όπως είπαμε, σε μια απλουστευμένη έκδοση ενός συστήματος εμπιστοσύνης που προκύπτει ως αποτέλεσμα ενός μηχανισμού «φήμης», σύμφωνα με τον οποίο οι συμμετέχοντες στο δίκτυο βαθμολογούνται μεταξύ τους. Πιο συγκεκριμένα, το σύστημα αυτό παρέχει στους χρήστες του εξατομικευμένες θετικές ή/και αρνητικές προτάσεις, οι οποίες στη συνέχεια μπορούν να χρησιμοποιηθούν για τον καθορισμό νέων εμπιστων ή μη αξιόπιστων σχέσεων στο δίκτυο. Η συγκεκριμένη έκδοση έχει ως στόχο το σχηματισμό των σχέσεων μεταξύ των χρηστών του, καθώς και τον υπολογισμό της φήμης τους.

Τέλος, η εργασία αυτή προσφέρει δυνατότητες μελλοντικής επέκτασής της για αφενός πρόβλεψη της επόμενης κίνησης των χρηστών, και αφετέρου δημιουργία προτάσεων απόκτησης νέων φίλων.

Abstract

Social relations that develop among the users of a social network use information that come from both the physical and the virtual world. In any case, these relations are considered to be a crucial component of a user's profile; thus, making good use out of them offers personalized services to end-users, as well as upgrades from providers, in terms of network functioning. The following Diploma Thesis focuses on both calculating the average time that intervenes between successive meetings of the users in a modern offline telecommunication environment and implementing a simplified version of the proposed algorithm of a recommender system in an online social community. Our goal is to implement a new model that will combine both types of data and be the first step in assessing user mobility.

In order for this average time calculation to be implemented, we must first and foremost identify each user's needs, according to data deriving from their past mobility patterns, groups of friends, locations visited, as well as many other data of such kind. We consider all this information to constitute the user's end profile inside the network, which helps classify the latter one in groups, accordingly. By collecting all this information, we follow with the development of the algorithm, as mentioned above.

The second part concerns a simplified version of a recommender trust-aware system that results from a reputation mechanism, according to which, the network participants rate each other. In particular, this system provides its users with personalized positive and/or negative suggestions, which can then be used to define new trusted or unreliable relationships in the network. This particular version targets users' relationships formulations and calculation of their reputation, as well.

In conclusion, this thesis offers the possibility of a future expansion, as calculating the average time that intervenes between successive meetings of the users can be extended to the degree of predicting their future mobility patterns along the network. On the other hand, in terms of the online community, the aforementioned recommender algorithm can be extended to the degree of friend suggestions.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω την αναπληρώτρια καθηγήτρια του τμήματος Μηχανικών Πληροφορικής και Τηλεπικοινωνιών, του Πανεπιστημίου Δυτικής Μακεδονίας, κυρία Μαλαματή Λούτα, για την επίβλεψη και την πολύτιμη βοήθεια και καθοδήγησή της καθ' όλη τη διάρκεια εκπόνησης της παρούσας Διπλωματικής Εργασίας. Αντίστοιχα, θα ήθελα να ευχαριστήσω και τον κύριο Νικόλαο Μαντά για την πολύτιμη συμβολή και βοήθειά του στην υλοποίηση.

Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα την οικογένειά μου για όλα αυτά τα χρόνια στήριξης, καθοδήγησης και συμπαράστασης.

Πίνακας περιεχομένων

Περίληψη.....	5
Abstract	6
Ευχαριστίες	7
Πίνακας Εικόνων	9
Πίνακας Πινάκων	9
1. Εισαγωγή.....	10
1.1. Κατηγορίες Δημιουργίας Προφίλ Χρηστών.....	11
2. Συστήματα Συστάσεων και Εφαρμογές.....	12
2.1. Πρόβλεψη Φυσικής Κίνησης	14
3. Κοινωνική Πληροφορία Χρήστη	16
3.1. Αξιοποίηση Πληροφορίας στο Φυσικό Κόσμο	17
3.1.1. Μοντέλο Πρόβλεψης βάσει Συμπεριφοράς	18
3.2. Αξιοποίηση Πληροφορίας στον Εικονικό Κόσμο.....	21
3.2.1. Πρωτόκολλο Δρομολόγησης SocialCLWPR.....	25
3.2.2. Σύστημα Εμπιστοσύνης για Εξατομικευμένες Προτάσεις.....	27
3.3. Συνδυαστικές Προσεγγίσεις των Δύο Κόσμων	30
3.3.1. Offline D2D Δρομολόγηση.....	33
3.3.2. Μη-γραμμική Πρόβλεψη Χρονοσειρών Πολλών Μεταβλητών	34
3.3.3. Σειρά Μεταβλητών Μαρκοβιανού Μοντέλου (VOMM).....	38
3.3.4. Επόμενο Check-in	42
3.3.5. Επόμενη Θέση Χρήστη	43
4. Υλοποίηση.....	45
4.1. Προσομοίωση του Μοντέλου Grid Manhattan	45
4.2. Προσομοίωση του Συνόλου Δεδομένων Erinions	51
4.3. Υλοποίηση Συνδυαστικής Γραφικής Απεικόνισης	55
5. Συμπεράσματα	59
6. Μελλοντική Αξιοποίηση	59
Βιβλιογραφία.....	60

Πίνακας Εικόνων

Εικόνα 1.	18
Εικόνα 2.	19
Εικόνα 3.	19
Εικόνα 4.	29
Εικόνα 5.	39
Εικόνα 6.	39
Εικόνα 7.	48
Εικόνα 8.	49
Εικόνα 9.	50
Εικόνα 10.	51
Εικόνα 11.	54
Εικόνα 12.	54
Εικόνα 13.	56
Εικόνα 14.	56
Εικόνα 15.	57
Εικόνα 16.	57
Εικόνα 17.	58

Πίνακας Πινάκων

Πίνακας I.....	46
----------------	----

1. Εισαγωγή

Ως προφίλ ενός χρήστη ορίζεται μια συλλογή προσωπικών δεδομένων που σχετίζονται μ' ένα συγκεκριμένο χρήστη. Ιδανικά, το προφίλ ενός χρήστη λειτουργεί ως μια αποκλειστική ψηφιακή αναπαράσταση αυτού του ατόμου. Ωστόσο, ο καθένας μπορεί να έχει αρκετά διαφορετικά προφίλ, ανάλογα με το σύστημα με το οποίο αλληλεπιδρά. Για παράδειγμα, από την οπτική ενός Η/Υ, το προφίλ του χρήστη αποτελεί ένα σύνολο ρυθμίσεων που θα κάνουν τον υπολογιστή να δείχνει και να λειτουργεί με τον τρόπο που θέλει ο εκάστοτε χρήστης. Για ένα σύστημα ηλεκτρονικού εμπορίου, όπως είναι ένα ηλεκτρονικό βιβλιοπωλείο, το προφίλ ενός χρήστη μπορεί να περιλαμβάνει πληροφορίες για την ταυτότητά του, όπως τα διαπιστευτήρια σύνδεσής του, πληροφορίες πληρωμής, δεδομένα της πιστωτικής κάρτας, ή ακόμα κυριότερα τις προτιμήσεις του χρήστη όσον αφορά στα είδη των βιβλίων που τον ενδιαφέρουν, πρόσφατες αγορές κ.λπ. Μερικά από τα δεδομένα αυτά προστίθενται στο σύστημα από τον ίδιο το χρήστη, ενώ άλλα προκύπτουν ύστερα από την ανάλυση παλαιότερων συναλλαγών. Απ' αυτήν την οπτική, το προφίλ προκύπτει από την παρακολούθηση του κάθε μοτίβου χρήσης, οπότε και είναι σχετικό μόνο με τη συμπεριφορά του χρήστη απέναντι σε μια συγκεκριμένη υπηρεσία. Φυσικά, τα δεδομένα αυτά μπορεί να αλληλεπιδρούν και με δεδομένα από διαφορετικές βάσεις δεδομένων, οπότε και βγάζουν ένα πιο συγκεκριμένο συμπέρασμα για τις προτιμήσεις του καθενός. Προφίλ χρήστη μπορούν να βρεθούν σε πολλά λειτουργικά συστήματα, ενδεικτικά μερικά εκ των οποίων είναι εφαρμογές ηλεκτρονικού εμπορίου, μέσα κοινωνικής δικτύωσης, συστήματα συστάσεων, συστήματα έξυπνης μάθησης κ.ά.

Η βασική ιδέα πίσω από τα προφίλ των χρηστών είναι ότι η συμπεριφορά που παρουσίασε ο χρήστης στο παρελθόν μπορεί να συσσωρευτεί, και έτσι να δημιουργηθεί ένα προφίλ χρήστη, ή ένα «λεξικό χρήστη», ή μια «απογραφή χρήστη», που θα περιλαμβάνει αναμενόμενες τιμές της συμπεριφοράς του. Στην απλούστερη μορφή του, το προφίλ αυτό είναι ένας φορέας που περιλαμβάνει απλές αριθμητικές περιλήψεις κάποιας πτυχής της συμπεριφοράς του χρήστη ή κάποιου είδους ποικίλου μοτίβου συμπεριφοράς. Ένα τέτοιο προφίλ μπορεί επίσης να περιλαμβάνει κατηγορικά, λογοκρίμενα ή άλλα μη-νουμερικά δεδομένα.

Κάθε προφίλ χρήστη ακολουθεί τη λογική των καταγεγραμμένων δεδομένων, με έμφυτες παραμέτρους στην επιστήμη των υπολογιστών. Το γεγονός αυτό, δηλαδή ο εγγενής αναγωγικός χαρακτήρας του προφίλ, είναι σημαντικό λόγω του ότι τα προφίλ αυτά μπορεί να επηρεάσουν την ασφάλεια και την ταυτοποίηση σ' ένα πολύ μεγάλο βαθμό. Εφόσον τα προφίλ αυτά επηρεάζουν τις ζωές όλων των χρηστών ενός οποιουδήποτε δικτύου αρκετά συχνά, είναι πολύ σημαντικό να επεξηγούνται κάθε φορά η βάση, αλλά και οι τρόποι που την επηρεάζουν.

Μετά την κατασκευή του προφίλ, αυτό μπορεί να χρησιμοποιηθεί κατάλληλα. Ένα λειτουργικό σύστημα αποθηκεύει τα προφίλ των χρηστών που περιλαμβάνουν τις ρυθμίσεις του καθενός για το φόντο του υπολογιστή του, τις προτιμήσεις του δείκτη του ποντικιού και πολλά άλλα χαρακτηριστικά. Ένα σύστημα συστάσεων, όπως αυτό θα αναλυθεί περισσότερο και παρακάτω, χρησιμοποιεί τις προτιμήσεις που είχε δείξει ο χρήστης στο παρελθόν, προκειμένου να προβλέψει πράγματα τα οποία ο χρήστης δεν έχει σκεφτεί ακόμα, όπως για παράδειγμα ένα βιβλίο, ένα τραγούδι, ή ακόμα και ένα φίλο. Παραδοσιακά στην ασφάλεια των υπολογιστών τα προφίλ των χρηστών κατασκευάζονται βασισμένα σε οποιαδήποτε βασική χρήση

συμπεριφορά του χρήστη προβλέπεται από την ανάλυση των διαθέσιμων δεδομένων του. Αυτά τα δεδομένα προκύπτουν συνήθως από τη συμπλήρωση μιας διαδικτυακής φόρμας ή από μελέτες κ.λπ. Ονομάζεται επίσης και Στατική ή Πραγματική Δημιουργία Προφίλ. Ωστόσο, παρατηρούνται κάποια προβλήματα σ' αυτή τη διαδικασία, στην αποκλειστική χρήση, δηλαδή, «υπονοούμενων» προφίλ. Αυτό οφείλεται στο γεγονός ότι ο χρήστης είναι επιφυλακτικός και δεν ενδιαφέρεται να αποκαλύψει προσωπικές του πληροφορίες σε κανέναν, είτε λόγω ασφάλειας, είτε επειδή πρόκειται για μια ανιαρή διαδικασία την οποία και θα θέλει να αποφύγει. Συνεπώς, η ακρίβεια της χρήσης αυτής της προσέγγισης για τη δημιουργία προφίλ υποβαθμίζεται ανάλογα με τη χρονική περίοδο.

2. *Υπονοούμενη Δημιουργία Προφίλ Χρήστη.* Η δυναμική προσέγγιση δημιουργίας προφίλ βασίζεται περισσότερο στην εκμάθηση των μελλοντικών κινήσεων του χρήστη, αντί να επικεντρώνεται στις τρέχουσες πληροφορίες που έχει μοιραστεί ο ίδιος. Αυτού του τύπου το σύστημα ονομάζεται επίσης και Δημιουργία Προφίλ με βάση τη Συμπεριφορά του Χρήστη, ή Προσαρμοστική ή ακόμα και Ορθολογική Δημιουργία Προφίλ. Σ' αυτή την προσέγγιση χρησιμοποιούνται και διαφορετικές τεχνικές φιλτραρίσματος των πληροφοριών.
3. *Υβριδική Δημιουργία Προφίλ Χρήστη.* Αυτή η προσέγγιση συνδυάζει τα πλεονεκτήματα των δύο παραπάνω τεχνικών, καθώς αφενός λαμβάνει υπόψη τα στατικά χαρακτηριστικά ενός χρήστη και αφετέρου ανακτά πληροφορίες αναφορικά με τη συμπεριφορά του. Με αυτή τη διαδικασία προκύπτουν αποδοτικότερα προφίλ χρηστών, αλλά παράλληλα διατηρείται και η εγκυρότητα των προσωρινών δεδομένων, καθώς τα δεδομένα ανανεώνονται ανά τακτά χρονικά διαστήματα.

2. Συστήματα Συστάσεων και Εφαρμογές

Τα Συστήματα Συστάσεων μπορούν να θεωρηθούν ως ο κληρονόμος της Διαδικασίας Δημιουργίας Προφίλ Χρήστη. Η διαδικασία για τη δημιουργία προφίλ χρηστών εφαρμόζεται ολοένα και περισσότερο σε διάφορες εφαρμογές, όπως είναι για παράδειγμα Συστήματα Διαχείρισης Σχέσεων Πελατών. Μια γενική εικόνα των κλάδων όπου χρησιμοποιούνται προφίλ χρηστών είναι:

1. *Έρευνα.* Μεγαλύτερη ανάπτυξη της διαδικασίας δημιουργίας προφίλ παρουσιάζεται σε εφαρμογές για προτάσεις έρευνας. Το σύστημα που δημιουργήθηκε από τον Jie Tang [2] έχει οδηγήσει στην υποδιαίρεση της διαδικασίας αυτής σε τρεις υποκατηγορίες, οι οποίες είναι εξαγωγή προφίλ, ενσωμάτωση και ανακάλυψη ενδιαφέροντος. Σε μια παρόμοια προσέγγιση [3] από τον Stuart Middleton, χρησιμοποιήθηκε ένα επιπλέον βήμα, η οπτικοποίηση του προφίλ, προκειμένου να αναπαρασταθεί η δημιουργία ενός προφίλ από το σύστημα, το οποίο χρησιμοποίησε ορθολογική προσέγγιση.

2. *Κινητές Υπηρεσίες*. Μια άλλη εφαρμογή που μπορεί να ευεργετηθεί από τα προφίλ των χρηστών είναι και οι ιστοσελίδες που βασίζονται στο e-Tourism [4]. Το σύστημα αυτό είναι ικανό να παρέχει προσωποποιημένες πληροφορίες με βάση την τοποθεσία του χρήστη. Καθώς ο τουρισμός βασίζεται εξ ολοκλήρου σε δημογραφικές πληροφορίες, όπως είναι η γεωγραφική τοποθεσία, το σύστημα είναι ικανό να παρέχει, επιπλέον, προτάσεις κοντινότερων τουριστικών τοποθεσιών σε κάθε νέο χρήστη ανάλογα με την τοποθεσία του.
3. *Διαχείριση Ενέργειας*. Ένας σημαντικός τομέας είναι και η διαχείριση ενέργειας. Υπάρχουν πολλές μεγάλες εταιρίες που αντιμετωπίζουν την πρόκληση της αποδοτικής και βελτιστοποιημένης διαχείρισης ενέργειας. Ένα σύστημα έξυπνης διαχείρισης ενέργειας [5], που αναπτύχθηκε από τον Mario Caruso, αποδείχθηκε ιδιαίτερα αποδοτικό. Στο σύστημα αυτό χρησιμοποιήθηκαν προφίλ χρηστών και μικροοικονομική λογιστική για έξυπνη διαχείριση ενέργειας.
4. *Εύρεση Εργασίας*. Η εύρεση εργασίας αποτελεί μια ανιαρή διαδικασία που κάθε άτομο θα πρέπει να κάνει στη ζωή του. Έτσι, δημιουργήθηκε η ανάγκη για ανάπτυξη ενός συστήματος, το οποίο θα προτείνει αυτοματοποιημένα θέσεις εργασίας στο χρήστη, ανάλογα με την εμπειρία και τα προσόντα του. Το σύστημα αυτό ονομάζεται CASPER (Case-Based Profiling for Electronic Recruitment) [6] και λαμβάνει υπόψη πληροφορίες από τα προφίλ των χρηστών, ενώ παράλληλα προτείνει και κατάλληλες θέσεις εργασίας για τον καθένα.
5. *Εξατομικευμένες Υπηρεσίες Τηλεόρασης*. Έχει παρατηρηθεί σημαντική πρόοδος στον τομέα των προτάσεων εξατομικευμένων προγραμμάτων και διαφημίσεων για την τηλεόραση (π.χ. IPTV, iDTV). Ένα υβριδικό σύστημα προτάσεων τηλεοπτικών προγραμμάτων, το οποίο ονομάστηκε gueveo.tv [7] λειτουργεί καλά επειδή και οι δυο μέθοδοι αλληλοσυμπληρώνουν η μία την άλλη με τέτοιο τρόπο, ώστε η μέθοδος που βασίζεται στο περιεχόμενο να προτείνει συνήθη προγράμματα και η συνεργατική μέθοδος να παρέχει την ανακάλυψη νέων προγραμμάτων. Στη μελέτη αυτή, κάθε χρήστης αναπαριστανόταν από ένα προφίλ που έδειχνε τις προτιμήσεις του και το οποίο περιείχε δύο τύπους πληροφοριών, προτιμήσεις τομέα (λίστες από διαθέσιμα κανάλια, ώρες που προτιμώνται από τους χρήστες για να δουν τηλεόραση) και προτιμήσεις προγραμμάτων (λέξεις-κλειδιά). Οι πληροφορίες αυτές, στη συνέχεια, συγκεντρώνονταν τόσο με σαφείς (παρακολούθηση των ωρών προβολής), όσο και με υπονοούμενες μεθόδους (συμπλήρωση ερωτηματολογίου). Στην περίπτωση του gueveo.tv, χρησιμοποιήθηκε Χαρτογράφηση Προστιθέμενης Αξίας (VSM), προκειμένου να δημιουργηθεί μια διανυσματική αναπαράσταση των προφίλ των χρηστών σε σχέση με τα προγράμματα που παρακολουθούν, και μετρήσεις συνημιτόνων για τον υπολογισμό της ομοιότητας μεταξύ των προγραμμάτων και των προφίλ των χρηστών.
6. *Αλγόριθμοι Ταξινόμησης και Ομαδοποίησης για τα προφίλ των χρηστών*. Αλγόριθμοι αυτής της κατηγορίας χρησιμοποιούνται ευρέως σε μεγάλα κοινωνικά δίκτυα, όπως είναι το MySpace [8]. Πραγματοποιήθηκε σύγκριση γνωστών αλγορίθμων μηχανικής μάθησης, προκειμένου να γίνει διαχωρισμός των ψεύτικων προφίλ από τα νόμιμα. Σ' αυτή τη μελέτη, κάθε χρήστης αναπαριστανόταν μ' ένα προφίλ, το οποίο περιλάμβανε δύο ειδών δεδομένα: κατηγορηματικά δεδομένα (φύλο, ηλικία, κοινωνική κατάσταση) και δεδομένα φόρμας (πληροφορίες κειμένου, όπως ενδιαφέροντα). Σε μια άλλη μελέτη [9], οι Paireekreng και Wong ερεύνησαν τη σημασία της χρήσης ταξινόμησης και ομαδοποίησης των προφίλ των χρηστών, από την πλευρά του χρήστη στα κινητά. Επικεντρώθηκαν στην εξατομίκευση περιεχομένου για να βοηθήσουν χρήστες κινητών τερματικών να ανακτήσουν

αποτελεσματικά πληροφορίες και υπηρεσίες. Στο προτεινόμενο πλαίσιο των δύο σταδίων, η ομαδοποίηση χρησιμοποιήθηκε για τη δημιουργία προφίλ χρήστη, ενώ η ταξινόμηση για να ταξινομηθούν τα προφίλ αυτά με βάση τις πληροφορίες της κάθε κλάσης, όπως αυτή προέκυπτε από την ταξινόμηση.

2.1. Πρόβλεψη Φυσικής Κίνησης

Γενικά, η πρόβλεψη της φυσικής κίνησης μπορεί να διακριθεί σε δύο βασικές υποκατηγορίες, οι οποίες αφορούν αφενός την πρόβλεψη της κίνησης των χρηστών του δικτύου, και αφετέρου τη βραχυπρόθεσμη πρόβλεψη της κίνησης του ίδιου του τηλεπικοινωνιακού δικτύου.

Όσον αφορά στην πρόβλεψη της κίνησης των χρηστών του δικτύου, όπως αναφέρθηκε και παραπάνω, αυτή επιτυγχάνεται με τα εκάστοτε προφίλ των χρηστών, που κατασκευάζονται σύμφωνα με τη θεωρία της μοντελοποίησης χρηστών, η οποία με τη σειρά της στοχεύει στην περιγραφή των χαρακτηριστικών και των προτιμήσεων των χρηστών αυτών. Γενικά, τέτοιου είδους τεχνικές μοντελοποίησης των χρηστών των δικτύων χρησιμοποιούνται σε μεγάλο βαθμό για την πρόβλεψη των μελλοντικών επιλογών τους, καθώς και για την κατάλληλη προσαρμογή της συμπεριφοράς του ίδιου του συστήματος στις ανάγκες των χρηστών του.

Πληροφορίες που μπορούν να αντληθούν από τα λεγόμενα αυτά προφίλ των χρηστών είναι πολλές και διάφορες και ποικίλουν ανάλογα με τις ανάγκες του εκάστοτε προβλήματος. Το προφίλ που διαχειριζόμαστε κάθε φορά περιέχει ένα σύνολο συντελεστών που αντιπροσωπεύουν το βαθμό ενδιαφέροντος ενός χρήστη προς ένα συγκεκριμένο στοιχείο μιας μετάδοσης πολυμέσων. Οι πληροφορίες σχετικά με το προφίλ χρήστη αξιοποιούνται για να προσδιοριστούν εκείνα τα στοιχεία πολυμέσων που εμφανίζονται ως τα πιο σημαντικά για το χρήστη. Με τον τρόπο αυτό, σε κάθε περίπτωση διαφορετικού προβλήματος, είναι δυνατόν να αποφασιστεί ποια συστατικά μπορούν να χειριστούν καλύτερα την κατάσταση και ποια άλλα θα μπορούσαν να υποβαθμιστούν, δηλαδή τελικά να μη χρησιμοποιηθούν, για μεγαλύτερη ασφάλεια. Με αυτόν τον τρόπο, η πολιτική διανομής των πόρων του συστήματος γίνεται «προσαρμοστική στο χρήστη», δηλαδή προσπαθεί πάντα να στοχεύει στην ικανοποίηση του χρήστη.

Ωστόσο, είναι δεδομένο ότι ένα σύστημα ή ένα δίκτυο απαιτούν με τη σειρά τους ένα μεγάλο «σύνολο εκπαίδευσης», δηλαδή την εξέταση ενός μεγάλου αριθμού συνεδριών, τόσο για την παραγωγή ενός προφίλ, ικανού να περιγράψει χαρακτηριστικά των χρηστών, όσο και για την πρόβλεψη των μελλοντικών ενεργειών του με μεγάλο βαθμό ακρίβειας. Κατά συνέπεια, ο χρόνος που απαιτείται για ένα σωστό λειτουργικό σύστημα μπορεί να γίνει ιδιαίτερα μακρύς. Δύο βασικές προσεγγίσεις [10] μιας απλής και κατανοητής λύσης αυτού του προαναφερθέντος προβλήματος είναι:

- *Συνεργατική Προσέγγιση.* Αυτή η προσέγγιση βασίζεται στην υπόθεση ότι τα άτομα που ανήκουν στην ίδια «ομάδα» (π.χ. άτομα ίδιας ηλικίας ή φύλου, που ζουν στην ίδια πόλη και

ανήκουν στην ίδια κοινωνική τάξη) συμπεριφέρονται παρόμοια και έτσι έχουν παρόμοια προφίλ.

- *Προσέγγιση Βάσει Περιεχομένου.* Αυτή η προσέγγιση προϋποθέτει ότι εάν ένας χρήστης παρουσιάσει μια συγκεκριμένη συμπεριφορά κάτω από ένα δεδομένο σύνολο περιστάσεων, τότε αυτό επαναλαμβάνεται υπό παρόμοιες συνθήκες.

Ωστόσο, μια βασική προϋπόθεση είναι και η ενημέρωση αυτών των προφίλ των χρηστών, η οποία μπορεί να υλοποιηθεί με τη χρήση μιας «ωβριδικής στρατηγικής» που θα συνδυάζει τις δύο αυτές προσεγγίσεις και θα δρα με τη σειρά της σε δύο φάσεις: αρχικά, παρέχεται ένας νέος χρήστης με προεπιλεγμένο προφίλ, που καθορίζεται με μια τεχνική συνεργασίας, και στη συνέχεια, το προεπιλεγμένο αυτό προφίλ «βελτιώνεται» μέσω τεχνικών περιεχομένου. Το πλεονέκτημα της συνδυαστικής αυτής τεχνικής, συνεπώς, είναι ότι εγγυάται την άμεση διαθεσιμότητα ενός προφίλ για κάθε χρήστη. Με την πάροδο του χρόνου, το σύστημα συλλέγει ένα μεγάλο σύνολο πληροφοριών και παρέχει μια ακριβέστερη περιγραφή των συνηθειών των χρηστών.

Αντίστοιχα, η πρόβλεψη της κίνησης ενός ολόκληρου δικτύου, όπως και η πρόβλεψη της κίνησης χρηστών, αποτελείται από ένα κομμάτι υλοποίησης με τη χρήση κατάλληλων αλγορίθμων πρόβλεψης και ένα κομμάτι αποτίμησης της απόδοσης των αλγορίθμων που χρησιμοποιήθηκαν.

Για την υλοποίηση πρόβλεψης της συμπεριφοράς ενός τηλεπικοινωνιακού δικτύου, χρησιμοποιείται κάθε φορά μια ομάδα, ένα σύνολο κατάλληλων αλγορίθμων πρόβλεψης. Αυτό συμβαίνει γιατί ένα δίκτυο έχει πολλά και διαφορετικά χαρακτηριστικά, επομένως για την πρόβλεψη της συμπεριφοράς του θα χρειαστούν διαφορετικοί αλγόριθμοι, καθένας απ' τους οποίους θα στοχεύει και στην πρόβλεψη ενός διακριτού χαρακτηριστικού και όλοι μαζί θα στοχεύουν στην όσο το δυνατόν καλύτερη πρόβλεψη του συνολικού δικτύου.

