



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΒΙΟΜΗΧΑΝΙΚΗΣ
ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ
ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

**Βιβλιογραφική Μελέτη και Σύγκριση Αποτελεσμάτων σε
Συστήματα Συστάσεων Υλοποιημένα βάσει της Διαχεόμενης
Πληροφορίας**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Βασιλική Μ. Καρακωνσταντή

Επιβλέπων : Συμεών Χρ. Παπαβασιλείου

Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2020



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΒΙΟΜΗΧΑΝΙΚΗΣ
ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ
ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

Βιβλιογραφική Μελέτη και Σύγκριση Αποτελεσμάτων σε Συστήματα Συστάσεων Υλοποιημένα βάσει της Διαχεόμενης Πληροφορίας

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Βασιλική Μ. Καρακωνσταντή

Επιβλέπων : Συμεών Χρ. Παπαβασιλείου

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 6^η Ιουλίου 2020.

.....

Συμεών Παπαβασιλείου

Καθηγητής

.....

Γιώργος Ματσόπουλος

Καθηγητής

.....

Αθανάσιος Παναγόπουλος

Αναπληρωτής Καθηγητής

Αθήνα, Ιούλιος 2020

.....

Βασιλική Μ. Καρακωνσταντή

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

Copyright © Βασιλική Καρακωνσταντή, 2020.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Βιβλιογραφική Μελέτη και Σύγκριση Αποτελεσμάτων σε Συστήματα Συστάσεων Υλοποιημένα βάσει της Διαχεόμενης Πληροφορίας

Στη γιαγιά μου, Βασιλική

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στον τομέα Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Ε.Μ.Π., στα πλαίσια των μεταπτυχιακών σπουδών μου στο Δ.Π.Μ.Σ «Τεχνο-οικονομικά Συστήματα». Με την ευκαιρία που μου δίνεται μέσω αυτής της διπλωματικής εργασίας, θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή του Ε.Μ.Π. κ. Συμεών Παπαβασιλείου και τον κ. Βασίλειο Καρυώτη για την εμπιστοσύνη που μου έδειξαν για την ανάληψη της συγκεκριμένης διπλωματικής και την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα.

Στη συνέχεια, ιδιαίτερες ευχαριστίες οφείλω στους Υποψήφιους Διδάκτορες Μαργαρίτα Βιτοροπούλου και Κωνσταντίνο Τσιτσεκλή, για τη συνεχή καθοδήγηση και την αδιάκοπη υποστήριξή τους για την ολοκλήρωση της μελέτης και τη συγγραφή του παρόντος τόμου. Η ευστοχία των υποδείξεων και των συμβουλών τους υπήρξε καταλυτικής σημασίας για την επιτυχή ολοκλήρωση της εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για όλη την αγάπη, την υποστήριξη και την υπομονή που μου έδειξαν καθ' όλη τη διάρκεια των σπουδών μου, καθώς και όλους εκείνους που συνέβαλαν με τον τρόπο τους στην επιτυχή ολοκλήρωση της προσπάθειάς μου.

Βιβλιογραφική Μελέτη και Σύγκριση Αποτελεσμάτων σε Συστήματα Συστάσεων Υλοποιημένα βάσει της Διαχεόμενης Πληροφορίας

Περίληψη

Τα Συστήματα Συστάσεων έχουν ως στόχο την πρόβλεψη των ενδιαφερόντων ενός χρήστη μιας διαδικτυακής πλατφόρμας ώστε να παράγουν προτάσεις προϊόντων που είναι πολύ πιθανόν να του αρέσουν και θα είναι ικανοποιημένος εάν τα επιλέξει. Έχουν ευρεία εφαρμογή στις μέρες μας και συνήθως χρησιμοποιούνται από ηλεκτρονικά καταστήματα εμπορίου με σκοπό να προωθήσουν τα προϊόντα τους και να αυξήσουν τις πωλήσεις τους. Τα δεδομένα που απαιτούνται για τα Συστήματα Συστάσεων προέρχονται είτε από αξιολογήσεις χρηστών, είτε έμμεσα από αναζητήσεις στις μηχανές αναζήτησης, καθώς και από το ιστορικό των αγορών ή από άλλες πληροφορίες σχετικά με τους ίδιους τους χρήστες και τα αντικείμενα που επιλέγουν. Οι διαδικτυακές πλατφόρμες χρησιμοποιούν αυτά τα δεδομένα για την δημιουργία συστάσεων, ανάλογα με το αντικείμενο του εκάστοτε ιστοτόπου, οι οποίες δεν είναι κοινές σε όλους τους χρήστες, αλλά βασίζονται στις προσωπικές προτιμήσεις του καθενός. Οι συστάσεις συνήθως βοηθούν τις αναζητήσεις του χρήστη και τον διευκολύνουν να έχει πρόσβαση στο περιεχόμενο που τον ενδιαφέρει ενώ, παράλληλα, μπορεί και να τον εκπλήξουν με προϊόντα που δεν θα είχε αναζητήσει ποτέ.

Στην παρούσα διπλωματική εργασία μελετώνται θεωρητικά τα Συστήματα Συστάσεων με ανάλυση των δυνατών και αδύναμων σημείων τους, περιγράφοντας τις κυριότερες μεθόδους με τις οποίες υλοποιούνται: της σύστασης μέσω συνεργασίας, της σύστασης βάσει περιεχομένου και του υβριδικού μοντέλου. Στη συνέχεια, παρουσιάζονται τα Κοινωνικά Συστήματα Συστάσεων στη γενική τους μορφή ώστε να διαπιστωθεί, μέσα από έρευνες της βιβλιογραφίας, η χρησιμότητα των αλληλεπιδράσεων μεταξύ των χρηστών ενός μέσου κοινωνικής δικτύωσης προς βελτιστοποίηση της ακρίβειας των συστάσεων έναντι των κλασικών Συστημάτων Συστάσεων. Περαιτέρω ανάλυση ακολουθεί σχετικά με αυτά τα συστήματα, τα οποία όμως λαμβάνουν υπόψη στο μηχανισμό τους την διάχυση πληροφορίας που παρατηρείται από τις αλληλεπιδράσεις των χρηστών, με αποτέλεσμα να μη χρειάζεται το Σύστημα Συστάσεων να προτείνει προϊόντα σε χρήστες τα οποία θα τα λάβουν έτσι κι αλλιώς από την ανταλλαγή πληροφορίας μεταξύ τους. Στο πλαίσιο αυτό

εξετάζονται δύο τεχνικές αυτής της κατηγορίας και γίνεται σύγκριση των αποτελεσμάτων τους ώστε να αποφανθούμε ποια είναι πιο κατάλληλη για εφαρμογή σε μέσα κοινωνικής δικτύωσης.

Λέξεις-κλειδιά: Συστήματα Συστάσεων, Μέσα Κοινωνικής Δικτύωσης, Διάχυση Πληροφορίας, Άμεση Σύσταση, Έμμεση Σύσταση.

Abstract

Recommender Systems aim to predict the interest of a user in a website in order to increase the efficiency of marketing and create added value for other users. The data required for a Recommender System is provided either by users' ratings, or indirectly from searches in search machines and, also, from market history and other information about the users and the objects they choose. Recommendations speed up users' searches and make it easy for them to access the content they are interested in, while they may also surprise them with items they would never have sought by themselves.

This thesis presents theoretically the Recommender Systems by analyzing their advantages and weaknesses, and also, by describing the main methods used for their implementation: collaborative, content-based and hybrid filtering. Next, the Social Recommender Systems are presented, in their general form, in order to determine, through literature review, the usefulness of the interactions between the users of a social network aiming to optimize the accuracy of the recommendations in comparison to classic Recommender Systems. Further analysis follows regarding these systems, while taking into account the information diffusion, which is observed when users interact with each other. Therefore, the Recommender System does not have to recommend items to users that they will receive anyway by information exchange between them. In this context, two techniques are analyzed, and their results are compared so as to decide which is a better choice to use in social networks.

Keywords: Recommender Systems, Online Social Network, Information Diffusion, Repost, Explicit Recommendation, Implicit Recommendation.

