

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΗΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

Τμήμα Πληροφορικής



Πτυχιακή Εργασία

Μελέτη Τεχνολογιών Έξυπνης Διαφήμισης
Στα Κοινωνικά Δίκτυα

Ονοματεπώνυμο: Νικόλαος Καλλονάς
Αριθμός Μητρώου: Π2013088
Επιβλέπων: Αν. Καθηγητής Μυλωνάς Φοίβος

Σεπτέμβριος 2021

Επιβλέπων Καθηγητής

Φοίβος Μυλωνάς, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Φοίβος Μυλωνάς, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Θεόδωρος Ανδρόνικος, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Παναγιώτης Κουρουθανάσης, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Περίληψη

Ο 21ος αιώνας είναι άρρηκτα συνυφασμένος με την άνθηση των Μέσων Κοινωνικής Δικτύωσης (ΜΚΔ) και την ενσωμάτωσή τους στην καθημερινή ζωή των χρηστών, ανεξαρτήτως φύλου, ηλικίας και καταγωγής. Από επιστήμονες της ψυχολογίας συγκαταλέγονται πλέον στις καθημερινές ανθρώπινες ανάγκες, ενώ αποτελούν σημαντικό παράγοντα διαμόρφωσης της συλλογικής ψυχολογίας, με δεδομένη τη μαζική επίδρασή τους στο σύγχρονο χρήστη.

Κατ' επέκτασιν, η ανάμειξη των ΜΚΔ στην επίτευξη διαφημιστικών σκοπών μπορεί να θεωρηθεί αναπόφευκτη, τόσο χάριν στη δημοφιλία τους και τον ολοένα αυξανόμενο αριθμό των χρηστών τους, όσο και στην αποδεδειγμένη αποτελεσματικότητά τους στην προώθηση και αγορά προϊόντων και υπηρεσιών. Τρανό παράδειγμα αποτελεί το γεγονός ότι όλο και περισσότερες επιχειρήσεις διατηρούν προφίλ σε περισσότερο από ένα ΜΚΔ, ενώ τα προτιμούν, ώστε να αυξήσουν τις πωλήσεις τους μέσω της προσωποποιημένης προβολής και προώθησης αγαθών.

Με στόχο την μεγιστοποίηση του πλήθους των προβολών των προϊόντων, υπηρεσιών και κάθε είδους αντικειμένων που διαφημίζονται σε ένα ΜΚΔ, αναπόσπαστο κομμάτι της υποδομής μιας τέτοιας πλατφόρμας μπορεί να θεωρηθεί ένα Recommendation System (RS). Άλλωστε, τα RS έχουν γνωρίσει μεγάλη άνθηση τις τελευταίες δεκαετίες, αφού εφαρμόζονται σε πάσης φύσεως ψηφιακές εμπορικές εφαρμογές, με χαρακτηριστικό παράδειγμα το Spotify, το Netflix, κ.ά. Πραγματοποιώντας μία σύντομη βιβλιογραφική ανασκόπηση και παρουσιάζοντας τα αποτελέσματα αυτής, φαίνεται πως τα παραδοσιακά συστήματα RS λαμβάνουν υπόψιν ένα σχετικά περιορισμένο πλήθος χαρακτηριστικών που αφορούν τους χρήστες και τις αξιολογήσεις τους προς αντικείμενα, τα οποία αξιοποιούνται για τη διαμόρφωση ενός μηχανισμού πρόβλεψης μελλοντικών αξιολογήσεων.

Παρά ταύτα, τα ΜΚΔ είναι ένα ιδιαίτερο περιβάλλον, αφού υπάρχει ένας τεράστιος όγκος πληροφορίας που διαδίδεται συνεχώς χάριν στις αλληπάλληλες αλληλεπιδράσεις μεταξύ των χρηστών. Τα παραδοσιακά RS δεν είναι σε θέση να αξιοποιήσουν τέτοιου είδους πληροφορία, με αποτέλεσμα να υστερούν σε σχέση με άλλες προσεγγίσεις στον τομέα της εξαγωγής των εξατομικευμένων συστάσεων. Για το λόγο αυτό, εντοπίζονται και παρουσιάζονται διάφορες προσεγγίσεις που έχουν προταθεί κατά καιρούς και επεκτείνουν την πληροφορία που λαμβάνεται υπόψη από ένα RS, ώστε να αντικατοπτρίζονται καλύτερα οι σύνθετες σχέσεις μεταξύ των χρηστών.

Έτσι και η παρούσα πτυχιακή εργασία εστιάζει στην εφαρμογή των RS στο χώρο της διαφήμισης στα ΜΚΔ. Για την επίτευξη αυτού, προτείνεται μία προσέγγιση που βασίζεται στην έννοια της εμπιστοσύνης, η οποία εκφράζεται και διαμορφώνεται σε κάποιο βαθμό μέσω των σχέσεων μεταξύ των χρηστών. Η εμπιστοσύνη είναι ένα χαρακτηριστικό των ανθρώπινων

σχέσεων που επηρεάζει σε μεγάλο βαθμό τις επιλογές και τις απόψεις των ατόμων και για αυτό εδώ χρησιμοποιείται για την αποδοτικότερη πρόβλεψη των προτιμήσεων ενός ατόμου. Το χαρακτηριστικό αυτό μπορεί να προκύπτει είτε άμεσα είτε έμμεσα από τις αλληλεπιδράσεις των χρηστών.

Το προτεινόμενο RS αξιολογείται πειραματικά χρησιμοποιώντας ένα ιδιαίτερα διαδομένο σύνολο δεδομένων, ενώ συγκρίνεται με δύο μεθοδολογίες που συναντώνται στη βιβλιογραφία. Κατά την αξιολόγηση χρησιμοποιείται η τεχνική cross validation για την εξαγωγή πιο αξιόπιστων συμπερασμάτων. Η αξιολόγηση εστιάζει στην ακρίβεια των RS να προβλέψουν τις αξιολογήσεις των αντικειμένων από χρήστες. Από την αξιολόγηση προκύπτει πως το προτεινόμενο RS είναι σε θέση να παράξει αρκετά ακριβείς προβλέψεις, ξεπερνώντας τη μία εκ των δύο προηγούμενων μεθοδολογιών.

Συνεπώς, συμπεραίνεται πως η προτεινόμενη προσέγγιση που ενσωματώνει το στοιχείο της εμπιστοσύνης μεταξύ των χρηστών για την πρόβλεψη αξιολογήσεων είναι σε θέση να εξάγει αρκετά ακριβείς προβλέψεις για μία πλατφόρμα ΜΚΔ. Ταυτόχρονα, εντοπίζονται ελλείψεις και περιορισμοί στην προσέγγιση αυτή που αναλύονται εκτενώς, με στόχο την καταγραφή ιδεών για την πιθανή μελλοντική βελτίωση του συστήματος.

Abstract

The 21st century is inextricably linked to the flourishing of social media and their integration into the daily lives of users, regardless of gender, age and origin. Psychological studies have shown that social media may be considered a part of daily human needs, while they are an important factor in shaping group intelligence, given their massive impact on the modern user.

Consequently, the involvement of social media in the achievement of advertising purposes can be considered inevitable, both because of their popularity and the growing number of their users, as well as their proven effectiveness in promoting the purchasing of products and services. A clear indication of this is the fact that more and more companies maintain profiles in more than one social media platforms, while preferring them, in order to increase their sales through personalized promotion and spotlighting of goods.

In order to maximize the number of views of products, services and all kinds of objects that are advertised in social media, a Recommendation System (RS) can be considered an integral part of the infrastructure of such a platform. After all, RS have flourished in recent decades, as they are applied to all kinds of digital commercial applications, with the typical examples being Spotify, Netflix, etc. In a brief literature review, the results of which are presented, it appears that traditional RS take into account a relatively limited number of features concerning users and their ratings of objects, which are used to form a mechanism for predicting future ratings.

Nevertheless, social media is a challenging environment, as there is a huge amount of information that is constantly disseminated due to the non-stop interactions between users. Traditional RS are not able to utilize this kind of information, so they lag behind other approaches to generating personalized recommendations. For this reason, various approaches that have been proposed in the literature are identified and presented, which extend the information taken into account by an RS, in order to better reflect the complexity of relationships between users

Thus, the present thesis focuses on the application of RS in the field of advertising in social media. To achieve this, an approach based on the concept of trust is proposed, which is expressed and shaped to some extent through the relationships between users. Trust is a characteristic of human relationships that greatly influences the choices and opinions of individuals and that is why it is used here to more effectively predict an individual's preferences. This feature can occur either directly or indirectly from user interactions.

The proposed RS is evaluated experimentally using a very commonly used dataset, while it is compared with two methodologies found in the literature. The evaluation uses the cross validation technique to draw more reliable conclusions. The rating focuses on RS accuracy in

predicting user ratings. The evaluation shows that the proposed RS is able to produce fairly accurate predictions, surpassing one of the two previously proposed methodologies.

Therefore, it is concluded that the proposed approach that incorporates the element of trust between users for predicting ratings is able to derive fairly accurate forecasts for a social media platform. At the same time, shortcomings and limitations are identified in this approach that are extensively analyzed, in order to provide directions for the future improvement of the system.

Ευχαριστίες

Με την πραγματοποίηση της πτυχιακής μου εργασίας, ολοκληρώνεται ουσιαστικά ο κύκλος των σπουδών μου στο Ιόνιο Πανεπιστήμιο. Θα ήθελα λοιπόν, να ευχαριστήσω όσους με βοήθησαν, καθ'όλη την διάρκεια της σταδιοδρομίας μου.

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κύριο Φοίβο Μυλωνά, που με τις συμβουλές του, την καθοδήγηση του, και την άριστη και άμεση επικοινωνία, με βοήθησε να φέρω εις πέρας αυτήν την εργασία. Επίσης θέλω να ευχαριστήσω, τους φίλους μου, τους εργοδότες και τους συναδέλφους μου, για την υπομονή και την κατανόηση που έδειξαν στο πρόσωπό μου, καθ'όλη την διάρκεια της υλοποίησης της εργασίας. Τέλος θέλω να ευχαριστήσω την οικογένεια μου για την υπομονή που έδειξε και την στήριξη που μου παρείχε όλα αυτά τα χρόνια.

Περιεχόμενα

1. Εισαγωγή	13
1.1 Συνεισφορά Εργασίας	16
1.2 Δομή Εργασίας	18
2. Συστήματα Συστάσεων	20
2.1 Τύποι Συστημάτων Συστάσεων	20
2.1.1 Collaborative Filtering	21
2.1.2 Content-based	21
2.1.3 Demographic-based	22
2.1.4 Knowledge-based	22
2.1.5 Utility-based	23
2.1.6 Hybrid	23
2.2 Εφαρμογές	24
2.3 Μετρικές Αξιολόγησης	27
3. Συστήματα Συστάσεων ως Μέσα Διαφήμισης στα Μέσα Κοινωνικής Δικτύωσης	31
4. Ανάπτυξη Συστήματος Συστάσεων για Διαφήμιση στα Μέσα Κοινωνικής Δικτύωσης	34
4.1 Μεθοδολογία Ανάπτυξης Συστήματος Συστάσεων	36
5. Αξιολόγηση Συστήματος Συστάσεων	39
5.1 Σύνολο Δεδομένων	39
5.2 Υλοποίηση Συστήματος Συστάσεων	41
5.3 Μετρικές Αξιολόγησης Προτεινόμενου Συστήματος	42
5.4 Αποτελέσματα	43
6. Συμπεράσματα	47
Βιβλιογραφία	49
Παραρτήματα	53

Κατάλογος Σχημάτων

4.1 Το πλήθος των ratings ανά αντικείμενο.....	41
4.2 Το πλήθος των ratings ανά χρήστη.....	42
4.3 Η τιμή του MAE για τον Trust-Based Algorithm συναρτήσει της παραμέτρου διάχυσης	46
4.4 Η τιμή του RMSE για τον Trust-Based Algorithm συναρτήσει της παραμέτρου διάχυσης p.....	47

Κατάλογος Εικόνων

1.1 Γράφημα ενηλίκων χρηστών σε σχέση με το σύνολο των χρηστών του διαδικτύου στο διάστημα 2005-2015.....	14
1.2 Ερωτηματολόγιο έρευνας και απαντήσεις για τα ΜΚΔ σε πανεπιστήμιο των Η.Π.Α.....	16
3.1 Μαθηματικός τύπος MSE.....	29
3.2 Προτάσεις συστήματος συστάσεων και αποτελέσματα μετρικής της ακρίβειας.....	29
3.3 Σύνολο σχετικών προτάσεων αντικειμένων σχετικών με τις επιθυμίες των χρηστών σε σχέση με τις επιλεγμένες προτάσεις.....	30
3.4 Μαθηματικός τύπος μέσης ακρίβειας.....	30

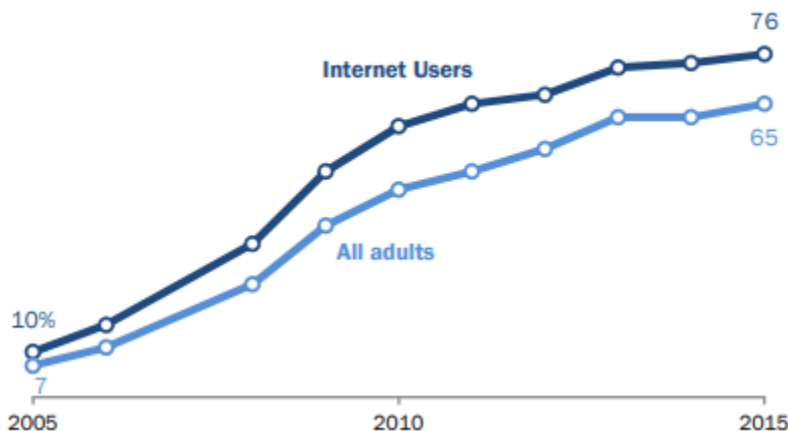
Κατάλογος Πινάκων

4.1 Η απόδοση των αλγορίθμων για τις μετρικές αξιολόγησης MAE και RMSE.....	45
---	----

1. Εισαγωγή

Τα Μέσα Κοινωνικής Δικτύωσης (στο εξής ΜΚΔ) αποτελούν πλέον ένα αναπόσπαστο κομμάτι της καθημερινότητας του ανθρώπου. Η αρχική χρήση των ΜΚΔ αφορούσε την επικοινωνία των ανθρώπων αλλά και την αλληλεπίδραση τους γενικότερα. Τα τελευταία χρόνια όμως η εξέλιξη τους είναι ραγδαία και η χρήση τους θεωρείται πλέον αναγκαία. Οι χρήστες πλέον δεν χρησιμοποιούν τα social media μόνο για την επικοινωνία τους ή για να μοιράζονται φωτογραφίες, βίντεο και γενικότερα στιγμές αλλά μπορούν μέσω της διαφήμισης σε αυτά να έχουν και κάποιο χρηματικό κέρδος. Στην συνέχεια θα παρουσιάσουμε κάποιες έρευνες που δείχνουν πόσο πολύ επηρεάζουν πλέον τους ανθρώπους τα ΜΚΔ.

Σύμφωνα με μια έρευνα που πραγματοποιήθηκε από τον (Perrin, 2015) στην Αμερική, οι ενήλικες που διαθέτουν λογαριασμό σε τουλάχιστον ένα ΜΚΔ αυξήθηκαν από το 7% που ήταν το ποσοστό το 2005, στο 65%. Τα στοιχεία αυτά συλλέχθηκαν από το “Pew Research Center”¹. Το ποσοστό αυτό συνεχίζει να ανεβαίνει καθώς τα ΜΚΔ γίνονται όλο και πιο επιδραστικά στις ζωές των ανθρώπων. Στην εικόνα 1.1. βλέπουμε την αύξηση που υπάρχει στους ενήλικες χρήστες στο διάστημα μεταξύ 2005-2015.



Εικόνα 1.1: Γράφημα ενηλίκων χρηστών σε σχέση με το σύνολο των χρηστών του διαδικτύου στο διάστημα 2005-2015, Πηγή: Pew Research Center.

Βέβαια η χρήση των ΜΚΔ διαφέρει από χρήστη σε χρήστη καθώς όπως είναι φυσιολογικό δεν έχουν όλοι οι χρήστες τα ίδια ενδιαφέροντα ή δεν ανήκουν στην ίδια ηλικιακή ομάδα κ.α. Το 2013 οι (Whiting & Williams, 2013) πραγματοποίησαν μια έρευνα η οποία αφορά τους λόγους για τους οποίους οι χρήστες χρησιμοποιούν τα ΜΚΔ. Για την πραγματοποίηση αυτής σύλλεξαν στοιχεία και δεδομένα από έρευνες που είχαν γίνει στο παρελθόν από άλλους ερευνητές και

¹ <https://www.pewresearch.org/>

κατέληξαν στους δέκα σημαντικότερους λόγους για τους οποίους οι χρήστες χρησιμοποιούν τα ΜΚΔ. Αναλυτικά:

- Social Interaction - Κοινωνική αλληλεπίδραση
- Information Seeking - Αναζήτηση πληροφοριών
- Pass Time - Ελεύθερο χρόνο
- Entertainment - Διασκέδαση
- Relaxation - Χαλάρωση
- Expression of opinions - Έκφραση γνώμης
- Communicatory utility - Επικοινωνιακή χρησιμότητα
- Convenience utility - Ευκολία χρήσης
- Information sharing - Ανταλλαγή πληροφοριών
- Surveillance/knowledge about others - Επιτήρηση ή γνώση για άλλους

Τα συμπεράσματα που βγήκαν σύμφωνα με τα στοιχεία της παραπάνω έρευνας είναι τα εξής:

- Το 88% χρησιμοποιεί τα ΜΚΔ για την κοινωνική αλληλεπίδραση που τους παρέχουν, καθώς πολύ θεωρούν ότι αυτή είναι η πιο σημαντική χρήση τους, διότι αυτά τους δίνουν την δυνατότητα να διατηρούν τις σχέσεις τους με τους φίλους ή τα μέλη της οικογένειας του που δεν μπορούν να δουν από κοντά
- Το 80% χρησιμοποιεί τα ΜΚΔ για να κάνει αναζητήσεις πληροφοριών που αφορούν για παράδειγμα διάφορες πωλήσεις, προϊόντα, καθώς επίσης για να μάθουν για επερχόμενες εκδηλώσεις ή για να πληροφορηθούν πότε έχει γενέθλια κάποιος φίλος τους. Επίσης σημαντικές είναι και οι αναζητήσεις που αφορούν την εύρεση εργασίας.
- Το 76% τα χρησιμοποιεί για τον ελεύθερο χρόνο τους, δηλαδή όταν δεν έχουν κάτι άλλο να κάνουν, επιλέγουν να περάσουν αυτόν τον χρόνο τους στα ΜΚΔ.
- Το 64% απάντησε ότι διασκεδάζει μέσω των ΜΚΔ και αυτό διότι τα χρησιμοποιεί παρακολουθώντας κάποια βίντεο, εικόνες ή ακόμα και σχόλια που μπορούν να τους κάνουν να γελάσουν και να περάσει ευχάριστα ο χρόνος τους.
- Το 60% απάντησε ότι χαλαρώνει μέσω της χρήσης των ΜΚΔ καθώς κάπως έτσι ξεφεύγουν από την πραγματικότητα και την καθημερινότητα τους.
- Το 56% θεωρεί τα ΜΚΔ ως ένα μέσω έκφρασης των ιδεών τους, καθώς πιστεύουν ότι τους δίνουν το πάτημα να μοιραστούν με άλλους χρήστες τα πιστεύω ή τις γνώσεις τους. Επίσης αλληλεπιδρούν με άλλους χρήστες κατακρίνοντας ή συμφωνώντας με τις ιδέες τους.
- Επίσης το 56% πάλι θεωρεί ότι τα ΜΚΔ έχουν επικοινωνιακή χρησιμότητα καθώς μέσω αυτών ενημερώνονται για διάφορα γεγονότα και αυτό τους δίνει την δυνατότητα να επικοινωνήσουν καλύτερα με άλλους ανθρώπους έχουν πράγματα να μοιραστούν.