Το κομμάτι της αποτίμησης της απόδοσης των αλγορίθμων αποτελείται από τον υπολογισμό του σφάλματος σε κάθε περίπτωση. Ο υπολογισμός αυτός προκύπτει από την εύρεση του απόλυτου και του σχετικού σφάλματος πρόβλεψης, ενώ σε αρκετές περιπτώσεις υπολογίζεται και το μέσο σχετικό και απόλυτο σφάλμα του αλγορίθμου. Βέλτιστη περίπτωση πρόβλεψης θεωρείται αυτή που θα έχει και τα δύο σφάλματα ίσα με μηδέν, ενώ όσο μεγαλύτερες είναι οι τιμές του σχετικού σφάλματος παρατηρούνται φαινόμενα υπερεκτίμησης ή υποτίμησης της κίνησης του δικτύου.

Πιο συγκεκριμένα, μελέτες, που έχουν ως στόχο την αξιολόγηση βραχυπρόθεσμων αλγορίθμων πρόβλεψης κυκλοφορίας σ' ένα ασύρματο δίκτυο [11], εξετάζουν μια μεγάλη ασύρματη δομή, καθώς επίσης και τη μοντελοποίηση του φόρτου κατά την κίνηση στα σημεία πρόσβασης (Access Points, AP).

Γενικά, η λειτουργία ενός σημείου πρόσβασης είναι να προβλέπει τον φόρτο κίνησης του δικτύου για το επόμενο χρονικό διάστημα και να χρησιμοποιεί τις προβλέψεις αυτές κατά τον έλεγχο εισόδου (admission control) για καλύτερη συμβουλή των πελατών και διαχείριση των απαιτήσεών τους, ώστε αυτοί να σχετίζονται με τα κατάλληλα σημεία πρόσβασης για καλύτερη αξιοποίηση των τοπικών πόρων. Οι προβλέψεις αυτές μπορούν να χρησιμοποιηθούν

και για τη μείωση των δαπανών ενέργειας από πλευράς χρήστη, τη βελτίωση της αξιοποίησης της χωρητικότητας ενός ασύρματου LAN (WLAN) και εξισορρόπηση του φόρτου κατά την κίνηση.

Με βάση τα μοντέλα αυτά γίνεται σχεδίαση αλγορίθμων πρόβλεψης για να προβλεφθεί ο φόρτος κίνησης στα σημεία πρόσβασης, σε διαφορετικές κλίμακες χρόνου. Στη συνέχεια, αυτοί οι αλγόριθμοι εφαρμόζονται σε πραγματικά δεδομένα κίνησης που αποκτήθηκαν από τα περισσότερα αξιοποιήσιμα σημεία πρόσβασης και αξιολογείται κάθε φορά η απόδοσή τους. Ο πιο συνηθισμένος τρόπος αξιολόγησης ενός τέτοιου αλγορίθμου είναι με τον υπολογισμό τόσο του απόλυτου, όσο και του σχετικού σφάλματος πρόβλεψης. Το μεν απόλυτο σφάλμα πρόβλεψης ορίζεται ως η απόλυτη διαφορά της προβλεπόμενης από την πραγματική κίνηση, ενώ το σχετικό σφάλμα πρόβλεψης είναι ο λόγος αυτού του απόλυτου σφάλματος προς την πραγματική κυκλοφορία. Ένας τέλειος αλγόριθμος πρόβλεψης θεωρείται πως έχει απόλυτο και σχετικό σφάλμα πρόβλεψης ίσο με 0. Ωστόσο, το σχετικό σφάλμα πρόβλεψης είναι μια συντηρητική μετρική, δεδομένου ότι δεν λαμβάνει υπόψη τα διαστήματα με μηδενική πραγματική κυκλοφορία, τα οποία είναι συχνά στο ίχνος. Επιπλέον, μεγάλα σχετικά σφάλματα πρόβλεψης υποδεικνύουν μεγάλες υπερβολές ή υποτιμήσεις της κυκλοφορίας. Για καλύτερη εκτίμηση της απόδοσης των αλγορίθμων πρόβλεψης, επομένως, πρέπει να ληφθούν υπόψη τόσο τα σχετικά, όσο και τα απόλυτα σφάλματα πρόβλεψης. Ένας καλός αλγόριθμος πρόβλεψης θα πρέπει να έχει χαμηλά απόλυτα και σχετικά σφάλματα πρόβλεψης. Το μέσο σχετικό σφάλμα πρόβλεψης όλων των hotspots είναι ο μέσος όρος του μέσου σχετικού σφάλματος πρόβλεψης, λαμβάνοντας υπόψη όλα τα hotspots.

3. Κοινωνική Πληροφορία Χρήστη

Όπως έχει αναφερθεί ήδη, κάθε χρήστης ενός τηλεπικοινωνιακού δικτύου ταυτοποιείται από ένα προφίλ χρήστη, το οποίο περιλαμβάνει πληροφορίες σχετικές με αυτόν. Αυτές οι πληροφορίες, όπως είδαμε και παραπάνω, μπορεί να αφορούν την τοποθεσία στην οποία αυτός βρίσκεται κάθε φορά, την κατάταξή του σε ομάδες, φίλους, καθώς και πολλά άλλα χαρακτηριστικά.

Τα τελευταία χρόνια, το μεγαλύτερο μέρος των ατόμων επικοινωνούν ή αλληλεπιδρούν μεταξύ τους μέσω κοινωνικών δικτύων. Με τη συνεχή ανάπτυξη των Πληροφοριακών και Επικοινωνιακών Τεχνολογιών (Information and Communication Technologies, ICTs), οι χρήστες μπορούν να έρθουν σε επαφή με διάφορους τρόπους, είτε μιλάμε για χρήστες ενός φυσικού δικτύου, είτε ενός εικονικού δικτύου. Παρόλα αυτά, δύο είδη κοινωνικών δικτύων θεωρούνται ιδιαίτερα δραστήρια και δημοφιλή την τελευταία δεκαετία: τα εικονικά (online) δίκτυα, όπου οι χρήστες συνδέονται μεταξύ τους με κοινοποιήσεις περιεχομένου, απόψεων, εμπειριών που αντλούν από τα διάφορα μέσα κοινωνικής δικτύωσης (π.χ. Facebook, Flickr κ.λπ.), και τα φυσικά (offline) δίκτυα, τα οποία με τη σειρά τους εκμεταλλεύονται τις Ad-Hoc

συνδέσεις μεταξύ ζευγών συσκευών (π.χ. κινητών τηλεφώνων, οχημάτων κ.λπ.), προκειμένου να κοινοποιήσουν το περιεχόμενο και τους πόρους κάθε χρήστη του δικτύου. Με λίγα λόγια, πρόκειται για μια μορφή μίμησης του τρόπου κατά τον οποίο ένα άτομο ψάχνει πληροφορίες χρησιμοποιώντας κάποιο κοινωνικό δίκτυο μέσω άμεσων επαφών προσώπου-με-πρόσωπο.

Καθεμία από τις δύο αυτές κατηγορίες δικτύων έχει διακριτά χαρακτηριστικά και αρκετά πλεονεκτήματα. Αρχικά, αυτές οι δύο «κοινωνίες» έχουν διαφορετικά τεχνικά χαρακτηριστικά, που οδηγούν και σε διακριτά είδη αλληλεπιδράσεων, όπως είναι κάποια μοτίβα που παρατηρούνται σε σχόλια ή likes online δικτύων ή μοτίβα κίνησης ή θέσης που περιλαμβάνονται σ' ένα φυσικό δίκτυο. Παράλληλα, όμως, και οι υποδομές τους είναι διαφορετικές. Οι εικονικές κοινωνίες βασίζονται σε υπηρεσίες εντός του δικτύου, ενώ οι φυσικές κοινωνίες αναπτύσσονται βασιζόμενες σε κινητά και Ad-Hoc δίκτυα. Ενώ, όμως, παρατηρούνται αρκετές διαφορές μεταξύ των δύο, μπορούμε να εντοπίσουμε και αρκετά μεταξύ τους κοινά χαρακτηριστικά. Για παράδειγμα, και τα δύο δίκτυα διευκολύνουν την κοινοποίηση και τη διάδοση πληροφοριών μεταξύ των χρηστών τους. Χάρη στο συνδετικό ρόλο που παίζουν τα άτομα μιας κοινωνίας, οι ιδιότητες των διαφορετικών κοινωνιών είναι κατά κάποιο τρόπο έμμεσα συνδεδεμένες. Σ' ένα δίκτυο φυσικού κόσμου, για παράδειγμα, η κίνηση των χρηστών είναι άρρηκτα συνδεδεμένη με τις online επαφές τους.

Γενικά, ένα δίκτυο φυσικού κόσμου ακολουθεί διαφορετική γραμμή έρευνας σε σχέση μ' αυτή που ακολουθεί ένα εικονικό δίκτυο. Παρόλα αυτά, υπάρχουν αρκετές περιπτώσεις και των δύο αυτών τύπων δικτύου που χειρίζονται αλληλεπιδράσεις διαφορετικών τύπων κοινωνιών. Αυτό οφείλεται στο γεγονός ότι ένα άτομο μπορεί να ανήκει σε περισσότερες από μία ετερογενείς κοινότητες, τις οποίες και διατρέχει στην καθημερινότητά του. Επομένως, δεν είναι δυνατό να προκύψει ένα συνολικό, έγκυρο, μοντέλο κατανόησης και πρόβλεψης της κινητικότητας των χρηστών, χρησιμοποιώντας κάθε φορά μόνο έναν τύπο δεδομένων.

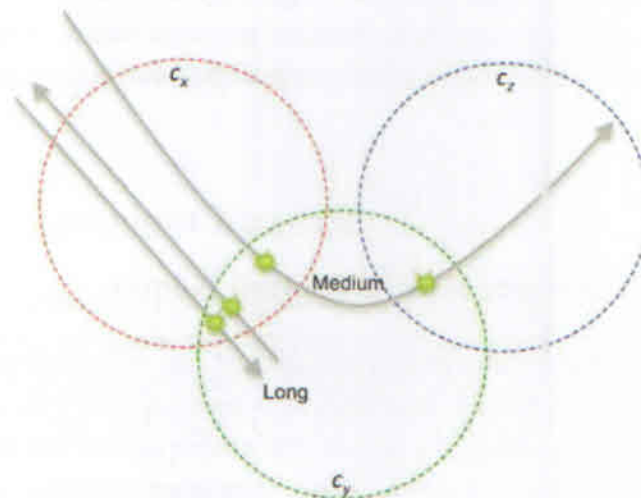
3.1. Αξιοποίηση Πληροφορίας στο Φυσικό Κόσμο

Γενικά, ένα δίκτυο φυσικού κόσμου ορίζεται ως μια offline κοινωνία ατόμων με κοινή τοποθεσία, η οποία προκύπτει από την εκμετάλλευση καθημερινών τους δραστηριοτήτων. Ένας μεγάλος αριθμός ερευνών, οι οποίες εξετάζουν την ποιότητα των σχέσεων των χρηστών ενός τηλεπικοινωνιακού δικτύου, έχουν καταλήξει στο συμπέρασμα πως τα επίπεδα εγγύτητας είναι το βασικό χαρακτηριστικό που επηρεάζει τις σχέσεις μεταξύ των χρηστών. Πιο συγκεκριμένα, φαίνεται πως όσο μικρότερη είναι η φυσική απόσταση μεταξύ ενός χρήστη και των υποψηφίων προς σύναψη χρηστών, τόσο πιθανότερο είναι και το να επέλθει μια σχέση φιλίας μεταξύ αυτών. Είναι εύκολα κατανοητό πως άτομα που βρίσκονται πιο κοντά γεωγραφικά, γίνονται και πιο εύκολα φίλοι. Αυτό δεν οφείλεται, όμως, μόνο στην ευκολότερη μεταξύ τους αλληλεπίδραση, αλλά και στο ότι τους δίνεται η δυνατότητα να ανταλλάσσουν πληροφορίες μεταξύ τους, να έχουν κοινές ομάδες, να συμμετέχουν σε κοινές δραστηριότητες. Όλα αυτά έχουν ως αποτέλεσμα μια πιθανότερη σύναψη φιλίας.

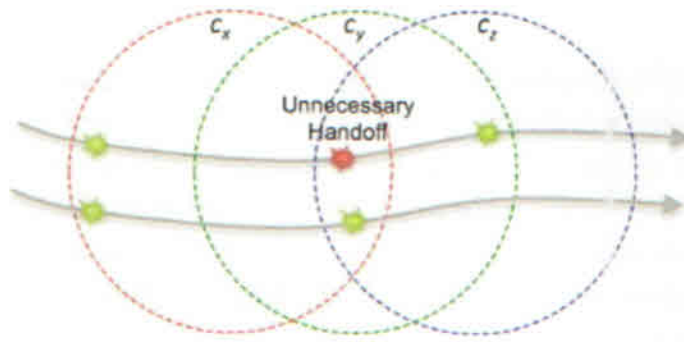
Ενδεικτικά, λοιπόν, και πέρα από τη μετρική που υπολογίζεται σ' αυτή την εργασία για την αξιοποίηση της φυσικής πληροφορίας, άλλη μια έρευνα που χρησιμοποιεί φυσική πληροφορία για την υλοποίηση ενός κατάλληλου αλγορίθμου που θα μπορέσει να ικανοποιήσει τις απαιτήσεις που αναφέρθηκαν, παρουσιάζεται στη συνέχεια.

3.1.1. Μοντέλο Πρόβλεψης βάσει Συμπεριφοράς

Πρόκειται για ένα προτεινόμενο μοντέλο που εστιάζει κυρίως στην πρόβλεψη κίνησης των χρηστών ενός δικτύου κινητής τηλεφωνίας βάσει της συμπεριφοράς τους [12]. Υλοποιεί τον αλγόριθμο πρόβλεψης της επόμενης κίνησης του χρήστη χρησιμοποιώντας (α) την τοποθεσία του, δηλαδή το ιστορικό των κινητικών μοτίβων που έχει ακολουθήσει στο παρελθόν, (β) την ομάδα, καθώς οι κινητοί σταθμοί συνήθως συμπεριφέρονται ως ομάδες, (γ) την ώρα-της-ημέρας, παράγοντας που δείχνει ότι η συμπεριφορά του χρήστη μεταβάλλεται ως συνάρτηση του χρόνου, και (δ) τη διάρκεια, η οποία διακρίνεται σε μικρή, μέτρια ή μεγάλη. Στόχος του παράγοντα της διάρκειας είναι, άμεσα, η αναπαράσταση της χρονικής διάρκειας κατά την οποία ο κινητός σταθμός είναι συνδεδεμένος μέσα σε μια κυψέλη και, έμμεσα, η αναπαράσταση της ταχύτητας με την οποία αυτός κινείται μέσα στην κυψέλη.



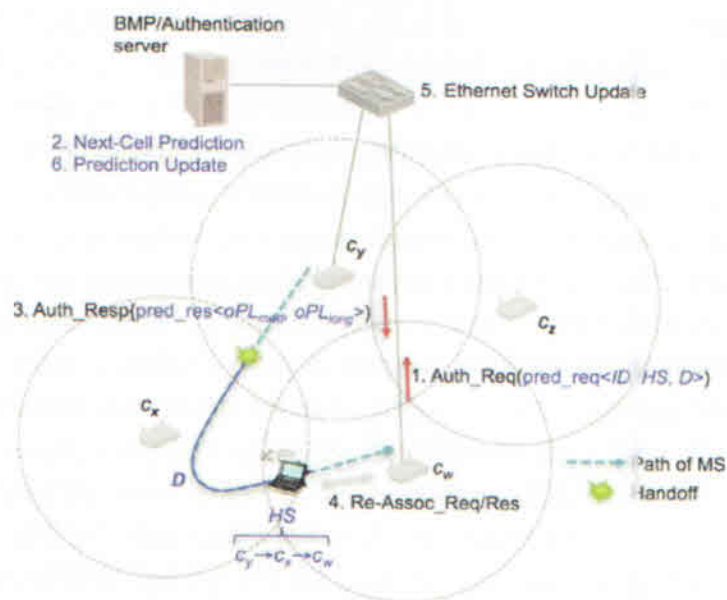
Εικόνα 1. Μέτρια διάρκεια, προεπιλογή της μετάδοσης ενός κινητού τερματικού μέσα στην κυψέλη. Μεγάλη διάρκεια, με παρουσία δραστηριότητας του κινητού τερματικού στην κυψέλη c_y .



Εικόνα 2. Μικρή διάρκεια. Αναπαριστά αχρείαστο ή ψευδές handoff.

Προκειμένου, λοιπόν, να προσομοιωθούν με ακρίβεια τα μοτίβα κινητικότητας και οι χειρισμοί σε μεγάλα δίκτυα, αναπτύχθηκε ένας προσομοιωτής που υλοποιεί ένα ραδιομοντέλο WLAN, δημιουργεί ρεαλιστικά μοντέλα κινητικότητας βασισμένα σε διατάξεις κτηρίων και πόλεων και υποστηρίζει τα πλαίσια διαχείρισης που απαιτούνται για την εφαρμογή σάρωσης, επαλήθευσης ταυτότητας και επανασύνδεσης. Επομένως, αυτό που κάνει είναι να παρακολουθεί την ακρίβεια πρόβλεψης της επόμενης κυψέλης και να αναπροσαρμόζει το μοντέλο πρόβλεψης που βασίζεται στη δυναμική συμπεριφορά των κινητών σταθμών.

Πιο συγκεκριμένα, η Εικόνα 3 απεικονίζει μια πιθανή διαμόρφωση της προτεινόμενης μεθόδου, όπου το σχήμα πρόβλεψης κίνησης υποτίθεται ότι υλοποιείται από ένα διακομιστή και είναι εγκατεστημένο μαζί με το διακομιστή ελέγχου ταυτότητας. Επιπλέον, η επόμενη κυψέλη πρόβλεψης c_w για το τρέχον handoff από c_x σε c_w , θεωρείται ότι είναι διαθέσιμη από το τελευταίο handoff από c_y σε c_x . Ως εκ τούτου, ένα κινητό τερματικό επικυρώνει/επανασυνδέει απευθείας το c_w , εξαλείφοντας την ανάγκη για εκτέλεση σάρωσης. Πλήρης σάρωση εκτελείται όταν το κινητό τερματικό εισέρχεται στο δίκτυο για πρώτη φορά.



Εικόνα 3. Πιθανή διαμόρφωση πρόβλεψης κίνησης βάσει συμπεριφοράς σε ένα WLAN περιβάλλον.

Το πρώτο βήμα της διαδικασίας ξεκινά όταν το κινητό τερματικό απαιτεί ένα handoff από c_x σε c_w , στέλνει $pred_req(ID, HS, D)$ στον εξυπηρετητή ως τμήμα της αίτησης επικύρωσης μέσω του c_w για την απόκτηση των προβλέψεων για το επόμενο handoff, όπου το ID είναι η ταυτότητα του κινητού τερματικού, HS η ακολουθία handoff για το τερματικό και D η χρονική διάρκεια που πέρασε το κινητό τερματικό στην τελευταία κυψέλη c_x . Στη συνέχεια, στο δεύτερο βήμα, ο εξυπηρετητής κάνει Πρόβλεψη Επόμενης Κυψέλης (Next-cell Prediction), η οποία περιλαμβάνει τις εξής λειτουργίες: Αρχικά, τα τελευταία k handoffs στην ακολουθία χρησιμοποιούνται για την εύρεση του πίνακα HS , για ένα σύνολο αντίστοιχων καταχωρήσεων που αντιπροσωπεύουν μια μη προσαρμοσμένη λίστα πρόβλεψης επόμενης κυψέλης (unordered Prediction List, uPL). Παράλληλα, το ID χρησιμοποιείται για την κατάταξη του πίνακα ομάδας, προκειμένου να βρει την ID ομάδα G_i στο σύνολο $\{G_0, G_1, \dots, G_{p-1}\}$, όπου το p αναπαριστά τον αριθμό των ομάδων. Στη συνέχεια, το G_i χρησιμοποιείται προκειμένου να επιλέξει ένα συγκεκριμένο ομαδικό χαρακτηριστικό της ώρας-της-ημέρας, το ToD_i . Όπως φαίνεται και από το όνομα, τα ToD μοντελοποιούν χαρακτηριστικά της ώρας-της-ημέρας διαφορετικών ομάδων χρηστών κινητών. Η προκαθορισμένη περίοδος για ένα ToD_i είναι T , που σημαίνει πως τα κινητά τερματικά που ανήκουν σ' αυτή την ομάδα, παρουσιάζουν παρόμοια handoff συμπεριφορά κατά τη συνολική διάρκεια της περιόδου T . Επιπλέον, μια ομάδα κινητών τερματικών που παρουσιάζει βραχυπρόθεσμες συμπεριφορές και συμπεριφορές περιοδικού handoff, μπορεί να μοντελοποιηθεί ξεχωριστά ως $ToD_i = \{\tau_0, \tau_1, \dots, \tau_{q-1}\}$, όπου το τ_j αναπαριστά ένα χρονικό τμήμα και το q αναπαριστά το συνολικό αριθμό των χρονικών αυτών των τμημάτων. Επομένως, ο τρέχων χρόνος, δηλαδή το ρολόι, προσδιορίζει το τ_j για μια συγκεκριμένη ομάδα G_i . Η σημασία του (ToD_i, s_j) είναι ότι ορίζει με μοναδικό τρόπο ένα σύνολο Ενσωματωμένων Κινητών Μέσων Όρων (Integrated Moving Average, IMA), προκειμένου να αναπαραστήσει τις συχνότητες των handoff ακολουθιών για μια συγκεκριμένη ομάδα κινητών τερματικών, με συγκεκριμένα χαρακτηριστικά ώρα-της-ημέρας και διάρκεια. Επομένως, το (ToD_i, s_j) μαζί με το HS_{match} προσδιορίζουν το σωστό υποσύνολο των προαναφερθέντων μέσων όρων, οι οποίοι θα χρησιμοποιηθούν στην εφαρμογή της σειράς προτεραιότητας για τις επόμενες προβλέψεις κυψελών (ορισμένες λίστες πρόβλεψης και για μεσαίας διάρκειας oPL_{med} και για μακράς διάρκειας oPL_{long}). Μόλις προσδιοριστεί η προτεραιότητα, στο τρίτο βήμα της διαδικασίας, ο διακομιστής πρόβλεψης κίνησης επιστρέφει $pred_resp(oPL_{med}, oPL_{long})$ ως κομμάτι της απάντησης επιβεβαίωσης. Μετά τη λήψη της απάντησης επικύρωσης, ακολουθεί το τέταρτο βήμα της διαδικασίας, στο οποίο το κινητό τερματικό επανασυνδέεται με το σημείο πρόσβασης στην κυψέλη c_w και ακολουθείται από τις Ενημερώσεις Ethernet Switch (Βήμα 5) και από τις λειτουργίες Ενημέρωσης Πρόβλεψης (Βήμα 6), οι οποίες ολοκληρώνουν το handoff. Κατά το επόμενο handoff, η πρώτη πρόβλεψη είτε στο oPL_{med} , είτε στο oPL_{long} , χρησιμοποιείται ως η πρόβλεψη της επόμενης κυψέλης, με βάση το αν η διάρκεια του κινητού τερματικού στην τρέχουσα κυψέλη c_w είναι μεσαία ή μακρά, αντίστοιχα. Η πρόβλεψη επόμενης κυψέλης έχει ως προεπιλογή το oPL_{med} , αν το κινητό τερματικό παρουσιάζει σύντομη διάρκεια στην τρέχουσα κυψέλη. Αν αποτύχει η πρώτη πρόβλεψη, δηλαδή αποτύχει η επικύρωση/επανασύνδεση, χρησιμοποιείται η δεύτερη πρόβλεψη κ.ο.κ. Όταν αποτύχουν όλες οι προβλέψεις στο $oPL_{med} = oPL_{long}$, πραγματοποιείται πλήρης σάρωση.

Ως τελικό αποτέλεσμα, η μελέτη αυτής της προσομοίωσης έδειξε ότι η πρόβλεψη κίνησης βάσει συμπεριφοράς (Behavior-based Mobility Prediction, BMP) παρέχει καλύτερες προβλέψεις, σε σχέση με τα συστήματα που βασίζονται σε εντολές, και οδηγεί σε πολύ χαμηλότερη καθυστέρηση μετάδοσης για όλους τους κινητούς σταθμούς.

3.2. Αξιοποίηση Πληροφορίας στον Εικονικό Κόσμο

Ένα δίκτυο online κόσμου απεικονίζει εικονικές κοινωνίες που αλληλεπιδρούν μεταξύ τους και μοιράζονται πληροφορίες μέσω διαφόρων μέσων κοινωνικής δικτύωσης ή ακόμα και πληθωριστικών σελίδων, όπως το Wiki. Στον τομέα των κινητών εφαρμογών, η πρόβλεψη μοτίβων κινητικότητας των χρηστών είναι ιδιαίτερα σημαντική για πολλούς λόγους. Για παράδειγμα, η πρόβλεψη της κινητικότητας θα μπορούσε να χρησιμοποιηθεί για να καθοριστεί το πού θα βρίσκεται ένα άτομο και ποιον θα συναντήσει. Πληροφορίες τέτοιου τύπου καλούνται μετρικές ανάλυσης και μπορούν να καταστήσουν κινητές εφαρμογές, βασισμένες στην τοποθεσία, ικανές να παρέχουν εξατομικευμένες υπηρεσίες, σχετικές με το συγκεκριμένο στο οποίο βρίσκεται ο χρήστης.

Πιο συγκεκριμένα, το σύνολο αυτό των απαραίτητων πληροφοριών μπορεί να διαφέρει από σύστημα σε σύστημα, ανάλογα με την περίπτωση και τα δεδομένα που πρέπει να μελετηθούν κάθε φορά για την υλοποίηση. Ωστόσο, μερικές από τις μετρικές αυτές που συμβάλλουν στην καλύτερη μελέτη ενός εικονικού δικτύου είναι ενδεικτικά:

- *Πυκνότητα (Density)*: Αναφέρεται στις συνδέσεις μεταξύ των χρηστών του συστήματος. Η πυκνότητα ορίζεται ως το πηλίκο του συνόλου των συνδέσεων που πραγματοποιεί ο κάθε χρήστης προς το σύνολο των πιθανών συνδέσεων που θα μπορούσε να είχε κάνει. Για παράδειγμα, αν ένα σύστημα αποτελείται από 20 χρήστες, και κάθε χρήστης θα μπορούσε να συνδεθεί ενδεχομένως και με τους υπόλοιπους 19, τότε η μέγιστη πυκνότητα του συστήματος αυτού είναι 100% (19/19). Αντίστοιχα, το 5% δείχνει ότι μόνο 1 στις 19 συνδέσεις είναι εφικτή.
- *Κεντρικό Σημείο (Hub)*: Θεωρείται ο κόμβος του οποίου το σύνολο των συνδέσεων υπερβαίνει το μέσο όρο συνδέσεων του συστήματος. Το σημείο αυτό έχει αρκετά μεγάλη σημασία αναφορικά με τις συνδέσεις ενός συστήματος και κάποια από τα βασικότερα χαρακτηριστικά του είναι αναφορικά (α) η «γήρανσή» τους, (β) η σύμπτυξη του μήκους των μονοπατιών του συστήματος, (γ) η αντοχή στις επιθέσεις, (δ) ο βαθμός συσχέτισης, καθώς και (ε) το φαινόμενο διάδοσης [13]. Αναφορικά με την αποκαλούμενη «γήρανση» των κόμβων αυτών, (α), πρόκειται για ένα φαινόμενο που παρουσιάζεται σ' ένα σύστημα όταν οι παλαιότεροι κόμβοι του υποσκιάζονται. Το φαινόμενο αυτό είναι, επίσης, υπεύθυνο και για τυχόν αλλαγές στην εξέλιξη και την τοπολογία του εκάστοτε συστήματος. Σχετικά με τη σύμπτυξη των μονοπατιών, (β), όσο πιο παρατηρήσιμο είναι ένα τέτοιο κεντρικό σημειοκόμβος μέσα στο σύστημα, τόσο περισσότερο συρρικνώνονται οι αποστάσεις μεταξύ των κόμβων. Στην περίπτωση δικτύου χωρίς κλίμακα, δηλαδή δικτύου όπου ο αριθμός των κόμβων του συνεχώς αυξάνεται, αυτά τα κεντρικά σημεία λειτουργούν ως γέφυρες μεταξύ

των υπολοίπων κόμβων, μικρότερου βαθμού. Όσον αφορά το (γ), την αντοχή δηλαδή στις επιθέσεις, διακρίνονται δύο περιπτώσεις επιθέσεων. Η πρώτη αποτελεί την τυχαία αποτυχία των κόμβων, όπου τα κεντρικά σημεία-κόμβοι είναι υπεύθυνα για την αντοχή του δικτύου. Η δεύτερη είναι η στοχευμένη επίθεση στα κεντρικά αυτά σημεία, κατά την οποία η ακεραιότητα του δικτύου μπορεί να καταρρεύσει πολύ εύκολα, καθώς όλοι οι κόμβοι του δικτύου είναι συνδεδεμένοι μ' ένα κεντρικό σημείο. Επομένως, μια επίθεση στα κεντρικά αυτά σημεία μπορεί να συμβάλλει στην κατάρρευση ολόκληρου του δικτύου. Τέλος, ο βαθμός (δ) συσχέτισης υποδεικνύει ότι κάθε κόμβος μπορεί να συνδεθεί μόνο με κόμβους του ίδιου βαθμού, ενώ το (ϵ) σημαίνει πως κάθε τέτοιο κεντρικό σημείο του δικτύου είναι υπεύθυνο και για την αποδοτική διάδοση υλικού μέσα στο δίκτυο.