Βιβλιογραφική Μελέτη και Σύγκριση Αποτελεσμάτων σε Συστήματα Συστάσεων Υλοποιημένα βάσει της Διαχεόμενης Πληροφορίας

Περιεχόμενα

1	Εισαγωγή.....	19
2	Εισαγωγή Στα Συστήματα Συστάσεων	21
	2.1 Συστήματα Συστάσεων	21
	2.2 Κατηγοριοποίηση των Συστημάτων Συστάσεων	25
	2.2.1 Σύσταση βάσει περιεχομένου (Content-based Filtering)	27
	2.2.2 Σύσταση μέσω συνεργασίας (Collaborative Filtering)	29
	2.2.3 Υβριδικά Συστήματα Συστάσεων (Hybrid Recommendation Systems).....	33
3	Κοινωνικά Συστήματα Συστάσεων (Social Recommender systems)	39
	3.1 Ανασκόπηση Δικτύου – Κοινωνικού Δικτύου	39
	3.1.1 Ανάλυση Σύνθετων Δικτύων (Complex Network Analysis)	39
	3.1.2 Ανάλυση Κοινωνικών Δικτύων	42
	3.2 Εισαγωγή στα Κοινωνικά ΣΣ (Social RS - SRS)	44
	3.2.1 Βιβλιογραφική Ανασκόπηση Κοινωνικών Συστημάτων Συστάσεων.....	45
	3.2.2 Συμπεράσματα για τα Κοινωνικά Συστήματα Συστάσεων	50
4	Συστήματα Συστάσεων και Διάχυση της πληροφορίας.....	53
	4.1 Διάχυση πληροφορίας σε Online Social Networks.....	53
	4.2 Διάχυση πληροφορίας σε Κοινωνικά ΣΣ (Information diffusion-aware recommendation systems).....	55
	4.2.1 Βιβλιογραφική ανασκόπηση εφαρμογών IDARS.....	56
5	Βασικές Μέθοδοι Συστημάτων Συστάσεων Με Ενσωμάτωση Της Διάχυσης Της Πληροφορίας.....	61
	5.1 Ανάλυση του αλγορίθμου για το πλαίσιο συστάσεων DifRec.....	62
	5.1.1 Αλγόριθμος «Reliable Set»	63
	5.1.2 Αλγόριθμος «Neighborhoodness Graph»	64

5.1.3	Αλγόριθμος «Independent Set»	65
5.1.4	Αλγόριθμος «Recommend»	67
5.1.5	Υπολογισμός του συνολικού Relevance Score του δικτύου.....	67
5.1.6	Σημαντικά συμπεράσματα για τον DifRec	69
5.2	Επίλυση προβλήματος κατανομής αντικειμένων με τη μέθοδο κάλυψης δικτύου	70
5.2.1	Εισαγωγή.....	70
5.2.2	Ανάλυση προβλήματος κάλυψης δικτύου.....	70
5.2.3	Ανάλυση του αλγορίθμου για το πλαίσιο συστάσεων CoveR.....	72
5.3	Αποτελέσματα των μεθόδων σε Scale-Free και Small-World δίκτυα	74
5.3.1	Αποτελέσματα για τον DifRec σε Scale-Free και Small-World Δίκτυα.....	77
5.3.2	Αποτελέσματα για τον CoveR σε Scale-Free και Small-World Δίκτυα	85
5.3.3	Σύγκριση αποτελεσμάτων των αλγορίθμων DifRec και CoveR.....	89
6	Σύνοψη & Συμπεράσματα	94
7	Βιβλιογραφία	99

Ευρετήριο Εικόνων

Εικόνα 2.1 Μέθοδοι Σύστασης μέσω Συνεργασίας και Σύστασης βάσει Περιεχομένου (πηγή: https://datacrunch.nl/2019/01/14/recommendation-engines-a-new-tool-for-effective-marketing/)... 30	30
Εικόνα 2.2 Παράδειγμα Weighted Υβριδικού Συστήματος Συστάσεων, όπου συνδυάζονται 2 διαφορετικά ΣΣ (content based, collaborative filtering). 34	34
Εικόνα 2.3 Κατηγορίες των Συστημάτων Συστάσεων..... 37	37
Εικόνα 3.1 Αναπαράσταση ενός απλοποιημένου Δικτύου. 40	40
Εικόνα 3.2 Σύγκριση τοπολογιών δικτύου βάσει της τυχειότητας [2] 41	41
Εικόνα 3.3 Παράδειγμα Scale-Free Δικτύου. Οι σκιαγραφημένοι κόμβοι είναι τα “hubs” [2]...... 42	42
Εικόνα 3.4 Παράδειγμα Κοινωνικού Δικτύου – Αλληλεπιδράσεις μεταξύ ανθρώπων σε μια μικρή κοινωνία. 43	43
Εικόνα 5.1 Παράδειγμα δημιουργίας ενός γράφου γειτνίασης..... 65	65
Εικόνα 5.2 Παράδειγμα δύο χρηστών a, b που δέχονται ως εξωτερική πρόταση το ίδιο προϊόν Π, αλλά δεν συνδέονται μεταξύ τους στον γράφο γειτνίασης. 66	66
Εικόνα 5.3 Χρήστες που δεν έλαβαν καμία σύσταση προϊόντος – μετρική $ V_0 $, με εφαρμογή του DifRec. 78	78
Εικόνα 5.4 Χρήστες στους οποίους παρατηρήθηκε capacity violation – μετρική $ V_c $, με εφαρμογή του DifRec..... 80	80
Εικόνα 5.5 Μέσος όρος άμεσων και έμμεσων συστάσεων για Scale-Free Δίκτυα, με εφαρμογή του DifRec. 81	81
Εικόνα 5.6 Μέσος όρος άμεσων και έμμεσων συστάσεων για Small-World Δίκτυα, με εφαρμογή του DifRec..... 82	82
Εικόνα 5.7 Αριθμός των προϊόντων που δέχθηκε ένας χρήστης με capacity violation – μετρική $ IV_c $, για Scale-Free δίκτυα, με εφαρμογή του DifRec. 83	83
Εικόνα 5.8 Αριθμός των προϊόντων που δέχθηκε ένας χρήστης με capacity violation – μετρική $ IV_c $, για Small-World δίκτυα, με εφαρμογή του DifRec..... 84	84
Εικόνα 5.9 Μέσο Relevance Score για Scale-Free και Small-World στο πλαίσιο συστάσεων DifRec. 85	85

Εικόνα 5.10 Χρήστες στους οποίους παρατηρήθηκε capacity violation – μετρική $ V_c $, με εφαρμογή του CoveR.	86
Εικόνα 5.11 Μέσος όρος άμεσων και έμμεσων συστάσεων για Scale-Free Δίκτυα, με εφαρμογή του CoveR.....	87
Εικόνα 5.12 Αριθμός των προϊόντων που δέχθηκε ένας χρήστης με capacity violation – μετρική $ V_c $, με εφαρμογή του CoveR.	88
Εικόνα 5.13 Μέσο Relevance δικτύου – μετρική S_C , για DifRec και CoveR.....	90
Εικόνα 5.14 Μέσος όρος άμεσων συστάσεων για DifRec και CoveR.....	91
Εικόνα 5.15 Χρήστες με capacity violation σε DifRec και CoveR.....	93

1 *Εισαγωγή*

Σε μια εποχή όπου το Διαδίκτυο έχει ενταχθεί ως αναπόσπαστο στοιχείο της ζωής μας, το σύνολο των πληροφοριών που λαμβάνουμε για πληθώρα γεγονότων είναι τάξεις μεγέθους μεγαλύτερο από αυτό που λαμβανόταν παραδοσιακά. Κατά συνέπεια, σε περιπτώσεις όπου χρειάζεται να λάβουμε αποφάσεις για κάποιο ζήτημα, είτε απλό, π.χ. η αγορά προϊόντος μέσω ηλεκτρονικού καταστήματος, είτε πιο σημαντικό, π.χ. η εύρεση εργασίας στον τομέα που επιθυμούμε, είναι πλέον επιτακτική η ανάγκη ύπαρξης μηχανισμού για φιλτράρισμα πληροφοριών βάσει διαφόρων κριτηρίων, όπως οι προτιμήσεις του χρήστη.