- Το 52% πιστεύει ότι χρησιμοποιεί κυρίως τα ΜΚΔ καθώς είναι εύκολη η πρόσβαση σε αυτά οποιαδήποτε στιγμή και σε οποιαδήποτε μέρος και επειδή τους δίνεται η δυνατότητα να μπορούν να επικοινωνούν με πολλούς ανθρώπους ταυτόχρονα
- Το 40% χρησιμοποιεί τα ΜΚΔ για να μπορέσει να προωθήσει τις επιχειρήσεις του ή να διαφημίσει τον ίδιο τον εαυτό του μέσω αυτών.
- Τέλος το 32% χρησιμοποιεί τα ΜΚΔ για να παρακολουθεί άλλους, για παράδειγμα πολλοί γονείς τα χρησιμοποιούν για να μπορούν να παρακολουθήσουν τα παιδιά τους κ.ο.κ.

Σε αυτή την παράγραφο θα ασχοληθούμε με μια έρευνα που διεξήχθη το 2015 σε ένα δημόσιο πανεπιστήμιο στις Η.Π.Α. Η έρευνα αφορούσε έξι διαφορετικά ΜΚΔ και στόχος της έρευνας ήταν να βρεθεί ποιο είναι το ΜΚΔ που χρησιμοποιούν περισσότερο οι φοιτητές. Στην εικόνα 1.2 φαίνονται οι ακριβείς ερωτήσεις του ερωτηματολογίου καθώς και τα συνολικά αποτελέσματα της έρευνας. Αναλυτικά, στην πρώτη θέση των προτιμήσεων βρίσκεται το Instagram με 29%, με δεύτερο να ακολουθεί το Snapchat(24%) και τρίτο το Facebook(23%), στην συνέχεια ακολουθούν το Twitter και το YouTube με 13% και 7% αντίστοιχα, ενώ στην τελευταία θέση βρίσκονται τα LinkedIn και Pinterest με 2%. Τα συμπεράσματα στα οποία κατέληξαν οι ερευνητές έχουν να κάνουν με το γεγονός ότι οι φοιτητές φαίνεται να προτιμούν περισσότερο τα ΜΚΔ που τους δίνουν την δυνατότητα να δημοσιεύσουν εικόνες και βίντεο. Ενώ σε αντίθετη περίπτωση φαίνεται ότι δεν τους ενδιαφέρουν σχεδόν καθόλου τα ΜΚΔ που αφορούν την επαγγελματική τους ζωή και δεν ενδιαφέρονται να προωθήσουν τα επαγγελματικά τους προσόντα στο διαδίκτυο. (Knight-McCord et al., 2016)

<i>Gender</i>			<i>Use social media?</i>		
	Male	59		Yes	97
	Female	41		No	3
<i>Age</i>			<i>Most preferred social media</i>		
	19-23	78		Instagram	29
	24+	22		Snapchat	24
<i>Year in school</i>				Facebook	23
	Freshman	14		Twitter	13
	Sophomore	29		YouTube	7
	Junior	32		LinkedIn	2
	Senior	25		Pinterest	2
<i>Race</i>			<i>Hours per day on social media</i>		
	Black	69		1-5	46
	White	13		6-10	29
	Asian	5		11-15	12
	Hispanic	6		16-20	7
	Other	7		>20	6
All numbers are percents.			<i>Use social media more on weekends?</i>		
				Yes	80
				No	20
All numbers are percents.			All numbers are percents.		

Εικόνα 1.2: Ερωτηματολόγιο έρευνας και απαντήσεις για τα ΜΚΔ σε πανεπιστήμιο των Η.Π.Α., Πηγή: Knight-McCord et al., 2016.

Τον τελευταίο χρόνο οι ζωές όλων των ανθρώπων σε όλων τον κόσμο έχουν αλλάξει λόγω της πανδημίας του Covid-19. Πολλοί άνθρωποι έχασαν τις δουλειές τους, άλλοι έχασαν αγαπημένα

τους συγγενικά πρόσωπα αλλά και φίλους, δεδομένο όμως είναι ότι σχεδόν όλοι χρειάστηκε να περάσουν περισσότερο χρόνο περιορισμένοι στο σπίτι τους, χάνοντας έτσι την φυσική επικοινωνία με τους γύρω τους. Στην συνέχεια θα παραθέσουμε μια έρευνα για την επίδραση που είχε η πανδημία του Covid-19 στα ΜΚΔ και γενικότερα στις διάφορες εφαρμογές λόγω όλων των παραπάνω γεγονότων. Ο Bakkar πραγματοποίησε μια έρευνα στο διάστημα μεταξύ 15 Μαρτίου 2020 μέχρι 31 Μαρτίου 2020, που σχεδόν όλος ο πλανήτης βρισκόταν σε καραντίνα. Σύλλεξε στοιχεία μέσω ερωτηματολογίων από 384 Αιγύπτιους φοιτητές διαφορετικών πανεπιστημίων και τα αποτελέσματα της έρευνας μας δείχνουν ότι σε αυτό το διάστημα υπήρξε ραγδαία αύξηση εγκατάστασης νέων εφαρμογών ΜΚΔ και μεγάλη χρήση αυτών από τους φοιτητές (Bakkar, 2020).

Στην συνέχεια θα αναφέρουμε κάποια δεδομένα για την χρήση των ΜΚΔ ως μέσα διαφήμισης. Όπως έχει αναφερθεί και παραπάνω τα ΜΚΔ έχουν όλα πάρα πολλούς εγγεγραμμένους χρήστες και η εξέλιξη τους είναι ραγδαία. Για αυτόν τον λόγο οι εταιρείες επιλέγουν πλέον τα ΜΚΔ για την διαφήμιση τους καθώς, όλοι αυτοί οι χρήστες είναι πιθανοί μελλοντικοί πελάτες τους. Σύμφωνα με έρευνα που έγινε από τους Appel και συν., τα ΜΚΔ θα είναι η νούμερο ένα επιλογή για διαφήμιση ξεπερνώντας την τηλεόραση και τις υπόλοιπες διαφημιστικές επιλογές των εταιρειών (Appel et al., 2020). Οι λόγοι για τους οποίους συμβαίνει αυτό είναι γιατί οι χρήστες είναι πάρα πολύ εύκολο να επικοινωνήσουν ή να δουν τα προϊόντα ή τις υπηρεσίες τους, καθώς επίσης είναι επίσης εύκολα να στείλουν αυτά και σε άλλους χρήστες μέσω ηλεκτρονικού μηνύματος. Οι εταιρίες λοιπόν τα τελευταία χρόνια επιλέγουν σαν πρόσωπα στις καμπάνιες τους άτομα με μεγάλη επιρροή στα ΜΚΔ καθώς έτσι θα μπορούν όλο και περισσότεροι άνθρωποι να δουν τα προϊόντα τους άρα θα υπάρξει μεγαλύτερος αριθμός υποψήφιων πελατών. Επίσης σύμφωνα με την ίδια έρευνα επειδή τα ΜΚΔ γνωρίζουν αυτήν την αλλαγή κατεύθυνσης των εταιριών και την επιλογή τους για διαφήμιση σε αυτά, προσπαθούν συνεχώς να βελτιώνουν τις υπηρεσίες και τις επιλογές διαφήμισης που παρέχουν στις εταιρίες που τα χρησιμοποιούν.

Τα τελευταία έτη τα συστήματα συστάσεων, γνωστά και ως Recommendation Systems (RS), έχουν γνωρίσει μεγάλη άνθηση στον χώρο της διαφήμισης, καθώς όπως υποδηλώνει και το όνομα τους πραγματοποιούν συστάσεις στους χρήστες. Αποτελούν πλέον ένα πολύ δυνατό μέσο διαφήμισης καθώς αξιοποιούν την ισχύ των ΜΚΔ και κάποια χαρακτηριστικά τους. Οι περισσότερες εταιρείες χρησιμοποιούν συστήματα συστάσεων για την διαφήμιση τους στα ΜΚΔ, καθώς έτσι μπορούν να στοχεύσουν καλύτερα το κοινό τους.

1.1 Συνεισφορά Εργασίας

Η χρήση των ΜΚΔ αποτελεί αναπόσπαστο κομμάτι στην καθημερινότητα των ατόμων και παρουσιάζει μια αυξητική τάση μέρα με τη μέρα σε άτομα πάσης φύσεως ηλικίας, φύλου, φυλής, κ.λπ. Χάρη στην πληθώρα των χρηστών αλλά και στην ποικιλομορφία των

προσωπικοτήτων συναντάται μια αλληπάλληλη και συνεχής ανταλλαγή διαφορετικών απόψεων και αντιδράσεων, καθώς και διαρκής εκδήλωση ενδιαφερόντων και προτιμήσεων (Correa et al., 2010 και O'Keeffe & Clarke-Pearson, 2011).

Ταυτόχρονα, έχει γίνει εμφανής η σημασία του ρόλου των ΜΚΔ στην προώθηση τοπικών και μη επιχειρήσεων, προϊόντων, ιδέων, δραστηριοτήτων, κ.ά. Η προώθηση αυτή μπορεί να γίνει με διάφορους τρόπους, όμως πρέπει να υποστηρίζεται κι από την υποδομή του εκάστοτε ΜΚΔ, με τη μορφή των αυτόματα εξαγόμενων συστάσεων. Οι συστάσεις αυτές συμβάλλουν στο να επεκταθεί περαιτέρω το πλήθος των ατόμων που πιθανώς θα συναντήσουν μία σελίδα ή μία δημοσίευση (Qualman, 2012).

Ο κλάδος των RS έχει αναπτυχθεί, ώστε να καλυφθεί η ανάγκη για εξατομικευμένες συστάσεις που προκύπτει σε διαφόρων ειδών πληροφοριακών συστημάτων τις τελευταίες δεκαετίες, ιδιαίτερα όσων παρουσιάζουν μία εμπορική διάσταση. Τα RS διακρίνονται σε διαφορετικές κατηγορίες και, κατά κόρον, αξιοποιούν ομοιότητες μεταξύ χρηστών και αντικειμένων (οτιδήποτε ενδέχεται αυτά να αναπαριστούν). Οι ομοιότητες αυτές συνήθως προκύπτουν από σχετικά απλές πληροφορίες, όπως είναι οι αξιολογήσεις των χρηστών για ένα αντικείμενο. Έτσι, συχνά διακρίνονται ομοιότητες μεταξύ χρηστών, εφόσον αυτοί αξιολογούν κοινά αντικείμενα, οι οποίες με κατάλληλες διαδικασίες τροφοδοτούν την εξαγωγή συστάσεων.

Παρά την επιτυχία αυτού του είδους των RS, αυτά δυσκολεύονται να ανταπεξέλθουν σε τόσο περίπλοκα περιβάλλοντα, όπως τα ΜΚΔ. Μία ιδιαίτερα σημαντική πρόκληση είναι πως αν το πλήθος των αντικειμένων και των χρηστών είναι αρκετά μεγάλο αλλά και αυξάνεται συνεχώς, τότε θα υπάρχει ένα σημαντικό πλήθος αυτών για τα οποία δε θα υπάρχουν αρκετά δεδομένα για την εξαγωγή συστάσεων. Το πρόβλημα αυτό συναντάται στη βιβλιογραφία ως cold start problem και συχνά αντιμετωπίζεται χρησιμοποιώντας άλλου είδους πληροφορίες, πέρα από χαρακτηριστικά που βασίζονται στο “ιστορικό” ενός χρήστη (π.χ. αξιολογήσεις).

Στην προκειμένη περίπτωση, στα ΜΚΔ υπάρχει μία μεγάλη ποικιλία πηγών από τις οποίες μπορούν να αντληθούν τα απαραίτητα δεδομένα. Συγκεκριμένα, το πιο αναπόσπαστο κομμάτι των ΜΚΔ που είναι αρκετά κοινό, ανεξάρτητα από την εκάστοτε πλατφόρμα, είναι οι σχέσεις που αναπτύσσονται μεταξύ των ατόμων. Η έννοια αυτή συχνά συναντάται ως σχέση φιλίας, “ακολουθού” ή συνεργάτη, αλλά σε κάθε περίπτωση χαρακτηρίζει την επαφή μεταξύ δύο ατόμων. Όπως έχει μελετηθεί στη βιβλιογραφία υπό διαφορετικές σκοπιές, οι σχέσεις μεταξύ ατόμων επηρεάζουν σε κάποιο βαθμό τις απόψεις και τις επιλογές μας (Alleva et al., 2021 και Banerjee, 1992). Έτσι, είναι μία αρκετά ικανή πηγή δεδομένων, η οποία μπορεί να χρησιμοποιηθεί από ένα RS στα πλαίσια ενός ΜΚΔ.

Συνεπώς, στα πλαίσια της παρούσας πτυχιακής εργασίας, προτείνεται και αναπτύσσεται ένα RS, το οποίο αξιοποιεί τις σχέσεις μεταξύ ατόμων σε ένα ΜΚΔ, ώστε να παρέχει εξατομικευμένες

συστάσεις για αντικείμενα. Ένα τέτοιο RS θεωρείται πως θα μπορούσε να προτείνει διαφημίσεις, δημοσιεύσεις και πάσης φύσεως αντικείμενα για προβολή σε μία πλατφόρμα MKΔ. Για την ακρίβεια, το εν λόγω RS υιοθετεί την έννοια της εμπιστοσύνης, η οποία εκφράζεται από τη σχέση μεταξύ δύο χρηστών σε ένα MKΔ και μπορεί να προκύπτει είτε ρητά (explicitly) είτε σιωπηρά (implicitly) από τις αλληλεπιδράσεις τους.

Χρησιμοποιώντας την έννοια της εμπιστοσύνης, διαμορφώνεται ένας τρόπος υπολογισμού ομοιότητας δύο χρηστών αναφορικά με ένα αντικείμενο. Ακόμη, για την αντιμετώπιση του cold start problem, προτείνεται ένας μηχανισμός διάδοσης των σχέσεων εμπιστοσύνης σε χρήστες που δεν είναι άμεσα συνδεδεμένοι μέσω μίας προσέγγισης βασισμένης στη θεωρία γράφων. Οι έμμεσες σχέσεις εμπιστοσύνης που διαμορφώνονται θεωρούνται πιο αδύναμες, καθώς απομακρύνονται από τις άμεσες σχέσεις, γεγονός που εκφράζεται μέσω του προτεινόμενου μηχανισμού διάδοσης. Με τον τρόπο αυτό, επεκτείνεται το πλήθος των ατόμων με κοινά χαρακτηριστικά που μπορεί να αξιοποιηθεί κατά την εξαγωγή συστάσεων.

Το προτεινόμενο RS αξιολογείται πειραματικά, χρησιμοποιώντας το σύνολο δεδομένων Epinions, το οποίο περιέχει ένα σύνολο αξιολογήσεων αντικειμένων από διαφορετικούς χρήστες, αλλά και τις σχέσεις εμπιστοσύνης που διαμορφώνονται μεταξύ αυτών. Το προτεινόμενο RS συγκρίνεται και με δύο κλασικούς αλγόριθμους RS που δεν αξιοποιούν τις σχέσεις εμπιστοσύνης: (α) μία User-Based Collaborative Filtering προσέγγιση που αξιοποιεί τον αλγόριθμο k-Nearest Neighbors και (β) το Singular Value Decomposition. Ο πειραματισμός εστιάζει στην αξιολόγηση της ικανότητας των RS να προβλέπουν με ακρίβεια τις αξιολογήσεις των χρηστών σε αντικείμενα. Τα εξαγόμενα αποτελέσματα έδειξαν πως το προτεινόμενο RS αποδίδει αρκετά καλά και, μάλιστα, λίγο καλύτερα από την Collaborative Filtering προσέγγιση, ενώ υστερεί ελάχιστα σε σχέση με το Singular Value Decomposition. Έτσι, κρίνεται πως το προτεινόμενο RS είναι σε θέση να παρέχει εξατομικευμένες συστάσεις για πλατφόρμες MKΔ με επαρκής ακρίβεια.

1.2 Δομή Εργασίας

Στο δεύτερο Κεφάλαιο αναφέρονται και παρουσιάζονται τα συστήματα συστάσεων. Αρχικά γίνεται μια αναφορά στην ραγδαία αύξηση που υπάρχει στην χρήση των RS, αλλά και στην μεγάλη χρησιμότητα και σχέση που έχουν με τα MKΔ. Στη συνέχεια παρουσιάζονται και αναλύονται ορισμένοι τύποι RS και πως αυτά λειτουργούν, δηλαδή τι στοιχεία χρησιμοποιεί το καθένα από αυτά για να εξάγει συμπεράσματα, τι πλεονεκτήματα και τι μειονεκτήματα υπάρχουν. Επίσης αναφέρονται και κάποιες γνωστές εφαρμογές που χρησιμοποιούν RS, και αναλύεται για ποιον λόγο το κάνουν αυτό, τι θέλουν να πετύχουν και πόσο πολύ χρήσιμα τους είναι πλέον. Τέλος αναφέρονται επίσης και κάποιες κατηγορίες μετρικών και το πως βοηθούν αυτές να γίνει αξιολόγηση των RS, δηλαδή που βασίζεται η αξιολόγηση για κάθε μια από αυτές.

Το τρίτο Κεφάλαιο αποσκοπεί στην μελέτη άλλων ερευνών που έχουν ίδιο στόχο με την δική μας έρευνα. Γίνεται μια παρουσίαση των ερευνών, δηλαδή για ποιον λόγο έγινε η κάθε έρευνα, τι μεθοδολογία και ποιοί αλγόριθμοι χρησιμοποιήθηκαν, ενώ τέλος αναφέρονται και τα συμπεράσματα στα οποία κατέληξε η κάθε έρευνα. Σε ορισμένες περιπτώσεις γίνεται και μια σύγκριση μεταξύ των διαφορετικών αποτελεσμάτων των ερευνών.