- *Κεντρικότητα Βαθμού (Degree Centrality)*: Πρόκειται για μια μετρική που υπολογίζει την έκταση στην οποία ένας κόμβος ενός δικτύου συνδέεται με όλους τους υπόλοιπους κόμβους του και εξαρτάται από το μέγεθος του ίδιου του δικτύου [14]. Με λίγα λόγια, όσο μεγαλύτερο είναι το δίκτυο, τόσο υψηλότερος είναι ο μέγιστος πιθανός βαθμός της κεντρικότητας. Αυτό σημαίνει ότι για μια συγκεκριμένη τιμή του βαθμού της κεντρικότητας, είτε ότι ένας κόμβος είναι συνδεδεμένος καλά μέσα σ' ένα μικρό δίκτυο, είτε ότι συνδέεται μόνο με ένα μικρό αριθμό κόμβων ενός μεγάλου δικτύου.
- *Κεντρικότητα Ενδιαμεσότητας (Betweenness Centrality)*: Η έννοια της ενδιαμεσότητας της κεντρικότητας αναφέρεται στον τρόπο με τον οποίο άλλοι κόμβοι χειρίζονται ή μεσολαβούν στις σχέσεις μεταξύ δυνάδων που δεν είναι άμεσα συνδεδεμένες [14]. Η συγκεκριμένη μετρική αποτελεί ένα σημαντικό δείκτη ελέγχου ανταλλαγής πληροφοριών ή ροής δεδομένων του δικτύου. Αν, για παράδειγμα, ένας κόμβος i πρέπει να περάσει από τον j για να επικοινωνήσει με τον k , τότε ο κόμβος i ελέγχει το περιεχόμενο και το συγχρονισμό της μετάδοσης των μηνυμάτων μεταξύ των κόμβων j και k . Όσο συχνότερα ο κόμβος i εντοπίζεται σ' ένα τέτοιο μονοπάτι δυνάδας, τόσο μεγαλύτερη είναι η πιθανότητα να ελέγχει και τις διεπαφές του δικτύου.
- *Επιρροή (Influence)*: Επικεντρώνεται σ' έναν μεμονωμένο κόμβο και την άμεση «γειτονιά» του μέσα στο δίκτυο [15]. Μια γενική εικόνα της κοινωνικής συμπεριφοράς μέσα σ' ένα εικονικό δίκτυο μπορούμε να δούμε, αν θεωρήσουμε πως οι χρήστες προσαρμόζουν τη συμπεριφορά, τις απόψεις και τις πεποιθήσεις τους στα πλαίσια της συμπεριφοράς άλλων μελών του κοινωνικού δικτύου στο οποίο συμμετέχουν. Η διαδικασία αυτή αποτελεί το μοντέλο κοινωνικής επιρροής του δικτύου. Ως ένα δίκτυο επαφών, το σύστημα είναι αυτό που ορίζει σε ποιο μέλος βρίσκεται εκτεθειμένος ένας χρήστης του. Οι άμεσες επαφές του χρήστη, δηλαδή οι γείτονές του, είναι συνήθως αυτές που επηρεάζονται περισσότερο από τη συμπεριφορά του. Με άλλα λόγια, το τοπικό περιεχόμενο του χρήστη (ego-network) είναι αυτό που πιθανότερα θα διαμορφώσει τη συμπεριφορά του.
- *Μεταβατικότητα (Transitivity)*: Εφαρμόζεται σ' ένα μη υπογεγραμμένο κατευθυνόμενο δίκτυο, στην περίπτωση που αυτό αποτελείται από τέτοιες ομάδες, έτσι ώστε οι ομάδες ίδιου βαθμού να μη σχετίζονται. Αντίστοιχα, ομάδες διαφορετικού βαθμού να σχετίζονται μεταξύ τους με βάση δυνάδες null ή ασυμμετρικές δυνάδες, που οδηγούν σε υψηλότερη βαθμίδα [15].

Είναι φανερό, επομένως, πως η συμπεριφορά ενός οποιαδήποτε χρήστη ενός εικονικού κοινωνικού δικτύου μπορεί να γίνει αρκετά εύκολα κατανοητή, λαμβάνοντας υπόψη διάφορες πτυχές της καθημερινής του ζωής. Η συμπεριφορά του χρήστη, λοιπόν, μέσα σ' ένα κοινωνικό

δίκτυο (Online Social Network (OSN) behavior) καλύπτει ποικίλες κοινωνικές δραστηριότητες που μπορούν να πραγματοποιήσουν οι χρήστες online, όπως είναι η δημιουργία φιλίας, η δημοσίευση περιεχομένου, οι περιηγήσεις (browsing) σε προφίλ άλλων χρηστών του δικτύου, η αποστολή/ λήψη μηνυμάτων κ.λπ.

Στη συνέχεια, περιλαμβάνεται μια σειρά από ήδη υπάρχουσες τεχνικές παρατήρησης των διεπαφών μεταξύ των χρηστών ενός συστήματος, καθώς επίσης και πρόβλεψης της κινητικότητάς τους [16].

Αναλυτικότερα, για το κομμάτι της σύνδεσης και των διεπαφών, οι λύσεις είναι οι εξής:

1. *Μη-κατευθυνόμενο γραφικό μοντέλο (Undirected Graph model)*: Σ' ένα γράφημα φιλίας, οι κόμβοι του γραφήματος παρουσιάζουν τους χρήστες του δικτύου και οι ακμές που ενώνουν τους κόμβους αυτούς, δείχνουν τις μεταξύ τους φιλίες. Παρατηρείται ότι σε πλατφόρμες κοινωνικής δικτύωσης κάθε χρήστης έχει επαφές μόνο μ' ένα μικρό σύνολο των φίλων του, ενώ με τους υπόλοιπους (οι οποίοι είναι πολλοί περισσότεροι από τους μισούς του «φίλους») δεν έχει καμία επαφή. Έτσι, λοιπόν, προέκυψε το γράφημα διεπαφής (interaction graph), όπου ακμή μεταξύ των κόμβων δημιουργείται μόνο στην περίπτωση εμφανούς διεπαφής μεταξύ δύο χρηστών. Το γεγονός αυτό οδηγεί και στο συμπέρασμα ότι ένα γράφημα διεπαφής προσφέρει καλύτερα αποτελέσματα σε σχέση μ' ένα γράφημα φιλίας.
2. *Κατευθυνόμενο γραφικό μοντέλο (Directed Graph model)*: Ως λανθάνουσες διεπαφές ορίζονται οι παθητικές διεπαφές των χρηστών online κοινωνικών δικτύων, οι οποίες δεν μπορούν να παρατηρηθούν από παραδοσιακές τεχνικές μετρήσεων. Σ' ένα κατευθυνόμενο λανθάνον γράφημα, μια γραμμή με φορά από το A στο B δείχνει ότι ο χρήστης A επισκέφθηκε το προφίλ του χρήστη B. Έτσι, ο αριθμός των ακμών που κατευθύνονται προς την κορυφή (in-degree) ενός κόμβου δείχνει τον αριθμό των επισκεπτών σ' εκείνο το προφίλ χρήστη και ο αριθμός των ακμών από την κορυφή του γραφήματος (out-degree) τον αριθμό των προφίλ που εκείνος επισκέφθηκε. Τα χαρακτηριστικά των υποβοσκομένων γραφημάτων κυμαίνονται μεταξύ των γραφημάτων ορατών διεπαφών και των κλασικών γραφημάτων φιλίας.
3. *Δειγματοληψία γραφήματος (Graph sampling)*: Οι τεχνικές αυτές χρησιμοποιούνται για να εξάγουν ένα μικρότερο, αλλά αντιπροσωπευτικό, στιγμιότυπο των κοινωνικών γραφημάτων (social graphs).

Στο κομμάτι της κινητικής δραστηριότητας, αντίστοιχα, οι ήδη υπάρχουσες λύσεις είναι:

1. *Παρακολούθηση κίνησης (Traffic monitoring)*: Μελέτη των online κοινωνικών δικτύων μπορεί να γίνει και καταγράφοντας την κίνηση του αντίστοιχου δικτύου. Η καταγραφή της κίνησης αυτής πραγματοποιείται μέσω δεδομένων clickstream, δεδομένων δηλαδή από τα κλικ που επιλέγει να κάνει ο χρήστης, που συλλέγονται από κοινωνικά δίκτυα. Επειδή, όμως, μπορεί να είναι «ημιτελή», περιορίζεται έτσι η χρήση και η απόδοσή τους. Καταρχάς, είναι περιορισμένα από τη διάρκεια συλλογής τους, ενώ δεν καταγράφεται ούτε η συμπεριφορά ανενεργών χρηστών. Επιπλέον, τα δεδομένα αυτά είναι περιορισμένα από τις τοποθεσίες καταγραφής, δηλαδή καταγράφεται μόνο η συμπεριφορά χρηστών που χρησιμοποιούν συγκεκριμένους παρόχους διαδικτυακών υπηρεσιών.

2. *Περιοχή ενδιαφέροντος (Locality of interest)*: Τα κέντρα δεδομένων κοινωνικών μέσων είναι τοποθετημένα στις ΗΠΑ, προκειμένου να παρέχουν συνεπείς υπηρεσίες σε όλους τους χρήστες ανά τον κόσμο. Επομένως, χρήστες εκτός των ορίων των ΗΠΑ αντιμετωπίζουν κάποια καθυστέρηση στην απόκριση, καθώς δημιουργείται αχρείαστη κίνηση στη ραχοκοκαλιά του διαδικτύου.
3. *Χαρακτηριστικά πλοήγησης (Navigation characteristics)*: Τα online κοινωνικά δίκτυα πλέον αποτελούν ένα σημαντικό κομμάτι της διαδικτυακής κίνησης σε σύγκριση με τις μηχανές αναζήτησης. Έχει παρατηρηθεί ότι χρήστες online κοινωνικών μέσων δικτύωσης είναι λιγότερο πιθανό να περιηγηθούν σε εξωτερικές ιστοσελίδες. Ωστόσο, όταν τις επισκέπτονται, παραμένουν για μεγαλύτερο χρονικό διάστημα σε σχέση μ' αυτούς που χρησιμοποιούν τις μηχανές αναζήτησης. Επιπλέον, τα κοινωνικά δίκτυα στέλνουν τους χρήστες σ' ένα μικρότερο υποσύνολο του διαδικτύου, σε σχέση με τις μηχανές αναζήτησης, και σε όχι και τόσο γνωστές σελίδες.

Οι δραστηριότητες των χρηστών ενός δικτύου, όμως, ενίοτε μπορεί να είναι έγκυρες, αλλά μπορεί να είναι και κακόβουλες. Επομένως, η κατανόηση της διαδικτυακής κοινωνικής συμπεριφοράς ενός χρήστη είναι σημαντική για διαφορετικές διαδικτυακές οντότητες από πολλές απόψεις:

- Για παρόχους διαδικτυακών υπηρεσιών (Internet Service Providers, ISPs): Καθώς η κίνηση της διαδικτυακής κοινωνικής συμπεριφοράς του χρήστη μεγαλώνει ραγδαία και γίνεται σημαντική, θέλουν να γνωρίζουν την εξέλιξη του μοτίβου κίνησης των online κοινωνικών δικτύων.
- Για παρόχους υπηρεσιών online κοινωνικής δικτύωσης: Τους βοηθά στην κατανόηση της συμπεριφοράς των πελατών τους απέναντι σε διαφορετικές συναρτήσεις, ειδικότερα όταν αυτές βρίσκονται ακόμα σε πειραματικό στάδιο. Επιπλέον, από άποψη επένδυσης στις υποδομές, όπως τοποθεσίες που θεωρούνται καλύτερες από άποψη κόστους (cost-effective) για την κατασκευή κέντρων δεδομένων ή ποιοι clusters δικτύου παράδοσης περιεχομένου (Content Delivery Network, CDN) θα συνεισφέρουν στην παράδοση των συχνότερα προσβάσιμων δεδομένων, είναι άκρως σημαντική η κατανόηση της γεωγραφικής κατανομής και της κινητικής δραστηριότητας των χρηστών.
- Για τους ίδιους τους χρήστες των online μέσων κοινωνικής δικτύωσης: Η μελέτη συμπεριφοράς είναι σημαντική για τη βελτίωση της εμπειρίας του χρήστη. Για παράδειγμα, υπάρχουν πολλοί κακόβουλοι λογαριασμοί στα κοινωνικά δίκτυα. Αυτοί οι λογαριασμοί δημιουργούν ανεπιθύμητα μηνύματα για τους έγκυρους χρήστες, οπότε η αναγνώριση και το μπλοκάρισμά τους είναι ιδιαίτερα σημαντικά για τη διασφάλιση της καλής εμπειρίας των υπολοίπων χρηστών του δικτύου.

Προκειμένου να γίνει λίγο πιο κατανοητή η διαδικασία που ακολουθείται για την εξαγωγή και αξιοποίηση των πληροφοριών ενός εικονικού δικτύου, ακολουθούν κάποιες ήδη υπάρχουσες προσεγγίσεις.

3.2.1. Πρωτόκολλο Δρομολόγησης SocialCLWPR

Μια πρώτη μελέτη υλοποίησης ενός αλγορίθμου πρόβλεψης συστάσεων μεταξύ χρηστών σ' ένα εικονικό δίκτυο, είναι αυτή του χειρισμού συμπεριφοράς των χρηστών ενός δικτύου οχημάτων (Vehicular Ad-Hoc Networks, VANET) [17]. Η συγκεκριμένη έρευνα προτείνει ένα βελτιστοποιημένο πρωτόκολλο δρομολόγησης βασισμένο σε θέσεις, για τα αστικά περιβάλλοντα VANET, εισάγοντας κοινωνικές πτυχές στο σχεδιασμό του, έτσι ώστε να αντιμετωπιστούν αποτελεσματικά κόμβοι που παρουσιάζουν κακή συμπεριφορά (εγωιστές ή κακόβουλοι), που σκοπίμως εγκαταλείπουν τα πακέτα και τελικά προοθούν τη συνεργασία στο δίκτυο. Το προτεινόμενο πρωτόκολλο αυτό ονομάζεται SocialCLWPR (Cross-Layer Weighted Position-based Routing) και ευνοεί τους κόμβους με στενή κοινωνική σχέση (μέλη της οικογένειας, φίλους, χρήστες με συνδέσεις εμπιστοσύνης) ως επόμενους κόμβους προώθησης. Αντίθετα, οι κόμβοι με ρητές συνδέσεις εμπιστοσύνης, όπως υποδεικνύονται από τα online κοινωνικά μέσα, δεν προτιμώνται στο δίκτυο.

Το SocialCLWPR, λοιπόν, έχει τα ακόλουθα βασικά χαρακτηριστικά: Πρώτα απ' όλα, είναι ένα πρωτόκολλο βασισμένο στη θέση που χρησιμοποιεί την απόσταση στο δρόμο (καμπύλη) ως μετρική, αντί της πραγματικής γεωγραφικής (ευκλείδειας) απόστασης. Επιπρόσθετα, χρησιμοποιεί την πρόβλεψη της θέσης του κόμβου, καθώς και πληροφορίες πλοήγησης, για τη βελτίωση της αποτελεσματικότητας του πρωτοκόλλου δρομολόγησης σε ένα δίκτυο οχημάτων. Αυτό επιτρέπει την αποτελεσματικότερη επιλογή της επόμενης μετακίνησης (next-hop) στις αστικές περιοχές υψηλών προδιαγραφών, όπου η διάδοση των μηνυμάτων ακολουθεί τα μοτίβα κίνησης των οχημάτων, μειώνοντας έτσι τις καθυστερήσεις από άκρο σε άκρο. Το SocialCLWPR παρακολουθεί, επίσης, την ποιότητα της σύνδεσης, χρησιμοποιώντας παραμέτρους όπως τους SNIR και MAC ρυθμούς σφάλματος πλαισίου. Επιπλέον, η αξιοπιστία των κόμβων εκτιμάται λαμβάνοντας υπόψη τον αριθμό των διαδοχικών μηνυμάτων "HELLO" που λαμβάνονται από έναν συγκεκριμένο κόμβο. Τέλος, ο μηχανισμός μεταφοράς και προώθησης χρησιμοποιείται σε αραιά δίκτυα, όταν ο κόμβος βρίσκεται αντιμέτωπος με το πρόβλημα τοπικών μεγίστων (διαπιστώνεται, δηλαδή, ότι είναι ο καλύτερος δρομολογητής στον κόμβο προορισμού). Από την άποψη αυτή, λαμβάνονται υπόψη οι πληροφορίες ουράς αναφορικά με τον αριθμό των πακέτων μεταφοράς και προώθησης, ώστε να παρέχεται κάποιο είδος εξισορρόπησης της κυκλοφορίας για καλύτερη Ποιότητα Υπηρεσιών και προκειμένου να αποφεύγονται τα τοπικά μέγιστα.

Αναλυτικότερα, προκειμένου να προβλέψει την επόμενη θέση του κόμβου, υπολογίζει το βάρος κάθε γειτονικού κόμβου, με βάση την επιλογή που χρησιμοποιείται, οπότε και προσφέρει μεγαλύτερο Λόγο Αποστολής Πακέτων (Packet Delivery Ratio) και χαμηλότερη καθυστέρηση από άκρη σε άκρη. Για την επιλογή του κόμβου, υπολογίζει την τιμή προτίμησης κάθε κόμβου, από τη λίστα με τους γειτονικούς κόμβους προς εκείνη την κατεύθυνση, συμπεριλαμβανομένου και του ίδιου, σύμφωνα με τον τύπο:

$$f = w_1 \times NormDistance + w_2 \times NormAngle + w_3 \times NormRoad + w_4 \times NodeReliability + w_5 \times MACinfo + w_6 \times CNFinfo + w_7 \times WeightedSNIR + w_8 \times SocialRelation,$$

και τις μεταβλητές να έχουν ως εξής:

1. *NormDistance*, κανονικοποιημένη απόσταση καμπύλης μεταξύ δυο κόμβων σ' ένα εύρος αναφοράς επικοινωνίας. Λαμβάνει υπόψη την προβλεπόμενη τοποθεσία του δεύτερου κόμβου, χρησιμοποιώντας ταχύτητα και πληροφορίες κατεύθυνσης που προσφέρονται από τα μηνύματα «HELLO».
2. *NormAngle*, αμοιβαία κατεύθυνση των δύο κόμβων, σε περίπτωση που αυτοί πλησιάζονται ή απομακρύνονται. $NormAngle = \cos\theta$.
3. *NormRoad*, παίρνει τιμή 0, αν οι κόμβοι κινούνται στον ίδιο δρόμο, αλλιώς παίρνει τιμή ίση με 1. $NormRoad = \begin{cases} 0, & \text{αν τα οχήματα είναι στον ίδιο δρόμο} \\ 1, & \text{αν τα οχήματα είναι σε διαφορετικούς δρόμους} \end{cases}$
4. *NodeReliability*, υπολογίζεται με βάση τον αριθμό των ληφθέντων «HELLO» μηνυμάτων, σύμφωνα μ' ένα δεδομένο κατώφλι, το οποίο μπορεί να ποικίλει ανάλογα με το ρυθμό εκπομπής των μηνυμάτων. $NodeReliability = \begin{cases} 1, & \text{αν } HelloCount \leq 2 \\ 0.5, & \text{αν } 2 < HelloCount \leq 4 \\ 0, & \text{αν } HelloCount > 4 \end{cases}$
5. *MACInfo*, επίπεδο διαμάχης στην περιοχή κοντά στο γειτονικό κόμβο. Υπολογίζει το μέσο αριθμό συγκρούσεων κατά τη χρονική περίοδο μεταξύ δύο μηνυμάτων «HELLO».
6. *CNFInfo*, επίπεδο χρήσης του κόμβου ως προς τον αριθμό των αποθηκευμένων πακέτων. Χρησιμοποιείται για να «τιμωρήσει» τους κόμβους που βρίσκονται σε τοπική μέγιστη κατάσταση.
7. *WeightedSNIR*, δείχνει την ποιότητα του καναλιού μεταξύ δύο κόμβων, δίνοντας προτεραιότητα σε κόμβους μακριά από τον κόμβο αποστολέα, αλλά όχι πολύ κοντά στην άκρη, όπου η πιθανότητα απόρριψης είναι υψηλή. Μεγαλύτερη προτεραιότητα δίνεται σε κόμβους με μικρότερο SNIR. $WeightedSNIR = \begin{cases} ax^2, & \text{αν } SNIR \leq SNIR_{th} \\ be^{-x}, & \text{αν } SNIR \geq SNIR_{th} \end{cases}$, με a, b παραμέτρους βελτιστοποίησης. Το εύρος στο οποίο δίνεται προτεραιότητα καθορίζεται από το κατώφλι του $SNIR_{th}$.
8. *SocialRelation*, αποτελεί τις κοινωνικές σχέσεις μεταξύ δυο κόμβων, με -1 να αναπαριστά μια φιλική/ αξιόπιστη σχέση, 1 μια αναξιόπιστη/ μη ευνοϊκή και 0 ουδέτερες σχέσεις, ή χρησιμοποιείται σε περιπτώσεις ανύπαρκτων σχέσεων. Οι σχέσεις αυτές βρίσκονται από τον πίνακα γειτνίασης. $SocialRelation = \begin{cases} -1, & \text{φιλική σχέση} \\ 0, & \text{ουδέτερες σχέσεις} \\ 1, & \text{αναξιόπιστη σχέση} \end{cases}$

Τέλος, τα βάρη w_i παρέχουν τη σχετική βαρύτητα των οκτώ παραγόντων στον υπολογισμό της τιμής προτίμησης, προκειμένου να υπάρξει μια αποτελεσματική απόφαση προώθησης.

Όσον αφορά την αξιολόγηση της απόδοσης του προτεινόμενου αυτού πρωτοκόλλου, φαίνεται πως από τα πειραματικά αποτελέσματα έχει πολύ καλύτερη απόδοση σε σχέση με το απλό CLWPR, λαμβάνοντας υπόψη ως μετρική το Λόγο Αποστολής Πακέτων. Η καλύτερη απόδοσή του οφείλεται στο γεγονός ότι το SocialCLWPR λαμβάνει υπόψη τόσο αξιόπιστες, όσο και αναξιόπιστες συνδέσεις, όπου οι αναξιόπιστες συνδέσεις αποτελούν το βασικότερο λόγο βελτίωσης της απόδοσης.

3.2.2. Σύστημα Εμπιστοσύνης για Εξατομικευμένες Προτάσεις

Πρόκειται για ένα σύστημα εμπιστοσύνης, το οποίο έχει ως στόχο την παροχή εξατομικευμένων προτάσεων χρηστών ενός εικονικού τηλεπικοινωνιακού δικτύου [18]. Η συγκεκριμένη μελέτη θα μας απασχολήσει και αργότερα στην υλοποίηση, καθώς ο απλοποιημένος αλγόριθμος που υλοποιείται στα πλαίσια της εργασίας αυτής βασίζεται σε μια απλουστευμένη έκδοση αυτού του συστήματος.

Πιο αναλυτικά, λοιπόν, και σε αντίθεση με ήδη υπάρχουσες προτάσεις, το σύστημα αυτό ενσωματώνει και την έννοια της εμπιστοσύνης μεταξύ των ατόμων, προκειμένου να βοηθήσει τα μέλη μιας ομάδας να πάρουν αποφάσεις που αφορούν τα μέλη μιας άλλης ομάδας. Πιο συγκεκριμένα, το σύστημα αυτό παρέχει στους χρήστες του εξατομικευμένες θετικές ή/και αρνητικές προτάσεις, οι οποίες στη συνέχεια μπορούν να χρησιμοποιηθούν για τον καθορισμό νέων έμπιστων ή μη αξιόπιστων σχέσεων στο δίκτυο. Προκειμένου να γίνει αυτό, όμως, θεωρήθηκε πως η έννοια της εμπιστοσύνης περιλαμβάνει κοινωνικό περιεχόμενο και από τους δύο χρήστες (φίλοι ή εχθροί), που εκφράζεται μέσω άμεσων συνδέσεων χρήστη-με-χρήστη, καθώς επίσης και από τα κοινά ενδιαφέροντα και τις επιθυμίες των χρηστών, τα οποία εισάγονται από άμεσες και έμμεσες συνδέσεις χρηστών-με-αντικείμενο.