Για τον λόγο αυτό έχουν δημιουργηθεί τα *Συστήματα Συστάσεων - ΣΣ (Recommender ή Recommendation Systems – RS)*. Τα συστήματα αυτά, με απλά λόγια, βοηθούν τον χρήστη να καταλήξει σε μια απόφαση προτείνοντας του προϊόντα/υπηρεσίες τα οποία προκύπτουν μέσα από διάφορους μηχανισμούς που θα αναλυθούν στα επόμενα κεφάλαια. Χαρακτηριστικό παράδειγμα είναι το Amazon [1]. Όταν ο χρήστης εισέρχεται στην συγκεκριμένη πλατφόρμα, αναζητά κάποιο προϊόν που τον ενδιαφέρει θέτοντας προαιρετικά κάποια κριτήρια (κόστος, μέγεθος, υλικό και άλλα). Τα ΣΣ του προτείνουν προϊόντα με παρόμοια χαρακτηριστικά προκειμένου να έχει τη δυνατότητα να επιλέξει έχοντας ποικιλία επιλογών. Ωστόσο, είναι εύλογο το σύνολο των προϊόντων που προτείνονται από την συγκεκριμένη διαδικασία να μην έχει μεγάλο μέγεθος διότι υπάρχει η πιθανότητα αρνητικού αποτελέσματος και να προκληθεί σύγχυση στο χρήστη. Δηλαδή, οι προτάσεις θα πρέπει να είναι λίγες σε πλήθος, αλλά να είναι και ικανοποιητικές.

Η δομή της διπλωματικής εργασίας έχει ως εξής: στο κεφάλαιο 2 παρουσιάζεται η θεωρητική προσέγγιση των ΣΣ και αναλύονται βασικές κατηγορίες στις οποίες διαχωρίζονται, καθώς επίσης και πλεονεκτήματα και μειονεκτήματα από την εφαρμογή τους. Στο κεφάλαιο 3 αναλύουμε και παραθέτουμε βιβλιογραφικές έρευνες για τα Κοινωνικά ΣΣ (Social Recommender Systems), τα οποία δέχονται ως πρόσθετη είσοδο στον μηχανισμό τους τις αλληλεπιδράσεις των χρηστών σε ένα μέσο κοινωνικής δικτύωσης, ώστε να ενισχυθούν και να βελτιστοποιηθούν τα κλασικά ΣΣ. Αντίστοιχα, στο κεφάλαιο 5 παρουσιάζουμε την έννοια της διάχυσης της πληροφορίας σε ένα μέσο κοινωνικής δικτύωσης και την αντίστοιχη υποκατηγορία των Κοινωνικών ΣΣ, που χρησιμοποιούν την έννοια αυτή στον μηχανισμό τους, τα Information Diffusion-Aware Recommender Systems. Γίνεται, επίσης, βιβλιογραφική ανασκόπηση για τα συστήματα αυτά. Δύο τέτοια συστήματα περιγράφονται στο κεφάλαιο 6 και, ύστερα, συγκρίνονται μέσω προσομοιώσεων σε συνθετικά δίκτυα Small-World και Scale-Free ώστε να διαπιστώσουμε τα πλεονεκτήματα και τα μειονεκτήματα του καθενός και να επιλέξουμε το πιο αποτελεσματικό. Στο τελευταίο κεφάλαιο υπάρχει μια σύνοψη των σημαντικών σημείων των προηγούμενων κεφαλαίων, καθώς επίσης και τα τελικά συμπεράσματα που προκύπτουν στην παρούσα διπλωματική εργασία.

2 *Εισαγωγή Στα Συστήματα Συστάσεων*

2.1 Συστήματα Συστάσεων

Η ψηφιακή επανάσταση των προηγούμενων δεκαετιών είχε ως άμεσο αποτέλεσμα τη σημαντική αύξηση της πληροφορίας που διακινείται ηλεκτρονικά, οδηγώντας σε μια κατάσταση που είναι γνωστή ως πληροφοριακή υπερφόρτωση [2]. Συνεπώς, υπήρξε άμεση ανάγκη για την κατασκευή συστημάτων τα οποία θα μπορούν να αναζητούν, να ταξινομούν και να κατηγοριοποιούν την διαθέσιμη πληροφορία.

Τα Συστήματα Συστάσεων (ΣΣ) έχουν προταθεί ως μια λύση που μπορεί να αντιμετωπίσει σε ένα βαθμό το πρόβλημα της υπερφόρτωσης της πληροφορίας στο Διαδίκτυο και τις επιμέρους πλατφόρμες. Όταν ένας χρήστης επισκέπτεται μία ιστοσελίδα, όπως ένα ηλεκτρονικό κατάστημα λιανεμπορίου, τα ΣΣ έχουν ως σκοπό να του προτείνουν προϊόντα/υπηρεσίες μέσα από διάφορους μηχανισμούς που μπορούν να εφαρμοστούν, με τέτοιο τρόπο ώστε να είναι «εύστοχα» και ο χρήστης να επιλέξει τελικά κάποιο εξ αυτών και να είναι ικανοποιημένος. Ένας δημοφιλής ορισμός των ΣΣ βασίζεται στη χρησιμότητα ενός προϊόντος για έναν χρήστη [2]:

Ορισμός: (Πρόβλημα συστάσεων) Έστω ένας χρήστης $u \in U$, όπου U το σύνολο όλων των χρηστών, και ένα προϊόν $i \in I$, όπου I το σύνολο όλων των προϊόντων που πρόκειται να προταθούν. Τότε, έστω ότι $f(u, i)$ η συνάρτηση που υπολογίζει τη χρησιμότητα του προϊόντος i στον χρήστη u , $f : U \times I \rightarrow R$, όπου R είναι ένα πλήρως διατεταγμένο σύνολο μη-αρνητικών πραγματικών ή ακέραιων αριθμών. Ο στόχος του προβλήματος

συστάσεων είναι να βρεθεί για κάθε χρήστη $u \in U$ το προϊόν $i' \in I$ από όλα τα στοιχεία του I , το οποίο μεγιστοποιεί την αναμενόμενη χρησιμότητα για το συγκεκριμένο χρήστη u , και συγκεκριμένα ένα προϊόν τέτοιο ώστε $\forall u \in U: i'_u = \operatorname{argmax}_{i \in I} f(u, i)$.

Με τα ΣΣ επιδιώκεται η καλύτερη δυνατή σύσταση που ταιριάζει στο χρήστη προκειμένου αφενός να επιτευχθεί πιο εύκολα και σίγουρα μια αγορά, αφετέρου ο χρήστης να μείνει ευχαριστημένος από την αγορά του προϊόντος ή της υπηρεσίας που του προτάθηκε ώστε να επανέλθει στη συγκεκριμένη πλατφόρμα που έκανε την αγορά, έχοντας στο ιστορικό του μία καλή εμπειρία αγοράς, αλλά και εμπιστοσύνη στο ΣΣ. Πολλές φορές, ένας παράγοντας που μπορεί να ληφθεί υπόψιν είναι οι προτιμήσεις του παρελθόντος. Είναι ένας καλός δείκτης για αγορές που μπορεί να γίνουν στο μέλλον από το συγκεκριμένο χρήστη. Εναλλακτικά, μπορούν να προτείνονται προϊόντα που βασίζονται σε συγκεκριμένες απαιτήσεις του χρήστη αντί να ληφθεί υπόψιν το ιστορικό του.

Σημαντικό ρόλο, παρόλα αυτά, διαδραματίζει η ευκολία που παρέχει το Διαδίκτυο στους χρήστες να εκφέρουν τη γνώμη τους με την επιλογή «Μου αρέσει» (like) και «Δε μου αρέσει» (dislike) ή με βαθμολόγηση σε κάποια κλίμακα. Σε περίπτωση όμως που κάποια αγορά δεν είναι καλή και βαθμολογηθεί αρνητικά από τον χρήστη είναι και αυτό σημαντικό, καθώς το σύστημα θα προχωρήσει σε διορθωτικές κινήσεις μέσω ενημέρωσης του προφίλ του χρήστη αλλά και του συσχετισμού του με νέα άτομα ή ομάδες ατόμων. Η αξιολόγηση αυτή των προϊόντων από πληθώρα χρηστών έχει βοηθήσει στη δημιουργία ομάδων από χρήστες με παρόμοια ενδιαφέροντα για προϊόντα. Τα ενδιαφέροντα και οι ενέργειες αυτών των ομάδων μπορούν να αξιοποιηθούν κάνοντας συστάσεις σε μεμονωμένα μέλη. Είναι εμφανές, λοιπόν, πόσο διαφορετικοί τρόποι υπάρχουν για να δημιουργηθούν μηχανισμοί συστημάτων συστάσεων, με τον καθένα να έχει διαφορετική τεχνική για το πως θα καταφέρει να προσελκύσει το ενδιαφέρον του χρήστη – πελάτη. Ένα βασικό κριτήριο επιλογής του πιο κατάλληλου μηχανισμού ΣΣ μπορεί να είναι η *συνάφεια* (Relevance Score) [3], η οποία, με απλά λόγια, δείχνει πόσο σχετικό είναι ένα προϊόν με κάποιον χρήστη. Είναι λογικό οι χρήστες να καταναλώνουν προϊόντα που βρίσκουν πιο

ενδιαφέροντα. Για αυτό το λόγο, το Relevance Score είναι ένας όρος που θα χρησιμοποιηθεί εκτενώς στην παρούσα εργασία.