Το τέταρτο Κεφάλαιο παρουσιάζει με λεπτομερή τρόπο τη μεθοδολογία ανάπτυξης του προτεινόμενου RS. Για την ακρίβεια, περιγράφεται αναλυτικά η γενική ιδέα στην οποία βασίζεται το RS, δηλαδή στην αξιοποίηση της σχέσης εμπιστοσύνης μεταξύ των χρηστών ενός ΜΚΔ. Στη συνέχεια, εισάγονται οι απαραίτητοι συμβολισμοί, ώστε να εκφραστεί με μαθηματική ακρίβεια αυτή η προσέγγιση και ο τρόπος με τον οποίο γίνονται προβλέψεις των αξιολογήσεων των αντικειμένων από τους χρήστες.

Στο πέμπτο Κεφάλαιο παρουσιάζεται η διαδικασία αξιολόγησης της προτεινόμενου συστήματος, αφού πρωτίστως προηγηθεί η περιγραφή του χρησιμοποιηθέντος συνόλου δεδομένων με τα αντίστοιχα χαρακτηριστικά του. Παράλληλα, αναλύεται και η διαδικασία υλοποίησης του συστήματος συστάσεων με τα απαιτούμενα εργαλεία. Επιπλέον, αναλύονται και οι μετρικές αξιολόγησης που απαιτούνται για την επαλήθευση της ορθότητας και της καταλληλότητας του, προβάλλοντας και τα εξαγόμενα από τον σχετικό πειραματισμό αποτελέσματα. Αξίζει να σημειωθεί πως η απόδοση της προτεινόμενης προσέγγισης συγκρίνεται με άλλες προσεγγίσεις που συναντώνται στη βιβλιογραφία.

Τέλος, στο έκτο Κεφάλαιο αποτυπώνονται τα συμπεράσματα που προκύπτουν από την εν λόγω εργασία, τα οποία σχετίζονται τόσο με τη βιβλιογραφία, όσο και με τη μεθοδολογία ανάπτυξης και αξιολόγησης του προτεινόμενου συστήματος συστάσεων. Επιπροσθέτως, αναφέρονται κάποιες μελλοντικές προεκτάσεις για βελτίωση της απόδοσης και της καταλληλότητας του προαναφερόμενου συστήματος.

2. Συστήματα Συστάσεων

Στο παρόν Κεφάλαιο, θα μελετήσουμε τα συστήματα συστάσεων. Συγκεκριμένα θα παρατεθούν οι τύποι συστημάτων συστάσεων που υπάρχουν, καθώς και πως αυτά εφαρμόζονται στα ΜΚΔ απο τις διάφορες εφαρμογές που υπάρχουν, αλλά και πόσο σημαντικά είναι για τις επιχειρήσεις και την διαφήμιση. Τέλος, θα αναφερθούν και κάποιες μετρικές αξιολόγησης αυτών.

Ο μεγάλος όγκος δεδομένων που υπάρχει πλέον στον χώρο του διαδικτύου είχε ως αποτέλεσμα την ραγδαία αύξηση χρήσης συστημάτων συστάσεων. Τα συστήματα συστάσεων είναι συστήματα τα οποία παρέχουν στους χρήστες εξατομικευμένες προτάσεις σε σχέση με προϊόντα ή υπηρεσίες που μπορεί να τους ενδιαφέρουν (Patty et al., 2018). Τα συστήματα συστάσεων ουσιαστικά ξεκίνησαν να χρησιμοποιούνται για να κάνουν πιο εύκολη την πλοήγηση των χρηστών. Οι εξατομικευμένες πληροφορίες που προσφέρουν βοηθούν και τους χρήστες να βρίσκουν πιο εύκολα τα δεδομένα που αναζητούν αλλά και τις εταιρείες να μπορούν να προωθήσουν τα προϊόντα τους ή τις υπηρεσίες τους πιο άμεσα. Επίσης όσο περισσότερο χρησιμοποιεί ο χρήστης μια εφαρμογή ή μια ιστοσελίδα, τόσο πιο πολλά δεδομένα μαζεύονται, άρα η λειτουργία των συστημάτων συστάσεων γίνεται ακόμα καλύτερη.

Ένα άκρως σημαντικό χαρακτηριστικό που αξιοποιούν πολλά συστήματα συστάσεων είναι οι χρήστες των ΜΚΔ και οι κινήσεις αυτών στο διαδίκτυο, έτσι ώστε να μπορούν να έχουν μια ανατροφοδότηση, γεγονός που θα τα βοηθήσει να κάνουν νέες προτάσεις στους ίδιους τους χρήστες ή στους φίλους τους. Με πιο απλά λόγια, τα συστήματα συστάσεων παρακολουθούν τα like για παράδειγμα που μπορεί να κάνει ένας χρήστης σε κάποιο προϊόν έτσι ώστε να του προτείνουν στην συνέχεια κάποιο άλλο παρόμοιο προϊόν από διαφορετικό προμηθευτή ή επιχείρηση. Η λειτουργία αυτή είναι πολύ σημαντική τόσο για τους χρήστες όσο και για τις εταιρίες, καθώς έτσι και ο χρήστης δέχεται μια βοήθεια στις αγορές του ή στις αναζητήσεις του αλλά και οι εταιρίες έχουν μια καλύτερη στόχευση στα προϊόντα που προωθούν στον κάθε χρήστη.

2.1 Τύποι Συστημάτων Συστάσεων

Στην παρούσα ενότητα θα ασχοληθούμε με τους τύπους συστημάτων συστάσεων. Συγκεκριμένα θα παραθέσουμε τις ιδιότητες και τα διαφορετικά χαρακτηριστικά του κάθε τύπου συστήματος συστάσεων. Αναλυτικά θα ασχοληθούμε με έξι διαφορετικές κατηγορίες συστημάτων συστάσεων:

- Collaborative Filtering - Συστήματα βασισμένα στην συνεργασία
- Content-Based - Συστήματα βασισμένα στο περιεχόμενο

- Demographic Based - Συστήματα βασισμένα σε δημογραφικά δεδομένα
- Knowledge Based - Συστήματα βασισμένα στην γνώση
- Utility Based - Συστήματα βασισμένα στην χρησιμότητα
- Hybrid - Υβριδικά συστήματα

2.1.1 Collaborative Filtering

Το Collaborative σύστημα συστάσεων είναι ένα σύστημα το οποίο θα προτείνει στον χρήστη αντικείμενα ή υπηρεσίες τις οποίες δείχνουν να προτιμούν άλλοι χρήστες με παρόμοιες προτιμήσεις με αυτόν (Adomavicius & Tuzhilin, 2015)

Η πρώτη κατηγορία συστήματος συστάσεων που θα παρουσιάσουμε είναι το Collaborative Filtering (CF). Το CF χρησιμοποιεί την συνεργατική δύναμη των βαθμολογιών που δίνονται από πολλούς χρήστες έτσι ώστε να κάνει προτάσεις. Το μεγαλύτερο πρόβλημα του συγκεκριμένου συστήματος είναι ότι οι βαθμολογίες που δίνονται είναι αραιές, λόγω των πολλών προϊόντων που μπορούν να υπάρχουν, μπορεί δηλαδή να υπάρχουν προϊόντα που δεν έχουν βαθμολογηθεί από πολλούς χρήστες έτσι ώστε να βρεθούν κοινά σημεία για να γίνουν προτάσεις. Για να λυθεί αυτό το πρόβλημα θεωρούμε ότι εφόσον για παράδειγμα δυο χρήστες έχουν πολλές κοινές βαθμολογήσεις, τότε είναι πολύ πιθανόν και κάποια προϊόντα που έχουν βαθμολογηθεί μόνο από τον έναν εκ των δυο, να αρέσουν και στον άλλον.

2.1.2 Content-based

Το Content-based σύστημα συστάσεων είναι ένα σύστημα στο οποίο στους χρήστες θα προταθούν προϊόντα ή υπηρεσίες παρόμοια με αυτά που έχουν δείξει να προτιμούν οι ίδιοι στο παρελθόν (Adomavicius & Tuzhilin, 2015).

Αυτό σύστημα συστάσεων επικεντρώνεται στα περιγραφικά χαρακτηριστικά του αντικειμένου έτσι ώστε να κάνει προτάσεις, δίνει δηλαδή βαρύτητα στις βαθμολογίες και τις επιλογές ενός χρήστη έτσι ώστε να προβεί σε περαιτέρω προτάσεις σε αυτόν. Το συγκεκριμένο σύστημα συστάσεων είναι πολύ θετικό για νέα αντικείμενα που δεν έχουν βαθμολογηθεί ακόμα αλλά μπορούν προταθούν σε έναν χρήστη καθώς αυτός έχει βαθμολογήσει παρόμοια αντικείμενα με κοινά χαρακτηριστικά στο παρελθόν. Από την άλλη πλευρά όμως το παρόν σύστημα έχει και δύο πολύ σημαντικά μειονεκτήματα. Πρώτον όσον αφορά του χρήστες που υπάρχουν ήδη στο σύστημα οι προτάσεις που μπορεί να κάνει είναι πολύ συγκεκριμένες καθώς χρησιμοποιεί μόνο την συμπεριφορά των χρηστών, έτσι αντικείμενα με διαφορετικά χαρακτηριστικά από αυτά που έχει ήδη επιλέξει ο χρήστης είναι πολύ δύσκολο να προταθούν. Επίσης το δεύτερο και πολύ σημαντικό μειονέκτημα που έχει αφορά τους νέους χρήστες, καθώς δεν έχει σχεδόν καμία χρησιμότητα σε αυτούς, γιατί εάν αυτοί δεν έχουν καμία αλληλεπίδραση με άλλα αντικείμενα

στο παρελθόν, τότε αυτό έχει ως αποτέλεσμα να μην μπορούν να τους γίνουν προτάσεις για άλλα αντικείμενα. (Adomavicius & Tuzhilin, 2015)

2.1.3 Demographic-based

Το Demographic-based συστήματα συστάσεων είναι τα συστήματα συστάσεων τα οποία για να εξάγουν αποτελέσματα και προτάσεις, χρησιμοποιούν τα δημογραφικά στοιχεία και χαρακτηριστικά των χρηστών στους οποίους θέλουν να κάνουν τις προτάσεις (Adomavicius & Tuzhilin, 2015)

Το παρόν σύστημα συστάσεων λοιπόν αφού συλλέξει τα δημογραφικά στοιχεία και χαρακτηριστικά του χρήστη, τότε ελέγχει τα δεδομένα και τις αναλύσεις που υπάρχουν για τις κατηγορίες στις οποίες ανήκει ο χρήστης, καθώς και τις αγοραστικές τάσεις οι οποίες μπορεί να υπάρχουν και σε συνδυασμό όλων των παραπάνω κάνει τις προτάσεις που σύμφωνα με τα συμπεράσματα στα οποία έχει καταλήξει θα ενδιαφέρουν τον χρήστη και θα είναι όσο γίνεται πιο κοντά στις δικές του προτιμήσεις.

2.1.4 Knowledge-based

Η συγκεκριμένη κατηγορία συστημάτων συστάσεων δεν αφορά αντικείμενα ή αγαθά που αγοράζονται ή επιλέγονται συχνά. Χρησιμοποιούνται δηλαδή κυρίως για αντικείμενα που ο χρήστης θα θέλει να γνωρίζει αρκετά στοιχεία για αυτά προτού προβεί στην αγορά ή την επιλογή του. Για παράδειγμα δεν θα χρησιμοποιηθούν για να γίνει μια πρόταση σε έναν χρήστη που θα αφορά την επιλογή μιας ταινία καθώς, οι χρήστες δεν θέλουν να γνωρίζουν πάρα πολλά δεδομένα για να την επιλέξουν, σε αντίθετη περίπτωση όμως είναι ιδανικά για να γίνει μια πρόταση στους χρήστες για την αγορά ενός σπιτιού, όπου θα θέλουν οι ίδιοι να γνωρίζουν αρκετές λεπτομέρειες και πληροφορίες πριν πάρουν την απόφαση τους. Πιο συγκεκριμένα η παραπάνω κατηγορία συστημάτων συστάσεων είναι κατάλληλη για τις παρακάτω επιλογές:

- Όταν οι χρήστες θέλουν να καθορίσουν ρητά τις απαιτήσεις τους, επομένως η διαδραστικότητα είναι ένα σημαντικό προσόν της κατηγορίας αυτής.
- Όταν είναι δύσκολο να λάβουν πληροφορίες οι χρήστες λόγω της πολυπλοκότητας που υπάρχει στον κλάδο για τον οποίο ενδιαφέρονται
- Σε ορισμένες κατηγορίες αντικειμένων που η αξία ή η κατάσταση του αντικειμένου μεταβάλλεται ανάλογα με τον χρόνο (Aggarwal, 2016).

Ένα πολύ σημαντικό πλεονέκτημα της συγκεκριμένης κατηγορίας είναι η μεγάλος έλεγχος που έχει ο χρήστης στην καθοδήγηση της διαδικασίας σύστασης ενός αντικειμένου, καθώς μπορούν να καθοριστούν λεπτομέρειες που θα έχουν σημαντικό ρόλο στην τελική επιλογή των προτάσεων.

2.1.5 Utility-based

Τα Utility-based συστήματα συστάσεων απαιτούν από το σύστημα να δημιουργήσει μια πλήρη λειτουργία χρησιμότητας σε όλα τα χαρακτηριστικά των αντικειμένων που εξετάζονται από αυτά (Burke, 2002).

Ένα σημαντικό όφελος αυτού του συστήματος συστάσεων είναι ότι μπορεί να ενσωματώσει πολλούς διαφορετικούς παράγοντες που συμβάλλουν στην αξία ενός αντικειμένου ή υπηρεσιών, όπως τον χρόνο παράδοσης και όχι μόνο τα χαρακτηριστικά του συγκεκριμένου αντικειμένου ή υπηρεσίας. Πολύ σημαντικό πλεονέκτημα σε αυτό το σύστημα συστάσεων είναι το γεγονός ότι, εκτός από τα χαρακτηριστικά του αντικειμένου το πότε θα μπορεί να παραδοθεί αυτό μπορεί να έχει πολύ μεγάλο ρόλο στην επιλογή του χρήστη, καθώς μπορεί ένας χρήστης να βρίσκεται αντιμέτωπος με κάποια προθεσμία και η ανάγκη του να είναι άμεση. Ένα σύστημα που βασίζεται σε βοηθητικά προγράμματα επιτρέπει στον χρήστη να εκφράσει όλες τις σκέψεις που πρέπει να υποβληθούν σε μια σύσταση (Aggarwal, 2016).

Η ευελιξία αυτή όμως των utility-based συστημάτων μπορεί επίσης να έχει και αρνητικές επιπτώσεις και αυτό γιατί, ο χρήστης θα πρέπει να κατασκευάσει μια πλήρη συνάρτηση προτίμησης και επομένως πρέπει να σταθμίσει ο ίδιος τη σημασία κάθε χαρακτηριστικού του αντικειμένου που έχει επιλέξει. Σε κάποιες περιπτώσεις με αντικείμενα που έχουν λίγα χαρακτηριστικά όπως η τιμή, η ποιότητα και ημερομηνία παράδοσης αυτό μπορεί να είναι εφικτό, όμως δεν ισχύει το ίδιο για πιο σύνθετους και υποκειμενικούς τομείς όπως ταινίες, ειδήσεις και άρθρα (Aggarwal, 2016).

2.1.6 Hybrid

Το Hybrid σύστημα συστάσεων είναι ένα σύστημα το οποίο για να κάνει κάποια πρόταση στον χρήστη χρησιμοποιεί έναν συνδυασμό δύο ή περισσότερων μεθόδων από αυτές που έχουμε αναφέρει παραπάνω (Adomavicius & Tuzhilin, 2015).

Για να δημιουργήσουμε ένα υβριδικό σύστημα συστάσεων, θα πρέπει πρώτα να γνωρίζουμε τα δεδομένα τα οποία διαθέτουμε, έτσι ώστε στην συνέχεια να προβούμε στην επιλογή των κατάλληλων συστημάτων συστάσεων. Για παράδειγμα τα CF συστήματα βασίζονται στις συνολικές βαθμολογίες των χρηστών για να βγάλουν τα αποτελέσματα τους, τα content-based βασίζονται αποκλειστικά στις βαθμολογίες των ίδιων των χρηστών που θα προτείνουν στην συνέχεια, ενώ τα knowledge-based βασίζονται στην αλληλεπίδραση με τον χρήστη, ενώ τέλος τα demographic-based συστήματα αναλύουν τα δημογραφικά στοιχεία που συλλέγουν από τους

χρήστες. Όπως καταλαβαίνουμε δηλαδή όλα αυτά τα συστήματα συστάσεων έχουν διαφορά δεδομένα τα οποία εισάγουν για να βγάλουν τα συμπεράσματα τους, ενώ όλα επίσης έχουν κάποιες αδυναμίες και κάποια πλεονεκτήματα. Στόχος λοιπόν των hybrid συστημάτων συστάσεων είναι ανάλογα με τα δεδομένα που διαθέτουν να επιλέξουν τον κατάλληλο συνδυασμό των παραπάνω RS έτσι ώστε να βγάλουν όσο το δυνατόν καλύτερα αποτελέσματα και τελικά προτάσεις για τους χρήστες ανάλογα με τις επιθυμίες τους και τα ενδιαφέροντα τους. (Aggarwal, 2016)

Σύμφωνα με τον (Aggarwal, 2016) τα Hybrid συστήματα συστάσεων χωρίζονται σε διαφορετικές κατηγορίες ανάλογα με τα δεδομένα που επιλέγουν, τον συνδυασμό των συστημάτων συστάσεων, αλλά και τα προβλήματα που θα πρέπει να αντιμετωπίσουν. Οι κατηγορίες αυτές είναι οι εξής:

- Weighted Hybrids
- Switching Hybrids
- Cascade Hybrids
- Feature Augmentation Hybrids
- Meta-Level Hybrids
- Feature Combination Hybrids
- Mixed Hybrids

2.2 Εφαρμογές

Στην παρούσα ενότητα θα αναφέρουμε πόσο σημαντικά είναι τα συστήματα συστάσεων σε όλες τις εφαρμογές αλλά και στις ιστοσελίδες στα ΜΚΔ. Πλέον οι εφαρμογές στα ΜΚΔ δεν παρέχουν μόνο το περιεχόμενο τους στους χρήστες, αλλά προσπαθούν να έχουν μεγαλύτερη αλληλεπίδραση με αυτούς κάνοντας οι ίδιες προτάσεις που θα μπορέσουν να βοηθήσουν το χρήστη να επιλέξει αυτό που θέλει ή να βρει αυτό που ψάχνει πιο εύκολα. Στην προηγούμενη ενότητα αναφέραμε τους τύπους RS και σε αυτήν θα παραθέσουμε κάποια παραδείγματα εφαρμογών στα ΜΚΔ και πόσο πολύ έχουν βελτιώσει την αλληλεπίδραση τους με τους χρήστες χρησιμοποιώντας τα RS.