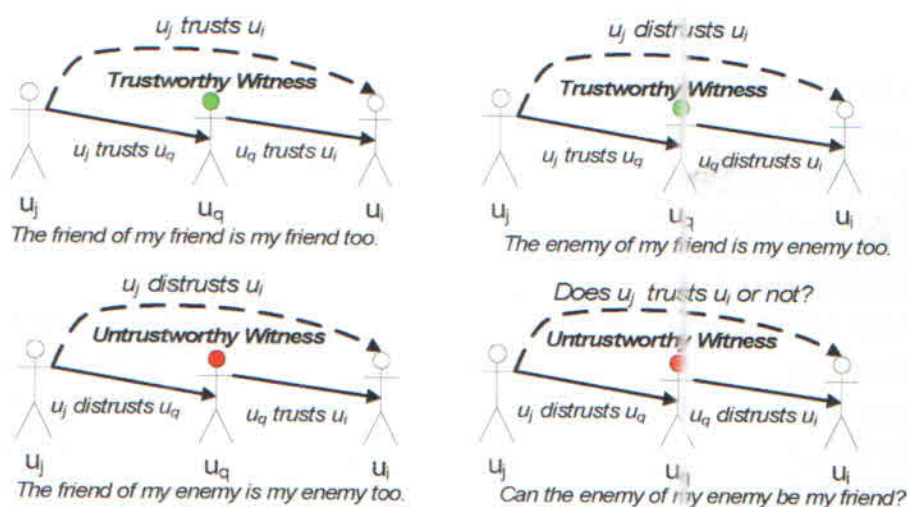
Το προτεινόμενο αυτό σύστημα, λοιπόν, βασίζεται σ' ένα μηχανισμό «φήμης», ο οποίος βαθμολογεί τους συμμετέχοντες με βάση παρατηρήσεις, παρελθούσες εμπειρίες και γνώμες άλλων χρηστών. Για τον υπολογισμό της φήμης κάθε μέλους, υιοθετήθηκαν διάφορες πτυχές της έννοιας της εμπιστοσύνης, όπως η μεταβατικότητα, η εξατομίκευση, το περιεχόμενο και διάφορες ιδέες από κοινωνικά αξιώματα. Γενικά, η εμπιστοσύνη δεν είναι ιδιαίτερα μεταβατική στα κοινωνικά δίκτυα, γι' αυτό και μειούται κατά μήκος του μονοπατιού μετάδοσής του, αλλά παρόλα αυτά θεωρείται πως μπορεί να αποτελέσει αντικείμενο επικοινωνίας μεταξύ ατόμων. Επιπλέον, η εμπιστοσύνη είναι εξατομικευμένη, δηλαδή είναι υποκειμενική και επηρεάζεται εύκολα από τις προσωπικές απόψεις και των υπολοίπων χρηστών, καθώς επίσης και εκείνων των μελών, τα οποία οι χρήστες εμπιστεύονται και σέβονται. Παρόλληλα, ενσωματώθηκε και η έννοια του χρόνου, λόγω του ότι η φήμη φθίνει με την πάροδο του χρόνου. Επομένως, η τιμή της φήμης ενός χρήστη τείνει στο μηδέν, εκτός αν νέα στοιχεία συμπάθειας ή αντιπάθειας προς το πρόσωπο αυτού του χρήστη βρίσκονται στο προσκήνιο.

Αναλυτικότερα, ο αλγόριθμος υλοποίησης αυτού του συστήματος αποτελείται από τρεις φάσεις, όπως αυτές φαίνονται παρακάτω:

1. Σχηματισμός των Συνδέσεων των Χρηστών (User Connection Formation): Ένας χρήστης μπορεί να εκφράσει άμεσα αν εμπιστεύεται ή όχι κάποιον άλλο, ή έμμεσα μέσω της άποψης του (ένα «like», ή ένα σχόλιο) σε μια δημοσιευμένη κατάσταση περιεχομένου. Οι σχέσεις εμπιστοσύνης μπορούν να κατηγοριοποιηθούν ως εξής: (α) άμεση σύνδεση χρήστη-προς-χρήστη, (β) άμεση σύνδεση χρήστη-προς-στοιχείο, (γ) έμμεση σύνδεση χρήστη-προς-στοιχείο και (δ) έμμεση σύνδεση χρήστη-προς-χρήστη. Στην (α) περίπτωση της άμεσης σύνδεσης των χρηστών, ένας χρήστης μπορεί να σχετίζεται άμεσα μ' έναν άλλο, σχηματίζοντας σχέσεις εμπιστοσύνης ή δυσπιστίας, όπως φιλίες, ή συνεργασίες στον πραγματικό κόσμο. Η μοντελοποίηση αυτών των δεδομένων έγινε υποθέτοντας πως κάθε μέλος έχει και ανανεώνει δύο λίστες, μια φιλική και μια εχθρική, για τους αξιόπιστους και

μη φίλους του, αντίστοιχα. Η λίστα των φίλων περιλαμβάνει μέλη τα οποία ο χρήστης ήδη εμπιστεύεται ή μπορεί να εμπιστευτεί και να αλληλεπιδράσει στο μέλλον. Η κύρια ιδέα πίσω από τον κατάλογο των εχθρών είναι ότι περιλαμβάνει μέλη που έχουν λάβει πολλές αρνητικές βαθμολογίες εμπιστοσύνης από τον χρήστη, τους συναδέλφους του ή άλλα μέλη του δικτύου (ανάλογα με το μοντέλο) και θεωρούνται αναξιόπιστοι για τον χρήστη. Η υπηρεσία κοινωνικής δικτύωσης μπορεί να χρησιμοποιήσει αυτόν τον κατάλογο για να ειδοποιήσει τον χρήστη όταν ένας εχθρός επιχειρεί να αλληλεπιδράσει μαζί του/ της. Στο (β) τύπο συνδέσεων, ο χρήστης παρέχει ένα σχόλιο επικρότησης ή αποδοκιμασίας σ' ένα συγκεκριμένο στοιχείο που έχει δημοσιευθεί από άλλον χρήστη, το οποίο, ανάλογα και με την εφαρμογή που χρησιμοποιείται σε κάθε περίπτωση, μπορεί να πάρει διάφορες μορφές. Ένας ελαφρώς διαφορετικός τύπος σύνδεσης μέσα σ' ένα κοινωνικό δίκτυο είναι η περίπτωση (γ), η οποία υλοποιείται μέσω στοιχείων περιεχομένου. Κάθε στοιχείο περιεχομένου που δημοσιεύεται από έναν χρήστη έχει ένα μοναδικό αναγνωριστικό και μια σφραγίδα χρόνου και μπορεί να περιέχει έναν ή περισσότερους υπερσυνδέσμους, που δείχνουν άλλα στοιχεία περιεχομένου μέσα στο κοινωνικό δίκτυο, ή στοιχεία Ενιαίου Εντοπισμού Πόρων (United Resource Location, URL) στον ιστό. Η προτίμηση σ' ένα στοιχείο εμφανίζεται έμμεσα, π.χ. μοιράζοντας ένα άρθρο, επαναλαμβάνοντας μια δημοσίευση, σχολιάζοντας θετικά ή αρνητικά την ανάρτηση ενός χρήστη κ.λπ. Τέλος, η έμμεση σύνδεση μεταξύ χρηστών (δ) μπορεί να εξαχθεί από τις άμεσες και έμμεσες συνδέσεις ενός χρήστη με τα στοιχεία άλλου χρήστη. Οι πληροφορίες χρήστη-προς-στοιχείο αντιστοιχίζονται στο επίπεδο χρήστη-προς-χρήστη και συγκεντρώνονται για να παρέχουν μια έμμεση σύνδεση χρήστη-προς-χρήστη.

2. Υπολογισμός της Κατάταξης Φήμης (Reputation Rating Estimation): Ο μηχανισμός φήμης ποσοτικοποιεί τις συνδέσεις εμπιστοσύνης που προσδιορίζονται στο κοινωνικό δίκτυο και παρέχει εξατομικευμένες αξιολογήσεις, οι οποίες εκφράζουν την τοπική πεποίθηση ενός χρήστη (χρήστης-αξιολογητής) σε σχέση με άλλα μέλη του δικτύου (χρήστες-στόχοι). Οι αξιολογήσεις φήμης δημιουργούνται συλλογικά, με την ενσωμάτωση της άποψης του αξιολογητή στο χρήστη-στόχο, καθώς και τη γνώμη κάποιων άλλων μελών του κοινωνικού δικτύου (μάρτυρες) σε σχέση με τον υπό αξιολόγηση χρήστη. Το δίκτυο παραπομπής των χρηστών, δηλαδή το σύνολο μαρτύρων, διαμορφώνεται μέσα σε συγκεκριμένους κύκλους εμπιστοσύνης και δυσπιστίας από τον αξιολογητή, βάσει εννοιών που αντλούνται από την κοινωνιολογία. Στην Εικόνα 4, για παράδειγμα, ο χρήστης u_j είναι ο αξιολογητής, ο u_i ο χρήστης-στόχος και ο u_q αποτελεί έναν μάρτυρα που μοιράζεται με τον u_j τις απόψεις του για τον u_i . Η αξιοπιστία ή αναξιοπιστία εκφράζονται με +1 και -1, αντίστοιχα, ή με πραγματικές τιμές του ίδιου εύρους, ενώ με 0 παρουσιάζεται η απουσία σύνδεσης μεταξύ δυο χρηστών.



Εικόνα 4. Μεταβατικότητα θετικών και αρνητικών εκφράσεων εμπιστοσύνης.

Ο προτεινόμενος μηχανισμός αξιολόγησης φήμης καταγράφει την επίδραση του χρόνου, διαμορφώνοντας το γεγονός ότι πιο πρόσφατα σχόλια θα πρέπει να έχουν μεγαλύτερο βάρος στη γενικότερη αξιολόγηση του χρήστη-στόχου από τον αξιολογητή. Η χρήση των πληροφοριών χρόνου επιτρέπει τη διάκριση μεταξύ των χρηστών που επιτυγχάνουν υψηλή φήμη για μια σύντομη χρονική περίοδο και των χρηστών που καταφέρνουν να διατηρήσουν τη φήμη τους σε υψηλό επίπεδο.

3. Παραγωγή Προτάσεων (Recommendations Generation): Βάσει των συνολικών εκτιμήσεων φήμης των μελών του κοινωνικού δικτύου, όπως εκτιμήθηκε από το χρήστη-αξιολογητή, το προτεινόμενο σύστημα δημιουργεί εξατομικευμένες θετικές ή/και αρνητικές συστάσεις χρηστών, οι οποίες μπορούν να χρησιμοποιηθούν για τη δημιουργία νέων συνδέσεων εμπιστοσύνης ή/και δυσπιστίας. Οι θετικές συστάσεις μπορούν να χρησιμοποιηθούν από τα μέλη, ώστε αυτά να μπορέσουν να συνδεθούν με νέα άτομα (σε ιστότοπους κοινωνικής δικτύωσης), να εγγραφούν σε νέα ιστολόγια, να μοιραστούν πόρους κ.λπ. Από την άλλη πλευρά, σε περίπτωση αρνητικών συστάσεων, το μοντέλο ουσιαστικά δημιουργεί έναν κατάλογο αναξιόπιστων χρηστών. Αυτή η εξατομικευμένη «μαύρη λίστα» μπορεί να αξιοποιηθεί από το σύστημα, προκειμένου αυτό να προειδοποιήσει τους χρήστες του όταν δημοσιεύονται στοιχεία περιεχομένου από τέτοιους αναξιόπιστους χρήστες και, επομένως, να τους αποθαρρύνει από τη σύνδεση ή την περιήγησή τους σε τέτοιο περιεχόμενο ή να τα φιλτράρει από τη ροή περιεχομένου τους.

Για την πειραματική αξιολόγηση του προτεινόμενου συστήματος, χρησιμοποιήθηκαν σύνολα δεδομένων που καλύπτουν τα περισσότερα από τα επιθυμητά χαρακτηριστικά ενός κοινωνικού δικτύου, όπως περιγράφηκε προηγουμένως. Μία ενότητα αφορά στην παρουσίαση αποτελεσμάτων στο εκτεταμένο σύνολο δεδομένων Epinions. Αυτό το σύνολο δεδομένων περιέχει τόσο άμεσες, όσο και έμμεσες δηλώσεις εμπιστοσύνης μεταξύ χρηστών. Στη δεύτερη ενότητα πραγματοποιήθηκε αξιολόγηση της ικανότητας του συστήματος να συστήνει αξιόπιστες συνδέσεις στα μέλη του δικτύου του, χρησιμοποιώντας αποκλειστικές συνδέσεις χρήστη-προς-χρήστη. Για το σκοπό αυτό, χρησιμοποιήθηκε το σύνολο δεδομένων Advogato, το οποίο περιέχει άμεσες δηλώσεις εμπιστοσύνης μεταξύ των χρηστών της κοινότητάς του. Τέλος,

στην τρίτη και τελευταία ενότητα αξιολογήθηκε η απόδοση του μοντέλου όσον αφορά στην πρόβλεψη θετικών ή αρνητικών ορίων σε δίκτυα εμπιστοσύνης με διαφορετικά χαρακτηριστικά, ενώ επίσης συγκρίθηκε και με αλγόριθμους τελευταίας τεχνολογίας (State-of-the-Art, SoA) στα διευρυμένα σύνολα δεδομένων Epinions και Wikipedia.

Αναλυτικότερα, στην ενότητα της αξιολόγησης με χρήση του συνόλου δεδομένων Epinions, και αναφορικά με τη λίστα που περιέχει τους φίλους του χρήστη του δικτύου, αποδεικνύεται η ικανότητα του μηχανισμού να βρίσκει χρήστες εμπιστοσύνης στην εκτεταμένη γειτονιά ενός χρήστη, και να εμπλουτίζει τον κύκλο φίλων του. Μέλη με πολλούς φίλους μπορούν να επωφεληθούν από το σύστημα, αφού μπορούν να διαγράψουν ήδη υπάρχοντες φίλους τους και να βρουν πρόσθετους φίλους που τους ενδιαφέρουν περισσότερο, σύμφωνα με τις προτάσεις του συστήματος. Αναφορικά με το κομμάτι της «μαύρης λίστας», αυτό με τη σειρά του δείχνει ότι το σύστημα συστήνει ως εχθρικούς χρήστες, μέλη με λίγες ομοιότητες με το χρήστη (σε βαθμολογίες άρθρων). Για χρήστες με μια μακρά λίστα εχθρών, το σύστημα μπορεί να παρέχει συστάσεις που θα μειώσουν περαιτέρω αυτόν τον κατάλογο. Αντίστοιχα, για την πρόβλεψη της φήμης των χρηστών, έγιναν πειράματα πάνω στο σύνολο δεδομένων Advogato, το οποίο μπορεί να κατηγοριοποιήσει τους χρήστες του σε τέσσερις κατηγορίες: (α) παρατηρητής, (β) μαθητευόμενος, (γ) ταξιδιώτης και (δ) αυθεντία. Τα αποτελέσματα της αξιολόγησης δείχνουν πως η μετρική που αφορά την εμπιστοσύνη, που δίνεται σ' ένα χρήστη u_i από έναν οποιοδήποτε χρήστη u_j και ισούται με το μέσο βαθμό εμπιστοσύνης που ανατίθεται στον u_i από τους χρήστες που τον εμπιστεύονται, έχει το χαμηλότερο βαθμό σφάλματος στην πρόβλεψη του βαθμού εμπιστοσύνης. Ωστόσο, ένα βασικό μειονέκτημά της είναι ότι πρόκειται για μια καθολική μετρική, δηλαδή απαιτεί γνώση από όλες τις μετρήσεις εμπιστοσύνης που έχουν ανατεθεί σ' ένα χρήστη, οπότε δεν είναι εύκολο να εισαχθεί σ' ένα καταναμημένο περιβάλλον ή σ' ένα μεγάλο δίκτυο. Γενικότερα, φαίνεται πως το Advogato είναι αργότερο του προτεινόμενου συστήματος και πως ο βαθμός σφάλματος στην πρόβλεψη του βαθμού εμπιστοσύνης είναι μεγαλύτερος. Σε περιπτώσεις όπου υπάρχουν πολλαπλά μονοπάτια μεταξύ δύο χρηστών, η επιλογή του μονοπατιού με τη μέγιστη τιμή δίνει και καλύτερα αποτελέσματα πρόβλεψης.

3.3. Συνδυαστικές Προσεγγίσεις των Δύο Κόσμων

Το συγκεκριμένο κομμάτι ερευνών μελετά τους δύο κόσμους, offline και online, μαζί, ως αλληλένδετους, παρά τα διαφορετικά χαρακτηριστικά τους. Αυτό το είδος ερευνών ονομάζεται Διακοινοτική Ανίχνευση και Εξόρυξη δεδομένων (Cross-community Sensing and Mining, CSM) και έχει ως στόχο την ενοποίηση ετερογενών κοινωνιών, αποκαλύπτοντας τη σύνθετη σύνδεση και αλληλεπίδραση μεταξύ των ιδιοτήτων τους (π.χ., χωρικών ή κοινωνικών) και προσδιορίζοντας τα ανθρώπινα πρότυπα συμπεριφοράς, αναλύοντας τα δεδομένα που συλλέχθηκαν από περιβάλλοντα πολλαπλών κοινοτήτων [19]. Το γεγονός αυτό οδηγεί, κατ' επέκταση, στη γεφύρωση του χάσματος ανάμεσα στις αλληλεπιδράσεις των χρηστών σ' ένα φυσικό ή σ' ένα εικονικό δίκτυο. Το πεδίο αυτό αποτελεί τη βασική μελέτη της παρούσας εργασίας, καθώς ο βασικός στόχος της υλοποίησης εμπίπτει σ' αυτόν τον τομέα.

Αναλυτικότερα, τα μοναδικά χαρακτηριστικά αυτής της ομάδας ερευνών μπορούν να ενσωματωθούν στις ακόλουθες κατηγορίες:

- **Δεδομένα:** Οι πηγές δεδομένων, γενικά, θεωρούνται ποικίλες και ετερογενείς, ανάλογα με την κάθε περίπτωση, και συλλέγονται είτε από κάποιο φυσικό, είτε από κάποιον εικονικό χώρο. Διαφορετικές ιδιότητες ατόμων και κοινωνιών μπορούν να εξαχθούν από ανεπεξέργαστα δεδομένα, τα οποία με τη σειρά τους συλλέγονται από τις offline ή online κοινωνίες, αντίστοιχα. Οι ιδιότητες διαφορετικών κοινωνιών μπορούν να είναι συσχετισμένες, ενώ επίσης μπορεί να πραγματοποιηθεί και ενσωμάτωσή τους.
- **Τεχνολογία:** Οι τεχνολογίες πυρήνα πολλών δικτύων περιέχουν ανιχνεύσεις κινητών τηλεφώνων, αναλύσεις κοινωνιών, εξαγωγή δεδομένων κ.λπ. Στόχος των ανιχνεύσεων αυτών και της επεξεργασίας δεδομένων κυμαίνεται μεταξύ της συλλογής ποικίλων δεδομένων από τις αλληλεπιδράσεις των χρηστών και της εξαγωγής φυσικών ή κοινωνικών χαρακτηριστικών για τη διασταύρωση ανάλυσης και συσχετισμού.
- **Εφαρμογές:** Σκοπός τους είναι να διευκολύνουν καινοτόμες υπηρεσίες σε κοινωνικό επίπεδο, όπως για παράδειγμα κοινωνικές προτάσεις, στοχευμένες διαφημίσεις, εντοπισμό κοινωνικών εκδηλώσεων.

Από αναλύσεις διαφόρων εικονικών κοινωνιών, έχει παρατηρηθεί σημαντική αύξηση των online κοινωνικών υπηρεσιών, όπως είναι τα μηνύματα, e-mails κ.λπ., τα οποία έχουν αλλάξει τον τρόπο με τον οποίο τα άτομα μοιράζονται και λαμβάνουν πληροφορίες και επικοινωνούν μεταξύ τους. Με την εκμετάλλευση αυτών των προαναφερθεισών υπηρεσιών, έχει αναπτυχθεί ένα μεγάλο κομμάτι ερευνών που σχετίζεται με την ανάλυση της ανθρώπινης αλληλεπίδρασης. Πιο συγκεκριμένα, και ταυτόχρονα με την ανάπτυξη του διαδικτύου, οι ερευνητές έχουν στρέψει την προσοχή τους στις online κοινωνίες, όπως είναι οι σελίδες online κοινωνικής δικτύωσης, μπλογκς κ.ά.

Αντίστοιχα, στο κομμάτι των κοινωνικών αλληλεπιδράσεων εντός μιας φυσικής κοινωνίας, η ευρεία χρήση κινητών τηλεφώνων με ενσωματωμένους αισθητήρες έχει προσφέρει σημαντικές ευκαιρίες που βοηθούν στον εντοπισμό και την απόκτηση επίγνωσης πάνω στις κοινωνικές συμπεριφορές μιας τέτοιας κοινωνίας [20]. Ένας από τους σημαντικότερους τομείς έρευνας του φυσικού κόσμου είναι αυτός της ευκαιριακής κοινωνικής δικτύωσης (Opportunistic Social Networking, OPSN) [21]. Άτομα που βρίσκονται στον ίδιο χώρο συχνά στερούνται της ευκαιρίας να εκμεταλλευτούν κοινωνικές σχέσεις για άμεση αλληλεπίδραση. Η ευκαιριακή δικτύωση, ωστόσο, διευκολύνει τις προτάσεις φίλων μεταξύ ατόμων που χρησιμοποιούν Ad-Hoc τεχνικές δικτύωσης, όπως είναι τα Bluetooth, Wi-Fi κ.λπ. Το Find & Connect [22] είναι ένα τέτοιο σύστημα που υποστηρίζει εφήμερες αλληλεπιδράσεις μεταξύ χρηστών βάσει των προφίλ των χρηστών και της φυσικής τους εγγύτητας. Για παράδειγμα, πολλά άτομα θέλουν να ενημερώνονται για κοντινά γεγονότα που τους αφορούν (π.χ. πληροφορίες κυκλοφοριακής κίνησης) ή χρειάζεται να μεταφέρουν πληροφορίες σχετικές με την τοποθεσία τους σε κοντινά τους γεωγραφικά άτομα (π.χ. η πώληση αχρησιμοποίητου εισιτηρίου κοντά σ' ένα σταθμό τρένων). Τέτοιες πληροφορίες είναι καλύτερο να διαδίδονται μεταξύ χρηστών που ανήκουν σε κοντινό εύρος, χωρίς την εκμετάλλευση του διαδικτύου.

Παράλληλα, και με την ανάπτυξη τεχνικών ανίχνευσης των κινητών τηλεφώνων, η κινητή κοινωνική δικτύωση (Mobile Social Networking, MSN), η οποία είναι ένα χαρακτηριστικό παράδειγμα της συγχώνευσης της κοινωνικής δικτύωσης με διάφορα ανιχνευμένα φυσικά χαρακτηριστικά, έχει αναπτυχθεί αρκετά ως ένα νέο είδος υπηρεσίας κοινωνικών αλληλεπιδράσεων. Σε αντίθεση με την ευκαιριακή κοινωνική δικτύωση, η οποία βασίζεται σε δίκτυα χωρίς υποδομές, οι υπηρεσίες κινητής δικτύωσης βασίζονται συνήθως σε χρήστες συστημάτων υποστήριξης, καθώς και σε συνδέσεις διαδικτύου. Επιπρόσθετα, με την εισαγωγή φυσικού περιεχομένου στα παραδοσιακά μέσα κοινωνικής δικτύωσης, η κινητή δικτύωση ενώνει offline συμπεριφορές με online αλληλεπιδράσεις. Στα σημερινά δεδομένα, η πιο πρόσφατη μορφή μιας τέτοιας υπηρεσίας είναι τα LBSN, τα οποία περιλαμβάνουν τόσο online κοινωνικές αλληλεπιδράσεις, όσο και offline πληροφορίες κοινοποιήσεων τοποθεσίας, όπως για παράδειγμα τοποθεσία που μπορεί να αποκτηθεί από τους ενσωματωμένους αισθητήρες ενός GPS. Από τη στιγμή που η τοποθεσία παίζει το σημαντικότερο ρόλο στην καθημερινότητα των χρηστών, η προσθήκη της τοποθεσίας στα παραδοσιακά μέσα κοινωνικής δικτύωσης βοηθάει στη γεφύρωση του φυσικού με τον εικονικό κόσμο. Ταυτόχρονα, εφόσον η τοποθεσία είναι ένα από τα βασικότερα χαρακτηριστικά του προφίλ ενός χρήστη, η εκτεταμένη γνώση των ενδιαφερόντων ενός χρήστη, της συμπεριφοράς του, ακόμα και των σχέσεών του με άλλα άτομα μπορεί να προκύψει από την τοποθεσία του. Χαρακτηριστικά παραδείγματα των LBSN αποτελούν το Foursquare, Facebook κ.ά. Το Twitter επίσης επιτρέπει στους χρήστες του να προσθέτουν την τοποθεσία τους στα tweets τους. Η ανάπτυξη των LBSN ενεργοποιεί και πολλές άλλες εφαρμογές που αλλάζουν τον τρόπο ζωής μας, όπως είναι οι προτάσεις που αφορούν νέους φίλους και τοποθεσίες. Για παράδειγμα, γειτονικές κοινοποιήσεις τοποθεσίας έχουν γίνει αντικείμενο εκμετάλλευσης για την επισήμανση πιθανών κοινωνικών συνδέσεων [23].

Τέλος, πέρα από τα LBSN, υπάρχουν και αρκετές άλλες μορφές κινητής δικτύωσης που προσπαθούν να συνδέσουν το φυσικό με τον εικονικό κόσμο. Μια από αυτές είναι και το EBSN, το οποίο με τη σειρά του προσπαθεί να δημιουργήσει ένα νέο σύνδεσμο μεταξύ των φυσικών γεγονότων ή των offline κοινωνικών αλληλεπιδράσεων και των online αλληλεπιδράσεων [24]. Από την άλλη μεριά, πέρα από τα EBSN, τα ίδια τα άτομα μπορεί να προτείνουν κοινωνικές εκδηλώσεις (π.χ. εργαστήρια, έξοδοι για φαγητό κ.λπ.) και να τα μοιραστούν σε online εγκαταστάσεις, οι οποίες μπορούν να προωθήσουν κοινωνικές αλληλεπιδράσεις πρόσωπο-με-πρόσωπο. Και η συμμετοχή, όμως, των χρηστών στο ίδιο γεγονός μπορεί να αποτυπωθεί. Το βασισμένο στο αντικείμενο κοινωνικό δίκτυο (Object-based Social Network, OBSN) είναι άλλη μια μορφή κινητής δικτύωσης, η οποία προσφέρει μια νέα δυνατότητα ενσωμάτωσης φυσικών χαρακτηριστικών στον τομέα της κοινωνικής δικτύωσης. Τα κοινωνικά δίκτυα αντικειμένου προσπαθούν να συνδέσουν άτομα και να δυναμώσουν τους δεσμούς τους μέσω της κοινής τους εμπειρίας με φυσικά αντικείμενα, υπό την υποστήριξη των τεχνικών του Internet of Things [25].

Όπως είδαμε, λοιπόν, αυτές οι συνδυαστικές μελέτες λαμβάνουν ως βασικά χαρακτηριστικά τους τόσο τον ανθρώπινο παράγοντα, όσο και την ανάλυση της κοινωνικής αλληλεπίδρασης. Ωστόσο, υπερβαίνουν όλους αυτούς τους τομείς όσον αφορά στην εστίαση και τις ερευνητικές προκλήσεις. Σε αντίθεση με τους τομείς που επικεντρώνονται σε μια ενιαία κοινότητα, φυσικού ή εικονικού δικτύου, οι συνδυαστικές μελέτες επικεντρώνονται ιδιαίτερα στην αλληλεπίδραση και τη σχέση μεταξύ διαφορετικών μορφών κοινοτήτων, προσπαθώντας

να γεφυρώσουν το χάσμα ανάμεσα στις ηλεκτρονικές συμπεριφορές και τις κοινωνικές αλληλεπιδράσεις εκτός σύνδεσης. Είναι φανερό, επομένως, πως κοινωνίες τέτοιου είδους διαθέτουν και φυσικά, αλλά και online χαρακτηριστικά.

Μερικές προσεγγίσεις και παραδείγματα που ασχολήθηκαν με το συγκεκριμένο τομέα ερευνών παρουσιάζονται, συνοπτικά, ακολούθως.

3.3.1. Offline D2D Δρομολόγηση

Ο όρος Συσκευή-προς-Συσκευή (Device-to-Device, D2D) αναφέρεται στην απευθείας μετάδοση περιεχομένου μεταξύ συσκευών, χωρίς την αποστολή όλων των δεδομένων του σταθμού, όπως συμβαίνει στα παραδοσιακά κινητά δίκτυα [26].