Ωστόσο, ίσως κάποιες φορές να μην αρκεί να εφαρμοστεί αυτό το κριτήριο μεμονωμένα [3]. Ένα δευτερεύον κριτήριο που μπορεί να ληφθεί είναι η *καινοτομία* (novelty). Πιο συγκεκριμένα, τα ΣΣ κάποιες φορές μπορεί να είναι χρήσιμα όταν το προϊόν που προτείνουν δεν το έχει δει ξανά στο παρελθόν ο χρήστης. Για παράδειγμα, οι δημοφιλείς ταινίες ενός προτιμώμενου είδους είναι σπάνιο να μην είναι ήδη γνωστές στο χρήστη. Η επαναλαμβανόμενη σύσταση δημοφιλών προϊόντων μπορεί, επίσης, να οδηγήσει στη μείωση της ποικιλίας των προϊόντων προς πώληση. Αξίζει, όμως, να σημειωθεί ότι η αναγκαιότητα και η αποτελεσματικότητα αυτού του κριτηρίου διαφέρει ανά περίπτωση και ανάλογα με τις επιθυμίες του χρήστη, καθώς υπάρχει περίπτωση να επιφέρει αρνητικά αποτελέσματα και να προκαλέσει δυσαρέσκεια στο χρήστη. Παραδείγματος χάριν, κάποιος χρήστης πιθανώς πράγματι να επιθυμεί να του προταθούν οι δημοφιλείς ταινίες στο Netflix καθώς, μπορεί, για παράδειγμα, η έλλειψη ελεύθερου χρόνου στην καθημερινότητά του να μην του επιτρέπει να κάνει εκτενή αναζήτηση και, τότε, οι δημοφιλείς ταινίες είναι μία ασφαλής επιλογή με πολύ μεγάλη πιθανότητα να έχει μια ευχάριστη εμπειρία. Από την άλλη πλευρά, υπάρχουν χρήστες που είναι λάτρεις του κινηματογράφου και αφιερώνουν σημαντικό μέρος του χρόνου τους σε αυτή τη δραστηριότητα. Έχουν καταφέρει να παρακολουθήσουν έναν μεγάλο αριθμό ταινιών, συμπεριλαμβανομένου και των δημοφιλών, όποτε η πρόταση μιας λιγότερο γνωστής ταινίας θα ήταν ιδανική σε αυτήν την περίπτωση. Σε άλλο τομέα, όπως η αγορά μιας ηλεκτρικής συσκευής, πρέπει να υπάρχει διαφορετική αντιμετώπιση και η έννοια ενός δημοφιλούς προϊόντος διαφοροποιείται. Από τη στιγμή που το προϊόν που αγοράζεται μπορεί να πάθει κάποια βλάβη, η επιλογή αγοράς πρέπει να είναι αρκετά προσεκτική, μελετώντας τα τεχνικά χαρακτηριστικά και την αξιοπιστία της εταιρείας που τα κατασκευάζει, αφού μπορεί να οδηγήσει σε υλική ζημία (μια όχι τόσο καλή ταινία θα αφήσει απλώς αίσθημα δυσαρέσκειας). Συνεπώς, είναι εμφανές ότι εξαρτάται από τις εκάστοτε συνθήκες η χρησιμότητα του κριτηρίου της καινοτομίας, και, μάλιστα, θα ήταν συνετό να εφαρμοστεί συνδυαστικά με άλλα κριτήρια.

Επίσης, ένα άλλο κριτήριο είναι η *έκπληξη* (serendipity) [3], βάσει του οποίου οι συστάσεις που γίνονται στον χρήστη είναι μη αναμενόμενες, αλλά δημιουργώντας του θετική έκπληξη για αυτά τα προϊόντα – αυτή είναι και η διαφορά από το κριτήριο της καινοτομίας. Υπάρχει πιθανότητα ένας χρήστης να επιλέγει προϊόντα συγκεκριμένου τύπου, παρόλο που μπορεί τελικά να ενδιαφερθεί για διαφορετικά προϊόντα. Παραδείγματος χάριν, αν κάποιος χρήστης προτιμάει να καταναλώνει ένα συγκεκριμένο είδος φαγητού, έστω ινδικό, και του γίνεται σύσταση για ένα νέο ινδικό εστιατόριο που βρίσκεται κοντά του, τότε η σύσταση χαρακτηρίζεται από καινοτομία και όχι απαραίτητα από έκπληξη. Από την άλλη πλευρά, όταν στον ίδιο χρήστη προτείνεται διαφορετική κουζίνα, έστω ελληνική, και του ήταν άγνωστο ότι μπορεί να τον προσελκύσει και να του αρέσει, τότε πρόκειται για έκπληξη. Το κριτήριο αυτό έχει το πλεονέκτημα ότι αυξάνει την ποικιλότητα των προτεινόμενων προϊόντων, καθώς, επίσης, δίνει κίνητρο στον χρήστη να ξεκινήσει να επιλέγει νέα και διαφορετικά ενδιαφέροντα. Παρόλα αυτά, οι αλγόριθμοι που χρησιμοποιούν τέτοιου είδους συστάσεις, συχνά έχουν την τάση να προτείνουν προϊόντα που απέχουν αρκετά από τα ενδιαφέροντα του χρήστη, για αυτό και προτείνεται να χρησιμοποιούνται σε συνδυασμό με άλλα κριτήρια.

Τέλος, παρεμφερή σκοπό με τα προηγούμενα κριτήρια έχει η *ποικιλία* (diversity) των συστάσεων. Τα ΣΣ τυπικά προτείνουν μία λίστα από κάποια κορυφαία προϊόντα, που μπορεί να αρέσουν στον χρήστη. Όταν αυτά τα προτεινόμενα προϊόντα είναι όμοια, αυξάνεται το ρίσκο να μην αρέσει στον χρήστη κανένα από αυτά. Από την άλλη πλευρά, αν η προτεινόμενη λίστα περιέχει προϊόντα διαφορετικού τύπου, υπάρχει μεγαλύτερη πιθανότητα να αρέσει στον χρήστη τουλάχιστον ένα.

Αξίζει σε αυτό το σημείο να παρατεθούν τα ευεργετικά σημεία των Συστημάτων Συστάσεων [4]. Η εφαρμογή τους μπορεί να προσφέρει πληθώρα πλεονεκτημάτων, όπως η αύξηση των πωλήσεων με την ταυτόχρονη βελτίωση της εμπειρίας και της ικανοποίησης του πελάτη. Επιπροσθέτως, μέσα από τις διάφορες τεχνικές σύστασης, υπάρχει η δυνατότητα πώλησης πολλών διαφορετικών προϊόντων. Μέσα από το μηχανισμό αυτό προτείνονται στον χρήστη όχι μόνο τα πιο δημοφιλή προϊόντα, αλλά και άλλα λιγότερο

δημοφιλή τα οποία ίσως να ήταν δύσκολο να εντοπίσει κανείς με άλλον τρόπο. Ωστόσο, όπως εξηγήσαμε παραπάνω, η ποικιλία σε προϊόντα ίσως σε κάποιες περιπτώσεις να επιφέρει και αρνητικές συνέπειες, ανάλογα βεβαίως με τις συνθήκες και τις ανάγκες του εκάστοτε χρήστη. Επίσης, εφόσον ο χρήστης είναι ικανοποιημένος από μία ιστοσελίδα και από τις προτάσεις που του γίνονται μέσα από αυτή, θα ανατρέξει ξανά σε αυτήν σε μελλοντική περίοδο, θα έχει δηλαδή εμπιστοσύνη σε αυτήν και στα προϊόντα-υπηρεσίες που του προσφέρει.