Η πρώτη εφαρμογή που θα μελετήσουμε είναι η πλατφόρμα online παρακολούθησης ταινιών Netflix. Η συγκεκριμένη πλατφόρμα τα τελευταία χρόνια έχει γίνει κορυφαία εφαρμογή για παρακολούθηση ταινιών online καθώς εκτός από ήδη υπάρχουσες ταινίες, έχει δημιουργήσει και την δική της εταιρία παραγωγής ταινιών. Η συγκεκριμένη εφαρμογή λειτουργεί με μηνιαία συνδρομή και ο κάθε χρήστης μπορεί να φτιάξει το δικό του προφίλ. Στο προφίλ του ο κάθε χρήστης μπορεί να πατήσει like στις αγαπημένες του ταινίες, μπορεί να βάλει κάποιες ταινίες στην λίστα του ώστε να μπορεί να τις παρακολουθήσει στο μέλλον ενώ του δίνεται η δυνατότητα να βαθμολογήσει τις ταινίες που υπάρχουν στην πλατφόρμα. Όλα τα παραπάνω

δίνουν την δυνατότητα στην εφαρμογή να χρησιμοποιήσει τις πληροφορίες που της δίνονται έτσι ώστε να κάνει προτάσεις στον χρήστη ανάλογα με τις προτιμήσεις του. Το αποτέλεσμα που προκύπτει από τα παραπάνω είναι ότι η πλοήγηση του χρήστη γίνεται πολύ πιο εύκολη στην συγκεκριμένη εφαρμογή καθώς σε ένα αρχείο χιλιάδων ταινιών μέσω των προτάσεων της εφαρμογής μπορεί πιο εύκολα να βρει ο χρήστης μια ταινία ή σειρά που θα του αρέσει.(Aggarwal, 2016).

Η επόμενη εφαρμογή που θα αναλύσουμε είναι το Skroutz.gr. Το skroutz.gr είναι ένα online site μέσα απο το οποίο οι χρήστες μπορούν να αγοράσουν διάφορα προϊόντα ή να συγκρίνουν τιμές από διάφορες άλλες ιστοσελίδες. Τα προϊόντα του ξεκινούν από είδη ρουχισμού και προϊόντα τεχνολογίας και καταλήγουν σε είδη σπιτιού και επαγγελματικά προϊόντα. Το skroutz.gr δίνει πολλές δυνατότητες στους χρήστες μέσω της ιστοσελίδας του. Αρχικά οι χρήστες μπορούν να δημιουργήσουν το δικό τους προφίλ στην ιστοσελίδα και στη συνέχεια μπορούν να αναζητήσουν οποιοδήποτε προϊόν επιθυμούν, ενώ τους γίνεται πιο εύκολη η πλοήγηση καθώς τα αντικείμενα είναι χωρισμένα σε κατηγορίες. Επίσης οι χρήστες έχουν την δυνατότητα τόσο να βαθμολογήσουν τα προϊόντα τα οποία έχουν αγοράσει όσο και να αφήσουν σχόλια για τις ιστοσελίδες μέσω των οποίων τα αγόρασαν έτσι ώστε να βοηθήσουν μελλοντικές αγορές άλλων χρηστών.

Η συγκεκριμένη ιστοσελίδα επίσης δίνει την δυνατότητα στους χρήστες να μην χρειάζεται να αναζητήσουν κάποια συγκεκριμένη μάρκα ή μοντέλο που ψάχνουν, καθώς μπορούν να επιλέξουν για παράδειγμα απλώς το χρώμα ή το μέγεθος μιας μπλούζας που επιθυμούν και οι επιλογές που θα τους εμφανιστούν θα έχουν όλες τις διαθέσιμες εταιρείες ρουχισμού που διαθέτουν μπλούζες με τα συγκεκριμένα χαρακτηριστικά. Τα RS που χρησιμοποιεί το skroutz.gr λαμβάνουν όλα τα παραπάνω στοιχεία των βαθμολογήσεων των χρηστών και τις πληροφορίες των αναζητήσεων τους έτσι ώστε να παρέχουν πιο εξατομικευμένες υπηρεσίες. Για παράδειγμα εάν ο χρήστης αναζητήσει κάποια συγκεκριμένη μάρκα ρουχισμού και κάποιο συγκεκριμένο στυλ μπλούζας, τότε οι προτάσεις του skroutz θα αφορούν ρούχα της συγκεκριμένης εταιρείας ή μπλούζες του ίδιου στυλ, αλλά διαφορετικής μάρκας ρουχισμού (Stefani et al., 2019).

Στην συνέχεια θα αναλύσουμε τα RS της εφαρμογής YouTube. Η εφαρμογή αυτή είναι μια πλατφόρμα μέσω της οποίας οι χρήστες μπορούν να παρακολουθήσουν ή να δημοσιεύσουν βίντεο. Στο Youtube πλέον υπάρχει μεγάλη πληθώρα βίντεο διαφορετικών κατηγοριών. Καθημερινά στην συγκεκριμένη εφαρμογή γίνεται παρακολούθηση δισεκατομμυρίων βίντεο από εκατομμύρια διαφορετικούς χρήστες, ενώ ταυτόχρονα κάθε λεπτό γίνεται upload χιλιάδων βίντεο. Όπως καταλαβαίνουμε από τα παραπάνω οι επιλογές των βίντεο που μπορεί να παρακολουθήσει ένας χρήστης είναι πάρα πολλές. Για αυτόν τον λόγο λοιπόν το YouTube χρησιμοποιεί τα RS έτσι ώστε να παρέχει στους χρήστες μια πιο εξατομικευμένη πλοήγηση. Αρχικά δίνεται η δυνατότητα στους χρήστες να κάνουν αναζήτηση στην μπάρα αναζήτησης που

διαθέτει η εφαρμογή με δύο διαφορετικούς τρόπους. Αρχικά μπορούν να αναζητήσουν κάποιο βίντεο με συγκεκριμένο τίτλο, μπορούν όμως επίσης να κάνουν μια πιο γενική αναζήτηση και η εφαρμογή θα εμφανίσει βίντεο που αφορούν το συγκεκριμένο θέμα που ενδιαφέρει τον χρήστη.

Επίσης δίνεται η δυνατότητα στον χρήστη να πατήσει εάν του αρέσει ή όχι ένα βίντεο καθώς και να σχολιάσει σε αυτό. Ο χρήστης μπορεί ακόμα να εγγραφεί σε κάποιο κανάλι ή δημιουργό που του αρέσει έτσι ώστε στην αρχική σελίδα του να εμφανίζονται πιο εύκολα βίντεο από τον συγκεκριμένο δημιουργό. Το YouTube λοιπόν χρησιμοποιεί όλες τις παραπάνω πληροφορίες που του προσφέρει ο κάθε χρήστης και έτσι εξατομικεύει τις προτάσεις του. Για παράδειγμα εάν κάποιος χρήστης αναζητήσει ή κάνει like σε κάποιον συγκεκριμένο καλλιτέχνη ή τραγούδι στην συνέχεια η εφαρμογή θα του εμφανίζει άλλα προτεινόμενα τραγούδια του καλλιτέχνη ή επίσης παρόμοια τραγούδια ή καλλιτέχνες που ταιριάζουν με τις προτιμήσεις του χρήστη (Davidson et al., 2010).

Στην συνέχεια θα αναλύσουμε την εφαρμογή Facebook. Για την συγκεκριμένη εφαρμογή όμως δεν θα μιλήσουμε για τα RS που χρησιμοποιεί για να προτείνει στους χρήστες της διαφημίσεις προϊόντων που μπορεί να τους αφορούν, αλλά θα δούμε μια διαφορετική πλευρά των RS που αφορούν τις προτάσεις φιλίας με άλλους χρήστες στην εφαρμογή. Αρχικά το Facebook είναι μια εφαρμογή στην οποία δημιουργείς έναν λογαριασμό και μπορείς να προσθέσεις διάφορα στοιχεία και πληροφορίες που αφορούν τον χρήστη. Για παράδειγμα μπορεί να προσθέσει ένας χρήστης την ηλικία του, τον τόπο διαμονής του, τα ενδιαφέροντα που μπορεί να έχει καθώς και να προσθέσει φωτογραφίες ή να κοινοποιήσει διάφορα βίντεο ή link. Η διαφορά με τις εφαρμογές που αναλύσαμε παραπάνω είναι η εξής, στην συγκεκριμένη εφαρμογή μπορείς να προσθέσεις φίλους όμως σε αντίθεση ας πούμε με το YouTube που μπορείς να ακολουθήσεις έναν άλλον χρήστη χωρίς να υπάρξει κάποια επιβεβαίωση από την πλευρά του, σε αυτή θα πρέπει ο χρήστης που προσθέτεις σαν φίλο σου στην εφαρμογή να δεχθεί την πρόταση φιλίας σου. Στην περίπτωση που δύο χρήστες είναι φίλοι στην εφαρμογή τότε εμφανίζεται στην αρχική σελίδα του καθενός η δραστηριότητα του άλλου χρήστη.

Το Facebook λοιπόν έχει εφαρμόσει ένα RS που αφορά τις προτάσεις φιλίας. Αρχικά ενώ οι χρήστες μπορούν να αναζητήσουν κάποιον χρήστη μέσω της μπάρας αναζήτησης που υπάρχει στην εφαρμογή, η ίδια η εφαρμογή χρησιμοποιεί τα στοιχεία που έχουν δοθεί από τους χρήστες έτσι ώστε να εμφανίζονται στην αρχική τους σελίδα κάποιοι χρήστες που η εφαρμογή θεωρεί ότι μπορεί να τους ενδιαφέρουν ή να τους γνωρίζουν και να μην ξέρουν ότι διαθέτουν προφίλ σε αυτήν. Τα στοιχεία που χρησιμοποιεί η εφαρμογή μπορεί να αφορούν κοινά ενδιαφέροντα, ίδιο μέρος καταγωγής λαμβάνει όμως υπόψη ότι μπορεί δύο χρήστες να έχουν πολλούς κοινούς φίλους στην εφαρμογή όμως οι ίδιοι να μην είναι φίλοι. Εν κατακλείδι προχωρήσαμε στην ανάλυση του συγκεκριμένου RS της εφαρμογής καθώς σε αντίθεση με τις προηγούμενες εφαρμογές δεν στοχεύει στο άμεσο κέρδος ή στην άμεση παροχή εξατομικευμένης πλοήγησης στον χρήστη όπως στις παραπάνω εφαρμογές αλλά μέσω αυτού προσπαθεί η ίδια να μεγαλώσει

στα κοινωνικά δίκτυα διαθέτοντας μεγαλύτερα προφίλ με περισσότερα κοινά στοιχεία μεταξύ των χρηστών που θα την βοηθήσουν με τα υπόλοιπα RS που χρησιμοποιεί (Aggarwal, 2016).

2.3 Μετρικές Αξιολόγησης

Στο συγκεκριμένο κεφάλαιο θα παρουσιάσουμε και θα αναλύσουμε τις μετρικές αξιολόγησης που χρησιμοποιούνται για να αξιολογηθούν όλα τα παραπάνω συστήματα συστάσεων που έχουν αναφερθεί.

Αρχικά το πιο σημαντικό κομμάτι στην αξιολόγηση ενός συστήματος συστάσεων είναι η ακρίβεια του. Για αυτόν τον λόγο οι ερευνητές που θέλουν να συγκρίνουν διαφορετικά συστήματα συστάσεων θα πρέπει να λάβουν υπόψη τους μια ή περισσότερες μετρικές αξιολόγησης. Όμως όπως θα δούμε παρακάτω η ακρίβεια δεν είναι το μόνο κριτήριο που χρειάζεται για την αξιολόγηση ενός RS, υπάρχουν πολλές μετρικές που χρησιμοποιούν άλλους παράγοντες ως βασικούς για να εξάγουν συμπεράσματα.

Οι μετρικές λοιπόν χωρίζονται σε διάφορες κατηγορίες, υπάρχουν μετρικές που βασίζονται σε μαθηματικούς τύπους για να εξάγουν τα συμπεράσματά τους και έχουν ως βασικά κριτήρια την ακρίβεια και το λάθος. Η κατηγορία αυτή ονομάζεται “Accuracy and Error Based Methods”. Η επόμενη κατηγορία ονομάζεται “Decision Support Methods”, η κατηγορία αυτή αφορά μετρικές οι οποίες μας βοηθούν να καταλάβουμε πόσο χρήσιμες ήταν για τον χρήστη οι προτάσεις που του δόθηκαν και εάν τελικά είχε γίνει σωστά η επιλογή των αντικειμένων. Στην συνέχεια θα παρουσιάσουμε την κατηγορία “Ranking Based Methods”. Σε αυτήν την κατηγορία δεν βλέπουμε τα συνολικά αποτελέσματα των RS αλλά μας δείχνει ουσιαστικά την σχετικότητα και την κατάταξη που είχαν οι επιλογές των RS με τις πραγματικές επιθυμίες των χρηστών. Τέλος θα μελετήσουμε κάποιες πιο γενικές έννοιες και ερμηνείες που μπορούν να βοηθήσουν τους ερευνητές αν τα RS προτείνουν αντικείμενα ασυνήθιστα ή αν επικεντρώνονται μόνο σε παλιότερες επιλογές των χρηστών. (Herlocker et al., 2004)

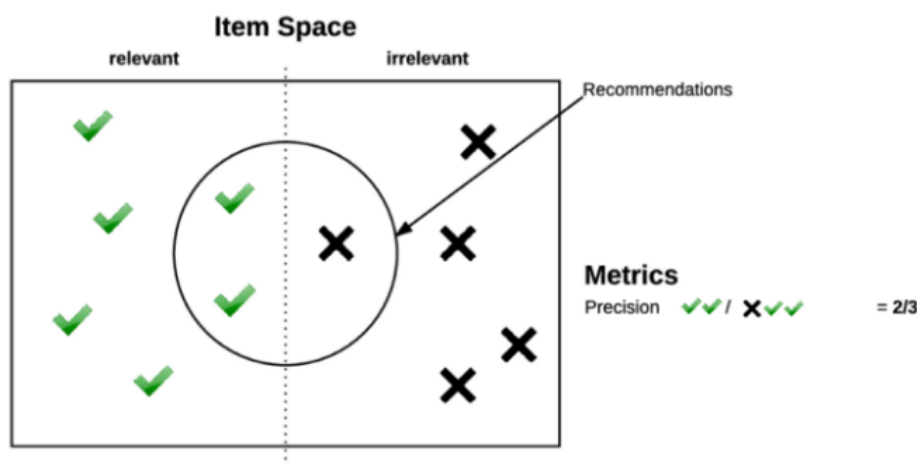
Αρχικά στην κατηγορία “Accuracy and Error Based Methods” υπάρχουν οι πολύ γνωστές μετρικές Mean Absolute Error(MAE) και Root Mean Squared Error(RMSE) που όμως θα τις παραλήψουμε, διότι αναφέρονται εκτενώς σε επόμενο κεφάλαιο. Μια πολύ σημαντική επίσης μετρική σε αυτήν την κατηγορία είναι η Mean Squared Error(MSE). Η MSE πρόκειται για μια μετρική παρόμοια με την MAE, όμως εδώ αντί να υπολογίζουμε την απόλυτη τιμή του λάθους για να αποφύγουμε το αρνητικό αποτέλεσμα, το τετραγωνίζουμε. Πολύ σημαντικό για την MSE είναι, πως όσο πιο κοντά τείνουν προς το μηδέν τα αποτελέσματα της τόσο περισσότερο σωστά είναι καθώς λόγω της ύπαρξης του τετραγώνου στην εξίσωση, τα αποτελέσματα λάθους μεγαλώνουν. (Herlocker et al., 2004)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

Εικόνα 3.1: Μαθηματικός τύπος MSE, Πηγή: Wikipedia

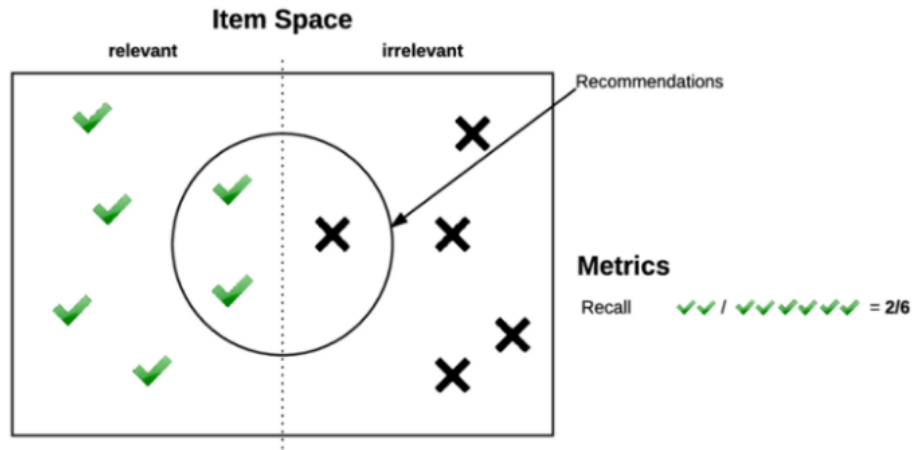
Η επόμενη κατηγορία “Decision Support Methods”, πρόκειται για μια κατηγορία μετρικών που αφορούν μετρήσεις αποφάσεων. Ουσιαστικά αυτές οι μετρικές μας βοηθούν να κατανοήσουμε πόσο χρήσιμα ήταν τα RS και πόσο βοήθησαν αυτά των χρήστη να επιλέξει τις σωστές προτάσεις για αυτόν, αποφεύγοντας τις λανθασμένες προτάσεις που του έγιναν.

Η πρώτη μετρική αυτής της κατηγορίας που θα ασχοληθούμε ονομάζεται “Precision”. Η μετρική αυτή ουσιαστικά αφορά το ποσοστό των προτάσεων που ήταν σχετικές με τις επιθυμίες των χρηστών σε σχέση με το σύνολο των προτάσεων που έγιναν στους χρήστες (Bellogin et al., 2011). Στην εικόνα 3.2, βλέπουμε ένα σύνολο προτάσεων που είναι σχετικές και μη με τις επιθυμίες του χρήστη. Στον κύκλο υπάρχουν οι τρεις προτάσεις που επιλέχθηκαν να γίνουν τελικά σε έναν χρήστη και στα δεξιά της εικόνας βλέπουμε ότι σύμφωνα με την μετρική της ακρίβειας, μόνο δύο εξ αυτών ήταν σχετικές με τις προτιμήσεις του χρήστη.