Ο συνδυασμός, λοιπόν, των χωρικών χαρακτηριστικών και της κοινωνικής ευαισθητοποίησης, τα οποία καταγράφονται στο ίχνος του NCCU [27], στοχεύουν στη βελτίωση της οικονομικής αποδοτικότητας της ευκαιριακής προώθησης. Περιγράφονται δύο χωρικά και δύο κοινωνικά χαρακτηριστικά και χαρακτηρίζονται στο ίχνος NCCU. Παράλληλα, χρησιμοποιούνται τέτοιες ιδιότητες για το σχεδιασμό του SAMPLER (Social-Aware, Mobility, and PoI Routing), που αποτελεί την πρώτη στρατηγική ευκαιριακής δρομολόγησης για το συνδυασμό κινητικότητας, σημείων ενδιαφέροντος και κοινωνικής ευαισθητοποίησης, ώστε να μπορεί να παρέχεται οικονομικά αποδοτική διανομή περιεχομένου σε διαλείποντα δίκτυα. Η αξιολόγηση του SAMPLER έγινε με τη χρήση δύο ιχνών: του NCCU και του μοντέλου συνθετικής κινητικότητας SWIM [28]. Το τελευταίο επιτρέπει την αξιολόγηση της απόδοσης ενός μεγάλου δικτύου.

Το SAMPLER λειτουργεί με την προώθηση μηνυμάτων σε κόμβους υψηλότερης κινητικότητας, μέχρι το μήνυμα να φτάσει σ' ένα στατικό σημείο αναμετάδοσης. Στατικά σημεία αναμετάδοσης αναπτύσσονται στρατηγικά στους πιο δημοφιλείς προορισμούς και διαβιβάζουν το περιεχόμενο που λήφθηκε σε κόμβους που ανήκουν στην κοινότητα της οποίας ο κόμβος προορισμού είναι επίσης μέλος. Σε αυτήν την κοινότητα, το μήνυμα προωθείται στους πιο δημοφιλείς κόμβους, μέχρι να φτάσει στον κόμβο προορισμού. Τα πειράματα της μελέτης έδειξαν ότι, με τη διερεύνηση του συνδυασμού χωρικών και κοινωνικών χαρακτηριστικών, το SAMPLER βελτιώνει τον λόγο παράδοσης, μειώνει το γενικό κόστος του δικτύου και επιτρέπει την ταχύτερη παράδοση μηνυμάτων. Αυτά τα αποτελέσματα ενισχύουν την παραδοχή ότι η καλύτερη κατανόηση των πραγματικών ιχνών κινητικότητας μπορεί να προσφέρει πολύτιμες ιδέες στο σχεδιασμό της ευκαιριακής δρομολόγησης D2D.

Αναλυτικότερα, η συγκεκριμένη μελέτη συνδυάζει δεδομένα από τις εξής πηγές: από το ίχνος NCCU, τη δημοτικότητα του κόμβου, τις κοινωνίες και, τέλος, την ίδια την κινητικότητα του χρήστη. Το προτεινόμενο SAMPLER έχει την ικανότητα να χρησιμοποιεί δεδομένα κοινοτήτων και μετρήσεις δημοτικότητας, όπως αυτές εισήχθησαν στο αρχικό σύστημα Bubble Rap [29], και παράλληλα να προσθέτει σ' αυτά τις ιδιότητες μεμονωμένης κινητικότητας και των σημείων ενδιαφέροντος με έξυπνο τρόπο, συμβάλλοντας στην επίτευξη υψηλού ποσοστού παράδοσης και στη μείωση των γενικών εξόδων του δικτύου.

Με λίγα λόγια, το SAMPLER λειτουργεί ως εξής:

- Το GlobalRank ορίζεται ίσο με την ακτίνα περιστροφής των κόμβων (ευαισθησία κινητικότητας).
- Το LocalRank ορίζεται όσο η δημοτικότητα των κόμβων (C-Window metric) στην κοινότητα προορισμού (κοινωνική ευαισθητοποίηση).
- Το μήνυμα προωθείται πάντα, αν ο κόμβος που συνάντησε είναι ο προορισμός.
- Κάθε φορά που ένας κινητός κόμβος συναντά ένα στατικό σημείο αναμετάδοσης, που δεν έχει ακόμα το μήνυμα, τότε το μήνυμα αυτό προωθείται (ευαισθητοποίηση του σημείου ενδιαφέροντος).
- Σε περίπτωση συνάντησης ενός κόμβου που ανήκει στην κοινότητα προορισμού, το σημείο αναμετάδοσης προωθεί το μήνυμα (κοινωνική ευαισθητοποίηση).
- Εκτός της κοινότητας προορισμού, το μήνυμα προωθείται εάν το GlobalRank του κόμβου που συναντάται είναι υψηλότερο από το GlobalRank του κόμβου που έχει το μήνυμα (ευαισθητοποίηση σχετικά με την κινητικότητα). Αυτό το βήμα εκτελείται το πολύ μία φορά από κάθε κόμβο.
- Μέσα στην κοινότητα προορισμού, το μήνυμα προωθείται εάν το LocalRank του κόμβου που συναντάται είναι υψηλότερο από το LocalRank του κόμβου που έχει το μήνυμα (κοινωνική ευαισθητοποίηση).

Αναφορικά με το πειραματικό κομμάτι, οι αλγόριθμοι εξετάστηκαν για διαφορετικό αριθμό σημείων αναμετάδοσης. Για κάθε περίπτωση, λήφθηκε υπόψη κάθε πιθανός συνδυασμός για κάθε αλγόριθμο, καθώς και για κάθε αριθμό σημείων αναμετάδοσης.

Τέλος, τα αποτελέσματα της πειραματικής μελέτης έδειξαν πως προκύπτουν αρκετοί περιορισμοί, λαμβάνοντας υπόψη το συνδυασμό φυσικών και κοινωνικών τομέων. Η συλλογή και η δημοσίευση μεγάλων ιχνών κινητικότητας θα επέτρεπε την περαιτέρω διερεύνηση των συνδυασμένων στρατηγικών κοινωνικού και γεωγραφικού χαρακτήρα, καθώς επίσης και την καλύτερη επικύρωση των στρατηγικών αυτών, υπό διαφορετικά όμως σενάρια.

3.3.2. Μη-γραμμική Πρόβλεψη Χρονοσειρών Πολλών Μεταβλητών

Ο τομέας της έρευνας που επικεντρώνεται στην αλληλεξάρτηση και την προβλεψιμότητα τόσο της ανθρώπινης κινητικότητας, όσο και των κοινωνικών αλληλεπιδράσεων [30], έχει ως στόχο να δείξει πως είναι δυνατή η βελτίωση της πρόβλεψης της κινητικότητας με την εκμετάλλευση της συσχέτισης των κινήσεων μεταξύ των χρηστών ενός δικτύου. Αυτό με τη σειρά του θα οδηγήσει στην πρόβλεψη και τη συμπερίληψη πτυχών που σχετίζονται με τη συμπεριφορά των χρηστών, δηλαδή με τις κινήσεις και τις κοινωνικές αλληλεπιδράσεις τους (φυσικές και απομακρυσμένες). Συγκεκριμένα, στην προτεινόμενη ανάλυση έγινε εκμετάλλευση και προσαρμογή της έννοιας της αμοιβαίας πληροφορίας, προκειμένου να ποσοτικοποιηθεί η συσχέτιση και να παραχθεί μια πρακτική μέθοδος για την επιλογή πρόσθετων δεδομένων, για τη βελτίωση της ακρίβειας των προβλέψεων κίνησης. Παρά το γεγονός ότι σ' αυτό το σύνολο δεδομένων δεν είναι δυνατό να χρησιμοποιηθούν κοινωνικοί δεσμοί,

διαπιστώθηκε πως σε περίπτωση επιλογής μοτίβων κινητικότητας των ταξί των οποίων η αμοιβαία ενημέρωση είναι υψηλή, η απόδοση του προγνωστικού βελτιώνεται δραστικά.

Αναλυτικότερα, το κομμάτι που αφορά το φυσικό κόσμο περιλαμβάνει ένα σύνολο δεδομένων της εφαρμογής Cabspotting με ίχνη κίνησης από 500 ταξί της ίδιας περιοχής. Η πρώτη φάση της μελέτης αυτής αποτελείται από την εφαρμογή μεθόδων πρόβλεψης με μη-γραμμικές χρονοσειρές, οποίες φαίνεται να έχουν ιδιαίτερα υψηλά ποσοστά θορύβου, γι' αυτό και η μοντελοποίηση τέτοιων κινήσεων στο χώρο έγινε με μια απλή διαδικασία πολλών παραγόντων $AR(p) + noise$. Όσο για τη σειρά p του πολυπαραγοντικού αυτορρυθμιστικού μοντέλου, χρησιμοποιήθηκε αυτή που ελαχιστοποιούσε ένα κριτήριο πληροφορίας. Ωστόσο, παρόλο που οι χρονοσειρές δε λαμβάνονται συχνά ως δείγματα, διαπιστώθηκε πως παρουσιάζουν ορισμένα χαρακτηριστικά της αιτιοκρατικής δυναμικής, επιβαρυνμένα με θόρυβο. Στην πραγματικότητα, ο προκαταρκτικός έλεγχος της ανακατασκευής του χώρου φάσης, μέσω του θεωρήματος ενσωμάτωσης του Takens, δείχνει μια υποκείμενη δομή, χαρακτηριστική των αιτιοκρατικών δυναμικών συστημάτων.

Όσον αφορά την εκτίμηση της διάστασης ενσωμάτωσης και της καθυστέρησης χρόνου, θεωρήθηκε η ίδια χρονική καθυστέρηση τ για όλες τις σειρές. Στην πραγματικότητα, για ένα συγκεκριμένο χρήστη, διαπιστώθηκε ότι η χρονική καθυστέρηση τ_{min} που αντιστοιχεί στο πρώτο τοπικό ελάχιστο της μέσης αμοιβαίας πληροφορίας, που γενικά υιοθετήθηκε για την εκτίμηση του τ σε μια περίπτωση μιας μεταβλητής, έχει την ίδια τάξη μεγέθους για οποιαδήποτε συνιστώσα. Το γεγονός αυτό έχει και πρακτικές συνέπειες, καθώς απλοποιεί την εφαρμογή αυτής της μεθοδολογίας για την ανάλυση των δεδομένων. Η βέλτιστη διάσταση ενσωμάτωσης υπολογίζεται με την εκμετάλλευση της μεθόδου των πλησιέστερων γειτόνων, στην περίπτωση της ενσωμάτωσης πολλών μεταβλητών. Για οποιοδήποτε σημείο των δεδομένων εξετάζεται ένας χώρος m *-διαστάσεων φάσης και υπολογίζεται ο αριθμός των ψευδών πλησιέστερων γειτόνων, π.χ. σημεία που γειτνιάζουν στο χώρο των m *-διαστάσεων, αλλά όχι στον $(m + 1)$. Η επιθυμητή διάσταση ενσωμάτωσης m είναι τέτοια, ώστε το ποσοστό ψευδών πλησιέστερων γειτόνων να είναι μικρό ($< 5\%$). Για τον υπολογισμό των πλησιέστερων γειτόνων υιοθετήθηκε η μέθοδος που εφαρμόζεται στο λογισμικό TISEAN.

Στο κομμάτι της ανάλυσης υποκατάστατων πολλών μεταβλητών, παράγεται ένα σύνολο $\{\widehat{s}_n^{(i)}\}$, $i = 1, 2, \dots, N$, δεδομένα χρονοσειρών πολλών μεταβλητών $\{s_n\}$ και N υποκατάστατα πολλών μεταβλητών του $\{s_n\}$. Τα υποκατάστατα είναι συνθετικές χρονοσειρές, κατασκευασμένες από το $\{s_n\}$, οι οποίες διατηρούν και γραμμικά και στατιστικά χαρακτηριστικά των αρχικών χρονοσειρών ως κατανομή πιθανοτήτων και αυτοσυσχέτιση, ενώ αφαιρούν τα αποτελέσματα της μη-γραμμικότητας και μη-στασιμότητας, αν υπάρχουν. Πιο συγκεκριμένα, για την παραγωγή των υποκατάστατων, χρησιμοποιήθηκε το σχήμα Επαναληπτικού Μετασχηματισμού Fourier Ρυθμιζόμενου Πλάτους (Iterative Amplitude-Adjusting Fourier Transform, IAAFT). Επομένως, επιλέχθηκε το κλάσμα των ψευδών κοντινότερων γειτόνων ως διάκριση για την εξέταση της μηδενικής υπόθεσης ότι τα ίχνη κινητικότητας μπορούν να περιγραφούν από ένα γραμμικό σταχαστικό μοντέλο. Ως $q(m)$ ορίζεται η τιμή της διάκρισης, όπως απαιτήθηκε για μια διάσταση ενσωμάτωσης m από τις παρατηρούμενες χρονοσειρές πολλών μεταβλητών, και $\widehat{q}^{(i)}(m)$ οι τιμές της διάκρισης όταν

αυτή ορίζεται από τα υποκατάστατα. Τα αριθμητικά πειράματα της μελέτης έδειξαν ότι η κατανομή του $\widehat{q}^{(i)}(m)$ περιγράφεται με μια λογική προσέγγιση από μια γκαουσιανή (Gaussian) συνάρτηση με μέσο όρο $\mu_{\widehat{q}}(m)$ και διακύμανση $\sigma_{\widehat{q}}^2(m)$. Έτσι, λοιπόν, ορίστηκε το μέτρο της σημαντικότητας ως $\Sigma(m) = \frac{|q(m) - \mu_{\widehat{q}}(m)|}{\sigma_{\widehat{q}}(m)}$. Στην περίπτωση αυτή, όταν η αρχική (null) υπόθεση είναι αληθής, τότε η τιμή p παρατήρησης ίσης ή μεγαλύτερης σημαντικότητας από $\Sigma(m)$ δίνεται από τον τύπο: $p(m) = \text{erfc} \left[\frac{\Sigma(m)}{\sqrt{2}} \right]$.

Δεδομένης αυτής της διαδικασίας, τέθηκε $\alpha = 0.05$ και παρατηρήθηκε πως αν $p(m) < \alpha$, τότε η αρχική υπόθεση από την οποία η κινητικότητα μπορεί να εκφραστεί από ένα γραμμικό στοχαστικό μοντέλο, απορρίπτεται με 95% επίπεδο εμπιστοσύνης (Confidence Level, CL).

Από το Cabspotting, επομένως, προέκυψε ως συμπέρασμα πως η ανθρώπινη κινητικότητα παρουσιάζει μια ισχυρή μη-γραμμική δυναμική. Επιπλέον, η ύπαρξη των βραχυπρόθεσμων συσχετίσεων, όπως φαίνεται από τη μέση ανάλυση αμοιβαίων πληροφοριών και τη μείωση του κλάσματος των ψευδών πλησιέστερων γειτόνων, δείχνει ότι μια τέτοια δυναμική θα πρέπει να έχει ένα αιτιοκρατικό συστατικό πιθανά επιβαρυνόμενο από μια στοχαστική δυναμική.

Αντιμετωπίζοντας μη-γραμμικά δυναμικά συστήματα με πιθανή αιτιοκρατική συνιστώσα, υιοθετήθηκε μια μέθοδος που εκμεταλλεύεται τέτοια χαρακτηριστικά για την πρόβλεψη των μελλοντικών μετακινήσεων ταξί στο Cabspotting. Η πολυδιάστατη μη-γραμμική πρόβλεψη (multivariate nonlinear prediction, MNP) εκτελείται προσεγγίζοντας τη δυναμική τοπικά στο χώρο φάσης, με μια σταθερά. Στο χώρο καθυστέρησης ενσωμάτωσης λαμβάνονται υπόψη όλα τα σημεία της γειτονιάς U_n του κράτους v_n , προκειμένου να προβλεφθούν οι συντεταγμένες στην ώρα $n+k$. Επομένως, η πρόβλεψη \widehat{v}_{n+k} για v_{n+k} δίνεται από τον τύπο: $\widehat{v}_{n+k} = \frac{1}{|U_n|} \sum_{v_j \in U_n} v_j + k$, δηλαδή το μέσο όρο των κρατών που αλληλεπιδρούν με μετρήσεις k βημάτων μπροστά από τους γείτονες v_j . Το καθολικό σφάλμα στην πρόβλεψη χρονοσειρών υπολογίζεται ξεχωριστά για κάθε συστατικό με: $e_j = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{s}_{j,n} - s_{j,n})^2}$. Το συνολικό σφάλμα μεταξύ της προβλεπόμενης θέσης και της πραγματικής δίνεται από τη γεωδесική απόσταση.

Απ' την άλλη μεριά, η δεύτερη πτυχή της μελέτης, επικεντρώνεται στο πώς ίχνη κινητικότητας ατόμων, με σχετιζόμενα γεωγραφικά μοτίβα και κοινωνικούς δεσμούς, μπορούν να χρησιμοποιηθούν για να βελτιώσουν την ακρίβεια της πρόβλεψης κίνησης. Η προσέγγιση, λοιπόν, αυτής της μελέτης συνοψίζεται ως εξής: θεωρώντας πως πρέπει να προβλεφθεί η κίνηση ενός ατόμου (οντότητα A), αντί να υπάρχει μόνο ένα διάνυσμα που θα περιγράφει την τοποθεσία του A ως είσοδο, να λαμβάνεται επίσης υπόψη και το ιστορικό κίνησης ενός άλλου ατόμου (οντότητα B), το οποίο χαρακτηρίζεται από τα μοτίβα κινητικότητας που σχετίζονται ιδιαίτερα μ' αυτά που ο χρήστης επιθυμεί να προβλέψει.

Ωστόσο, υπάρχουν αρκετά θέματα όσον αφορά στο κομμάτι της εφαρμογής και ανάπτυξης της πρόβλεψης. Αναλυτικότερα, ο υπολογισμός κοινών πληροφοριών σ' ένα

πολυδιάστατο περιβάλλον (χρήστες > 2) είναι ακριβός και δεν κλιμακώνεται αποδοτικά. Εξετάζοντας αυτή τη φορά την περίπτωση του Nokia Mobility Data Challenge (NMDC) με 39 χρήστες και πληροφορίες σχετικές με ίχνη GPS, που χρησιμοποιούνται για την ανάλυση της κίνησης των χρηστών, τηλεφωνικών αριθμών, ιστορικό κλήσεων και γραπτών μηνυμάτων (sms), ιστορικό Bluetooth και WLAN, η υπολογιστική πολυπλοκότητα είναι $O(N^n)$, όπου N είναι το υποσύνολο των χρηστών και n η απόλυτη τιμή του tuple που λαμβάνεται υπόψη κάθε φορά. Παρόλα αυτά, το πρόβλημα που υπάρχει, συνήθως, δεν περιλαμβάνει παραπάνω από 100 ίχνη κίνησης, π.χ. μέγεθος του κύκλου των πιο σημαντικών φίλων ενός ατόμου. Γι' αυτό το λόγο, μπορούν να αξιολογηθούν τιμές αμοιβαίων πληροφοριών για οποιοδήποτε ζεύγος ίχνους, το οποίο είναι $O(N^2)$. Εντούτοις, η αναπαράσταση πολλαπλών μεταβλητών ενσωματώσεων δεν είναι εφικτή για ένα φάσμα μεγαλύτερο από 40 διαστάσεις. Ακόμη και για ένα σήμα δύο συντεταγμένων που αντιπροσωπεύει ένα ίχνος κινητικότητας, δεν είναι ασυνήθιστο να υπάρχει μεγάλη ανακατασκευή ενσωμάτωσης λόγω θορυβωδών δεδομένων. Ως εκ τούτου, δεν πρέπει να λαμβάνονται υπόψη ταυτόχρονα περισσότερα από τρία ίχνη κινητικότητας. Επιπλέον, ο αλγόριθμος για την αναζήτηση του καταλληλότερου επιπρόσθετου συνόλου σημείων κινητικότητας για τη βελτίωση της πρόβλεψης κλιμακώνεται ως $O(N^2)$, όπου N είναι ο αριθμός των χρηστών που λαμβάνονται υπόψη στην εφαρμογή.

Συνοψίζοντας, η πρώτη φάση της συγκεκριμένης μελέτης αφορά στην εφαρμογή μεθόδων πρόβλεψης με μη-γραμμικές χρονοσειρές για την πρόβλεψη των μελλοντικών GPS συντεταγμένων των χρηστών, δεδομένης της παρελθούσας τους κίνησης ως είσοδο. Επιπλέον, λαμβάνεται υπόψη ως είσοδος στο πρόβλημα και η κίνηση άλλων χρηστών, και πιο συγκεκριμένα φίλων τους. Η μοντελοποίηση της θέσης ενός χρήστη επιτυγχάνεται μ' ένα χρονικά μεταβαλλόμενο διάνυσμα τεσσάρων διαστάσεων s_n με τις ακόλουθες διαστάσεις: ώρα της ημέρας h_n , γεωγραφικό πλάτος φ_n , γεωγραφικό μήκος λ_n και ύψος ξ_n . Η πρόβλεψη των μελλοντικών καταστάσεων του διανύσματος s_n επιτυγχάνεται με την ανακατασκευή του χώρου φάσης του s_n μέσω του θεωρήματος ενσωματωμένης καθυστέρησης. Ουσιαστικά, λοιπόν, αυτή η μέθοδος αναζητά το ιστορικό παρελθούσας κίνησης του χρήστη για να βρει και να εξάγει ακολουθίες αξιών που είναι πολύ παρόμοιες με το πρόσφατο ιστορικό. Υποθέτοντας έναν ορισμένο βαθμό αιτιοκρατίας στο σύστημα, η υπόθεση είναι ότι, δεδομένης μιας συγκεκριμένης κατάστασης (γεωγραφικές συντεταγμένες), υπάρχει μεγάλη πιθανότητα ότι αυτό θα ακολουθηθεί από την ίδια επόμενη κατάσταση. Στη συνέχεια, εφαρμόζονται μοντέλα γραμμικής πρόβλεψης σ' αυτές τις χρονοσειρές, καθώς επίσης και ανάλυση και πρόβλεψη σφαλμάτων, ακολουθώντας ακριβώς την ίδια διαδικασία, όπως αυτή παρουσιάστηκε στο κομμάτι ανάλυσης του συνόλου δεδομένων Cabspotting [30].

Αξίζει να σημειωθεί ότι πολλοί παράγοντες θα μπορούσαν να θεωρηθούν ως σήματα κοινωνικών δεσμών, σύμφωνα με το πλαίσιο των σεναρίων ανάπτυξης. Κατά συνέπεια, η ποιότητα των προβλέψεων μπορεί να επηρεάζεται σημαντικά, είτε θετικά είτε αρνητικά, από τα κριτήρια που χρησιμοποιούνται για την ανίχνευση των κοινωνικών δεσμών. Στο σύνολο δεδομένων NMDC δεν υπάρχουν πληροφορίες σχετικά με τους κοινωνικούς δεσμούς μεταξύ ατόμων, ούτε με πραγματικό ούτε με εικονικό χαρακτήρα. Στο διαθέσιμο σύνολο δεδομένων, η παρουσία ενός ατόμου στο βιβλίο διευθύνσεων ενός άλλου αντιπροσωπεύει στην πραγματικότητα την ισχυρότερη έννοια της κοινωνικής ισοπαλίας. Επιπλέον, δύο άτομα χωρίς κοινωνικούς δεσμούς ενδέχεται να παρουσιάζουν παρόμοια πρότυπα κινητικότητας, με

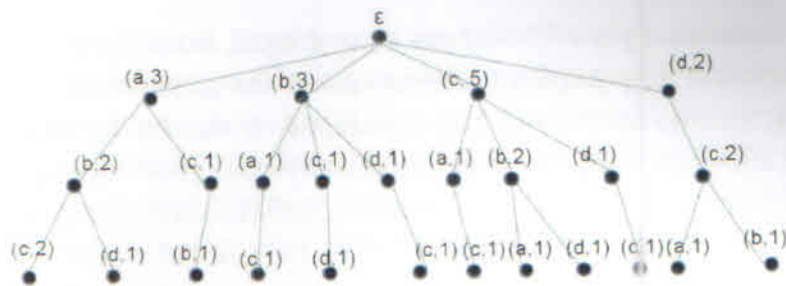
αποτέλεσμα την υψηλή αξία της αμοιβαίας πληροφόρησης. Είναι πιθανό τα άτομα με ισχυρούς κοινωνικούς δεσμούς, φοιτητές, φίλοι, συνεργάτες κλπ., να συμπεριφέρονται παρόμοια και τα ίχνη κινητικότητάς τους να χαρακτηρίζονται από πρότυπα με υψηλή αξία αμοιβαίας πληροφόρησης. Ως εκ τούτου, η ακρίβεια του προγνωστικού αλγορίθμου θα βελτιωθεί στην περίπτωση που η δυναμική των ιχνών συσχετίζεται σε μεγάλο βαθμό, ακόμη και αν δεν υπάρχει κοινωνική ισοδυναμία.

Παρατηρήθηκε, λοιπόν, με τη χρήση του συνδυαστικού προτεινόμενου αλγορίθμου, βελτίωση τουλάχιστον μιας τάξης μεγέθους σε σχέση με την πρόβλεψη που βασίζεται μόνο σε μεμονωμένα ίχνη. Επιπλέον, φαίνεται πως οι κοινωνικοί δεσμοί μπορούν επίσης να χρησιμοποιηθούν για την επιλογή του χρήστη για τα πρόσθετα ίχνη ως είσοδο. Στην πραγματικότητα, διαπιστώθηκε πως η απόδοση της πρόβλεψης βελτιώνεται δραματικά με την επιλογή μοτίβων κινητικότητας ατόμων που βρίσκονται στο βιβλίο διευθύνσεων του χρήστη.

3.3.3. Σειρά Μεταβλητών Μαρκοβιανού Μοντέλου (VOMM)

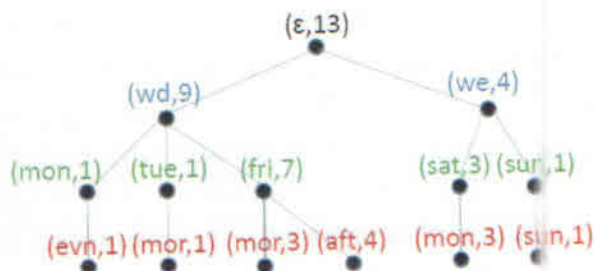
Αντίστοιχα, η πρόταση των χωροχρονικών ερευνών [31] προτείνει μια μέθοδο πρόβλεψης της κίνησης των χρηστών του δικτύου, με μεθόδους όπως αυτές προαναφέρθηκαν, σε συνδυασμό όμως και με την ενσωμάτωση μιας προσαρμοσμένης *σειράς μεταβλητών του Μαρκοβιανού μοντέλου (Variable Order Markov Model, VOMM)*, η οποία είναι μια παραλλαγή του Μπαεσιανού δικτύου (context specific Bayesian network). Γενικά, σε δυναμικά μπαεσιανά δίκτυα (dynamic Bayesian networks, DBN) μια υπόθεση Markov συνήθως περιορίζει την τάξη των υποθετικών εξαρτήσεων τόσο στο μοντέλο μετάβασης, ορίζοντας υποθετικές πιθανότητες σε μεταβλητές εσωτερικού επιπέδου X_t , όσο και στο μοντέλο αισθητήρα, ορίζοντας υποθετικές πιθανότητες που περιλαμβάνουν μεταβλητές εσωτερικού επιπέδου X_t και μεταβλητές απόδειξης $E(t)$. Αντίστοιχα, σε μοντέλα κίνησης, οι μεταβλητές επιπέδου μπορεί να αλληλεπιδρούν με τοποθεσίες και μεταβλητές.