2.2 Κατηγοριοποίηση των Συστημάτων Συστάσεων

Υπάρχουν τρεις βασικοί τύποι συστάσεων [5], [6], όπως θα αναλύσουμε εκτενέστερα σε παρακάτω υποενότητες. Επιγραμματικά, η πρώτη κατηγορία είναι τα **ΣΣ βάσει περιεχομένου** (Content-based RS), στην οποία ο μηχανισμός δημιουργίας συστάσεων εστιάζει στην δημιουργία ενός προφίλ για κάθε χρήστη αναλύοντας τα χαρακτηριστικά των προϊόντων που οι χρήστες έχουν βαθμολογήσει στο παρελθόν. Η δεύτερη κατηγορία είναι τα **ΣΣ μέσω συνεργασίας** (Collaborative-filtering RS), όπου σε αντίθεση με τα ΣΣ βάσει περιεχομένου, επικεντρώνονται στην ομοιότητα μεταξύ προϊόντων, καθώς επίσης και στην ομοιότητα των αξιολογήσεων των χρηστών σε προϊόντα, δημιουργώντας «κοινότητες» χρηστών με παρεμφερή ενδιαφέροντα. Έτσι, για κάποιο συγκεκριμένο χρήστη επιλέγεται να προταθούν προϊόντα τα οποία κάποιος άλλος χρήστης από την «κοινότητα κοινών ενδιαφερόντων» έχει ήδη επιλέξει, όποτε είναι πολύ πιθανό να αρέσει και στον πρώτο χρήστη αφού έχουν παρόμοιες προτιμήσεις.

Επιπρόσθετα, στο σημείο αυτό μπορούμε να πούμε ότι υπάρχουν και άλλες δύο κατηγορίες, όχι τόσο διαδεδομένες όσο οι προηγούμενες. Συγκεκριμένα, είναι τα **ΣΣ βάσει γνώσης** (knowledge-based RS), στα οποία οι συστάσεις βασίζονται σε απαιτήσεις καθορισμένες από τον χρήστη και όχι από το ιστορικό αξιολογήσεών του, ενώ συνήθως αφορά προϊόντα που δεν είναι συχνή η αγορά τους [3]. Ένα παράδειγμα είναι η αγορά κατοικίας, όπου ο χρήστης θέτει κάποιους περιορισμούς, όπως μέγιστη τιμή, διαρρύθμιση του χώρου,

τοποθεσία και άλλα. Άλλη μία κατηγορία είναι τα **δημογραφικά ΣΣ** (Demographic RS), τα οποία κατηγοριοποιούν τους χρήστες με βάση τα δημογραφικά χαρακτηριστικά τους, προσπαθώντας να αντιστοιχίσουν συγκεκριμένα δημογραφικά στοιχεία σε αξιολογήσεις και καταναλωτική συμπεριφορά.

Τέλος, υπάρχουν τα **υβριδικά ΣΣ** (hybrid filtering RS), τα οποία συνδυάζουν διάφορες προσεγγίσεις ΣΣ εκμεταλλευόμενα τα πλεονεκτήματά τους και ελαχιστοποιώντας, όσο είναι δυνατόν, τα αρνητικά τους σημεία.

Τα βασικά μοντέλα για ΣΣ βασίζονται σε δύο είδη δεδομένων, τα οποία είναι: α) οι αλληλεπιδράσεις μεταξύ χρηστών και αντικειμένων, όπως οι αξιολογήσεις ή η καταναλωτική συμπεριφορά, και β) οι πληροφορίες χαρακτηριστικών των χρηστών και των προϊόντων, όπως λέξεις-κλειδιά, που παραθέτει ο χρήστης στο προφίλ του, και μπορούν να αντιστοιχιστούν με περιγραφές προϊόντων ώστε να προταθούν στον χρήστη. Η προσέγγιση με τις λέξεις-κλειδιά μπορεί να αποδειχθεί χρήσιμη καθώς δε χρησιμοποιεί αξιολογήσεις προϊόντων και, σε περίπτωση που ένας χρήστης είναι νέος σε μια πλατφόρμα, συνεπώς, δεν διατίθενται πληροφορίες για παρελθοντικές του αγορές, διευκολύνεται η διαδικασία συστάσεων προϊόντων [3]. Σε όλες τις περιπτώσεις, ο κύριος στόχος είναι η μεγιστοποίηση των πωλήσεων, αλλά με ταυτόχρονη ικανοποίηση του χρήστη-καταναλωτή. Ωστόσο, αξίζει να αναφέρουμε ότι δεν είναι πάντα άμεσο κίνητρο το κέρδος σε ένα ΣΣ. Ένα τέτοιο παράδειγμα είναι οι προτάσεις φιλίας σε ένα μέσο κοινωνικής δικτύωσης. Ενώ μια πρόταση προϊόντος αυξάνει το εμπορικό κέρδος διευκολύνοντας τις πωλήσεις προϊόντων, η αύξηση του αριθμού των κοινωνικών διασυνδέσεων ενός χρήστη βελτιώνει την εμπειρία του σε ένα μέσο κοινωνικής δικτύωσης, ενθαρρύνοντας την ανάπτυξη του τελευταίου και, κατά συνέπεια, την αύξηση των διαφημιστικών εσόδων τους. Έτσι, δημιουργείται κέρδος μέσα από τις διαφημίσεις για την εταιρεία που δημιούργησε τον ιστότοπο, καθώς, επίσης, και κέρδος από πιθανές πωλήσεις για τις εταιρείες που προβάλλουν τα προϊόντα τους.

2.2.1 Σύσταση βάσει περιεχομένου (Content-based Filtering)

Πρόκειται για ένα σύνολο από τεχνικές που βασίζονται στο προφίλ του χρήστη χρησιμοποιώντας χαρακτηριστικά των αντικειμένων που ο χρήστης έχει βαθμολογήσει στο παρελθόν. Με αυτόν τον τρόπο, αντικείμενα που είναι περισσότερο σχετικά με τα θετικά βαθμολογημένα αντικείμενα, προτείνονται στον χρήστη [7]. Χρησιμοποιούνται σε πληθώρα τομέων δραστηριότητας, όπως για να προτείνουν ιστοτόπους, νέα άρθρα, εστιατόρια, τηλεοπτικά προγράμματα και προϊόντα προς πώληση.

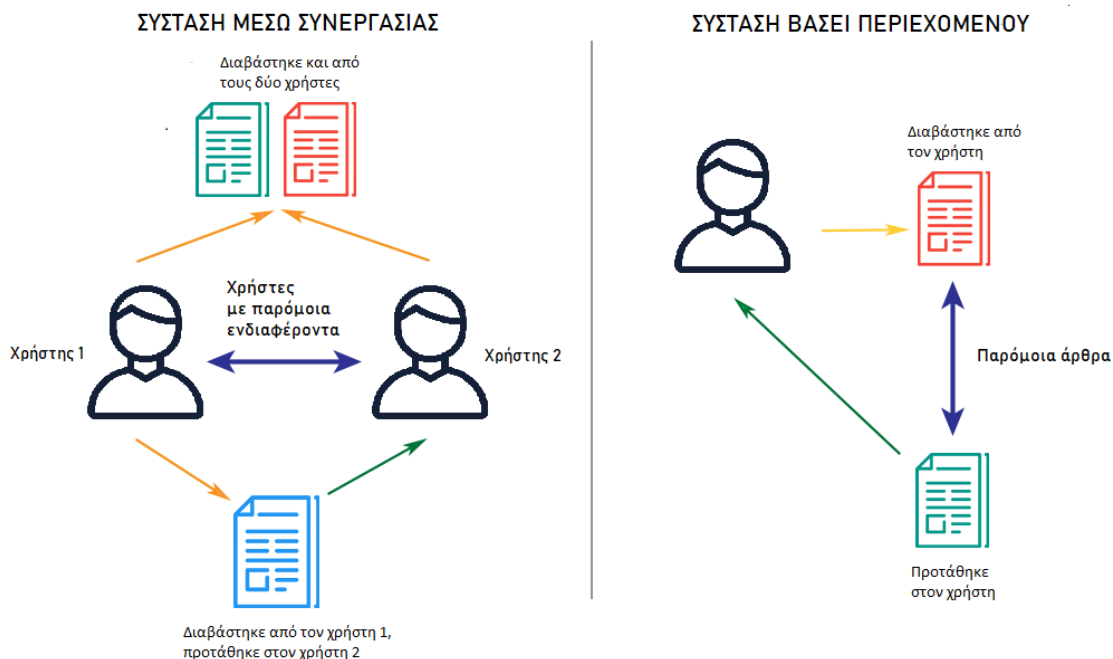
Η τεχνική αυτή δεν προϋποθέτει να υπάρχουν πληροφορίες για τα προφίλ άλλων χρηστών, καθώς, στην περίπτωση αυτή, η σύσταση δεν επηρεάζεται από αυτά. Επίσης, αυτή η προσέγγιση είναι χρήσιμη όταν ένα προϊόν είναι καινούργιο και, ως εκ τούτου, υπάρχουν ελάχιστες αξιολογήσεις διαθέσιμες για αυτό. Αυτό συμβαίνει διότι άλλα προϊόντα με παρόμοια χαρακτηριστικά ίσως έχουν βαθμολογηθεί στο παρελθόν από τον χρήστη. Ως εκ τούτου, το μοντέλο αυτό θα αξιοποιήσει αυτές τις αξιολογήσεις σε συνδυασμό με τα χαρακτηριστικά του προϊόντος ώστε να εξάγει ακριβή σύσταση, ακόμα κι αν δεν υπάρχει καταχωρημένο ιστορικό αξιολογήσεων για το προϊόν. Επίσης, αν αλλάξουν οι προτιμήσεις αυτές, ο αλγόριθμος έχει την ικανότητα να προσαρμόσει τις συστάσεις σε μικρό χρονικό διάστημα.