Εικόνα 3.2: Προτάσεις συστήματος συστάσεων και αποτελέσματα μετρικής της ακρίβειας, Πηγή: researchgate.net.

Η επόμενη μετρική αυτής της κατηγορίας ονομάζεται “Recall”. Η μετρική αυτή αφορά το ποσοστό των αντικειμένων που ήταν σχετικά με τις επιθυμίες των χρηστών και επιλέχθηκαν να προταθούν σε σχέση με το σύνολο των αντικειμένων που ήταν σχετικά με τις επιθυμίες των χρηστών (Schroeder et al., 2011). Για παράδειγμα στην εικόνα 3.3 βλέπουμε ότι από το σύνολο των έξι αντικειμένων που ήταν σχετικά με τις επιθυμίες των χρηστών τελικά μόνο δύο από αυτά επιλέχθηκαν να προταθούν από το σύστημα συστάσεων. Ο βασικός στόχος αυτής της μετρικής είναι ουσιαστικά να μην χάνονται από τις προτάσεις, αντικείμενα που είναι σχετικά με τις επιθυμίες των χρηστών.



Εικόνα 3.3: Σύνολο σχετικών προτάσεων αντικειμένων σχετικών με τις επιθυμίες των χρηστών σε σχέση με τις επιλεγμένες προτάσεις, Πηγή: researchgate.net.

Η τρίτη κατηγορία μετρικών με την οποία θα ασχοληθούμε ονομάζεται “Ranking Based Methods”. Στις μετρικές που αναφέραμε μέχρι τώρα μας ενδιέφερε εάν το σύνολο των προτάσεων που έγιναν από τα RS ήταν σχετικά με τις επιθυμίες των χρηστών, στην συγκεκριμένη κατηγορία μετρικών μας ενδιαφέρει και η σειρά με την οποία γίνονται οι προτάσεις, καθώς μπορεί οι αρχικές προτάσεις να μην είναι σχετικές με τις επιθυμίες του χρήστη και αυτός να πρέπει να συνεχίσει να πλοηγείται και να αναζητά στο διαδίκτυο, ότι θα έκανε δηλαδή και στην περίπτωση που δεν θα υπήρχαν τα RS.

Η μετρική που θα αναλύσουμε σε αυτήν την κατηγορία ονομάζεται “Average Precision”. Η διαφορά της συγκεκριμένη μετρικής με την μετρική “Precision” που έχουμε αναφέρει παραπάνω είναι ότι σε αυτήν την μετρική ελέγχουμε αν είναι η σειρά με την οποία εμφανίζονται οι προτάσεις σωστή και σύμφωνες με τις επιθυμίες των χρηστών (Jensen et al., 2009). Στην εικόνα 3.4 βλέπουμε τον μαθηματικό τύπο μέσα από τον οποίο βγάζουμε την μέση ακρίβεια.

$$Average\ Precision = \frac{\sum_r P@r}{R}$$

Εικόνα 3.4: Μαθηματικός τύπος μέσης ακρίβειας, Πηγή: springerlink.

Τέλος υπάρχουν και άλλες μετρικές που μπορούν να βοηθήσουν στην αξιολόγηση των προτάσεων που γίνονται. Στην συνέχεια θα παρουσιάσουμε ορισμένες εξ αυτών. Αναλυτικά:

Coverage: Το συγκεκριμένο εργαλείο μας βοηθά να μετρήσουμε τον αριθμό των αντικειμένων που κατάφεραν να προταθούν από το RS σε σχέση με μια συνολική βάση αντικειμένων. Μπορεί επίσης να χρησιμοποιηθεί και για άλλους λόγους, για παράδειγμα με το εργαλείο αυτό μπορούμε να δούμε την αναλογία γνωστών και μη γνωστών αντικειμένων που κατάφεραν να προταθούν από το RS. (Kaminskas & Bridge, 2016)

Novelty: Το εργαλείο αυτό μας βοηθάει να μην γίνονται συνεχώς οι ίδιες προτάσεις στον χρήστη, έτσι ώστε να μην γίνει για αυτόν η περιήγησή του μονότονη. Προσπαθεί δηλαδή εκτός από τις συνηθισμένες προτάσεις να παρουσιάσει και κάποιες διαφορετικές προτάσεις που θα τραβήξουν την προσοχή του χρήστη και θα του κινήσουν την περιέργεια (Kaminskas & Bridge, 2016).

Diversity: Το εργαλείο αυτό έχει παρόμοια στοιχεία με το εργαλείο που αναφέραμε παραπάνω καθώς, προσπαθεί στις προτάσεις του να έχει μια μεγαλύτερη ποικιλία επιλογών έτσι ώστε ο χρήστης να μην έχει μια μονότονη περιήγηση. Για παράδειγμα δεν θα πρέπει να εμφανίζει συνεχώς τις ίδιες προτάσεις τραγουδιών στον χρήστη κατά την επίσκεψή του σε μια πλατφόρμα, αλλά αυτές να μην προκύπτουν από τις προτιμήσεις του αλλά να είναι όσο γίνεται διαφορετικές κάθε φορά (Kaminskas & Bridge, 2016).

3. Συστήματα Συστάσεων ως Μέσα Διαφήμισης στα Μέσα Κοινωνικής Δικτύωσης

Το παρόν κεφάλαιο έχει ως στόχο να μελετήσει να αναλύσει και να παρουσιάσει έρευνες που συναντώνται στην βιβλιογραφία και συνάδουν με το σκοπό της παρούσας πτυχιακής εργασίας.

Η πρώτη έρευνα που θα παρουσιάσουμε πραγματοποιήθηκε από τους (Massa & Avesani, 2007). Αυτοί πιστεύουν ότι πολλές φορές τα RS αποτυγχάνουν να βρουν χρήστες με παρόμοια ενδιαφέροντα λόγω έλλειψης στοιχείων, για αυτόν τον λόγο αποφάσισαν να χρησιμοποιήσουν μια μετρική εμπιστοσύνης και έναν αλγόριθμο για την εκτίμηση του trust weight, το οποίο θα αντικαταστήσει το similarity weight. Τα δεδομένα τους τα σύλλεξαν από το Epinions.com, χρησιμοποίησαν 49.290 διαφορετικούς χρήστες, οι οποίοι είχαν αξιολογήσει σχεδόν 140.000 διαφορετικά αντικείμενα τουλάχιστον μια φορά. Για να πραγματοποιήσουν όλα τα παραπάνω χρησιμοποίησαν CF, πραγματοποιώντας κάποιες μετατροπές, και τελικά προτείνουν μια δική τους αρχιτεκτονική της οποίας ονομάζουν, Trust-aware Recommender System. Τα συμπεράσματα στα οποία καταλήγουν είναι πως χρησιμοποιώντας βαθμολογίες μόνο από directly trusted users πετυχαίνουμε το μικρότερο δυνατό αριθμό λαθών, έχοντας την αποδεκτή κάλυψη. Η τεχνική αυτή είναι πολύ χρήσιμη για έναν νέο χρήστη. Πιστεύουν επίσης ότι δεν είναι δεδομένο, ότι όλες οι παρόμοιες μέθοδοι βασισμένες σε trust metrics θα είναι επιτυχημένες, λόγω των πολλών διαφορετικών χαρακτηριστικών των αντικειμένων που επηρεάζουν τα αποτελέσματα των αλγορίθμων.

Σύμφωνα με τους (O'Donovan & Smyth, 2005), δίνεται πολύ μεγάλη έμφαση στο user similarity και ίσως είναι λίγο υπερεκτιμημένο όλο αυτό. Οι ίδιοι πιστεύουν ότι πρέπει να λαμβάνεται περισσότερο υπόψη το trustworthiness μεταξύ των χρηστών. Σκοπεύουν να προτείνουν λοιπόν δύο υπολογιστικά μοντέλα εμπιστοσύνης και πως μπορούν αυτά να ενσωματωθούν εύκολα σε απλά CF, αλλά και πως μπορούν να βελτιώσουν την ακρίβεια των προβλέψεων. Η έρευνα τους βασίστηκε κυρίως σε trust-based RS, αναλυτικά χρησιμοποίησαν, trust-based weighting, trust-based filtering, αλλά και έναν συνδυασμό αυτών των δύο. Τα δύο διαφορετικά μοντέλα που δημιούργησαν αφορούσαν δύο διαφορετικά χαρακτηριστικά το ένα επικεντρώνεται στο προφίλ γενικά ενώ το δεύτερο επικεντρώνεται στα αντικείμενα τα οποία υπάρχουν μέσα στο προφίλ. Για την έρευνα χρησιμοποίησαν την MovieLens dataset, που βρίσκονται 943 βαθμολογημένων ταινιών. Ο μέσος όρος των ταινιών που βρίσκονται σε κάθε προφίλ είναι 105 ταινίες. Τα τελικά αποτελέσματα τα οποία προκύπτουν από την έρευνα τους είναι, ότι τα συνηθισμένα CF έχουν βασιστεί αρκετά στις ομοιότητες μεταξύ των προφίλ αξιολόγησης των χρηστών. Στην δική τους όμως έρευνα επικεντρώνονται στην παρακολούθηση της ακρίβειας ενός προφίλ. Επίσης θεωρούν ότι υπάρχουν πολύ τρόποι για να ενσωματωθούν τα δύο αυτά μοντέλα σε ένα απλό CF, για καλύτερα αποτελέσματα. Τέλος έχουν διαπιστώσει ότι η χρήση trust value έχει θετικό αντίκτυπο στην καλύτερη απόδοση καθώς μειώνει το μέσο σφάλμα πρόβλεψης κατά 22%

Σύμφωνα με τους (Ray & Mahanti, 2010) τα trust aware συστήματα συστάσεων είναι πολύ έξυπνα και βοηθούν πολύ στο να κάνουν σωστές και εξατομικευμένες προβλέψεις, ενώ θεωρούν επίσης ότι είναι πολύ καλύτερα από τα παραδοσιακά RS. Στην δική τους έρευνα αυτό που θέλουν να πετύχουν είναι μια προσέγγιση για την βελτίωση της ακρίβειας των προβλέψεων. Για να το πετύχουν αυτό, ανακατασκευάζουν το trust network, πιο συγκεκριμένα αφαιρούν τα trust links μεταξύ των χρηστών που έχουν συντελεστή συσχέτισης κάτω από ένα ποσοστό που έχουν ορίσει. Τα δεδομένα για την έρευνα τους τα έχουν συλλέξει από το Epinions.com. Χρησιμοποίησαν 50.000 διαφορετικούς χρήστες και 140.000 διαφορετικά αντικείμενα, ενώ το 53% των χρηστών που χρησιμοποιήθηκαν είχαν βαθμολογήσει τουλάχιστον 5 αντικείμενα. Επίσης χρησιμοποίησαν την MAE για να καθορίσουν τον επιθυμητό συντελεστή συσχέτισης. Τα συμπεράσματα στα οποία κατέληξαν είναι ότι η επιθυμητή παράμετρος είναι 0.5. Επίσης πιστεύουν ότι ενώ η προσέγγιση τους για ανασυγκρότηση του δικτύου αποδίδει σημαντικά καλύτερα, έχει και αυτή κάποιους περιορισμούς. Πρώτον μπορεί να δώσει προβλέψεις μόνο για εκείνους τους χρήστες που έχουν βαθμολογήσει μερικά αντικείμενα και έχουν περάσει κάποια trust statements, και δεύτερον το coverage της δικιάς του προσέγγισης του αριθμού των αντικειμένων που μπορούν να προβλεφθούν είναι μικρότερο από το αρχικό coverage.

Στην δική τους έρευνα οι (Massa & Bhattacharjee, 2004), αναφέρουν ότι με την επιλογή των trust-aware RS μπορούμε να μειώσουμε τα μειονεκτήματα που υπάρχουν με την χρήση των παραδοσιακών RS που βασίζονται στο user similarity. Πιστεύουν επίσης ότι είναι πολύ πιθανό να βρεθούν χρήστες με παρόμοια ενδιαφέροντα αλλά αυτό δεν σημαίνει κάτι, διότι συμβαίνει λόγω του μεγάλου όγκου χρηστών, κάτι το οποίο μπερδεύει τις προτάσεις των κλασσικών RS. Αυτοί θέλουν να αποδείξουν ότι εκμεταλλευόμενοι τα webs of trust, είναι πολύ πιθανό να διαδοθεί η εμπιστοσύνη και να προστεθούν weight για τους άλλους χρήστες. Επίσης θέλουν να δείξουν πως αυτή η μεταβλητή μπορεί να υπολογιστεί για μεγαλύτερο αριθμό χρηστών. Τα στοιχεία τους βασίστηκαν και σε αυτήν την έρευνα στο Epinions.com. Το σύνολο των δεδομένων αποτελείται από 49.290 χρήστες οπότε έχουν εκφράσει την κριτική τους σε αντικείμενα (με βαθμολογία), ενώ έχουν καθορίσει το web of trust τους, δηλαδή αναφερόμαστε σε χρήστες που έχει ελεγχθεί ότι οι κριτικές τους είναι χρήσιμες. Χρησιμοποίησαν στην έρευνα τους επίσης CF συστήματα συστάσεων τα οποία προσπάθησαν να βελτιώσουν με την χρήση trust-aware RS. Αυτό που έδειξαν με την έρευνα τους τελικά είναι ότι ο τρόπος με τον οποίο δουλεύουν τα παραδοσιακά RS, μπορεί να λειτουργήσει μόνο σε ένα μικρό τμήμα, και τα στοιχεία που βγαίνουν πολλές φορές όπως είπαμε ξανά παραπάνω εξαρτώνται στην τυχαιότητα των χρηστών. Ενώ τα trust-based RS, μπορούν να βγάλουν καλύτερα αποτελέσματα για μεγαλύτερες ποσότητες χρηστών. Και σε αυτήν την έρευνα όμως αναφέρεται το γεγονός ότι επειδή ένα trust-based RS λειτουργεί αυτό δεν σημαίνει ότι και όλα τα υπόλοιπα θα έχουν την ίδια δεδομένη επιτυχία.

Η επόμενη έρευνα που θα μελετήσουμε ανήκει στους (Montaner et al., 2002), και ο σκοπός αυτής είναι η δημιουργία ενός μοντέλου εμπιστοσύνης στον συνεργατικό κόσμο, σαν μια προσέγγιση βελτίωσης των recommender agents. Σε αυτό το μοντέλο, αναγνωρίζεται η εξέλιξη της εμπιστοσύνης, δηλαδή η δυναμική της. Επίσης το μοντέλο που θέλουν να προτείνουν δίνει έμφαση στην προληπτικότητα ο agent, αναζητά άλλους παράγοντες σε έλλειψη πληροφοριών, αντί να παραμείνει παθητικός ή να παρέχει είτε αρνητική είτε κενή απάντηση στον χρήστη όπως συμβαίνει σε άλλες περιπτώσεις. Για την έρευνα τους χρησιμοποιούν opinion-based filtering methods, οποίες είναι βασισμένες σε social model of trust. Προτείνουν τελικά έναν δικό τους τύπο, για να εξάγουν αποτελέσματα, όπου η μεγάλη διαφορά σε σχέση με άλλους τύπους είναι ότι ο υπολογισμός της αξίας εμπιστοσύνης βασίζεται στη συσχέτιση person-person του Glance. Στο μοντέλο τους, απαιτείται διαφάνεια των δεδομένων των χρηστών ενώ στο σύστημά επικρατεί το απόρρητο. Τέλος καταλήγουν στο συμπέρασμα ότι εάν δώσουμε μεγαλύτερη βαρύτητα στην εμπιστοσύνη μπορούν να βελτιωθούν οι αποδόσεις των RS.

Στην παρούσα έρευνα η (Golbeck, 2006), θέλει να κάνει μια παρουσίαση του FilmTrust, μια ιστοσελίδα που χρησιμοποιεί την εμπιστοσύνη, σε web-based social networks, με σκοπό να κάνει σωστές προτάσεις ταινιών στους χρήστες. Χρησιμοποιώντας το σύστημα FilmTrust ως βάση, θέλει να δείξει ότι οι συστάσεις είναι πιο ακριβείς από άλλες τεχνικές που χρησιμοποιούνται όταν οι απόψεις του χρήστη για μια ταινία αποκλίνουν από τον γενικό μέσο όρο. Επίσης θεωρεί ότι μπορούν να χρησιμοποιηθούν τα αποτελέσματα της εφαρμογής αυτής και από άλλα RS, με σκοπό την βελτίωση των αλγορίθμων των CF. Η συγκεκριμένη εφαρμογή χρησιμοποιεί Recommender Rating, και στόχος της έρευνας είναι να δείξει εάν τα trust-based ratings έχουν κάποιο όφελος τελικά προς τους χρήστες. Στην έρευνα συμμετείχαν 500 διαφορετικοί users, ενώ έγιναν 11.250 διαφορετικές βαθμολογήσεις σε περισσότερες από 1.250 διαφορετικές ταινίες. Για να υπάρχουν ταινίες με κοινές βαθμολογίες οι χρήστες ενημερώθηκαν ότι θα πρέπει να βαθμολογήσουν όσες ταινίες είχαν δει από ένα συγκεκριμένο top 50 που τους είχε αποσταλεί. Μετά την πραγματοποίηση της έρευνας, τα αποτελέσματα που προέκυψαν ήταν ότι τα trust-based προγνωστικά αξιολογήσεων έχουν καλύτερη ακρίβεια σε σχέση με έναν μέσο όρο βαθμολογιών που απλά εκχωρείται σε μια βαθμολογία. Επίσης το trust system έχει καλύτερα αποτελέσματα από ότι ένα σύστημα προτεινόμενων βαθμολογιών που προτείνονται με βάση την συσχέτιση.

4. Ανάπτυξη Συστήματος Συστάσεων για Διαφήμιση στα Μέσα Κοινωνικής Δικτύωσης

Σε αυτό το Κεφάλαιο, περιγράφεται το σύστημα συστάσεων που αναπτύσσεται στα πλαίσια της παρούσας πτυχιακής εργασίας και αποσκοπεί στην εξατομικευμένη σύσταση αντικειμένων/προϊόντων στα πλαίσια ενός κοινωνικού δικτύου.

Όπως θα έχει ήδη γίνει φανερό, το κύριο χαρακτηριστικό που διαφοροποιεί τα RS στα ΜΚΔ από αυτά που συναντώνται σε άλλου είδους περιβάλλοντα είναι το γεγονός ότι υπάρχουν και διαμορφώνονται ουσιώδεις σχέσεις μεταξύ των ατόμων, οι οποίες εκφράζονται μέσω αλληλεπιδράσεων στο εκάστοτε ΜΚΔ. Οι σχέσεις αυτές χαρακτηρίζονται από πλούσια πληροφορία, που μπορεί σε μεγάλο βαθμό να αξιοποιηθεί από ένα RS, ώστε να παρέχει πιο ακριβείς συστάσεις. Το κυριότερο ζήτημα σε αυτή την περίπτωση είναι να προσδιοριστούν και να ποσοτικοποιηθούν τα χαρακτηριστικά της σχέσης μεταξύ δύο ή περισσότερων ατόμων, με τρόπο που να είναι αξιοποιήσιμος από τεχνικές και μεθόδους για ανάπτυξη RS.