Γενικά, μια προσέγγιση VOMM χρησιμοποιεί δομές δένδρων, προκειμένου να αντιμετωπίσει το πρόβλημα της αραίωσης του πίνακα μετάδοσης. Αν η μέγιστη τάξη του είναι N , τότε το δένδρο έχει ένα μέγιστο βάθος $N + 1$ και κάθε μονοπάτι μέσα σ' αυτό ορίζει μια υποακολουθία συμβόλων που εμφανίζονται στην ακολουθία κατάρτισης. Κάθε κόμβος του δένδρου περιλαμβάνει ένα σύμβολο q από την αλφάβητο Σ και έναν μετρητή c για την καταγραφή του αριθμού των περιστατικών του πλαισίου, που κατασκευάστηκαν με τη σύζευξη όλων των συμβόλων από τη ρίζα στον κόμβο εκείνο.



Εικόνα 5. Δέντρο πρόθεμα (trie) που κατασκευάστηκε από το PPM για μια ακολουθία εκπαίδευσης «abcdcacbdcbac». Σημάδια κόμβων: $(q \in \Sigma, c)$

Αναλυτικότερα, η προσέγγιση αυτή χωρίζεται σε δύο μέρη τα οποία είναι η εισαγωγή του χρονικού περιεχομένου και η εισαγωγή του κοινωνικού περιεχομένου, αντίστοιχα. Για την πρόβλεψη τοποθεσίας, τα σύμβολα αλληλεπιδρούν με τις τοποθεσίες. Προκειμένου να εισαχθεί, επομένως, χρονική περιοδικότητα ως χρονικό περιεχόμενο, πραγματοποιήθηκε μετατροπή του υπάρχοντος δένδρου και επέκταση κάθε χωρικού κόμβου μ' ένα υποδένδρο, κατασκευασμένο από τους χρονικούς σχολιασμούς που καθορίστηκαν από τα χρονικά σήματα των περιεχομένων του πλαισίου s , που αντιστοιχούν σε αυτόν τον κόμβο. Έστω $\lambda = (D_1, D_2, \dots, D_i)$ το σύνολο των χρονικών χαρακτηριστικών. Στη προκειμένη περίπτωση, χρησιμοποιήθηκαν τρία χρονικά χαρακτηριστικά $\lambda = (ts, wo, dw)$, όπου ts είναι ο διαχωρισμός της ημέρας (πρωί, μεσημέρι, βράδυ), το wo δείχνει αν είναι καθημερινή ή σαββατοκύριακο και το dw δείχνει τη μέρα της εβδομάδας. Για κάθε σύμβολο q που εμφανίζεται στην αλληλουχία καθορίζεται ο χρονικός σχολιασμός του $(\tau_1, \tau_2, \dots, \tau_i)_q$ για $\forall j \in D_j$. Η Εικόνα 6 δείχνει ένα παράδειγμα χρονικού υποδένδρου για ένα χωρικό κόμβο. Στο χρονικό αυτό υποδένδρο, τα χρονικά χαρακτηριστικά $(\tau_1, \tau_2, \dots, \tau_i)$ ταξινομούνται, έτσι ώστε το χρονικό χαρακτηριστικό τ_x στο επίπεδο x του δένδρου να είναι εννοιολογικά πιο συγκεκριμένο από το χρονικό χαρακτηριστικό τ_{x-1} στο επίπεδο $x - 1$. Η εννοιολογική ιδιαιτερότητα μπορεί να εφαρμοστεί από μια σημασιολογία συμπερίληψης, η οποία υποδηλώνει ότι τα χρονικά διαμερίσματα στο επίπεδο x υποδιαίρονται σε χρονικά διαμερίσματα στο επίπεδο $x + 1$.



Εικόνα 6. Παράδειγμα χρονικού περιεχομένου ενός υποδένδρου.

Η αρχή κατασκευής του VOMM και ο μηχανισμός διαφυγής θα διασφαλίσουν και στις δύο περιπτώσεις ότι οι μετρητές διατηρούνται σωστά και χρησιμοποιούνται κατά τη διάρκεια της εκπαίδευσης και της πρόβλεψης. Στη συγκεκριμένη περίπτωση, επιλέχθηκε η ακολουθία $w_0 \leq dw \leq ds$ που δίνεται από την απλή σημασιολογία της χρονικής συμπερίληψης. Όπως και στη γενική περίπτωση που περιγράφηκε παραπάνω, κάθε κόμβος συνδέεται με ένα μετρητή που αποθηκεύει τις επισκέψεις που ταιριάζουν με το συγκεκριμένο χρονικό πλαίσιο. Κατά τη διάρκεια της φάσης πρόβλεψης, το χρονικό υποδένδρο μπορεί να μειωθεί και να ταιριάξει με το τρέχον υποδένδρο και να ξεφύγει σε πιο γενικά χρονικά χαρακτηριστικά, όποτε δεν υπάρχουν αρκετά διαθέσιμα στοιχεία για το πιο συγκεκριμένο πλαίσιο. Εκτός από αυτά τα φυσικά χρονικά μοτίβα, είναι επίσης δυνατό να συμπεριληφθούν περισσότερα χρονικά μοτίβα δεδομένων, τα οποία μπορούν να ανακαλυφθούν ή να προσαρμοστούν στο πολιτισμικό υπόβαθρο του χρήστη (π.χ. σχετικά με τις ημέρες που θεωρούνται «Σαββατοκύριακο»). Όταν χρησιμοποιούνται μεμονωμένα χρονικά χαρακτηριστικά, η ενσωμάτωση των κοινωνικών πλαισίων γίνεται πιο περίπλοκη. Ως εκ τούτου, χρησιμοποιήθηκαν κοινά στατικά χρονικά χαρακτηριστικά για όλους τους χρήστες. Αν, για παράδειγμα, μια χωρική ακολουθία sq έχει χρονικό πλαίσιο (mor, wd, mon) , πρέπει να αυξηθούν όλοι οι μετρητές κατά μήκος αυτής της πορείας με τον ίδιο ακριβώς τρόπο, όπως στη χωρική περίπτωση που περιγράφηκε παραπάνω. Μέσω αυτής της κράτησης βήμα-προς-βήμα αφαίρεσης, μπορεί να χρησιμοποιηθεί ο πρωτότυπος PPM αλγόριθμος, με την ικανότητά του να εξετάζει διάφορα μήκη αλληλουχίας χωρίς τροποποίηση. Αν, για ένα πλαίσιο s , δεν έχει εντοπιστεί αυτό το ιδιαίτερο χρονικό πλαίσιο (ts, w_0, dw) , τότε χρησιμοποιείται το γενικότερο (w_0, dw) , και αν δεν εντοπίζεται ούτε αυτό, τότε χρησιμοποιείται το dw . Εάν πάλι δεν έχει εντοπιστεί ούτε το dw , γίνεται χρήση καθαρά χωρικών σχέσεων. Στην πράξη, χρησιμοποιείται το πρότυπο PPM εκτός από την περίπτωση της εισαγωγής του χρονικού υποδένδρου μετά από κάθε κόμβο στο δένδρο, δηλαδή $\Sigma = \Sigma_{loc} \cup \Sigma_{temp}$, όπου Σ_{loc} είναι τα σύμβολα θέσης και $\Sigma_{temp} = \bigcup_{j=1}^2 D_j$, που στην προκειμένη περίπτωση είναι $\Sigma_{temp} = \{mor, afn, evn, we, wd, mon, tue, \dots, sun\}$.

Στο κομμάτι της εισαγωγής κοινωνικού περιεχομένου, ορίστηκαν τα κοινωνικά hotspots (socializable hotspots, SHs) ως hotspots όπου τουλάχιστον δύο χρήστες βρίσκονταν στην ίδια τοποθεσία σε μια δεδομένη στιγμή. Για ένα συγκεκριμένο χρήστη, γίνεται διάκριση μεταξύ των κοινωνικών hotspots που συμμετείχε (κλάση I) και εκείνων των κοινωνικών hotspots, όπου συμμετείχαν «φίλοι» (κλάση II). Οι «φίλοι» μπορούν να αποκτηθούν είτε μέσω άμεσης δήλωσης, είτε μέσω έμμεσης εξόρυξης. Τα μεμονωμένα σημεία ενδιαφέροντος μπορούν να αντιμετωπιστούν ως μια ειδική μορφή ενός κοινωνικού hotspot, όπου ο αριθμός των ατόμων που επισκέπτονται τον τόπο είναι ίσος με 1, ιδέα που επιτρέπει και την ενσωμάτωση του κοινωνικού hotspot στο δένδρο. Στο δένδρο που περιέχει χωρική και χρονική γνώση, σημειώνεται κάθε κόμβος με μετρητές για κάθε σύνολο ατόμων που έχουν παρατηρηθεί στο αντίστοιχο περιβάλλον. Η παρουσία άλλων χρηστών μπορεί να αναγνωριστεί από την ανίχνευση της κοινωνικής κατάστασης [32] χρησιμοποιώντας την επικοινωνία συσκευής-προς-συσκευή. Αυτός ο τύπος επικοινωνίας χρησιμοποιείται, επίσης, και για την ανταλλαγή χωροχρονικών υποδένδρων VOMM μεταξύ φίλων, κατά τη διάρκεια συναντήσεων για τον καθορισμό των κοινωνικών hotspots κλάσης II. Η ανταλλαγή hotspots κλάσης I επιτρέπει την έγχυση τεράστιου αριθμού πρόσθετων γνώσεων που συλλέγονται από άλλους χρήστες του συγκεκριμένου δένδρου, η χρήση του οποίου παρεμβάλλεται κατάλληλα για πρόβλεψη μέσω της σύγκρισής του

με το σημερινό κοινωνικό πλαίσιο. Επιπλέον, τα κοινωνικά hotspots κλάσης II απαιτούν τη σύγκριση των σημείων πρόσβασης των διαφορετικών χρηστών, για την ανίχνευση ζωντανών σημείων πρόσβασης. Αυτό μπορεί να επιτευχθεί είτε συγκρίνοντας τις ετικέτες που έχει εκχωρήσει ο χρήστης στα hotspots του, είτε με την ομαδοποίηση αυτών σε μικρή ακτίνα μεταξύ όλων των χρηστών. Τα κοινωνικά hotspots κλάσης I ανιχνεύονται από τον ίδιο τον χρήστη. Κάθε φορά που ανιχνεύεται ένας φίλος στην ίδια θέση, αυξάνονται και οι μετρητές για τις επιμέρους επισκέψεις στο τρέχον χωροχρονικό πλαίσιο, καθώς επίσης και ο μετρητής του συνόλου των φίλων που είναι παρόντες. Τα πρόσφατα ανακοινωθέντα κοινωνικά hotspots της κλάσης II πρέπει να ταιριάζουν με τις διαδρομές στο υπάρχον χωροχρονικό δένδρο του χρήστη. Αν η διαδρομή υπάρχει ήδη, μπορεί απλά να προστεθεί ένας επιπλέον μετρητής, που αντιστοιχεί στη νέα κοινωνική κατάσταση της κλάσης II (σύνολο φίλων), που έχει γίνει γνωστή από το φιλικό δένδρο. Αν η διαδρομή δεν υπάρχει, τότε εισάγεται, μαζί με τις τιμές του μετρητή που έχουν αποκτηθεί και αντιστοιχούν σε κοινωνικά hotspots, στο δένδρο του φίλου. Ως εκ τούτου, η συμπερίληψη πληροφοριών κοινωνικού περιεχομένου απαιτεί μόνο κάποιες προσαρμογές στο δένδρο πρόβλεψης. Για κάθε χρονική στιγμή, η κοινωνική κατάσταση του χρήστη θεωρείται γνωστή [32]. Αν για ένα δεδομένο χωροχρονικό πλαίσιο ενός χρήστη u_{i1} , είναι παρόντες και άλλοι χρήστες u_{i2}, \dots, u_{im} και το δένδρο για το δεδομένο πλαίσιο περιέχει μετρητές για κάποιες προηγούμενες επισκέψεις, ο αλγόριθμος πρέπει να ζυγίζει ποια συμπεριφορά, που παρατηρήθηκε προηγουμένως, είναι η πιο σχετική γι' αυτή την κατάσταση. Ένα σταθμισμένο άθροισμα όλων των μετρητών $C(sq)_{U'}^{soc}$ ενός κόμβου q , στο χρονικό υποδένδρο, υπολογίζεται μ' ένα μέτρο αλληλοεπικάλυψης μεταξύ U' και U ως βάρος: $\overline{C}_U^{soc}(sq) = \sum_{U'} \overline{C}_{U'}^{soc} * Jacc(U', U)$, όπου $Jacc$ ο συντελεστής Jaccard. Επειδή το σύνολο των ατόμων που εμπλέκονται σε κοινωνικές καταστάσεις/κοινωνικά hotspots δεν επιτρέπει ερμηνεία του «πιο συγκεκριμένου», όπως στο χρονικό υποδένδρο, ή «συνέβη πριν», όπως στο χωρικό δένδρο, δεν είναι δυνατή η χρήση ενός αντίστοιχου διαφυγής υποδένδρων. Η χρήση του συντελεστή Jaccard παρέχει διάφορα οφέλη. Κάθε φορά που ο χρήστης είναι μόνος, τα μονοπάτια που αντλήθηκαν από την ατομική του συμπεριφορά έχουν βάρος 1, ενώ κάθε άλλος κόμβος έχει μέγιστο βάρος 0.5 (κοινωνικά hotspots κλάσης I με τον χρήστη και άλλο άτομο).

Τέλος, και όσον αφορά την αποτίμηση της απόδοσης αυτού του συνδυασμού μεθόδων, τα αποτελέσματα έδειξαν ότι η συμπερίληψη χρονικού περιεχομένου βελτιώνει σημαντικά την ακρίβεια πρόβλεψης. Αναλυτικότερα, στις ομάδες δεδομένων που χρησιμοποιήθηκαν, παρατηρήθηκε ότι η απόδοση του σταθερού μοντέλου Markov μειώνεται σημαντικά, ενώ τα αποτελέσματα του VOMM παραμένουν σταθερά. Τα μοντέλα σταθερής σειράς Markov απαιτούν τεράστιο όγκο δεδομένων εκπαίδευσης, ώστε να αποδίδουν καλύτερα, με αποτέλεσμα όμως να υποφέρουν από προβλήματα ψυχρής εκκίνησης. Αυτό είναι και ένα από τα βασικά πλεονεκτήματα της προσέγγισης αυτής. Επιπρόσθετες πληροφορίες που προκύπτουν ως αποτελέσματα της έρευνας, είναι ότι μεγαλύτερο ιστορικό κίνησης και μικρότερος αριθμός θέσεων, καθώς επίσης και χρήστες οι οποίοι ακολουθούν έναν σχετικά «αναμενόμενο» τρόπο ζωής, αυξάνουν την ακρίβεια της πρόβλεψης.

3.3.4. Επόμενο Check-in

Συνεχίζοντας, η εύρεση λύσης στο πρόβλημα της πρόβλεψης της επόμενης κοινοποίησης τοποθεσίας των χρηστών ενός τηλεπικοινωνιακού δικτύου (Next Check-in Problem) [33] επιτεύχθηκε χάρη στην ανάλυση του ιστορικού των παρελθουσών τοποθεσιών στις οποίες αυτοί βρέθηκαν. Συλλέχθηκε και αναλύθηκε ένα ευρύ φάσμα δεδομένων από κοινοποιήσεις τοποθεσίας των ίδιων των χρηστών μέσω της εφαρμογής FourSquare, ενώ ταυτόχρονα ορίστηκε και αξιολογήθηκε μια σειρά από στρατηγικές πρόγνωσης, που κυμαίνονται από τις παρελθούσες επισκέψεις των χρηστών ως τις λειτουργίες εξόρυξης παγκόσμιων πληροφοριών σχετικά με το σύστημα, όπως η γεωγραφική απόσταση μεταξύ των περιοχών, η δημοτικότητά τους ή τα χρονικά μοτίβα που ακολουθούν όταν επισκέπτονται τις προτιμήσεις τους. Τα αποτελέσματα της απόδοσης αυτής της προσέγγισης, ωστόσο, δείχνουν ότι οι προσεγγίσεις που βασίζονται μόνο σε μια συγκεκριμένη πτυχή των δεδομένων συμπεριφοράς των χρηστών, στην προκειμένη περίπτωση μόνο στα δεδομένα ιστορικού από παρελθούσες τοποθεσίες τους, δεν επιτυγχάνουν κορυφαία απόδοση. Προτάθηκε, λοιπόν, και αξιολογήθηκε μια προσέγγιση που συνδυάζει διάφορα χαρακτηριστικά κινητικότητας και αντιμετωπίζει την επόμενη εργασία πρόβλεψης θέσης ως πρόβλημα εποπτευόμενης μάθησης, με στόχο την ταυτόχρονη εκμετάλλευση της προγνωστικής δύναμης πολλών παραγόντων. Η αξιολόγηση αφορούσε δύο διαφορετικούς εποπτευόμενους ταξινομητές: ένα μοντέλο γραμμικής παλινδρόμησης και δέντρα μοντέλου M5. Τα τελευταία επιτυγχάνουν αισθητά υψηλότερη απόδοση από οποιαδήποτε άλλη προσέγγιση μεμονωμένων χαρακτηριστικών, βελτιώνοντας τις βαθμολογίες πρόβλεψης σε διαφορετικές μετρήσεις αξιολόγησης.

Αναλυτικότερα, αρχικά ορίζεται ένα σύνολο από χρήστες U και ένα σύνολο τοποθεσιών L . Κάθε κοινοποίηση παρουσίας c ενός χρήστη u ορίζεται ως ένα tuple $\{l, t\}$, όπου το $l \in L$ αναπαριστά το χώρο και t είναι η χρονοσφραγίδα της κοινοποίησης. Το γενικό σύνολο των κοινοποιήσεων ορίζεται ως C και το σύνολο των κοινοποιήσεων ενός χρήστη ως C_u . Στη συνέχεια, σχηματίζεται το πρόβλημα της πρόβλεψης της επόμενης κοινοποίησης τοποθεσίας ως εξής: Δεδομένου ενός χρήστη u , του οποίου τα δεδομένα κοινοποίησης είναι c (στο χώρο l' και σε χρόνο t'), σκοπός είναι να γίνει μια κατάταξη του συνόλου των τοποθεσιών L , έτσι ώστε ο επόμενος χώρος επίσκεψης να καταταχθεί στην υψηλότερα πιθανή θέση στη λίστα. Σύμφωνα με την παραπάνω τοποθέτηση, το πρόβλημα αυτό είναι ουσιαστικά μια εργασία κατάταξης, όπου γίνεται υπολογισμός μιας βαθμολογίας κατάταξης \hat{r} για όλες τις τοποθεσίες στο L . Η επιλογή των υποψήφιων τοποθεσιών μεταβάλλεται στα πλαίσια μιας δεδομένης πόλης. Σε κάθε περίπτωση, θεωρούνται t' και l' ο χρόνος και η τοποθεσία της δεδομένης κοινοποίησης, αντίστοιχα. Ως t' είναι ο χρόνος της δεδομένης πρόβλεψης και η βαθμολογία κατάταξης όλων των χαρακτηριστικών υπολογίζεται λαμβάνοντας υπόψη γνώση από εκείνο το χρόνο και έπειτα. Τα χαρακτηριστικά του προβλήματος πρόβλεψης, λοιπόν, υπολογίζουν μια αριθμητική τιμή \hat{r}_k για κάθε υποψήφιο χώρο k , η οποία στη συνέχεια χρησιμοποιείται για την παραγωγή μιας εξατομικευμένης κατάταξης των τοποθεσιών. Στη συνέχεια, ορίζεται ως $rank(k)$ η κατάταξη των χώρων k , που αποκτώνται μετά τη διαλογή κατά αύξουσα σειρά όλων των χώρων του L , σύμφωνα με το \hat{r}_k . Στόχος του προβλήματος είναι η μέτρηση της έκτασης του χώρου τον οποίο θα επισκεφθεί ο χρήστης και είναι υψηλά στην κατάταξη, σύμφωνα με τον αλγόριθμο που χρησιμοποιείται.

Για τη μέτρηση της απόδοσης του αλγορίθμου χρησιμοποιήθηκαν δύο μετρικές. Αρχικά, η *Εκατοστιαία Κατάταξη (Percentile Rank, PR)* του χώρου k που επισκέφθηκε ο χρήστης, η οποία ορίζεται ως $PR = \frac{|L| - rank(k) + 1}{|L|}$. Το αποτέλεσμα της εκατοστιαίας κατάταξης είναι ίσο με 1, όταν ο επόμενος χώρος επίσκεψης είναι πρώτος στη σειρά κατάταξης, και μειώνεται γραμμικά ως το 0 καθώς η τοποθεσία υποβιβάζεται κατά μήκος της λίστας. Η *Μέση Εκατοστιαία Κατάταξη (Average Percentile Rank, APR)* ορίζεται ως ο μέσος όρος της πρόβλεψης κατά μήκος όλων των κοινοποιήσεων που έχει πραγματοποιήσει ο χρήστης. Ως αποτέλεσμα έχει τη μέση κανονικοποιημένη θέση της σωστής περίπτωσης από τη λίστα κατάταξης των περιπτώσεων. Χρησιμοποιήθηκε, επίσης, και *ακρίβεια πρόβλεψης* για την αξιολόγηση της απόδοσης όταν χρησιμοποιείται διαφορετικό μέγεθος λίστας πρόβλεψης N . Σ' αυτή την περίπτωση, η πρόβλεψη του χώρου όπου θα γίνει η επόμενη κοινοποίηση είναι επιτυχής, αν ένας χώρος μπει στην κατάταξη με τις κορυφαίες N τοποθεσίες. Η *μέση ακρίβεια* είναι το πηλίκο των επιτυχημένων περιπτώσεων προς το συνολικό αριθμό των έργων πρόβλεψης.

3.3.5. Επόμενη Θέση Χρήστη

Αντίστοιχα, μια άλλη μέθοδος πρόβλεψης, που αξίζει να σημειωθεί, είναι αυτή που βασίζεται στις ατομικές και συλλογικές γεωγραφικές προτιμήσεις των χρηστών [34]. Η μελέτη αυτή, λοιπόν, προτείνει έναν μεμονωμένο δείκτη πρόβλεψης κίνησης (mobility predictor), ο οποίος συνδυάζει τις παρελθούσες επιλογές κίνησης του χρήστη και τη συνολική του συμπεριφορά, όσον αφορά (α) στην τάση για αλλαγή τοποθεσίας και (β) στον τύπο των γεωγραφικών περιοχών που θεωρούνται ενδιαφέρουσες ως προς τη συλλογικότητα για μια δοθείσα χρονική στιγμή, στα πλαίσια της χρήσης τοποθεσίας, σημείων ενδιαφέροντος και απόστασης των ταξιδιών. Το χαρακτηριστικό αυτό θεωρείται ότι επηρεάζει τις επιλογές κίνησης ως διακομιστής των δραστηριοτήτων.

Πιο συγκεκριμένα, το μοντέλο πρόβλεψης που προτείνεται έχει ως εξής: έστω ότι ο χώρος διαιρείται σε n πλέγματα $i \in N_n$ και με το k υποδηλώνεται ο χρόνος, ο οποίος μπορεί να επιλεγεί για παράδειγμα ως ημέρες, ώρες, λεπτά. Έστω ένας πληθυσμός ατόμων με γνωστές τις προηγούμενες τροχιές τους και έστω καθένα από τα m θεωρούμενα άτομα ως u , και η θέση ενός ατόμου u κατά το χρόνο k ως $x(u)_k = i$. Το επιθυμητό αποτέλεσμα είναι η πρόβλεψη της επόμενης τοποθεσίας του ατόμου $x(u)_{k+1}$, δεδομένων ιστορικών δεδομένων. Για να γίνει αυτό, λοιπόν, ακολουθήθηκε μια προσέγγιση πιθανότητας και καθορίστηκε μια πιθανότητα για κάθε κελί πλέγματος j να είναι η επόμενη θέση του ατόμου, ως συνάρτηση του ατόμου και των συλλογικών συμπεριφορών του παρελθόντος. Η κυψέλη με τη μεγαλύτερη πιθανότητα επιλέγεται στη συνέχεια, βάσει θέσης του προσώπου. Έστω ότι η συμπεριφορά είναι περιοδική με την πάροδο του χρόνου, με μια περίοδο T , π.χ. εβδομαδιαία συμπεριφορά. Στη συνέχεια, για την πρόβλεψη της τοποθεσίας του ατόμου την ώρα k , χρησιμοποιήθηκαν πληροφορίες για προηγούμενες επιλογές σε στιγμές $k - T, k - 2T, \dots$

Στη συνέχεια, πραγματοποιήθηκε η μελέτη αυτού του προτεινόμενου μοντέλου, χρησιμοποιώντας ένα τεράστιο δίκτυο κινητής τηλεφωνίας, όσον αφορά (α) στη θέση των κινητών, (β) στη χρήση τοποθεσίας και (γ) στα σημεία ενδιαφέροντος. Για τη συλλογική

συμπεριφορά λαμβάνονται υπόψη τόσο οι αποστάσεις που ταξιδεύουν οι χρήστες, όσο και οι τύποι των τοποθεσιών που γίνονται αντικείμενα επίσκεψης. Οι τύποι των τοποθεσιών χρησιμοποιούνται επειδή είναι αδύνατη η εύρεση των δραστηριοτήτων των χρηστών μόνο με δεδομένο το ίχνος κίνησης, οπότε και χρησιμοποιούνται πληροφορίες σχετικά με τους πόρους που έχει να προσφέρει κάθε περιοχή. Με άλλα λόγια, γίνεται σχεδιασμός μιας πιθανότητας για να επιλεγεί ένας δεδομένος προορισμός ως συνάρτηση της απόστασης του προορισμού, της παρουσίας σημείων ενδιαφέροντος, παρόμοιων μ' αυτά που έχουν επισκεφθεί οι χρήστες συλλογικά, και του τύπου της περιοχής που επιλέγεται.

Η μεθοδολογία που ακολουθήθηκε έχει ως εξής:

1. Έγινε εισαγωγή μιας σειράς μετρήσεων $M_s = m_q, m_{q+1}, \dots, m_z \in M^{z-q-1}$, όπου ο χρήστης πραγματοποιεί συνδέσεις κατά τη διάρκεια ενός συγκεκριμένου χρονικού διαστήματος $\Delta T = t_{m_z} - t_{m_q} > 0$, σε μια περιοχή ακτίνας ΔS , σύμφωνα με τον τύπο: $\max_{distance}(p_{m_i}, p_{m_j}) < \Delta S \forall q \leq i, j \leq z$.
2. Τα σημεία $M_s = m_q, m_{q+1}, \dots, m_z \in M^{z-q-1}$ συνενώνονται, προκειμένου μια μεμονωμένη γεωγραφική περιοχή $p_s = (z - q)^{-1} \sum_{i=q}^z p_{m_i}$ (που περικυκλώνει τα σημεία) να μπορεί να θεωρηθεί ως μια εικονική τοποθεσία που χαρακτηρίζεται από μια ομάδα συνεχών μετρήσεων τοποθεσίας. Αυτή η τοποθεσία γίνεται η προέλευση ή ο προορισμός του ταξιδιού.
3. Μόλις εντοπιστούν οι εικονικές τοποθεσίες, μπορεί να γίνει αξιολόγηση των στάσεων και των ταξιδιών ως διαδρομές μεταξύ των θέσεων των χρηστών, στις συνεχείς αυτές εικονικές τοποθεσίες.