Το μειονέκτημα, ωστόσο, είναι η ανάγκη πολύ καλής γνώσης των χαρακτηριστικών ενός προφίλ. Απαιτείται πλούσια περιγραφή των αντικειμένων και ένα πολύ καλά οργανωμένο προφίλ χρήστη προτού γίνει η σύσταση αντικειμένων. Ιδιαίτερα όταν ένας χρήστης είναι νέος στο σύστημα, το τελευταίο δεν θα είναι ικανό να παρέχει αξιόπιστες προτάσεις. Παράλληλα, απαιτείται πολύ καλή γνώση του τομέα τον οποίο αφορά η σύσταση. Παραδείγματος χάριν, για πρόταση ταινιών, το σύστημα πρέπει να έχει πληροφορίες σχετικά με σημαντικές έννοιες του κινηματογράφου, όπως να γνωρίζει τους ηθοποιούς, τους σκηνοθέτες, κατηγορίες, αλλά και συνεχώς να ανανεώνεται το περιεχόμενο καθώς τα δεδομένα αλλάζουν πολύ συχνά.

Σε μια συγκεκριμένη μελέτη [8] επισημαίνονται αναλυτικά τα θετικά και τα αρνητικά σημεία που χαρακτηρίζουν τα συστήματα αυτά. Παρουσιάζονται, επίσης, διάφοροι αλγόριθμοι εκμάθησης των προφίλ των χρηστών, όπως δέντρα αποφάσεων, αλγόριθμος του κοντινότερου γείτονα, καθώς και στοχαστικές μέθοδοι. Οι αλγόριθμοι αυτοί επιλέγονται ανάλογα με τον τρόπο που παρουσιάζονται τα προϊόντα, καθώς υπάρχουν πολλές εναλλακτικές με τις οποίες μπορούν να προβληθούν. Για παράδειγμα, υπάρχουν προϊόντα που προτείνονται σε έναν χρήστη που συχνά είναι «τακτοποιημένα» και αποθηκευμένα σε πίνακα βάσης δεδομένων, όπως ένας πίνακας που περιέχει διάφορα εστιατόρια κι έχει καταγεγραμμένα όχι μόνο τα ονόματα των εστιατορίων, αλλά και άλλα χαρακτηριστικά τους (είδος φαγητού, τρόπος εξυπηρέτησης, και κοστολόγιο). Επίσης, για κάθε ονομασία προϊόντων αντιστοιχίζεται ένας μοναδικός αριθμός, ώστε να μην υπάρχει σύγχυση όταν κάποια προϊόντα έχουν πανομοιότυπο όνομα. Ωστόσο, υπάρχουν και περιπτώσεις όπου τα προϊόντα δεν έχουν δομημένα και καλώς ορισμένα χαρακτηριστικά που θα μπορούσαν να αντιστοιχιστούν εύκολα με τις προτιμήσεις του χρήστη ώστε να προταθούν σε αυτόν, καθιστώντας πιο περίπλοκη τη διαδικασία. Ένα τέτοιο παράδειγμα είναι τα άρθρα ειδήσεων, τα οποία είναι κείμενα χωρίς περιορισμούς και δεν ορίζονται ξεκάθαρα χαρακτηριστικά που θα μπορούσαν να τα συμπεριλάβουν σε υποκατηγορίες άρθρων που ορισμένοι χρήστες επιθυμούν να διαβάσουν. Δηλαδή, ένα άρθρο ειδήσεων είναι μια γενική κατηγορία και δεν μπορεί να εξαχθεί εύκολα από το περιεχόμενό του συμπέρασμα για το είδος των ειδήσεων που παραθέτει (πολιτικές, διεθνείς, τοπικές, περιβαλλοντικές και άλλα). Επίσης, υπάρχει πολυπλοκότητα στις λέξεις (η ίδια λέξη μπορεί να έχει πολλές διαφορετικές έννοιες ή πολλές διαφορετικές λέξεις να έχουν την ίδια σημασία). Συνεπώς, είναι δύσκολο να εντοπιστούν χαρακτηριστικές λέξεις από το ΣΣ ώστε να καταφέρει να προτείνει επιτυχώς το άρθρο σε κάποιον χρήστη. Για αυτό το λόγο, από τους διάφορους μηχανισμούς που προτείνονται, κάθε φορά επιλέγεται ο πιο κατάλληλος βάσει του τρόπου αναπαράστασης των προϊόντων.

2.2.2 Σύσταση μέσω συνεργασίας (Collaborative Filtering)

Η τεχνική αυτή βασίζεται στη δημιουργία ενός πίνακα χρηστών-αντικειμένων όσον αφορά την προτίμηση των αντικειμένων από τους χρήστες. Με απλά λόγια, αξιοποιούνται οι αξιολογήσεις από πολλούς χρήστες σε ένα κλίμα συνεργασίας ώστε να γίνει πρόβλεψη αξιολογήσεων που δεν είναι διαθέσιμες. Στην Εικόνα 2.1 μπορεί κανείς να διακρίνει εύκολα τη μέθοδο αυτή (αφορά σε μία από τις υποκατηγορίες της που θα αναλύσουμε παρακάτω), καθώς και την προηγούμενη και να κατανοήσει τη διαφορά τους. Ο Schafer [9] ορίζει ως πηγή δημιουργίας της τεχνικής αυτής μια συνήθεια των ανθρώπων, που υφίσταται για αιώνες – η ανταλλαγή απόψεων μεταξύ τους. Επίσης, υποδεικνύονται βασικές μέθοδοι με τις οποίες εφαρμόζονται αυτά τα συστήματα. Επισημαίνεται, τέλος, ότι με την πάροδο των ετών και όσο ο όγκος των πληροφοριών αυξάνεται, η σύσταση μέσω συνεργασίας θα είναι αναγκαία και θα υπάρχει σε όλες σχεδόν τις τεχνικές που θα εφαρμόζονται για σύσταση προϊόντων. Η σύσταση βάσει περιεχομένου είναι αποτελεσματική σε ορισμένες περιπτώσεις, που εξηγήσαμε παραπάνω όπου υπερτερεί της σύστασης μέσω συνεργασίας. Αυτό συμβαίνει καθώς, πλέον, οι διαδικτυακές πλατφόρμες χρησιμοποιούνται από πληθώρα χρηστών που αυξάνονται συνεχώς με γεωμετρική πρόοδο. Συνεπώς, η δημιουργία κοινοτήτων από χρήστες με παρεμφερή ενδιαφέροντα θα μειώνει σημαντικά τους πόρους που απαιτούνται για αποθήκευση πληροφοριών στη βάση δεδομένων κάθε συστήματος, σε σύγκριση με την αποθήκευση πληροφοριών για κάθε χρήστη ξεχωριστά. Επιπλέον, είναι πιθανό να μειώνεται σημαντικά ο χρόνος ανταπόκρισης του συστήματος για να εξάγει αποτελέσματα. Τέλος, λόγω της ταχέως αυξανόμενης χρήσης των μέσων κοινωνικής δικτύωσης είναι αναπόφευκτη η εισαγωγή των κοινωνικών σχέσεων και αλληλεπιδράσεων στην διαδικασία σύστασης προϊόντων. Όπως θα δούμε και στο επόμενο κεφάλαιο με τα κοινωνικά ΣΣ, η συντριπτική πλειοψηφία ερευνών χρησιμοποιεί σύσταση μέσω συνεργασίας ως βασικό άξονα για την ανάπτυξη της έρευνάς τους.