Τα ΜΚΔ επιτρέπουν τόσο την επικοινωνία μεταξύ των χρηστών, όσο και την ανάπτυξη και διαμόρφωση διαπροσωπικών σχέσεων μέσω των αλληλεπιδράσεων μεταξύ αυτών με τη χρήση αντιδράσεων, σχολίων και κοινοποιήσεις δημοσιευμάτων. Λαμβάνοντας, λοιπόν, υπόψη πως το σύνολο των ατόμων δείχνει να επηρεάζεται και να έλκεται από τις προτιμήσεις του κοινωνικού περίγυρου (Lerman, 2006), προκύπτει εύλογα το γεγονός πως από τους χρήστες τους οποίους διατηρεί και αλληλεπιδρά κανείς στο λογαριασμό του ΜΚΔ που διαθέτει, μπορούν να πηγάσουν αξιόπιστες και κατάλληλες συστάσεις για τον εν λόγω χρήστη (Λιανέρη & Στάμου, 2013).

Διάφορα χαρακτηριστικά έχουν διερευνηθεί για τη δημιουργία ενός RS, με κάποια εξ αυτών να ανήκουν στην κατηγορία των explicit και άλλα σε αυτή των implicit. Η πρώτη κατηγορία αφορά μια ενεργητική συλλογή πληροφορίας, η οποία προκύπτει από επιλογή του χρήστη. Πιο συγκεκριμένα, σε αυτή τη κατηγορία η πληροφορία μπορεί να προέρχεται από τη βαθμολόγηση ενός αντικειμένου από το χρήστη, από τη ταξινόμηση πλήθους αντικειμένων βάσει των προτιμήσεών του ή και τη δημιουργία μιας λίστας αυτών και πάλι βάσει όσων προτιμά (Γκοτζαμπουγιούκη, 2009).

Στην δεύτερη κατηγορία πραγματοποιείται μια παθητική συλλογή πληροφορίας, χωρίς να απαιτείται κάποια ενέργεια από την πλευρά του χρήστη. Για παράδειγμα, η μελέτη του κοινωνικού δικτύου ενός χρήστη για την εξαγωγή πληροφοριών από παρόμοιες αναζητήσεις και προτιμήσεις, η διατήρηση ιστορικού παραγγελιών προϊόντων και η αποθήκευση στοιχείων σχετικά με αντικείμενα στα οποία πλοηγείται σε ένα ηλεκτρονικό κατάστημα αποτελούν implicit χαρακτηριστικά (Γκοτζαμπουγιούκη, 2009).

Πιο στοχευμένα και αναφορικά με τα ΜΚΔ, τα ακριβή χαρακτηριστικά που αφορούν μία σχέση μεταξύ ατόμων μπορεί να προκύπτουν είτε σιωπηρά (implicitly), αν αυτά εκδηλώνονται μέσα από αλληλεπίδραση στο ΜΚΔ, είτε ρητά (explicitly), αν αυτά προκύπτουν από ρητή δήλωση του χρήστη. Για παράδειγμα, η υψηλή συχνότητα επικοινωνίας μεταξύ δύο ατόμων μπορεί να δείχνει ότι αυτοί είναι κοντινοί φίλοι, χωρίς όμως να το έχουν δηλώσει αναγκαστικά στην πλατφόρμα του ΜΚΔ, άρα αποτελεί implicit χαρακτηριστικό. Αντίθετα, αν η πλατφόρμα δίνει τη δυνατότητα να κατηγοριοποιήσει κανείς τις σχέσεις του με τους φίλους του, αυτό είναι ένα explicit χαρακτηριστικό, αφού η πληροφορία δηλώνεται ρητά και συνειδητά από το χρήστη.

Συνήθως, τα implicit χαρακτηριστικά απαιτούν κάποια προεπεξεργασία για να αξιοποιηθούν, ενώ η ερμηνεία τους ως προς τη σχέση μεταξύ των ατόμων είναι συχνά υποκειμενική σε κάποιο βαθμό, αφού μπορεί να εκφράζει διαφορετική συμπεριφορά από χρήστη. Για παράδειγμα, κάθε χρήστης μπορεί να έχει διαφορετική συχνότητα επικοινωνίας με τους κοντινούς φίλους του, με αποτέλεσμα να είναι δύσκολη η κατηγοριοποίηση της σχέσης τους με κάποιον απλό τρόπο καθολικά. Παρά ταύτα, τα implicit χαρακτηριστικά παρέχουν διάφορα οφέλη, καθώς δεν απαιτούν την πραγματοποίηση κάποιας συγκεκριμένης ενέργειας από το χρήστη, που κατά κανόνα πραγματοποιείται από ένα μικρό ποσοστό των χρηστών. Ως αποτέλεσμα, συχνά χρησιμοποιούνται για την αντιμετώπιση προβλημάτων, όπως το cold start, όπου η σύσταση παρουσιάζει δυσκολίες λόγω έλλειψης ratings σε ένα αντικείμενο ή χρήστη (Lika et al., 2014).

Ακόμη, υπάρχει διαφοροποίηση ως προς τον τύπο των σχέσεων από τη μία πλατφόρμα ΜΚΔ σε μία άλλη, ανάλογα με τη λειτουργικότητα που αυτή παρέχει. Για παράδειγμα, στο Facebook ένας χρήστης έχει τη δυνατότητα αποστολής αιτήματος φιλίας προς έναν άλλον, το οποίο ο δεύτερος θα πρέπει να αποδεχτεί και, εφόσον γίνει αυτό, οι χρήστες θεωρούνται φίλοι. Στην περίπτωση αυτή, η σχέση μεταξύ των χρηστών θεωρείται συμμετρική, αφού η σχέση είναι αμοιβαία, καθώς επίσης απαιτείται επιβεβαίωση από τη μεριά του δεύτερου χρήστη. Αντίθετα, σε περιπτώσεις, όπως το Twitter, ένας χρήστης μπορεί να κάνει follow έναν άλλον, χωρίς να απαιτείται επιβεβαίωση, ενώ η σχέση τους είναι μη συμμετρική. Αυτού του είδους τα χαρακτηριστικά, καθώς και άλλα, όπως η φύση των σχέσεων (π.χ. στο LinkedIn οι σχέσεις είναι κυρίως επαγγελματικές), θεωρούνται βασικά χαρακτηριστικά που πρέπει να ληφθούν υπόψη κατά το σχεδιασμό ενός συστήματος συστάσεων σε ΜΚΔ (Guy, 2015).

Ένα κύριο χαρακτηριστικό που έχει μελετηθεί αρκετά στη βιβλιογραφία είναι η έννοια της εμπιστοσύνης (trust), αφού συχνά οι επιλογές ενός ατόμου επηρεάζονται από άτομα που θεωρούν αξιόπιστα, είτε αυτά είναι κοντινοί φίλοι είτε ειδικοί σε κάποιον τομέα. Αν και ως έννοια είναι αρκετά διαδεδομένη στα RS, ο ορισμός της εμπιστοσύνης διαφέρει αρκετά από έρευνα σε έρευνα. Ταυτόχρονα, άλλες φορές γίνεται η υπόθεση πως η εμπιστοσύνη είναι ρητό χαρακτηριστικό, ενώ άλλοτε προκύπτει μέσω προεπεξεργασίας implicit χαρακτηριστικών.

Ανεξάρτητα από την εκάστοτε προσέγγιση, πάντα θεωρείται πως διαμορφώνεται ένα δίκτυο εμπιστοσύνης μεταξύ χρηστών, το οποίο διαφέρει ως προς τα χαρακτηριστικά του. Σε αυτό το δίκτυο η εμπιστοσύνη αναπαρίσταται ως βάρος σε κάθε ζεύξη μεταξύ των κόμβων, όπου κόμβοι θεωρούνται οι χρήστες του ΜΚΔ. Βασικές διαφορές συναντώνται στο αν η σχέση εμπιστοσύνης μεταξύ δύο χρηστών είναι απόλυτη, οπότε το βάρος είναι μία δυαδική τιμή (0 ή 1), ή σχετική, με την έννοια ότι το βάρος είναι μία συνεχής τιμή που ανήκει στο εύρος $[0, 1]$. Αντίστοιχα, οι σχέσεις εμπιστοσύνης μπορεί να είναι συμμετρικές, οπότε το δίκτυο είναι ένας μη κατευθυνόμενος γράφος, ή μη συμμετρικές, οπότε το δίκτυο είναι κατευθυνόμενος γράφος.

Μία ενδιαφέρουσα ιδιότητα που έχει μελετηθεί σε αυτού του είδους τις έρευνες είναι το πως μπορεί να διαχυθεί η εμπιστοσύνη κατά το δίκτυο, ώστε να γίνουν πιο αξιόπιστες οι συστάσεις, χρησιμοποιώντας ως μέτρο σύγκρισης περισσότερους χρήστες, σε σχέση με αυτούς που υπάρχει ρητή σχέση εμπιστοσύνης στο δίκτυο. Διαφορετικές προσεγγίσεις ενσωματώνουν μία ποικιλία μεθόδων που συνήθως συναντώνται σε γράφους, όπως τυχαίοι περίπατοι στο δίκτυο εμπιστοσύνης (Jamali & Ester, 2009), διάσχιση του δικτύου μέσω Breadth-First Search (Massa & Avesani, 2007 και Golbeck, 2005), κ.ά.

4.1 Μεθοδολογία Ανάπτυξης Συστήματος Συστάσεων

Συνεχίζοντας με τα στάδια της ανάπτυξης του συστήματος συστάσεων για διαφήμιση στα ΜΚΔ που αποτελεί και το επίκεντρο της παρούσας πτυχιακής εργασίας, πρέπει πρωτίστως να αναλυθούν οι κύριες παραδοχές που έλαβαν χώρα.

Αρχικά, θεωρείται πως είναι διαθέσιμο το δίκτυο εμπιστοσύνης που διαμορφώνεται από το εκάστοτε ΜΚΔ, το οποίο μπορεί να έχει προκύψει είτε από implicit είτε explicit χαρακτηριστικά. Η σχέση εμπιστοσύνης μεταξύ δύο χρηστών θεωρείται δυαδικό, οπότε το βάρος που ανατίθεται στη ζεύξη μεταξύ δύο κόμβων του δικτύου παίρνει τιμή 0 ή 1. Η σχέση εμπιστοσύνης θεωρείται μη συμμετρική, με αποτέλεσμα ο γράφος που αναπαριστά το δίκτυο να είναι μη κατευθυνόμενος.

Με στόχο την επέκταση των σχέσεων ενός χρήστη πέρα από αυτές που είναι άμεσα διαθέσιμες (δηλαδή η γειτονιά του κόμβου στο γράφο), προτείνεται μία μέθοδος διάχυσης της εμπιστοσύνης που βασίζεται στη μεταβατική ιδιότητα που μπορεί να θεωρηθεί ότι ισχύει σε μία σχέση εμπιστοσύνης. Με απλά λόγια, αν ένας χρήστης A εμπιστεύεται ένα χρήστη B και αυτός με τη σειρά του εμπιστεύεται το Γ, τότε μπορεί να θεωρηθεί ότι ο A εμπιστεύεται σε κάποιο βαθμό το Γ. Βέβαια, αυτή η σχέση εμπιστοσύνης που προκύπτει έμμεσα μεταξύ του A και του Γ μπορεί να θεωρηθεί ότι δεν είναι τόσο ισχυρή όσο αυτή μεταξύ του A και του B.

Για να αναπαρασταθούν τα παραπάνω πιο τυπικά, έστω $d(u, v)$ η απόσταση μεταξύ των κόμβων u και v . Έτσι, αν ο κόμβος A εμπιστεύεται ρητά το B, τότε $d(A, B) = 1$. Καθώς ο

γράφος είναι κατευθυνόμενος, το αντίστροφο δεν ισχύει αναγκαστικά, δηλαδή είναι πιθανό $d(A, B) \neq d(B, A)$. Διαισθητικά, όσο πιο μακριά είναι ένας κόμβος Γ από τον κόμβο A στο δίκτυο, τόσο πιο “αδύναμη” θα είναι και η σχέση εμπιστοσύνης μεταξύ αυτών. Με άλλα λόγια, η απόσταση είναι αντιστρόφως ανάλογη της εμπιστοσύνης μεταξύ δύο κόμβων. Για να αναπαρασταθεί αυτή η σχέση, έστω $t(u, v)$ η σχέση εμπιστοσύνης μεταξύ των κόμβων u, v . Τότε, ορίζεται πως η εμπιστοσύνη θα ισούται με

$$t(u, v) = p^{d(u,v) - 1},$$

όπου $p \in (0, 1]$ η παράμετρος διάχυσης. Διαισθητικά, για κάθε βήμα μακριά από τους άμεσα συνδεδεμένους κόμβους ενός κόμβου A , η τιμή της εμπιστοσύνης μεταξύ τους πολλαπλασιάζεται με την παράμετρο διάχυσης. Με τον τρόπο αυτό, καθώς απομακρύνεται κανείς από ένα χρήστη, η τιμή εμπιστοσύνης μειώνεται εκθετικά. Αν, όμως, η παράμετρος διάχυσης πάρει τιμή 1, τότε ανεξάρτητα από την απόσταση, η τιμή της εμπιστοσύνης θα παραμείνει 1.

Με βάση τα παραπάνω, ορίζεται η εμπιστοσύνη μεταξύ οποιωνδήποτε δύο χρηστών στο δίκτυο, εφόσον υπάρχει ένα μονοπάτι χρηστών από τον έναν στον άλλον, οι οποίοι συνδέονται με μία σχέση εμπιστοσύνης. Η τιμή αυτή μπορεί να χρησιμοποιηθεί κατά τη σύσταση αντικειμένων σε ένα χρήστη με τρόπο που περιγράφεται στη συνέχεια. Συγκεκριμένα, η τεχνική βασίζεται και επεκτείνει τη διαδεδομένη τεχνική Collaborative Filtering και μάλιστα την user-based υποκατηγορία.

Όπως και στα περισσότερα RS, θεωρείται πως είναι διαθέσιμος ένας πίνακας αξιολογήσεων R με διαστάσεις $M \times N$, όπου M το πλήθος των χρηστών και N το σύνολο των αντικειμένων. Κάθε θέση $r_{ij} \in [1, 5]$ αυτού του πίνακα περιέχει την αξιολόγηση του χρήστη i για το αντικείμενο j , η οποία μπορεί όμως να είναι και κενή, σε περίπτωση που το αντικείμενο δεν έχει αξιολογηθεί από το χρήστη. Στόχος του RS είναι να προβλέψει τις μελλοντικές προτιμήσεις κάθε χρήστη, δεδομένου του πίνακα R , ή αλλιώς να προβλέψει με επιτυχία τις κενές θέσεις του πίνακα αυτού.

Βάσει του user-based Collaborative Filtering, χρησιμοποιείται μία συνάρτηση ομοιότητας για τη σύγκριση δύο χρηστών. Στην προκειμένη περίπτωση, χρησιμοποιείται η cosine similarity η οποία ορίζεται ως εξής:

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I_{uv}} (r_{ui})^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi})^2}},$$

όπου I_{uv} το σύνολο των αντικειμένων που έχουν αξιολογήσει από κοινού οι χρήστες u, v . Χρησιμοποιώντας την ομοιότητα μεταξύ των χρηστών, η τελική πρόβλεψη ενός χρήστη u για ένα αντικείμενο i που συμβολίζεται ως $\hat{r}_{ui}$ προκύπτει ως:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in U_i} (r_{vi} - \bar{r}_v) w(u, v, i)}{\sum_{v \in U_i} |w(u, v, i)|},$$

όπου $\bar{r}_u$ ο μέσος όρος των αξιολογήσεων του χρήστη u , U_i το σύνολο των χρηστών που έχουν αξιολογήσει το αντικείμενο i και $w(u, v, i)$ ένα βάρος που υπολογίζεται μεταξύ των χρηστών u, v για το αντικείμενο i . Για την ακρίβεια, το βάρος αυτό υπολογίζεται ως εξής:

$$w(u, v, i) = \frac{2 \text{sim}(u, v) \cdot t(u, v)}{\text{sim}(u, v) + t(u, v)}.$$

Αξίζει να σημειωθεί πως οι εξισώσεις αυτές έχουν χρησιμοποιηθεί και στη δουλειά (O'Donovan & Smyth, 2005), όμως σε αυτή χρησιμοποιείται διαφορετικός ορισμός της εμπιστοσύνης μεταξύ των χρηστών.

5. Αξιολόγηση Συστήματος Συστάσεων

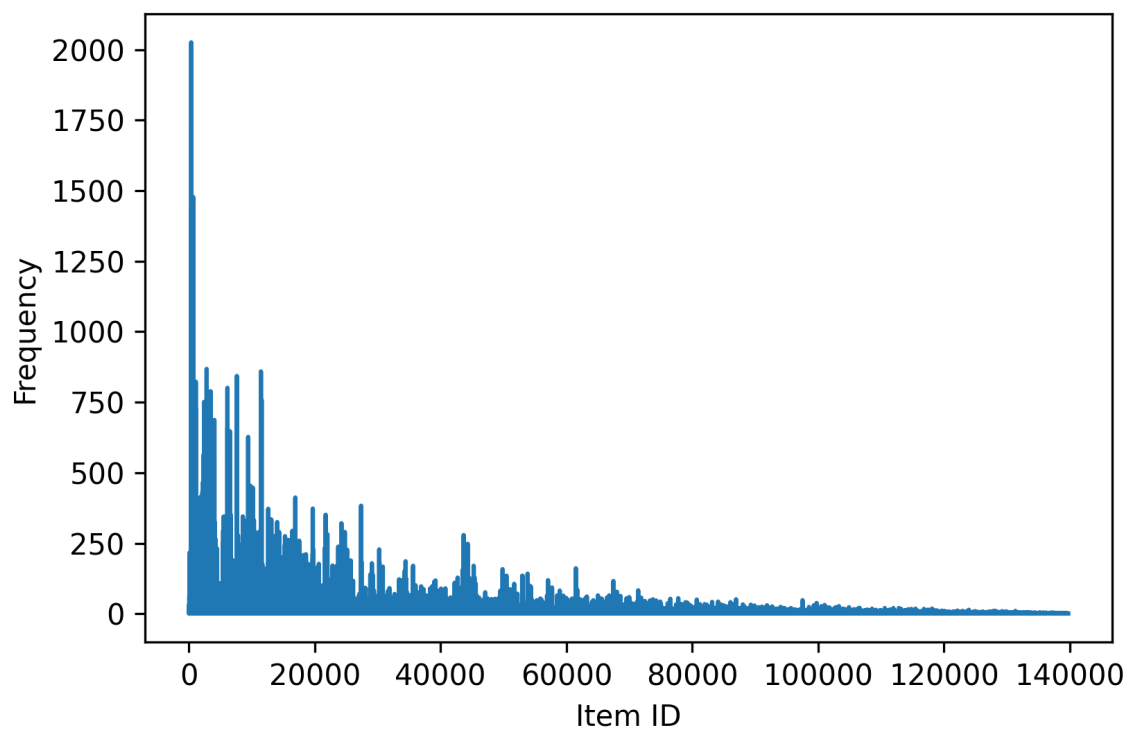
Στο παρόν Κεφάλαιο αναλύεται η διαδικασία αξιολόγησης του προτεινόμενου RS, όπως αυτό περιγράφηκε προηγουμένως. Για την ακρίβεια, περιγράφονται πλήρως λεπτομέρειες που αφορούν στην υλοποίηση του συστήματος, τα δεδομένα που χρησιμοποιούνται για την αξιολόγηση, τις μετρικές αξιολόγησης και, τέλος, τα αποτελέσματα από τα πειράματα που διεξήχθησαν σε σύγκριση με άλλους αλγορίθμους.