Τέλος, για την πρόβλεψη της ακρίβειας του προτεινόμενου μοντέλου πρόβλεψης, αυτό εφαρμόστηκε σ' ένα σύνολο από 2000 ίχνη και η ακρίβεια της πρόβλεψης της μεμονωμένης τοποθεσίας αξιολογήθηκε ως το σφάλμα μεταξύ της τοποθεσίας που προβλέφθηκε και αυτής που παρατηρήθηκε. Επιπλέον, έγινε σύγκριση των αποτελεσμάτων δεδομένης της συλλογικότητας που συνίσταται μόνο από τον ίδιο το χρήστη· επομένως, δεν αξιοποιήθηκε καμία πληροφορία από άλλους χρήστες. Τα αποτελέσματα αυτής της περίπτωσης έδειξαν αυξημένη τιμή σφάλματος λόγω της απουσίας καθολικών προτιμήσεων άλλων χρηστών.

Όπως αναφέρθηκε ήδη, ένα αποδοτικό μοντέλο πρόβλεψης θα πρέπει να εξάγει μοτίβα από διάφορους συνδυασμούς ακολουθιών χωρικού, χρονικού και κοινωνικού περιεχομένου. Καθαρά χωρικά μοτίβα, μοτίβα δηλαδή ακολουθιών παρελθοντικού χωρικού περιεχομένου (τοποθεσία), επηρεάζουν την πρόβλεψη της τοποθεσίας. Σ' αυτές τις περιπτώσεις, η μελλοντική τοποθεσία του χρήστη εξαρτάται μόνο από ακολουθίες προηγούμενων τοποθεσιών του. Καθαρά χρονικά μοτίβα, όπως «κάθε βράδυ ο χρήστης επιστρέφει στο σπίτι του ανεξαρτήτως προηγούμενης ή τωρινής τοποθεσίας», αφηφούν γενικά το χωρικό πλαίσιο. Χωροχρονικά μοτίβα εκφράζουν περισσότερο συγκεκριμένες συσχετίσεις χωρικού και χρονικού περιεχομένου, συνδέοντας χρονικές περιοδικότητες με τοποθεσίες. Το μοντέλο, λοιπόν, θα πρέπει να καθορίζει και να δημιουργεί επακριβώς έννοιες όπως «επίσκεψη», «τοποθεσία» κ.λπ. Κοινωνικά περιεχόμενα μπορούν επίσης να επηρεάσουν την πρόβλεψη της τοποθεσίας, π.χ. «αν με τους φίλους μου μετά το πάρκο ακολουθεί και μια βόλτα από το μπαρ», «αν με τα παιδιά μου ακολουθεί μια βόλτα από το ζαχαροπλαστείο». Κοινωνικά περιεχόμενα μπορούν να προκύψουν

από την ανάλυση συν-τοποθεσιών (co-locations). Επίσης, λαμβάνονται υπόψη και μακροπρόθεσμες πληροφορίες κοινωνικής δικτύωσης, όπως οι φιλικές σχέσεις στα μέσα κοινωνικής δικτύωσης.

4. Υλοποίηση

Όπως αναφέρθηκε και στην αρχή της εργασίας, στόχος της συγκεκριμένης υλοποίησης είναι η κατασκευή ενός συνδυαστικού μοντέλου κίνησης, το οποίο συνδυάζει πληροφορίες και από το φυσικό, αλλά και από τον εικονικό κόσμο. Το συνδυαστικό αυτό μοντέλο, με τη σειρά του, παρέχει πληροφορίες που θα βοηθήσουν στην καλύτερη εκτίμηση της κινητικότητας των χρηστών του συστήματος. Οι απαραίτητες αυτές πληροφορίες αντλήθηκαν για το μεν φυσικό κόσμο, από δεδομένα κίνησης μιας συγκεκριμένης περιοχής, όπως αυτά εξάγονται με βάση το μοντέλο Μανχάταν, για το δε εικονικό, από το σύνολο δεδομένων Erinions. Τα δεδομένα αυτά με τη σειρά τους αφορούν πληροφορίες σχετικές με τους χρήστες του συστήματος, τα αντικείμενα που κοινοποιούν, καθώς επίσης και τις αξιολογήσεις που αφορούν τόσο τους ίδιους τους χρήστες, όσο και το περιεχόμενο που κοινοποιούν.

Αναλυτικότερα, το κομμάτι της υλοποίησης διακρίνεται στην υλοποίηση μιας μετρικής για τον υπολογισμό του μέσου χρόνου που μεσολαβεί μεταξύ των διαδοχικών επαφών δύο κόμβων, στα πλαίσια της φυσικής πληροφορίας, από άποψη εγγύτητας των χρηστών. Αντίστοιχα, στα πλαίσια των συστάσεων της εικονικής πληροφορίας, υλοποιήθηκε μια απλοποιημένη μορφή του προτεινόμενου αλγορίθμου που παρουσιάστηκε στην παράγραφο 3.2.2. Τέλος, παρουσιάζεται η συνδυαστική γραφική απεικόνιση, όπως προκύπτει από πληροφορίες που αντλούνται από τις δύο αυτές υλοποιήσεις.

4.1. Προσομοίωση του Μοντέλου Grid Manhattan

Διάφορες μελέτες, που αφορούν στην πρόβλεψη της κίνησης ενός χρήστη που βρίσκεται μέσα στο δίκτυο, κατέληξαν στο συμπέρασμα ότι οι χρήστες έχουν διαφορετικά μοτίβα κίνησης, τα οποία εξαρτώνται από τη ρουτίνα της καθημερινότητάς τους. Ένας βασικός παράγοντας, ωστόσο, που θεωρείται από πολλές μελέτες πως επηρεάζει την ποιότητα των σχέσεων μεταξύ των χρηστών ενός δικτύου είναι τα επίπεδα εγγύτητας μεταξύ αυτών. Αναλυτικότερα, διάφορες μελέτες, που αφορούν σχέσεις φιλίας μεταξύ χρηστών ενός δικτύου του φυσικού κόσμου, κατέληξαν στο συμπέρασμα πως όσο μικρότερη είναι η φυσική απόσταση μεταξύ δύο χρηστών του ίδιου δικτύου, τόσο πιθανότερο είναι γι' αυτούς τους δύο χρήστες να συνάψουν μια σχέση φιλίας. Γενικότερα, άνθρωποι που είναι πιο κοντά γεωγραφικά είναι πολύ ευκολότερο να συνάψουν σχέσεις μεταξύ τους και άρα κατ' επέκταση να γίνουν φίλοι ευκολότερα. Το γεγονός

αυτό δεν οφείλεται, βέβαια, μόνο στη συνεχή και ευκολότερη μεταξύ τους αλληλεπίδραση, αλλά και στη δυνατότητα που έχουν να ανήκουν σε κοινές ομάδες και να συμμετέχουν σε κοινές δραστηριότητες, ανταλλάσσοντας, συνεπώς, πληροφορίες μεταξύ τους.

Προκειμένου, λοιπόν, να αξιοποιηθεί η παραπάνω βασική πληροφορία, αρχικά, λάβαμε ως είσοδο του προγράμματος μια ήδη υπάρχουσα ανάλυση των δεδομένων, όπως αυτά προέκυψαν από την κινητικότητα των κόμβων του δικτύου, βάσει του μοντέλου κίνησης Μανχάταν (Manhattan) [17]. Στη συνέχεια, και σε συνδυασμό με τις ήδη υπάρχουσες δύο μετρικές του προγράμματος, υπολογίστηκε και μια τρίτη. Καθεμία από αυτές αφορά αντίστοιχα και σ' ένα διαφορετικό χαρακτηριστικό, που τελικά μαζί συμβάλλουν στον επιθυμητό υπολογισμό του χρόνου που μεσολαβεί μεταξύ των διαδοχικών επαφών δύο κόμβων.

Το σύνολο των δεδομένων που χρησιμοποιήθηκε στην προκειμένη περίπτωση για ανάλυση, αποτελείται από 200 κόμβους, οι οποίοι με τη σειρά τους κινούνται στο χώρο για 200 δευτερόλεπτα βάσει του μοντέλου Μανχάταν. Πιο συγκεκριμένα, ο χώρος αυτός αποτελεί έναν πίνακα 2000×2000 μέτρων, χρησιμοποιώντας ένα οδικό δίκτυο με ένα 5×5 πλέγμα Μανχάταν, όπου οριζόντιοι και κάθετοι δρόμοι χωρίζονται σε δύο γραμμές/ λωρίδες. Στους κάθετους δρόμους, οι κόμβοι μπορούν να μετακινηθούν βόρεια ή νότια, ενώ στους οριζόντιους μπορούν να κατευθυνθούν ανατολικά ή δυτικά, αντίστοιχα. Οι 200 αυτοί κόμβοι κινούνται μέσα σ' αυτήν την προαναφερθείσα περιοχή, με μια ταχύτητα 10 m/sec, ενώ οι κινήσεις τους παράγονται με τη χρήση του εργαλείου BonnMotion. Υπάρχουν 10 ζεύγη πηγών δεδομένων που αποστέλλουν δεδομένα UDP (User Datagram Protocol) με ταχύτητα εφαρμογής ίση με 2048 kB/s. Το εύρος (range) του κάθε κόμβου έχει οριστεί στα 500 μέτρα, σύμφωνα με το πρωτόκολλο IEEE802.11p, απόσταση η οποία θεωρείται ότι επιτρέπει κάθε φορά σε δύο κόμβους να είναι συνδεδεμένοι επομένως, κόμβοι οι οποίοι βρίσκονται σε απόσταση μικρότερη των 500 μέτρων, θεωρούνται ότι βρίσκονται σε επαφή. Στο σημείο αυτό, να τονίσουμε ότι ως επαφή λογίζεται η δυνητική σύνδεση δύο κόμβων, εφόσον αυτοί βρίσκονται εντός του επιτρεπτού πλαισίου απόστασης, βάσει της παραπάνω ακτίνας, και όχι η πραγματική τους σύνδεση.

Μια συνοπτική εικόνα των δεδομένων που προαναφέρθηκαν απεικονίζεται στον πίνακα που ακολουθεί.

Πίνακας I.

Μοντέλο Κινητικότητας	Μοντέλο Grid Manhattan
Διάρκεια Προσομοίωσης	200 δευτερόλεπτα
Περιοχή Προσομοίωσης	2000×2000 μέτρα
Σύνολο Κόμβων	200
Ταχύτητα Κόμβων	10 m/sec
Ζεύγη Πηγών Δεδομένων	10

Ταχύτητα Δεδομένων UDP	2048 kB/sec
WiFi	6 Mbps (802.11p)

Όπως αναφέρθηκε ήδη, για τη συγκεκριμένη προσομοίωση χρειάστηκαν συνολικά τρεις μετρικές. Αναλυτικότερα, η πρώτη μετρική, αφορά στη συχνότητα επαφής μεταξύ των δύο κόμβων. Αυτό επιτυγχάνεται με τη μέτρηση του πλήθους των επαφών που έχουν γίνει μεταξύ ενός κόμβου προορισμού και ενός κόμβου παραλήπτη, μέχρι το τρέχον δευτερόλεπτο της προσομοίωσης, σε σχέση με το συνολικό πλήθος των επαφών που έχει ο συγκεκριμένος κόμβος με όλους τους υπόλοιπους. Στη συνέχεια, η δεύτερη μετρική σχετίζεται με το χρονικό διάστημα κατά το οποίο οι δύο κόμβοι βρίσκονται σε επαφή. Υπολογίζει, δηλαδή, το συνολικό χρονικό διάστημα κατά το οποίο ένας κόμβος προορισμός είναι σε επαφή με ένα κόμβο παραλήπτη, πάλι μέχρι το τρέχον δευτερόλεπτο της προσομοίωσης, σε σχέση με το συνολικό πλήθος των μεταξύ τους επαφών. Τέλος, και αναφορικά με το κομμάτι της υλοποίησης, η τρίτη και τελευταία μετρική υπολογίζει τη μέση χρονική διάρκεια που μεσολαβεί μεταξύ των διαδοχικών επαφών δύο κόμβων. Πιο συγκεκριμένα, υπολογίζει το μέσο χρόνο που μεσολαβεί μεταξύ αυτών των διαδοχικών επαφών μέχρι το τρέχον δευτερόλεπτο της προσομοίωσης.

Αναλυτικότερα, για τον υπολογισμό της τρίτης μετρικής του προγράμματος:

- Κατά τη διάρκεια της προσομοίωσης χρειάστηκε να συγκεντρώνουμε τα σημεία εναλλαγής κατάστασης, από μη επαφή σε επαφή, στον ήδη υπάρχοντα πίνακα mymatrix, με την τελευταία παράμετρο του ορισμένη σε 3.
- Κατά τη διάρκεια εξαγωγής των αποτελεσμάτων έχουμε τις μεταβλητές `time_total` και `time_total_buffer`. Σε κάθε δευτερόλεπτο, ας πούμε για παράδειγμα τώρα στο 5^ο δευτερόλεπτο, κάνουμε πρώτα μια επανάληψη (loop) από το αρχικό δευτερόλεπτο στο τρέχον. Κατά την διάρκεια της επανάληψης μαρκάρουμε τον χρόνο που οι κόμβοι δεν είναι σε επαφή, αυξάνοντας το `buffer` κατά 1 για κάθε δευτερόλεπτο. Σε περίπτωση που οι κόμβοι είναι σε επαφή σε κάποιο δευτερόλεπτο, προσθέτουμε την τιμή του `buffer` στην κύρια μεταβλητή (`time_total`) και μηδενίζουμε το `buffer`. Τέλος, κάνουμε άλλη μια επανάληψη από τον χρόνο που βρισκόμαστε (5 στο παράδειγμα μας) έως το τελευταίο δευτερόλεπτο. Στην επανάληψη αυτή, κάνουμε αναζήτηση για κάποια καινούργια επαφή που θα συμβεί στο μέλλον. Αν αυτή δεν υπάρξει, τότε η επανάληψη τερματίζεται και προχωράμε στο επόμενο δευτερόλεπτο.

Προκειμένου, όμως, να γίνει ευκολότερη η κατανόηση της λειτουργίας του προγράμματος, θα παρουσιάσουμε και μια γενική εικόνα της προσομοίωσης. Αρχικά, για το πρώτο δευτερόλεπτο (`second 1`), ξεκινάμε και ελέγχουμε τις επαφές του κόμβου 1 με όλους τους υπόλοιπους. Αφού ολοκληρωθεί αυτός ο έλεγχος, το πρόγραμμα συνεχίζει την ίδια διαδικασία για τον κόμβο 2, μετά για τον 3 κ.ο.κ., πάντα όμως στο ίδιο δευτερόλεπτο (`second 1`). Εφόσον ολοκληρωθεί ο έλεγχος των επαφών και για τους 200 κόμβους, συνεχίζουμε στο δεύτερο δευτερόλεπτο της προσομοίωσης, όπου και πάλι ακολουθείται η ίδια διαδικασία, δηλαδή έλεγχος επαφών του κόμβου 1 με όλους τους υπόλοιπους, μετά του κόμβου 2 κ.ο.κ. Τέλος, η προσομοίωση τερματίζει μόλις ολοκληρωθεί ο έλεγχος αυτός και για το 200^ο δευτερόλεπτό της.

Οι κατάλληλες πληροφορίες που χρησιμοποιήθηκαν και αφορούν στη γεωγραφική τοποθεσία των χρηστών του δικτύου, περιέχονται στο αρχείο output.txt, όπως αυτό προέκυψε από το εργαλείο BonnMotion, χρησιμοποιώντας πληροφορίες για τα ίχνη των κόμβων σύμφωνα με το προαναφερθέν μοντέλο Μανχάταν. Το αρχείο αυτό περιλαμβάνει μια λίστα με τη θέση όλων των κόμβων του δικτύου ανά δευτερόλεπτο, για όλα τα δευτερόλεπτα της προσομοίωσης, που είναι 200.

Όσον αφορά στα αποτελέσματα της προσομοίωσης, αυτά αποτυπώνονται σε 2 αρχεία, το adjacency.txt και το metrics.txt. Το adjacency.txt περιλαμβάνει ένα συμμετρικό πίνακα 200×200, για το σύνολο των κόμβων του δικτύου, με τις τιμές του αθροίσματος των τριών μετρικών του προγράμματος, όπως αυτές προκύπτουν σε κάθε περίπτωση από το ίδιο το πρόγραμμα. Πιο συγκεκριμένα, η αθροιστική αυτή τιμή προκύπτει σύμφωνα με τον τύπο $0.33 \times (\text{μετρική}1 + \text{κανονικοποιημένη_μετρική}2 + \text{κανονικοποιημένη_μετρική}3)$ (1), όπου οι κανονικοποιημένες τιμές υπολογίζονται ως το πηλίκο της τιμής της μετρικής το δεδομένο δευτερόλεπτο προς τη μέγιστη τιμή που έχει πάρει αυτή η μετρική μέχρι εκείνο το δευτερόλεπτο της προσομοίωσης στο οποίο βρισκόμαστε.

Αντίστοιχα, το αρχείο metrics.txt περιλαμβάνει μια λίστα με το σύνολο των αποτελεσμάτων της προσομοίωσης. Ένα ενδεικτικό κομμάτι της λίστας αυτής φαίνεται στην Εικόνα 7. Όπως φαίνεται και στην εικόνα, λοιπόν, κάθε γραμμή περιλαμβάνει 7 τιμές, χωρισμένες μεταξύ τους με κόμμα. Η πρώτη τιμή που αναγράφεται αναφέρεται στο δευτερόλεπτο της προσομοίωσης, στη συγκεκριμένη περίπτωση είναι το 50^ο, ενώ η δεύτερη και η τρίτη τιμή αναφέρονται στους κόμβους. Βλέπουμε πως η δεύτερη τιμή είναι πάντα 85, γεγονός που οφείλεται στο ότι κάθε φορά παίρνουμε και ελέγχουμε έναν-έναν τους κόμβους για κάθε δευτερόλεπτο, όπως είπαμε και προηγουμένως. Αντίστοιχα, η τρίτη τιμή αυξάνεται σε κάθε γραμμή κατά 1, καθώς γίνεται έλεγχος των επαφών του κόμβου 85 με όλους τους υπόλοιπους κόμβους του δικτύου. Τέλος, ακολουθούν άλλες τρεις τιμές, οι οποίες αφορούν τη μετρική 1, τη μετρική 2, τη μετρική 3 και την κανονικοποιημένη μετρική 3, αντίστοιχα, όπως αυτές παρουσιάστηκαν αναλυτικά για τον υπολογισμό τους παραπάνω. Γραμμές όπου οι τιμές της μετρικής 3 και της κανονικοποιημένης τιμής της είναι ίσες με μηδέν, σημαίνουν πως οι δύο κόμβοι δε βρίσκονται σε μεταξύ τους επαφή.

Time Step	Node	Metric 1	Metric 2	Metric 3	Normalized Metric 3	Normalized Metric 3
2017110	50,85,109	0.000	0.000	0.000	0.000	0.000
2017111	50,85,110	0.000	0.000	0.000	0.000	0.000
2017112	50,85,111	0.000	0.000	0.000	0.000	0.000
2017113	50,85,112	0.111	0.843	0.000	0.000	0.000
2017114	50,85,113	0.000	0.000	0.000	0.000	0.000
2017115	50,85,114	0.000	0.000	0.000	0.000	0.000
2017116	50,85,115	0.000	0.000	0.000	0.000	0.000
2017117	50,85,116	0.111	0.725	14.000	0.280	0.280
2017118	50,85,117	0.000	0.000	0.000	0.000	0.000
2017119	50,85,118	0.222	0.098	42.000	0.840	0.840
2017120	50,85,119	0.111	0.627	0.000	0.000	0.000
2017121	50,85,120	0.000	0.000	0.000	0.000	0.000
2017122	50,85,121	0.000	0.000	0.000	0.000	0.000
2017123	50,85,122	0.000	0.000	0.000	0.000	0.000
2017124	50,85,123	0.000	0.000	0.000	0.000	0.000
2017125	50,85,124	0.000	0.000	0.000	0.000	0.000
2017126	50,85,125	0.111	0.176	0.000	0.000	0.000

Εικόνα 7. Στιγμιότυπο του αρχείου metrics.txt.

Ολοκληρώνοντας, το τελικό κομμάτι της προσομοίωσης αποτελεί η γραφική απεικόνιση του δικτύου, η οποία δίνεται από τον πίνακα γειτνίασης που περιέχεται στο αρχείο adjacency.txt. Το εργαλείο που χρησιμοποιείται σε κάθε περίπτωση για την απεικόνιση είναι το Graph Online, με την εισαγωγή των στοιχείων που προκύπτουν σε κάθε περίπτωση από τον πίνακα adjacency. Αναλυτικότερα, μια ενδεικτική εικόνα των αποτελεσμάτων του προγράμματος παρουσιάζεται παρακάτω, όπου έχουμε μια προσομοίωση που αφορά το σύνολο των 200 κόμβων, για το χρονικό διάστημα των 100 δευτερολέπτων. Ενδεικτικά, για τη γραφική απεικόνιση επιλέχθηκε τυχαία ένα χρονικό διάστημα μέχρι και το 50^ο δευτερόλεπτο της προσομοίωσης.



Εικόνα 8. Γραφική απεικόνιση του μοντέλου έως το 50^ο δευτερόλεπτο της προσομοίωσης.



Εικόνα 9. Καθαρότερη εικόνα της γραφικής απεικόνισης μέχρι το 50^ο δευτερόλεπτο.

Όπως είναι φανερό από την Εικόνα 8, λοιπόν, προκύπτει ένας ιδιαίτερα πυκνός και δυσανάγνωστος γράφος. Το αποτέλεσμα αυτό, ωστόσο, είναι απολύτως λογικό αν λάβουμε υπόψη τον όγκο των δεδομένων που χρησιμοποιείται από το πρόγραμμα για την εξαγωγή των αποτελεσμάτων. Η εικόνα, επομένως, δίνει ένα ενδεικτικό παράδειγμα των επαφών που πραγματοποιούνται στα 50 πρώτα δευτερόλεπτα της προσομοίωσης, σύμφωνα με τη διαδικασία που αναλύθηκε παραπάνω και για τους 200 κόμβους του δικτύου, χωρίς αυτό να σημαίνει φυσικά πως όλοι οι κόμβοι απαραίτητα είναι συνδεδεμένοι μεταξύ τους.

Αν παρατηρήσουμε, όμως, λίγο προσεκτικότερα τη γραφική απεικόνιση του δικτύου στην Εικόνα 9, θα δούμε σε πρώτη φάση πως σ' αυτήν απεικονίζονται πολλές μικρές πράσινες τελείες πάνω στις οποίες αναγράφονται και κάποιοι αριθμοί. Οι τελείες αυτές φυσικά απεικονίζουν τους κόμβους του δικτύου, ενώ αντίστοιχα οι αριθμοί που αναγράφονται πάνω, δείχνουν και ποιος είναι ο κάθε κόμβος. Στο παράδειγμα στην Εικόνα 9, ξεχωρίζει ο κόμβος 152. Πέρα όμως από τους αναγραφόμενους κόμβους, διακρίνονται επίσης και πολλές ακμές μωβ χρώματος, οι οποίες με τη σειρά τους απεικονίζουν την ύπαρξη επαφής μεταξύ δύο κόμβων. Κάθε κόμβος μπορεί να βρίσκεται σε επαφή κάθε φορά με πολλούς άλλους από τους υπόλοιπους 199, γι' αυτό το λόγο άλλωστε και οι ακμές του γράφου είναι τόσο πυκνές. Τέλος, πάνω σε κάθε ακμή αναγράφεται και ένας αριθμός, ο οποίος προκύπτει από τον υπολογισμό του βάρους κάθε ακμής, σύμφωνα με τον τύπο που αναφέρθηκε στο αρχείο adjacency.txt.

1. Σχηματισμός των Συνδέσεων των Χρηστών (User Connection Formation): Ένας χρήστης μπορεί να εκφράσει άμεσα αν εμπιστεύεται ή όχι κάποιον άλλο, ή έμμεσα μέσω της άποψης του (ένα «like», ή ένα σχόλιο) σε μια δημοσιευμένη κατάσταση περιεχομένου. Οι σχέσεις εμπιστοσύνης μπορούν να κατηγοριοποιηθούν ως εξής: (α) άμεση σύνδεση χρήστη-προς-χρήστη, (β) άμεση σύνδεση χρήστη-προς-αντικείμενο. Στην (α) περίπτωση της άμεσης σύνδεσης των χρηστών, ένας χρήστης μπορεί να σχετίζεται άμεσα μ' έναν άλλο, σχηματίζοντας σχέσεις εμπιστοσύνης ή δυσπιστίας, όπως φίλιες, ή συνεργασίες στον πραγματικό κόσμο. Η μοντελοποίηση αυτών των δεδομένων έγινε υποθέτοντας πως κάθε μέλος έχει και ανανεώνει δύο λίστες, μια φιλική και μια εχθρική, για τους αξιόπιστους και μη φίλους του, αντίστοιχα. Η λίστα των φίλων περιλαμβάνει μέλη τα οποία ο χρήστης ήδη εμπιστεύεται ή μπορεί να εμπιστευτεί και να αλληλεπιδράσει στο μέλλον. Στο (β) τύπο συνδέσεων, ο χρήστης παρέχει ένα σχόλιο επικρότησης ή αποδοκιμασίας σ' ένα συγκεκριμένο στοιχείο που έχει εκδοθεί από άλλον χρήστη, το οποίο, ανάλογα και με την εφαρμογή που χρησιμοποιείται σε κάθε περίπτωση, μπορεί να πάρει διάφορες μορφές.
2. Υπολογισμός της Κατάταξης Φήμης (Reputation Rating Estimation): Ο μηχανισμός φήμης ποσοτικοποιεί τις συνδέσεις εμπιστοσύνης που προσδιορίζονται στο κοινωνικό δίκτυο και παρέχει εξατομικευμένες αξιολογήσεις, οι οποίες εκφράζουν την τοπική πεποίθηση ενός χρήστη (χρήστης-αξιολογητής) σε σχέση με άλλα μέλη του δικτύου (χρήστες-στόχοι).

Για την πειραματική αξιολόγηση του προτεινόμενου συστήματος, χρησιμοποιήθηκαν σύνολα δεδομένων που καλύπτουν τα περισσότερα από τα επιθυμητά χαρακτηριστικά ενός κοινωνικού δικτύου, όπως περιγράφηκε προηγουμένως. Πρόκειται για μια ενότητα που αφορά την παρουσίαση αποτελεσμάτων στο εκτεταμένο σύνολο δεδομένων Epinions, το οποίο περιλαμβάνει δεδομένα σχετικά με την ταυτότητα (id) του κάθε χρήστη, τα αντικείμενα (items) που αυτός δημοσιεύει, καθώς και τις αξιολογήσεις (ratings) τόσο των αντικειμένων αυτών, όσο και των σχέσεων μεταξύ των ίδιων των χρηστών.