Εικόνα 2.1 Μέθοδοι Σύστασης μέσω Συνεργασίας και Σύστασης βάσει Περιεχομένου (πηγή: <https://datacrunch.nl/2019/01/14/recommendation-engines-a-new-tool-for-effective-marketing/>)

Στην κατηγορία των ΣΣ μέσω συνεργασίας, εμπεριέχονται δύο υποκατηγορίες [10]: η μέθοδος **βάσει της μνήμης** (memory-based ή neighborhood-based) και η μέθοδος **βάσει μοντέλου** (model-based). Όσον αφορά την πρώτη, είναι πιο απλή καθώς δεν χρησιμοποιεί κανένα μοντέλο. Επιδιώκεται η πρόβλεψη αξιολογήσεων χρηστών σε προϊόντα με βάση τις «γειτονιές» που υπάρχουν. Αυτές οι «γειτονιές» μπορούν να ορισθούν με δύο διαφορετικούς τρόπους, δημιουργώντας έτσι δύο υποκατηγορίες και στις τεχνικές βάσει μνήμης [3].

1. «Γειτονιά» των χρηστών (**User-based**): αναζητούνται χρήστες με παρόμοιες βαθμολογίες για παρεμφερή προϊόντα και τότε προβλέπεται η βαθμολογία ενός συγκεκριμένου χρήστη για κάποιο προϊόν με το οποίο ακόμη δεν έχει αλληλοεπιδράσει, αλλά ένας άλλος χρήστης με παρόμοια ενδιαφέροντα το έχει

βαθμολογήσει. Αυτή είναι και η περίπτωση που απεικονίζεται στην Εικόνα 2.1 για τη Σύσταση μέσω Συνεργασίας. Εφόσον, η βαθμολογία είναι πάνω από ένα όριο ικανοποίησης, τότε το προϊόν θα προταθεί στον χρήστη.

2. «Γειτονιά» των προϊόντων (**Item-based**): για συγκεκριμένα προϊόντα εντοπίζεται αν έχουν βαθμολογηθεί στο παρελθόν από κάποιο συγκεκριμένο χρήστη. Εάν όχι, τότε αναζητούνται παρόμοια προϊόντα με τα προηγούμενα, τα οποία ο χρήστης έχει ήδη βαθμολογήσει. Έτσι, δίνεται ίδια βαθμολογία και στα νέα για το χρήστη προϊόντα. Εάν η βαθμολογία είναι ικανοποιητική, προτείνονται στο χρήστη.

Η item-based μέθοδος έχει πιο ικανοποιητικά αποτελέσματα καθώς η ομοιότητα μεταξύ προϊόντων είναι πιο «αξιόπιστη» από εκείνη μεταξύ χρηστών. Αυτό συμβαίνει διότι είναι σύνηθες κάποιοι χρήστες μετά από συγκεκριμένο χρονικό διάστημα να αλλάζουν τις προτιμήσεις τους και τις αξιολογήσεις τους, γεγονός που προκαλεί μείωση της αξιοπιστίας της εφαρμογής της user-based μεθόδου. Το βασικό μειονέκτημα των παραπάνω τεχνικών βάσει μνήμης είναι η αναγκαία διατήρηση μεγάλης ποσότητας πληροφοριών. Αυτό συμβαίνει για παράδειγμα όταν υπάρχει πολύ μεγάλος αριθμός χρηστών και, έτσι, ο πίνακας αξιολογήσεων προϊόντων είναι υπερβολικά μεγάλος [11]. Λόγω αυτής της υπερφόρτισης, το σύστημα δεν θα καταφέρει να ανταποκριθεί γρήγορα.

Η δεύτερη μέθοδος (model-based), η οποία βασίζεται σε κάποιο μοντέλο, επιδιώκει να επιλύσει τέτοιου είδους προβλήματα. Αυτές οι προσεγγίσεις εκμεταλλεύονται αξιολογήσεις χρηστών-προϊόντων, για τις οποίες παρατηρείται ότι ακολουθούν κάποιο συγκεκριμένο μοτίβο, προκειμένου να κατασκευάσουν μοντέλα πρόβλεψης βαθμολογήσεων με διάφορες τεχνικές εκμάθησης (Bayesian Δίκτυα, Ομαδοποίηση, Τεχνικές Συσχέτισης) [12]. Η τεχνική της *ομαδοποίησης* (clustering) προσπαθεί να χωρίσει τα δεδομένα για ένα δίκτυο σε υποσυστήματα ώστε να εντοπίσει σημαντικές ομάδες που υπάρχουν μέσα σε αυτά. Μόλις σχηματιστούν οι ομάδες, για κάποιον συγκεκριμένο χρήστη, οι απόψεις των άλλων χρηστών μέσα στην ίδια ομάδα μπορούν να υπολογιστούν κατά μέσο όρο και να γίνουν προτάσεις σε αυτόν. Εάν η μέθοδος αυτή εφαρμοστεί σωστά, θα παράγει ομάδες υψηλής ποιότητας στις οποίες η ομοιότητα μεταξύ των χρηστών εντός της ίδιας ομάδας είναι

υψηλή, ενώ η ομοιότητα μεταξύ των ομάδων θα είναι χαμηλή. Αναφορικά με τις *τεχνικές συσχέτισης* (association analysis), συνήθως ακολουθούν κάποιους κανόνες συσχέτισης που προβλέπουν την εμφάνιση ενός προϊόντος βάσει της παρουσίας άλλων προϊόντων μέσα σε μία συναλλαγή. Κάθε συναλλαγή αποτελείται από σύνολα προϊόντων. Οι κανόνες συσχέτισης μπορούν να σχηματίσουν αναπαράσταση δεδομένων με τέτοιο τρόπο ώστε να βελτιώνεται η επάρκεια χώρου αποθήκευσης καθώς και η απόδοση των ΣΣ. Άλλη μία δημοφιλής τεχνική βασίζεται στα *Bayesian Δίκτυα* [10]. Σε ένα Bayesian δίκτυο αντιστοιχίζονται όλα τα προϊόντα με τους κόμβους και, με την εφαρμογή ενός αλγορίθμου εκμάθησης για τέτοιου είδους δίκτυα, εξάγονται γονικά προϊόντα (parent items) για κάθε προϊόν, τα οποία μπορούν να προβλέψουν με βέλτιστο τρόπο τις βαθμολογίες των προϊόντων για τα οποία δεν έχουμε δεδομένα.

Συνοψίζοντας, παραθέτουμε εδώ τα θετικά και τα αρνητικά σημεία που μπορεί να εντοπίσει κανείς σε τεχνικές Σύστασης μέσω Συνεργασίας. Πιο συγκεκριμένα, υπάρχουν κάποια σημαντικά πλεονεκτήματα έναντι της σύστασης βάσει περιεχομένου [12]. Έχουν καλύτερη απόδοση σε τομείς όπου δεν υπάρχει αρκετό περιεχόμενο που να σχετίζεται με τα προϊόντα, καθώς, επίσης, όπου το περιεχόμενο είναι δύσκολο να αναλυθεί από ένα υπολογιστικό σύστημα (όπως απόψεις και ιδανικά). Επιπλέον, αυτές οι τεχνικές έχουν την ικανότητα να προτείνουν στο χρήστη προϊόντα, με τα οποία σχετίζεται, χωρίς το περιεχόμενό τους να αντιστοιχίζεται απαραίτητα με το προφίλ του (αφορά την ιδιότητα “serendipity”, που έχουμε εξηγήσει σε προηγούμενη ενότητα).