5.1 Σύνολο Δεδομένων

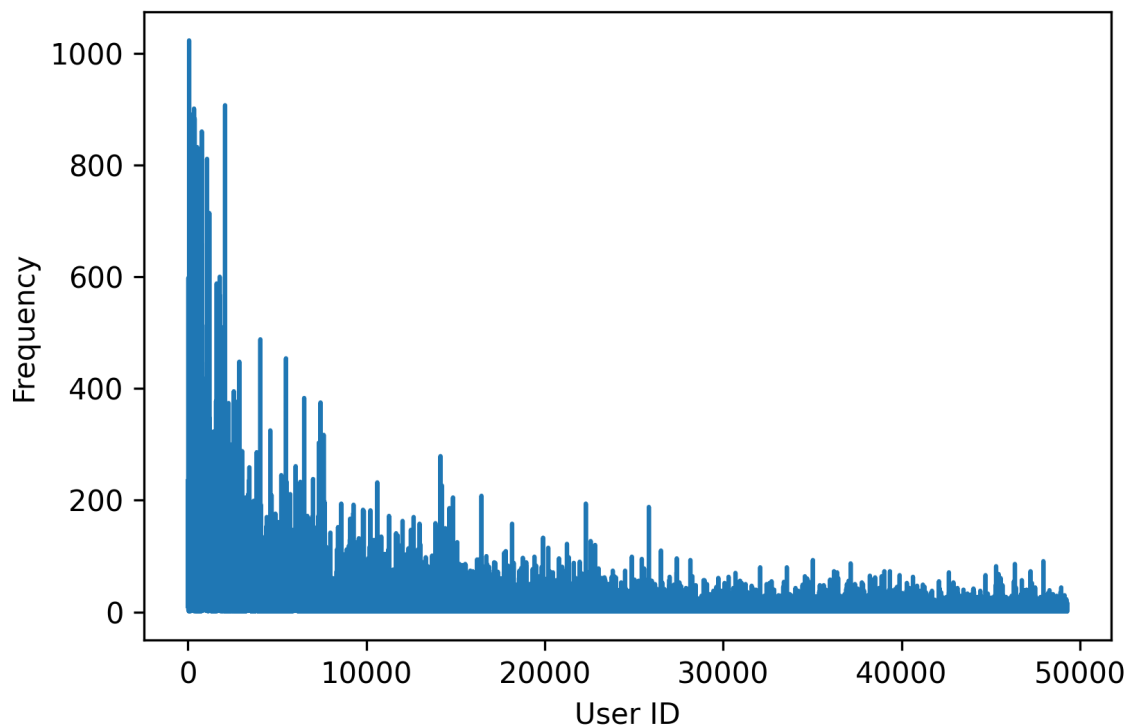
Το σύνολο δεδομένων που επιλέγεται για την αξιολόγηση του συστήματος έχει αντληθεί από το MKΔ Epinions (Massa & Avesani, 2007), το οποίο αποσκοπούσε στην αξιολόγηση αντικειμένων από πλήθος χρηστών. Οι χρήστες, με τη σειρά τους, είχαν τη δυνατότητα να δηλώσουν ρητά ότι εμπιστεύονταν άλλους, αν συμφωνούσαν με τις αξιολογήσεις τους. Αντίστοιχα, μπορούσαν και να αποκλείσουν χρήστες, αν και αυτή η πληροφορία δεν είναι διαθέσιμη στο σύνολο δεδομένων. Αξίζει να σημειωθεί πως το σύνολο δεδομένων είναι διαθέσιμο στο Διαδίκτυο².

Στο εν λόγω σύνολο δεδομένων συμπεριλαμβάνεται τόσο το δίκτυο που διαμορφώνεται από τις σχέσεις εμπιστοσύνης όσο και οι αξιολογήσεις των χρηστών. Έτσι, αποτελεί ένα από τα πιο δημοφιλή σύνολα δεδομένων για έρευνα στο συγκεκριμένο τομέα. Το δίκτυο αποτελείται συνολικά από 49.990 χρήστες (κόμβους) και 487.181 σχέσεις εμπιστοσύνης (ζεύξεις) μεταξύ αυτών. Για αυτούς τους χρήστες συμπεριλαμβάνονται και 664.824 αξιολογήσεις που αφορούν 137.738 διαφορετικά αντικείμενα. Οι αξιολογήσεις ανήκουν στο εύρος [1, 5]. Συνεπώς, κατά μέσο όρο κάθε χρήστης έχει δηλώσει ότι εμπιστεύεται περίπου 10 χρήστες, ενώ έχει αξιολογήσει περίπου 13,49 αντικείμενα, με αποτέλεσμα κάθε αντικείμενο να έχει κατά μέσο όρο 4,76 αξιολογήσεις. Η κατανομή των αξιολογήσεων ανά αντικείμενο και ανά χρήστη αντικατοπτρίζονται στα Σχήματα 4.1 και 4.2, αντίστοιχα.

² http://www.trustlet.org/downloaded_epinions.html



Σχήμα 4.1: Το πλήθος των ratings ανά αντικείμενο.



Σχήμα 4.2: Το πλήθος των ratings ανά χρήστη.

5.2 Υλοποίηση Συστήματος Συστάσεων

Η υλοποίηση του RS λαμβάνει χώρα σε γλώσσα προγραμματισμού Python 3.8, με τη βοήθεια δύο συμπληρωματικών βιβλιοθηκών:

1. Η Surprise³ που παρέχει διάφορα βοηθητικά εργαλεία για την ανάπτυξη RS, καθώς και διαθέτει υλοποιήσεις από διάφορους δημοφιλείς αλγορίθμους, ώστε να επιταχύνει την έρευνα γύρω από αυτά.
2. Η NetworkX⁴ που παρέχει υλοποιήσεις ενός μεγάλου πλήθους αλγορίθμων που σχετίζονται με γράφους. Έτσι, εδώ χρησιμοποιείται για τον υπολογισμό των αποστάσεων μεταξύ των κόμβων, ώστε να υπολογιστούν οι τιμές εμπιστοσύνης που αντιστοιχούν στη συνάρτηση t .

Κάτι που αξίζει να σημειωθεί για τη διαδικασία υλοποίησης είναι πως για να επιταχυνθεί η διαδικασία πειραματισμού, προϋπολογίζονται οι ομοιότητες μεταξύ όλων των χρηστών, που

³ <http://surpriselib.com/>

⁴ <https://networkx.org/>

προκύπτουν από τη συνάρτηση cosine similarity, όπως περιγράφεται σε προηγούμενη ενότητα, καθώς και οι αποστάσεις μεταξύ όλων των αντικειμένων. Έτσι, αντί να υπολογίζονται αυτές οι τιμές κατά την πρόβλεψη της αξιολόγησης κάθε αντικειμένου, γίνεται μία φορά κατά τη φάση εκπαίδευσης του εκάστοτε αλγορίθμου. Όμως, για να γίνει αυτό υπάρχει σημαντική αύξηση της κατανάλωσης μνήμης RAM, αφού απαιτείται ένας $|U| \times |U|$ πίνακας για τις αποστάσεις μεταξύ των χρηστών, όπου στην προκειμένη περίπτωση θα απαιτούνται $|U|^2 = 49.990^2 = 2.499.000.100 \simeq 2.5 \cdot 10^9$. Μέσω αυτού, προκύπτει και ο πίνακας για τις τιμές εμπιστοσύνης μεταξύ κάθε ζεύγους χρηστών και, δεδομένου ότι το σύνολο τιμών της συνάρτησης t ανήκει στο $[0, 1]$, κάθε τιμή αυτού θα αναπαρίσταται μέσω μίας float ή double μεταβλητής. Συνεπώς, ένας τέτοιος πίνακας θα απαιτούσε $4 \cdot 2.5 \cdot 10^9 = 10^{10}$ bytes = 10 GB ή $8 \cdot 2.5 \cdot 10^9 = 20$ GB, αντίστοιχα. Μιας και οι πόροι αυτοί δεν είναι εύκολα διαθέσιμοι και διαχειρίσιμοι, χρησιμοποιούνται μόνο οι πρώτοι 5.000 χρήστες του συνόλου δεδομένων, το οποίο ισχύει τόσο για το δίκτυο εμπιστοσύνης, όσο και για τις αξιολογήσεις. Συμπερασματικά, απομένουν 103.388 αντικείμενα με συνολικό πλήθος αξιολογήσεων ίσο με 281.255, ενώ το δίκτυο των 5.000 εναπομείναντων χρηστών έχει 205.935 ζεύξεις. Είναι φανερό πως, αν και διατηρήθηκε μόνο το $\sim 10\%$ των χρηστών, τα ποσοστά των αντικειμένων, αξιολογήσεων και ζεύξεων που απομένουν στο σύνολο δεδομένων παραμένει σχετικά υψηλό (της τάξης του 40%), λόγω της κατανομής αυτών.

5.3 Μετρικές Αξιολόγησης Προτεινόμενου Συστήματος

Οι κύριες μετρικές αξιολόγησης που χρησιμοποιούνται για την αξιολόγηση των αλγορίθμων είναι αποσκοπούν στην ποσοτικοποίηση του σφάλματος της πρόβλεψης της αξιολόγησης ενός αντικειμένου από ένα χρήστη. Για την επίτευξη αυτού, χρησιμοποιούνται δύο μετρικές, η Mean Absolute Error (MAE) και η Root Mean Squared Error (RMSE).

Η MAE υπολογίζει τη μέση απόλυτη διαφορά των προβλέψεων από τις πραγματικές τιμές και υπολογίζεται ως:

$$MAE = \frac{1}{|\hat{R}|} \sum_{r_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|,$$

όπου $\hat{R}$ το σύνολο των προβλέψεων, r_{ui} η πραγματική και $\hat{r}_{ui}$ η προβλεπόμενη αξιολόγηση του χρήστη u για το αντικείμενο i .

Αντίστοιχα, το RMSE ορίζεται ως:

$$RMSE = \sqrt{\frac{1}{|\hat{R}|} \sum_{r_{ui} \in \hat{R}} (r_{ui} - \hat{r}_{ui})^2}.$$

Το RMSE δίνει περισσότερο βάρος σε προβλέψεις που απέχουν πολύ από την πραγματική τιμή, λόγω του τετραγώνου της διαφοράς τους που χρησιμοποιείται κατά τον υπολογισμό του. Το RMSE θα είναι πάντα μεγαλύτερο ή ίσο του MAE και υποδηλώνει σε κάποιο βαθμό τη διακύμανση των σφαλμάτων, με την έννοια ότι όσο μεγαλύτερη η διαφορά μεταξύ των δύο τιμών, τόσο μεγαλύτερη η διακύμανση των σφαλμάτων. Φυσικά, αφού και οι δύο μετρικές εκφράζουν σφάλμα, όσο μικρότερες οι τιμές (και κοντινότερες στο 0), τόσο το καλύτερο.

5.4 Αποτελέσματα

Για την αξιολόγηση του προτεινόμενου RS, πραγματοποιείται εκτενής πειραματισμός στο προαναφερόμενο σύνολο δεδομένων και τα εξαγόμενα αποτελέσματα συγκρίνονται με άλλους δύο αλγόριθμους που συναντώνται συχνά στη βιβλιογραφία. Σε αυτό το σημείο αξίζει να αποσαφηνιστεί πως ο αλγόριθμος που προτείνεται και παρουσιάζεται στην παρούσα πτυχιακή εργασία στο εξής θα αποκαλείται ως “Trust-based Algorithm”.

Ο πρώτος αλγόριθμος με τον οποίο συγκρίνονται τα αποτελέσματα είναι μία παραλλαγή του γνωστού k-Nearest Neighbors (k-NN), σύμφωνα με την οποία η πρόβλεψη της αξιολόγησης του αντικείμενου i από το χρήστη u προκύπτει από την εξίσωση:

$$\hat{r}_{ui} = \frac{\sum_{v \in N_i^k(u)} \text{sim}(u, v) \cdot r_{vi}}{\sum_{v \in N_i^k(u)} \text{sim}(u, v)},$$

όπου $N_i^k(u)$ οι k (το πολύ) ομοιότεροι γείτονες του u ως προς το αντικείμενο i . Για την ακρίβεια, η γειτονιά αυτή προκύπτει, υπολογίζοντας την ομοιότητα μεταξύ χρηστών (π.χ. με cosine similarity) που έχουν αξιολογήσει το αντικείμενο i .

Ο δεύτερος αλγόριθμος, τα αποτελέσματα του οποίου συγκρίνονται με την προτεινόμενη προσέγγιση, είναι ο Singular Value Decomposition (SVD). Σύμφωνα με αυτόν, γίνεται παραγοντοποίηση του πίνακα αξιολογήσεων R , χρησιμοποιώντας τις ιδιοτιμές και τα ιδιοδιανύσματα αυτού. Όμως, ως αλγόριθμος δεν είναι ιδιαίτερα εύκολα ερμηνεύσιμος χωρίς επαρκή μαθηματική ανάλυση, αλλά ο ενδιαφερόμενος μπορεί να βρει περισσότερες πληροφορίες στη βιβλιογραφία (Lambers, 2011).

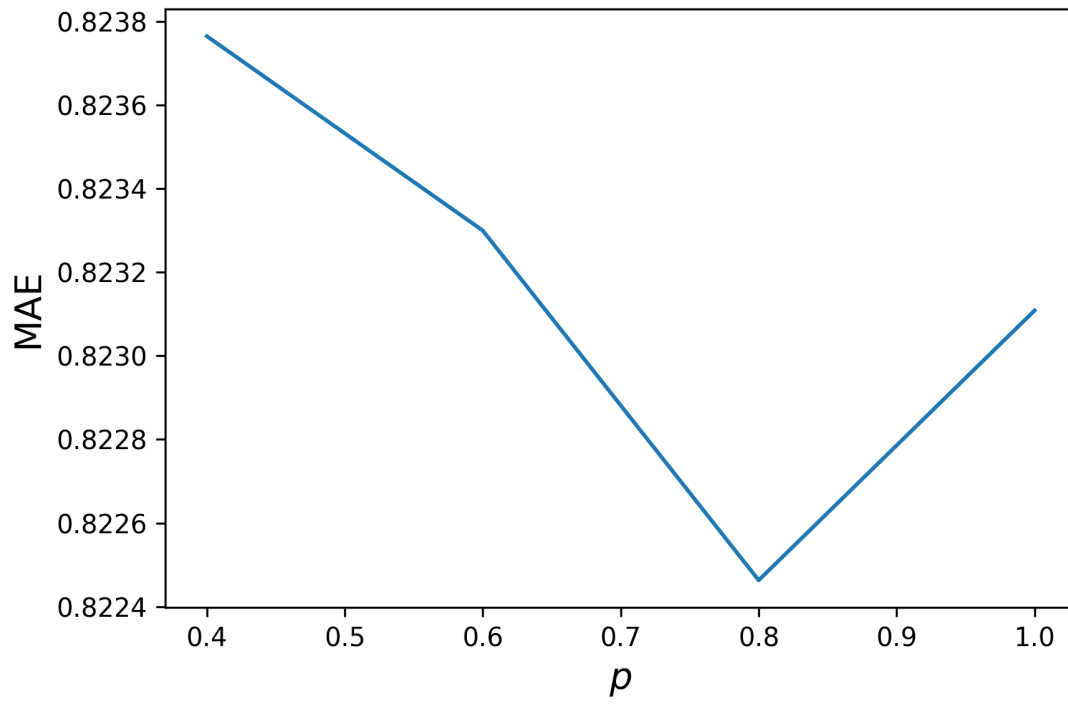
Ανεξάρτητα του αλγορίθμου, χρησιμοποιείται η τεχνική 10-fold cross validation για την αξιολόγησή του. Σύμφωνα με αυτή, το σύνολο δεδομένων μοιράζεται σε 10 ίσα τμήματα και η εκπαίδευση του αλγορίθμου γίνεται σε 9 εξ αυτών, ενώ η αξιολόγηση στο τελευταίο. Έτσι, οι τελικές μετρικές αξιολόγησης προκύπτει ως ο μέσος όρος των τιμών που προέκυψαν για κάθε μέρος του συνόλου δεδομένων, με αποτέλεσμα να παρέχεται μεγαλύτερη αξιοπιστία.

Οι μετρικές που προέκυψαν από τον πειραματισμό για κάθε αλγόριθμο παρουσιάζεται συνοπτικά στον Πίνακα 4.1. Όπως φαίνεται, ο αλγόριθμος που βασίζεται στον k-NN παρουσιάζει τη χειρότερη απόδοση, αφού παρουσιάζει τιμές 0,841 στην MAE και 1,128 στην RMSE για k=100, που φάνηκε να παρουσιάζει ικανοποιητική απόδοση, χωρίς να αυξάνεται σημαντικά ο χρόνος εκτέλεσης. Αντίθετα, ο SVD παρουσιάζει τα υψηλότερα αποτελέσματα με 0,812 MAE και 1,052 RMSE. Ο αλγόριθμος της προτεινόμενης προσέγγισης, που αναγράφεται ως Trust-based Algorithm, παρουσιάζει μία ενδιάμεση απόδοση με 0,822 MAE και 1,099 RMSE.