Ειδικότερα, λοιπόν, το κομμάτι που αφορά στην υλοποίηση του αλγορίθμου για τα πλαίσια της παρούσας εργασίας παρουσιάζεται ακολούθως. Το πρόγραμμα κατά την εκκίνηση ξεκινάει τη διαδικασία φόρτωσης των δεδομένων από το σύνολο Epinions, όπως αυτά δίνονται από δύο αρχεία τύπου κειμένου, στους πίνακες δομών Χρήστη (User), Χρήστη-προς-Χρήστη (User-to-User) και Χρήστη-προς-Αντικείμενο (User-to-Item). Πιο συγκεκριμένα:

- Ορίσαμε μια δομή `usr` (User), η οποία περιέχει την πραγματική ταυτότητα (id), καθώς και το εικονικό όνομα του χρήστη. Η αντιστοίχιση αυτή έγινε επειδή στα δεδομένα που χρησιμοποιήθηκαν υπάρχουν τυχαία id χρηστών, όπως είναι για παράδειγμα το 5887594372, τα οποία όμως θα πρέπει να αντιστοιχηθούν σε κάποιο σειριακό id μέσα στο εύρος 1 έως 200 του πλέγματος (grid) μας.
- Αντίστοιχα, δημιουργήσαμε τη δομή `u2u` (User-to-User), που περιέχει τα εικονικά id των δύο χρηστών του συστήματος, καθώς και τη μεταξύ τους αξιολόγηση (rating). Για την αξιολόγηση αυτή ισχύουν οι τιμές που αναφέρθηκαν και παραπάνω, δηλαδή αν οι δύο χρήστες είναι φίλοι, τότε η τιμή της αξιολόγησης είναι 1, ενώ σε περίπτωση που είναι εχθροί, η τιμή είναι -1, αντίστοιχα.

- Τέλος, ορίσαμε και μια δομή u2i (User-to-Item), η οποία με τη σειρά της περιέχει τις τιμές itemid, authorid και memberid, καθώς επίσης και την αξιολόγηση του χρήστη αυτού από το memberid.

Εφόσον τα δεδομένα μας φορτωθούν στις λίστες, το επόμενο στάδιο είναι να τρέξουμε ένα βρόγχο στον οποίο για κάθε σχέση Χρήστη-προς-Χρήστη θα καλούμε μια συνάρτηση αξιολόγησης (rating). Η συνάρτηση αυτή δέχεται ως παραμέτρους τα δύο εικονικά id των χρηστών και στη συνέχεια, υπολογίζει και επιστρέφει την επιθυμητή μετρική, ο υπολογισμός της οποίας αποτελείται από δύο μέρη. Υπολογίζουμε, λοιπόν, μια παράμετρο w1, που αφορά στη σχέση του ενός χρήστη με τον άλλο, αν δηλαδή αυτοί είναι φίλοι ή εχθροί, και μια παράμετρο w2, η οποία αποτελεί το πηλίκο των αξιολογήσεων ενός χρήστη στο αντικείμενο του άλλου προς το σύνολο των αξιολογήσεων του αντικειμένου αυτού. Οι τιμές των παραμέτρων αυτών παρουσιάζονται και αναλυτικότερα παρακάτω.

Όπως αναφέρθηκε και παραπάνω, η συγκεκριμένη έκδοση του προγράμματος που αφορά στο εικονικό δίκτυο βγάζει ως αποτελέσματα τις αξιολογήσεις των χρηστών, όπως αυτές παρέχονται από τα δύο αρχεία user_to_user_200.txt και user_to_item_200.txt. Αναλυτικότερα, το αρχείο user_to_user_200.txt περιλαμβάνει μια λίστα αποτελούμενη από τους χρήστες και τις τιμές που προσδιορίζουν τη μεταξύ τους σχέση, αν δηλαδή αυτή είναι αξιόπιστη, αναξιόπιστη ή ανύπαρκτη. Αντίστοιχα, το δεύτερο αρχείο user_to_item_200.txt περιλαμβάνει πάλι μια λίστα με χρήστες και το σύνολο των αξιολογήσεων που αυτοί έχουν κάνει πάνω στις δημοσιεύσεις των φίλων τους.

Αναφορικά με τα αποτελέσματα της προσομοίωσης, αυτά αποθηκεύονται σ' ένα τελικό αρχείο adjacency.txt, το οποίο έχει πάλι τη μορφή ενός πίνακα γειτνίασης 200×200 και περιλαμβάνει τις κανονικοποιημένες τιμές της μετρικής, και στο τέλος της προσομοίωσης τα δεδομένα αυτά εξάγονται σ' ένα αρχείο κειμένου. Η τελική κανονικοποιημένη τιμή της μετρικής που περιέχεται στο αρχείο αυτό δίνεται σύμφωνα με τον τύπο $0.5 \times (\text{παραμέτρος_w1} + \text{παραμέτρος_w2})$ (2).

Η μεταβλητή παράμετρος_w1 αποτελεί την αξιολόγηση που δίνεται από το αν η σχέση μεταξύ των δύο χρηστών είναι παρούσα, και κατ' επέκταση εφόσον υπάρχει, αν είναι φιλική ή εχθρική. Η μεταβλητή παράμετρος_w2, αντίστοιχα, προκύπτει από τον τύπο ως $\frac{\text{θετικές_αξιολογήσεις}(i \rightarrow j) - \text{αρνητικές_αξιολογήσεις}(i \rightarrow j)}{\text{θετικές_αξιολογήσεις}(j) + \text{αρνητικές_αξιολογήσεις}(j)}$ [18]. Ο αριθμητής του κλάσματος αναφέρεται στο σύνολο των αξιολογήσεων, για παράδειγμα, του χρήστη i για τα items του χρήστη j , ενώ ο παρονομαστής στις αξιολογήσεις για το σύνολο των items του χρήστη j .

Ολοκληρώνοντας, το τελικό κομμάτι της προσομοίωσης αποτελεί πάλι η γραφική απεικόνιση του δικτύου, όπως αυτή φαίνεται στην Εικόνα 11. Όπως είναι φανερό, και σ' αυτήν την περίπτωση ο γράφος είναι ιδιαίτερα πυκνός και δυσανάγνωστος, γεγονός που οφείλεται ομοίως στο μεγάλο όγκο δεδομένων. Αυτό εξηγείται αν αναλογιστούμε πως κάθε φορά το πρόγραμμα ξεκινάει από το χρήστη με ταυτότητα (ID) ίση με 1 και ελέγχει τις σχέσεις του με τους υπόλοιπους 200 χρήστες του δικτύου, συνεχίζει την ίδια διαδικασία για το χρήστη με ταυτότητα ίση με 2 κ.ο.κ., όπως ακριβώς γινόταν και στην προσομοίωση του φυσικού δικτύου.



Εικόνα 11. Γραφική απεικόνιση online δικτύου.



Εικόνα 12. Καθαρότερη εικόνα της γραφικής απεικόνισης του online δικτύου.

Παρατηρώντας μια πιο κοντινή απεικόνιση του γράφου του δικτύου στην παραπάνω εικόνα, βλέπουμε πάλι τους χρήστες να αναπαριστώνται με πράσινες τελείες, οι οποίες αναγράφουν και την ταυτότητά τους επάνω. Στην Εικόνα 12, διακρίνεται εύκολα ο χρήστης με ταυτότητα 62, καθώς και όλες οι επαφές και αξιολογήσεις του με τους υπόλοιπους χρήστες, όπως αυτές αναπαριστώνται με τις ακμές μωβ χρώματος. Τέλος, οι αριθμοί που αναγράφονται πάνω στις ακμές αποτελούν τα βάρη τους, όπως αυτά προκύπτουν από το τύπο της προηγούμενης σελίδας, που περιέχεται στον πίνακα adjacency. Τα βάρη αυτά αποτελούν μια ένδειξη προσωρινού ενδιαφέροντος από τον ένα χρήστη του συστήματος στον άλλο.

4.3. Υλοποίηση Συνδυαστικής Γραφικής Απεικόνισης

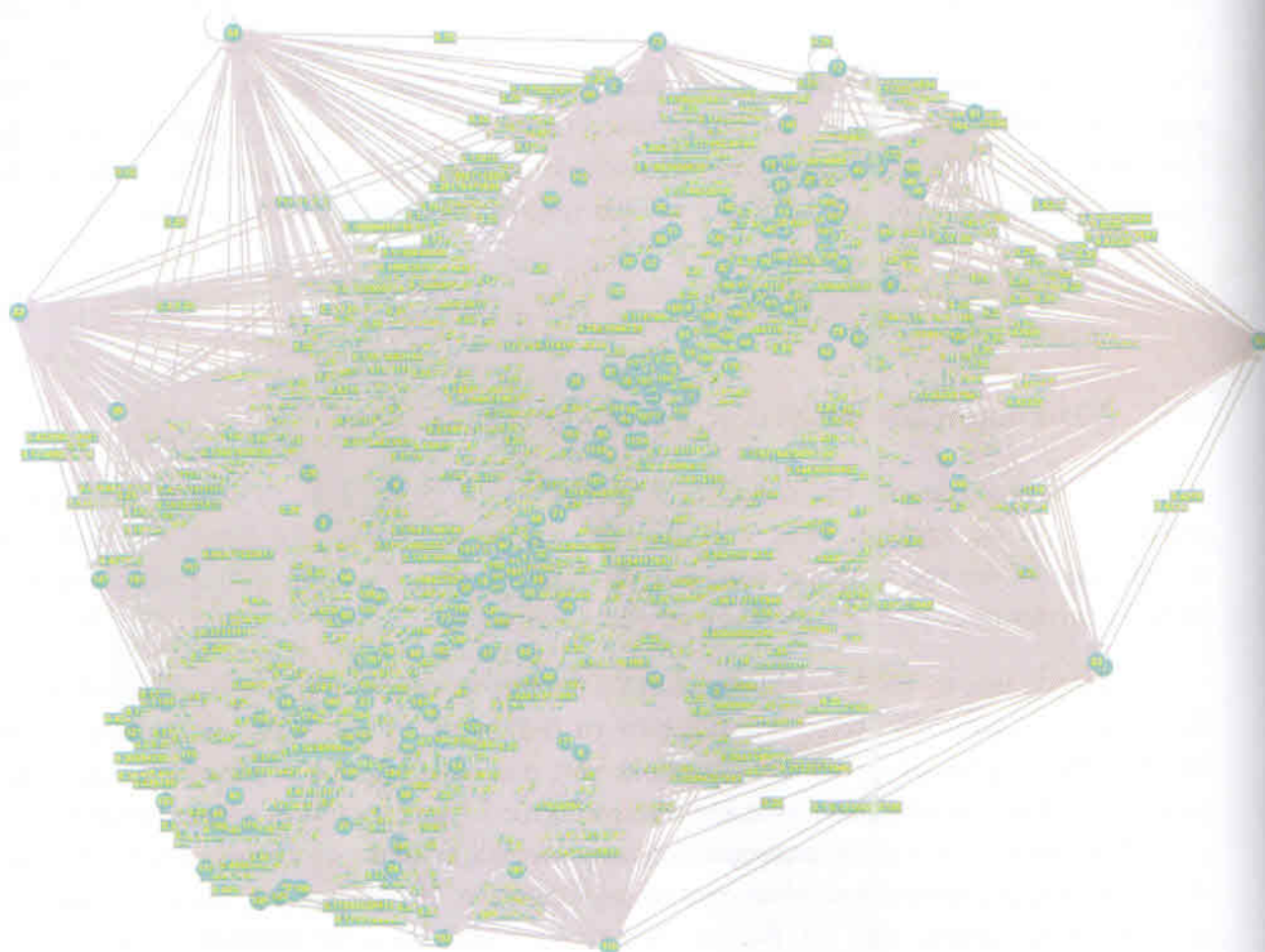
Το τελικό κομμάτι της παρούσας εργασίας αφορά τη δημιουργία ενός συνδυαστικού γράφου, ο οποίος περιέχει τιμές που δίνονται από τους πίνακες γειτνίασης και των δύο προγραμμάτων, και της προσομοίωσης του μοντέλου κίνησης Μανχάταν και της έκδοσης που αφορά στην προσομοίωση του συνόλου δεδομένων Erinions.

Στόχος μας σ' αυτό το σημείο είναι να βρούμε τη σχέση ανάμεσα στο φυσικό οδικό δίκτυο και στο εικονικό δίκτυο, προκειμένου να βγάλουμε ένα συνολικό αποτέλεσμα για τη διάδοση της πληροφορίας. Όπως αναφέρθηκε ήδη, στο online δίκτυο οι χρήστες μοιράζονται μεταξύ τους πληροφορίες μέσω ενός κοινωνικού δικτύου. Όμως, επειδή οι σχέσεις σύνδεσης σε μια online κοινωνία αυτής της μορφής αλλάζουν με αργούς ρυθμούς, το «κοινό» που καλύπτεται στις περισσότερες περιπτώσεις είναι σχετικά περιορισμένο. Η ενσωμάτωση, επομένως, και της φυσικής πληροφορίας από το δίκτυο οχημάτων συμβάλλει στην κάλυψη περισσότερων χρηστών. Μ' αυτό τον τρόπο, μάλιστα, αυξάνεται και η αποδοτικότητα του προτεινόμενου αλγορίθμου.

Για την υλοποίηση του συγκεκριμένου προγράμματος, λάβαμε ως είσοδο αφενός τον πίνακα γειτνίασης που βγαίνει ως αποτέλεσμα του offline προγράμματος για τα πρώτα 50 δευτερόλεπτα της προσομοίωσης, και αφετέρου το πίνακα γειτνίασης που βγαίνει ως αποτέλεσμα του online προγράμματος. Τα αρχεία αυτά είναι τα adjacency_offline.txt και adjacency_online.txt, αντίστοιχα. Επιπλέον, επειδή και οι δύο πίνακες που προκύπτουν είναι 200×200, ταιριάζουν δηλαδή οι διαστάσεις τους, είναι εφικτό να πούμε ότι για κάθε τέτοια θέση του τελικού 200×200 πίνακα θα έχουμε τα συνδυαστικά βάρη των δύο προγραμμάτων, σύμφωνα με τον τύπο $0.5 \times \text{offline_μετρική} + 0.5 \times \text{online_μετρική}$ (3).

Ως offline_μετρική ορίζεται το αποτέλεσμα που προκύπτει από το φυσικό κόσμο και εμπεριέχεται στον αντίστοιχο πίνακα γειτνίασης με τον τύπο (1). Ομοίως, η online_μετρική λαμβάνεται όπως υπολογίστηκε για τον εικονικό κόσμο, σύμφωνα με τον τύπο (2).

Η γραφική απεικόνιση, λοιπόν, του συνδυαστικού αυτού αποτελέσματος έχει ως εξής:



Εικόνα 13. Συνδυαστική γραφική απεικόνιση των αποτελεσμάτων των δύο δικτύων.



Εικόνα 14. Καθαρότερη εικόνα της γραφικής απεικόνισης του συνδυαστικού μοντέλου.

adjacency_avg.txt										
1	0.0000000000	0.2500000000	0.17896530612	0.0000000000	0.0000000000	0.2500000000	0.0000000000	0.17250000000	0.2500000000	0.17250000000
2	0.2500000000	0.0000000000	0.2500000000	0.0000000000	0.2500000000	0.2500000000	0.2500000000	0.0000000000	0.2500000000	0.0000000000
3	0.18770648259	0.2500000000	0.0000000000	0.0000000000	0.2500000000	0.0000000000	0.2500000000	0.0000000000	0.13620588235	0.0000000000
4	0.2500000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.2500000000	0.0000000000	0.41823529412	0.0000000000	0.0000000000
5	0.0000000000	0.2500000000	0.2500000000	0.0000000000	0.0000000000	0.2500000000	0.2500000000	0.0000000000	0.0000000000	0.0000000000
6	0.2500000000	0.2500000000	0.0000000000	0.0000000000	0.2500000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.2500000000
7	0.0000000000	0.2500000000	0.2500000000	0.17031512605	0.2500000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.2500000000
8	0.17160000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.17160000000
9	0.2500000000	0.0000000000	0.12887254902	0.0000000000	0.0000000000	0.0000000000	0.2500000000	0.0000000000	0.0000000000	0.0000000000
10	0.17160000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.2500000000	0.0000000000	0.17160000000	0.0000000000	0.0000000000
11	0.12941176471	0.0000000000	0.16823529412	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.41176470588	-0.07471176471	0.41176470588
12	0.2500000000	0.42769230769	0.2500000000	0.0000000000	0.2500000000	0.0000000000	0.2500000000	0.0000000000	-0.25000000000	0.0000000000
13	0.0000000000	0.0000000000	0.13972426471	0.0000000000	0.0000000000	0.2500000000	0.0000000000	0.25000000000	0.17531250000	0.0000000000
14	0.2500000000	0.2500000000	0.0000000000	0.0000000000	0.0000000000	0.17111111111	0.2500000000	0.0000000000	0.0000000000	0.0000000000
15	0.13601176471	0.2500000000	0.0000000000	0.0000000000	0.2500000000	0.2500000000	0.2500000000	0.42150000000	0.0000000000	0.42150000000

Εικόνα 17. Αποτελέσματα πίνακα average adjacency.

5. Συμπεράσματα

Πολλές σύγχρονες μελέτες, που αφορούν τη σχέση μεταξύ ενός δικτύου φυσικού κόσμου και ενός εικονικού, έδειξαν ότι οι δύο ετερογενείς αυτές περιοχές ενδιαφέροντος μπορούν να συνδεθούν μεταξύ τους. Η ανακάλυψη αυτή θεωρείται ιδιαίτερα σημαντική, καθώς μέσω αυτής της αλληλεπίδρασης γίνεται αντιληπτή μια πολύπλοκη σύνδεση και αλληλεπίδραση μεταξύ των ιδιοτήτων των δύο αυτών κοινωνιών, ενώ παράλληλα μπορούν να ταυτοποιηθούν ευκολότερα τα μοντέλα συμπεριφοράς των χρηστών, με την ανάλυση των δεδομένων που συλλέγονται από τέτοια παρόμοια πολυκοινωνικά περιβάλλοντα. Αντίστοιχα, κάτι τέτοιο υλοποιήθηκε και στη συγκεκριμένη εργασία, φανερώνοντας πώς μπορεί να υφίσταται σχέση εξάρτησης ανάμεσα σ' ένα φυσικό και ένα εικονικό δίκτυο. Απ' τη μία μεριά, έχουμε δεδομένα που αφορούν στη γεωγραφική τοποθεσία των χρηστών ενός δικτύου, ενώ απ' την άλλη μεριά, έχουμε μια συλλογή δεδομένων που αφορούν στις σχέσεις μεταξύ χρηστών ενός online κοινωνικού δικτύου. Ο συνδυασμός των δύο αυτών πληροφοριών μπορεί να οδηγήσει σε βελτίωση της δικτυακής λειτουργίας, καθώς επίσης και σε ευκολότερη και ταχύτερη εύρεση φίλων, από πλευράς χρηστών, αλλά και σε παροχή εξατομικευμένων υπηρεσιών από πλευράς παρόχων. Αυτό οφείλεται στην ύπαρξη μεγαλύτερης πληρότητας του αλγορίθμου πρόβλεψης, καθώς αυτός περιλαμβάνει δεδομένα που αφορούν το κοντινό φυσικό γεωγραφικό περιβάλλον του χρήστη, καθώς επίσης και μια λίστα εικονικών του φίλων και δημοσιεύσεων που του αρέσουν.

6. Μελλοντική Αξιοποίηση

Η παρούσα Διπλωματική Εργασία προσφέρει περιθώρια μελλοντικής αξιοποίησης αναφορικά με την πρόβλεψη της επόμενης κίνησης των χρηστών ενός τηλεπικοινωνιακού δικτύου. Με περαιτέρω εργασία πάνω στο κομμάτι της εύρεσης του χρόνου που μεσολαβεί μεταξύ των διαδοχικών συναντήσεων δύο κόμβων, θα είναι δυνατή και η πρόβλεψη και άρα ο καθορισμός της επόμενης κίνησης του κόμβου. Θα μπορεί, δηλαδή, ο εκάστοτε φορέας που χρησιμοποιεί το πρόγραμμα να προβλέψει, για τον εκάστοτε χρήστη του δικτύου που μελετά σε κάθε περίπτωση, με ποιους από τους υπόλοιπους χρήστες του δικτύου έχει τις περισσότερες πιθανότητες να έρθει σε επαφή και πότε.

Αντίστοιχα, το δεύτερο κομμάτι της υλοποίησης που αφορά σε μια απλουστευμένη έκδοση ενός μηχανισμού φήμης, μπορεί να συνεχιστεί με την ολοκλήρωσή της, ώστε ο μηχανισμός αυτός να μπορεί να δημιουργεί για τον κάθε χρήστη του προτάσεις φίλων και περιεχομένου.

Βιβλιογραφία

- [1] S. Kanoje, S. Girase και D. Mukhopadhyay, «User Profiling Trends, Techniques and Applications,» *International Journal of Advance Foundation and Research in Computer (IJAFRC)*, τόμ. 1, αρ. 1, January 2014.
- [2] J. Tang, L. Yao, D. Zhang και J. Zhang, «A Combination Approach to Web User Profiling,» *Transactions on Knowledge Discovery from Data*, τόμ. 5, pp. 54-88, December 2010.
- [3] N. R. S. D. C. D. R. Stuart E. Middleton, «Ontological User Profiling in Recommender Systems,» *ACM Transactions on Information Systems (TOIS)*, pp. 54-88, 2004.
- [4] M. Ouanaïm και al., «Dynamic User Profiling Approach for Services Discovery in Mobile Environments,» *In proceedings of the 6th IWC MC Conference*, pp. 550-554, 2010.
- [5] M. Caruso και al., «User Profiling and Micro-accounting for Smart Energy Management,» *In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, p. 42, 2013.
- [6] Bradley και al., «Case-based User Profiling for Content Personalisation,» *In Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 62-72, 2000.
- [7] A. B. B. Martinez, M. R. Lopez, E. C. Mantenegro, J. C. Brquillo, F. A. M. Fonte και A. Peleteiro, «A hybrid contentbased and item-based collaborative filtering to recommend TV programs enhanced with singular value decomposition,» *Elsevier Information Sciences: an International Journal*, pp. 4290-4311.
- [8] D. Irani, S. Webb και C. Pu, «Study of static classification of social spam profiles in MySpace,» *International Conference on Weblogs and Social Media*, pp. 82-89.
- [9] W. Paireekreng και K. W. Wong, «Client-side mobile user profile for content management using data mining techniques,» *International Symposium on Natural Language Processing*, pp. 96-100.
- [10] G. Araniti, P. De Meo, A. Iclera και D. Ursino, «Adaptively Controlling the QoS of Multimedia Wireless Applications Through “User Profiling” Techniques,» *Ieee Journal on Selected Areas in Telecommunications*, τόμ. 21, December 2003.
- [11] M. Papadopouli, E. Raftopoulos και H. Shen, «Evaluation of short-term traffic forecasting algorithms in wireless networks».

- [12] W. Wanalertlak, B. Lee, C. Yu, M. Kim και S.-M. Park, «Behavior-based Mobility Prediction for Seamless Handoffs in Mobile Wireless Networks,» *Electrical Engineering & Computer Science Faculty Publications*, 2011.
- [13] Wikipedia, «Hub (network science)».
- [14] D. Knoke και S. Yang, *Social Network Analysis*, Second Edition, τόμ. 154, SAGE Publications, 2008.
- [15] affiliation, «Social Network Analysis,» αρ. 6.99A.
- [16] L. Jin, Y. Chen, T. Wang, P. Hui και A. V. Vasilakos, «Understanding User Behavior in Online Social Networks: A Survey,» *IEEE Communication Magazine*, September 2013.
- [17] N. Mantas, M. Louta, K. Katsaros και S. Kraounakis, «Social CLWPR: A Socially enhanced Position based Routing Protocol for handling misbehavior in VANETs.»
- [18] M. Eirinaki, M. Louta και I. Varlamis, «A Trust-Aware System for Personalized User Recommendations in Social Networks.» *IEEE Transactions on Systems, Man and Cybernetics: Systems*.
- [19] B. Guo, Z. Yu, D. Zhang και X. Zhou, «Cross-Community Sensing and Mining (CSM)» *IEEE Communications Magazine*.
- [20] N. Eagle και al., «Inferring Social Network Structure using Mobile Phone Data,» *PNAS*, τόμ. 106, pp. 15274-15278, 2007.
- [21] B. Guo και al., «Hybrid SN: Interlinking Opportunistic and Online Communities to Augment Information Dissemination,» *Proc. of UIC'12 Conf.* pp. 188-195, 2012.
- [22] X. Zuo και al., «Connecting People at a Conference: A Study of Influence Between Offline and Online Using a Mobile Social Application,» *Proc. IEEE CPSCOM'12 Conf.*, pp. 277-284, 2012.
- [23] E. Cho, S. A. Myers και J. Leskovec, «Friendship and Mobility: user movement in location-based social networks,» *Proc. of KDD'11 Conf.*, pp. 1082-1090, 2011.
- [24] X. Liu και al., «Event-based social networks: linking the online and offline social worlds,» *Proc. of KDD'12 Conf.*, pp. 1032-1040, 2012.
- [25] B. Guo και al., «MemPhone: From Personal Memory Aid to Community Memory Sharing using Mobile Tagging,» *Proc. of PerCom Workshops*, 2013.
- [26] I. O. Nunes, C. Celes, I. Nunes, P. O. S. Vaz de Melo και A. A. F. Loureiro, «Combining Spatial and Social Awareness in D2D Opportunistic Routing.» *IEEE Communication Magazine*.

- [27] T.-C. Tsai και H.-H. Chan, «NCCU Trace: Social-network-aware Mobility,» *IEEE Communications Magazine*, τόμ. 53, αρ. 10, pp. 144-149, 2015.
- [28] S. Kosta, A. Mei και J. Stefa, «Large-Scale Synthetic Social Mobile Networks with SWIM,» *IEEE Transactions on Mobile Computing*, τόμ. 13, αρ. 1, pp. 116-129, 2014.
- [29] P. Hui, J. Crowcroft και E. Yoneki, «Bubble Rap: Social-based Forwarding in Delay-Tolerant Networks,» *IEEE Transactions on Mobile Computing*, τόμ. 10, αρ. 11, pp. 1576-1589, 2011.
- [30] M. De Domenico, A. Lima και M. Musolesi, «Interdependence and Predictability of Human Mobility and Social Interactions».
- [31] «Space, Time and Ambient Intelligence,» *International Joint on Artificial Intelligence*, July 2011.
- [32] G. Groh, A. Lehmann, J. Reimers, R. Friess και L. Schwarz, «Detecting social situations from interaction geometry,» *Proc. IEEE SocialCom 2010*, 2010.
- [33] A. Noulas, S. Scellato, N. Lathia και C. Mascolo, «Mining User Mobility Features for Next Place Prediction in Location-based Services».
- [34] C. Francesco, G. Di Lorenzo και C. Ratti, «Human Mobility Prediction Based on Individual and Collective Geographical Preferences,» *13th International IEEE Conference on Intelligent Transportation Systems*, September 2010.
- [35] H. Leo Chen, «User Clustering and Traffic Prediction in a Trunked Radio System,» 2005.
- [36] E. Cho, S. Myers και J. Leskovec, «Friendship and Mobility: User Movement in Location-Based Social Networks,» *Proc. KDD*, 2011.
- [37] S. Ioannidis, A. Chaintreau και L. Massoulie, «Optimal and Scalable Distribution of Content Updates over a Mobile Social Network,» *Proc. INFOCOM*, 2009.
- [38] H. Chen και L. Trajkovic, «Trunked Radio Systems: Traffic Prediction Based on User Clusters,» *Ieee Xplore*, October 2004.
- [39] Y. Hijikata, «Implicit User Profiling for on demand Relevance Feedback,» *9th International Conference on Intelligent User Interfaces*, pp. 198-205, 2004.
- [40] S. Moghaddam και A. Helmy, «Multi-modal Mining and Modeling of Big Mobile Networks Based on Users Behavior and Interest».