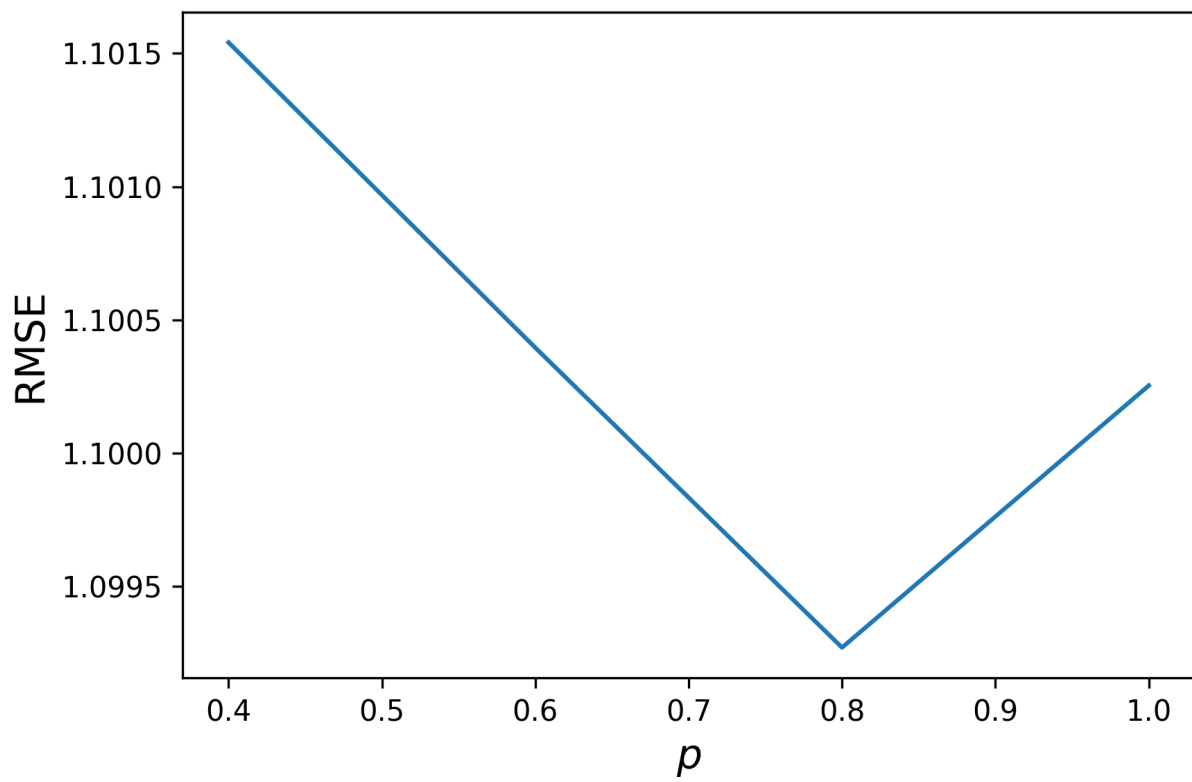
Πίνακας 4.1: Η απόδοση των αλγορίθμων για τις μετρικές αξιολόγησης MAE και RMSE.

Αλγόριθμος	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
Trust-based Algorithm	0,822	1,099
k-NN	0,841	1,128
SVD	0,812	1,052

Συνεχίζοντας, διερευνάται η απόδοση του Trust-based Algorithm καθώς μεταβάλλεται η παράμετρος διάχυσης p που καθορίζει το ρυθμό πτώσης της εμπιστοσύνης μεταξύ δύο χρηστών, καθώς η απόστασή μεταξύ τους αυξάνεται. Συγκεκριμένα, στο Σχήμα 4.3 παρουσιάζεται η απόδοση του αλγορίθμου σύμφωνα με το MAE, ενώ στο Σχήμα 4.4 σύμφωνα με το RMSE. Όπως είναι φανερό, ο αλγόριθμος παρουσιάζει καλύτερη απόδοση και στις 2 μετρικές για $p = 0.8$, αν και δεν υπάρχει μεγάλη απόκλιση στις μετρικές για διαφορετικές τιμές της παραμέτρου διάχυσης.



Σχήμα 4.3: Η τιμή του MAE για τον Trust-Based Algorithm συναρτήσει της παραμέτρου διάχυσης ρ .



Σχήμα 4.4: Η τιμή του RMSE για τον Trust-Based Algorithm συναρτήσει της παραμέτρου διάχυσης p .

6. Συμπεράσματα

Η παρούσα πτυχιακή εργασία πραγματεύεται το ζήτημα της έξυπνης διαφήμισης στα ΜΚΔ και για την ακρίβεια την ενσωμάτωση των RS σε αυτά. Η χρήση των ΜΚΔ είναι πλέον ευρέως διαδεδομένη σε πάσης φύσεως ηλικίες και ιδιαίτερα στις νεαρές. Η χρήση τους εστιάζει στην ανταλλαγή απόψεων, αντιδράσεων και προτιμήσεων, αλλά και στην αγορά και προώθηση προϊόντων και υπηρεσιών, προάγοντας τους κοινωνικούς και επιχειρηματικούς δεσμούς.

Τα πρωτόγνωρα γεγονότα που κλήθηκε να αντιμετωπίσει η ανθρωπότητα τα τελευταία έτη της πανδημίας COVID-19, επέφεραν ραγδαίες αλλαγές/μεταβολές στην καθημερινή ζωή των ατόμων ανά τον κόσμο. Μεταξύ αυτών, κορύφωσαν τη χρήση και την αλληλεπίδραση μέσω των ΜΚΔ, ενέτειναν τις διαδικτυακές αγορές κι έδωσαν κίνητρο για τη διεξαγωγή ερευνών σχετικά με τη μεθοδολογία επίτευξης προσωποποιημένης κι αποδοτικότερης διαφήμισης.

Έτσι, στα πλαίσια της παρούσας πτυχιακής, έγινε μία λεπτομερής παρουσίαση των διαφόρων ειδών RS, των εφαρμογών και το πώς αυτά θα μπορούσαν να αξιοποιηθούν από πλατφόρμες ΜΚΔ για την εξαγωγή προσωποποιημένων συστάσεων για την υποστήριξη και προβολή διαφημίσεων. Με επίκεντρο τη συγκεκριμένη θεματολογία, έχει λάβει χώρα και πληθώρα επιστημονικών δημοσιεύσεων που η μεθοδολογία και τα αποτελέσματα των οποίων συμπεριλαμβάνονται και παρουσιάζονται στην παρούσα εργασία.

Λαμβάνοντας υπόψη τα παραπάνω, κύρια συνεισφορά της εργασίας αποτελεί ο σχεδιασμός και η υλοποίηση ενός RS, το οποίο είναι σε θέση να αξιοποιήσει τις ιδιαιτερότητες των σχέσεων που διαμορφώνονται στα πλαίσια ενός ΜΚΔ. Για την ακρίβεια, το εν λόγω RS αξιοποιεί τις σχέσεις εμπιστοσύνης που μπορεί να εκφράζονται άμεσα ή έμμεσα από τις αλληλεπιδράσεις των χρηστών, ώστε να εξάγει ένα βαθμό ομοιότητας μεταξύ τους, η οποία αξιοποιείται για την πρόβλεψη αξιολογήσεων των διαφόρων αντικειμένων από τους χρήστες.

Το προτεινόμενο RS αξιολογείται πειραματικά και συγκρίνεται με δύο διαδεδομένες μεθοδολογίες στο χώρο των RS. Η αξιολόγηση λαμβάνει χώρα για ένα σύνολο δεδομένων που συναντάται συχνά στη βιβλιογραφία σε αντίστοιχες προσεγγίσεις. Ακόμη, χρησιμοποιούνται δύο μετρικές που ποσοτικοποιούν το σφάλμα των προβλέψεων των αξιολογήσεων του προτεινόμενου RS, ενώ χρησιμοποιείται και η τεχνική cross validation για την εξαγωγή πιο αξιόπιστων συμπερασμάτων. Μέσω της αξιολόγησης, φάνηκε πως η απόδοση του προτεινόμενου RS υπερτερεί της μίας εκ των δύο υπάρχουσών μεθοδολογιών, οπότε και κρίνεται επαρκές για τη χρήση σε μία πλατφόρμα ΜΚΔ.

Αν και τα εξαγόμενα αποτελέσματα είναι ενθαρρυντικά και εν μέρει υποσχόμενα, περαιτέρω αξιολόγηση κρίνεται απαραίτητη για τη βελτίωση αυτών. Η αξιολόγηση αυτή θα μπορούσε να πραγματοποιηθεί, λαμβάνοντας υπόψη διαφορετικά σύνολα δεδομένων, ώστε να φανεί αν οι

σχέσεις που διαμορφώνονται από το ένα ΜΚΔ σε άλλο επιφέρουν κοινά αποτελέσματα. Ακόμη, χρήσιμη κρίνεται η αποδοτικότερη υλοποίηση του αλγορίθμου, αφού φάνηκαν περιορισμοί στην πρακτική εφαρμογή του σε ένα αρκετά μεγάλο σύνολο δεδομένων, τόσο ως προς θέμα μνήμης όσο και χρόνου. Κάτι τέτοιο θα μπορούσε να γίνει αξιοποιώντας τεχνικές που βασίζονται στον k-NN, για τον περιορισμό των όμοιων χρηστών. Ακόμη, η αξιολόγηση θα μπορούσε να γίνει για διαφορετικές μετρικές που συχνά συναντώνται στον τομέα των RS, όπως είναι η ικανότητα του RS να προβλέπει ικανοποιητικά τα top-N αντικείμενα που θα ενδιέφεραν ένα χρήστη, όπου προκύπτουν μετρικές, όπως το diversity και το coverage μεταξύ άλλων.

Βιβλιογραφία

- Γκοτζαμπουγιούκη, Μ. (2009). *Συστήματα Σύστασης σε Ηλεκτρονικές αγορές*. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης. <http://ikee.lib.auth.gr/record/114415/files/ptuxiaki.pdf>
- Λιανέρη, Β. – Α., & Στάμου, Ο. (2013). *Χρήση Κοινωνικών Δικτύων ως Βάση ενός Συστήματος Συστάσεων*. Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.
- Adomavicius, G., & Tuzhilin, A. (2015). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6), 734-749.
- Aggarwal, C. C. (2016). *Recommender systems* (1st ed.). Cham: Springer International Publishing.
- Alleva, J. M., Medoch, M. M., Priestley, K., Philippi, J. L., Hamaekers, J., Salvino, E. N., Humblet, S., & Custers, M. (2021). “I appreciate your body, because...” Does promoting positive body image to a friend affect one’s own positive body image?. *Body Image*, 36, pp. 134-138.
- Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, 48(1), 79-95.
- Bakkar, M. (2020). The Role of COVID-19 Spread in increasing the use of electronic applications. *SSRN*.
- Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, 107(3), pp. 797-817.

- Bellogin, A., Castells, P., & Contador, I. (2011, October). Precision-oriented evaluation of recommender systems: an algorithmic comparison. *Proceedings of the fifth ACM conference on Recommender systems*, 333-336.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4), 331-370.
- Correa, T., Hinsley, A. W., & De Zuniga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in human behavior*, 26(2), pp. 247-253.
- Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How large US companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive*, 9(4).
- Davidson, J., Liebald, B., & Liu, J. (2010). The YouTube video recommendation system. *Proceedings of the fourth ACM conference on Recommender systems*, 293-296.
- Golbeck, J. (2006, May). Generating predictive movie recommendations from trust in social networks. *International Conference on Trust Management*, 93-104.
- Golbeck, J. A. (2005). *Computing and applying trust in web-based social networks*. University of Maryland, College Park.
- Guy, I. (2015). Social Recommender Systems. *Recommender Systems Handbook*, pp. 511-543.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5-53.
- Jamali, M., & Ester, M. (2009, June). Trustwalker: a random walk model for combining trust-based and item-based recommendation. *In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 397-406.

- Jensen, C. S., Snodgrass, R. T., Chomicki, J., Toman, D., Ferrari, E., & Vossen, G. (2009). *Average Precision*. Encyclopedia of Database Systems.
- Kaminskas, M., & Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(1), 1-42.
- Knight-McCord, J., Cleary, D., Grant, N., Herron, A., Lacey, T., Livingston, T., & Emanuel, R. (2016). What social media sites do college students use most. *Journal of undergraduate ethnic minority psychology*, 2(21), 21-26.
- Lambers, J. (2011). *The SVD Algorithm*. Stanford University. <https://web.stanford.edu/class/cme335/lecture6.pdf>
- Lerman, K. (2006). Social networks and social information filtering on digg. arXiv preprint cs/0612046
- Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4), 2065-2073.
- Massa, P., & Avesani, P. (2007, October). Trust-aware recommender systems. *In Proceedings of the 2007 ACM conference on Recommender systems*, 17-24.
- Massa, P., & Bhattacharjee, B. (2004, March). Using trust in recommender systems: an experimental analysis. *International conference on trust management*, 221-235.
- Montaner, M., Lopez, B., & de la Rosa, J. L. (2002, July). Developing trust in recommender agents. *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, 304-305.
- O'Donovan, J., & Smyth, B. (2005, January). Trust in recommender systems. *In Proceedings of the 10th international conference on Intelligent user interfaces*, 167-174.

- O'Keeffe, G. S., & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4), pp. 800-804.
- Patty, J. C., Kirana, E. T., & Giri, M. S. (2018). Recommendations System for Purchase of Cosmetics Using Content-Based Filtering. *International Journal of Computer Engineering and Information Technology*, 10(1), 1-5.
- Perrin, A. (2015). Social media usage. *Pew research center*, 52-68.
- Qualman, E. (2012). *Socialnomics: How social media transforms the way we live and do business*. John Wiley & Sons.
- Ray, S., & Mahanti, A. (2010, January). Improving prediction accuracy in trust-aware recommender systems. *2010 43rd Hawaii international conference on system sciences*, 1-9.
- Schroeder, G., Thiele, M., & Lehner, W. (2011, October). Setting goals and choosing metrics for recommender system evaluations. *UCERSTI2 workshop at the 5th ACM conference on recommender systems*, 53.
- Stefani, M. A., Stefanis, V., & Garofalakis, J. (2019, July). CFRS: A Trends-Driven Collaborative Fashion Recommendation System. *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1-4.
- Whiting, A., & Williams, D. (2013). Why people use social media: a uses and gratifications approach. *Qualitative market research: an international journal*.
- Zhou, X., Hu, Y., Li, Y., Josang, A., & Cox, C. (2012). The state-of-the-art in personalized recommender systems for social networking. *Artificial Intelligence Review*, 37(2), 119-132.

Παραρτήματα

Κώδικας Υλοποίησης Προτεινόμενου Συστήματος Συστάσεων

```
import pandas as pd
from surprise import Dataset, Reader, AlgoBase, PredictionImpossible, SVD
import networkx as nx
import numpy as np
from surprise.model_selection.split import train_test_split
import csv
from surprise.prediction_algorithms.knns import KNNBasic
from surprise.model_selection import cross_validate

MAX_NODE = 5_000

def get_user_item_ratings():
    with open('ratings_data.txt') as f:
        data = f.read().split('\n')

    users, items, ratings = [], [], []
    for line in data:
        if not line:
            continue
        user, item, rating = line.split()
        if int(user) > MAX_NODE:
            continue
        users.append(user)
        items.append(item)
        ratings.append(rating)

    ratings_dict = {'item': items,
                   'user': users,
                   'rating': ratings}
    df = pd.DataFrame(ratings_dict)
    reader = Reader(rating_scale=(1, 5))

    # The columns must correspond to user id, item id and ratings (in that
    order).
```

```

data = Dataset.load_from_df(df[['user', 'item', 'rating']], reader)
return data

def get_trust_data():
    with open('trust_data.txt') as f:
        content = f.read().split('\n')

        edges, nodes = {}, set()
        for line in content:
            frm, to, _ = line.split()
            frm, to = int(frm), int(to)
            if max([frm, to]) > MAX_NODE:
                continue
            if frm not in edges:
                edges[frm] = {}
            edges[frm][to] = {"weight": 1}

        G = nx.from_dict_of_dicts(edges, create_using=nx.DiGraph)
        trusts = dict(nx.algorithms.\
            shortest_paths.all_pairs_shortest_path_length(G))

    return trusts

class MyPredictor(AlgoBase):
    def __init__(self, sim_options={}, bsl_options={}, trusts=None,
prop_par=1.0):
        AlgoBase.__init__(self, sim_options=sim_options,
                           bsl_options=bsl_options)
        self.trusts = trusts
        self.prop_par = prop_par

    def fit(self, trainset):
        AlgoBase.fit(self, trainset)
        self.sim = self.compute_similarities()

        self.the_mean = np.mean([r for (_, _, r) in
                                   self.trainset.all_ratings()])
        self.user_means = {}
        for user in self.trainset.all_users():
            self.user_means[user] = np.mean([

```

```

        x[1] for x in self.trainset.ur[user]
    ])
    for k, v in self.trusts[
        int(self.trainset.to_raw_uid(user))
    ].items():
        self.trusts[
            int(self.trainset.\
                to_raw_uid(user))
        ][k] = self.prop_par**(v-1)

    self.item_means = {}
    for item in self.trainset.all_items():
        self.item_means[item] = np.mean(
            [x[1] for x in self.trainset.ir[item]]
        )

    return self

def estimate(self, u, i):
    if not self.trainset.knows_user(u):
        raise PredictionImpossible('User is unknown.')
    if not self.trainset.knows_item(i):
        raise PredictionImpossible('Item is unknown')
    user_mean = 3
    if u in self.user_means:
        user_mean = self.user_means[u]
    wgt_avg = 0
    wgt_sum = 0

    trusts = dict(self.trusts[int(self.trainset.to_raw_uid(u))])

    for user, rating in self.trainset.ir[i]:
        try:
            sim = self.sim[u, user]
        except IndexError:
            continue
        try:
            trust_score = trusts[int(self.trainset.to_raw_uid(user))]
        except KeyError:
            continue

```

```

        weight = 2 * sim * trust_score/(sim + trust_score)
        wgt_avg += (rating - self.user_means[user]) * weight
        wgt_sum += abs(weight)
    if wgt_avg == 0 and wgt_sum == 0:
        offset = 0
    else:
        try:
            offset = wgt_avg/wgt_sum
        except ZeroDivisionError:
            offset = 0
    return user_mean + offset

def get_topk(self, u, k=5):
    predictions = []
    for i in self.trainset.all_items():
        if i in {x[0] for x in self.trainset.ur[u]}:
            continue
        prediction = self.predict(u, self.trainset.to_raw_iid(i))
        predictions.append((prediction, i))
    return sorted(predictions, reverse=True, key=lambda x: x[0].est)[:k]

data = get_user_item_ratings()
trusts = get_trust_data()
for p in [0.2, 0.4, 0.6, 0.8, 1.0]:
    result = cross_validate(
        MyPredictor(sim_options={'name': 'cosine', 'user_based': True},
                    trusts=trusts, prop_par=p), data, measures=['RMSE',
'MAE'], cv=10)
    print(result)

```