Ωστόσο, η ευρεία χρήση των ΣΣ μέσω συνεργασίας αποκάλυψε ορισμένα πιθανά αδύναμα σημεία. Ένα μειονέκτημα της τεχνικής αυτής είναι η ανάγκη ύπαρξης βαθμολογίας από χρήστες για κάποιο προϊόν, ο οποίος έχουν παρόμοια ενδιαφέροντα με το χρήστη στο οποίο θα γίνει η πρόταση. Στην περίπτωση, οπότε, που ένα προϊόν είναι νέο και δεν έχει βαθμολογίες, η μέθοδος αυτή δεν μπορεί να εφαρμοστεί αποτελεσματικά. Το σύστημα δε θα το προτείνει έως ότου βαθμολογηθεί από έναν ικανοποιητικό αριθμό χρηστών. Αντίστοιχο πρόβλημα υπάρχει όταν κάποιος χρήστης είναι νέος σε μία πλατφόρμα, καθώς δεν είναι γνωστές οι προτιμήσεις για τον συγκεκριμένο χρήστη ώστε να ενταχθεί σε μία

«γειτονιά» με χρήστες παρεμφερών ενδιαφερόντων. Αξίζει, επίσης, να σημειωθεί ότι η σύσταση μέσω συνεργασίας δεν διασφαλίζει διαφάνεια των αποτελεσμάτων της καθώς η μοναδική επεξήγηση για την επιλογή πρότασης ενός προϊόντος είναι ότι άγνωστοι χρήστες με παρόμοιες προτιμήσεις το έχουν βαθμολογήσει θετικά. Αντίθετα, στη σύσταση βάσει περιεχομένου υπάρχει η δυνατότητα να παρέχονται επεξηγήσεις σχετικά με τον τρόπο λειτουργίας τους συστήματος με ρητή αναφορά χαρακτηριστικών ή περιγραφών που οδήγησαν στην εμφάνιση ενός προϊόντος στη λίστα συστάσεων. Τέλος, υπάρχουν όμοια προϊόντα για τα οποία χρησιμοποιούνται διαφορετικά ονόματα, δηλαδή *συνώνυμα* (για παράδειγμα οι λέξεις «αχλάδι» και «απίδι», οι οποίες αναφέρονται στο ίδιο προϊόν με διαφορετική ονομασία). Τα ΣΣ μέσω συνεργασίας συνήθως δεν μπορούν να αντιστοιχίσουν τους δύο διαφορετικούς όρους που αποδίδονται στο προϊόν ώστε να υπολογίσουν την ομοιότητά τους.

2.2.3 Υβριδικά Συστήματα Συστάσεων (Hybrid Recommendation Systems)

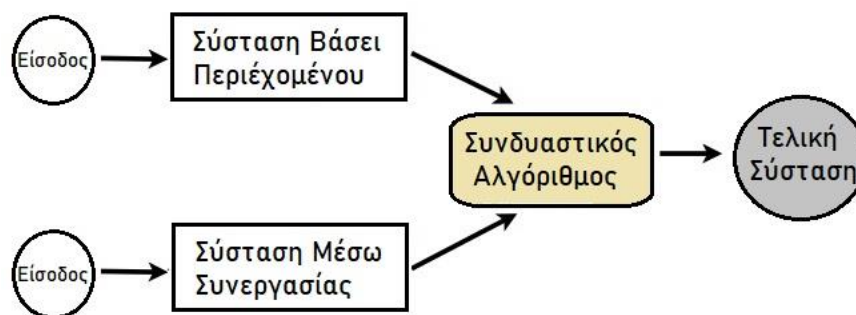
Τα συστήματα αυτά συνδυάζουν διαφορετικές τεχνικές συστάσεων ώστε να επιτυγχάνεται βέλτιστο αποτέλεσμα και να μην υπάρχουν τόσοι περιορισμοί όσοι στις προηγούμενες τεχνικές. Δηλαδή, η ιδέα για την υβριδική τεχνική είναι ο συνδυασμός περισσότερων του ενός τεχνικών που θα παρέχουν πιο ακριβείς και αποτελεσματικές συστάσεις σε σχέση με μία τεχνική, αφού τα μειονεκτήματα της μίας μπορούν να εξαλειφθούν από την άλλη. Για παράδειγμα, ένας συνδυασμός είναι αυτός της *Σύστασης βάσει περιεχομένου* και της *Σύστασης μέσω συνεργασίας* προκειμένου να μπορούμε να έχουμε πιο ακριβείς προτάσεις, όπως έχει αποδειχθεί από διάφορες μελέτες [2].

Παρακάτω θα παρουσιάσουμε συνοπτικά τις υποκατηγορίες στις οποίες μπορούν να διακριθούν τα Υβριδικά ΣΣ, καθώς και εφαρμογές τους στη βιβλιογραφία [13], [6] :

- **Weighted:** *Οι βαθμολογίες από διάφορες τεχνικές συστάσεων συνδυάζονται ώστε να παραχθεί μία τελική σύσταση προϊόντων.* Ένα απλό παράδειγμα τέτοιου συστήματος

παρουσιάζεται στην Εικόνα 2.2, όπου χρησιμοποιούνται οι δύο βασικές τεχνικές σύστασης – συνεργασίας και περιεχομένου. Ο Pazzani [14] χρησιμοποιεί τις συστάσεις που έχουν προκύψει από διαφορετικά ΣΣ (συνεργασίας, περιεχομένου και δημογραφικά ΣΣ) ως ένα σύνολο ψήφων, οι οποίες ύστερα συνδυάζονται σε ένα σύστημα και εξάγουν την σύσταση που θα λάβει ο χρήστης.

Υβριδικά Συστήματα Συστάσεων



Εικόνα 2.2 Παράδειγμα *Weighted* Υβριδικού Συστήματος Συστάσεων, όπου συνδυάζονται 2 διαφορετικά ΣΣ (*content based, collaborative filtering*).

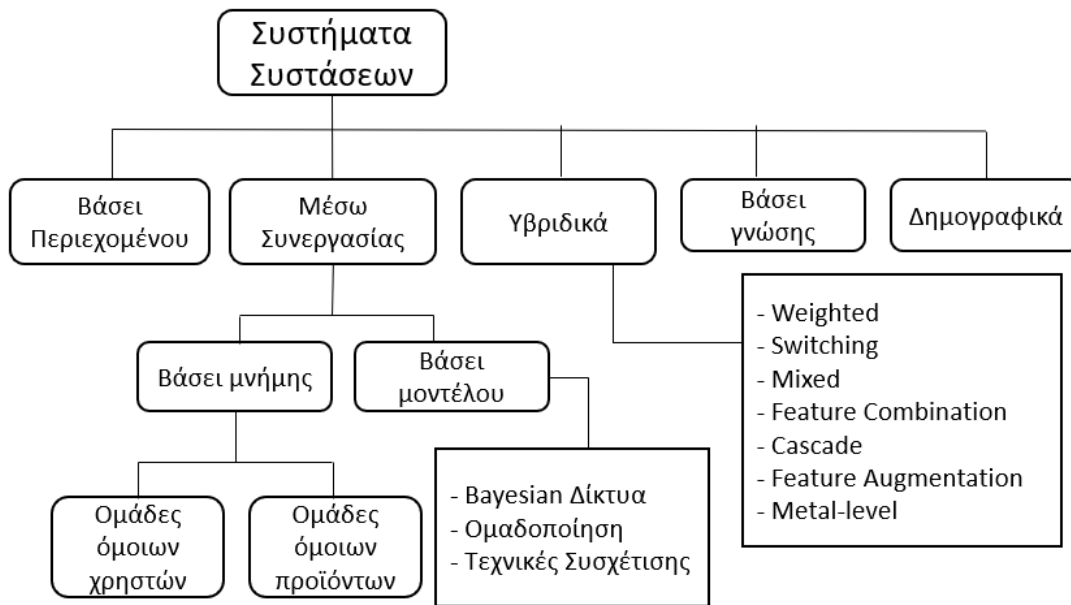
- **Switching:** Το σύστημα εναλλάσσεται μεταξύ διαφορετικών τεχνικών σύστασης ανάλογα με την τρέχουσα κατάσταση που προκύπτει. Για παράδειγμα, το σύστημα “DailyLearner” [15] χρησιμοποιεί μία υβριδική μέθοδο που συνδυάζει τις τεχνικές βάσει περιεχομένου και μέσω συνεργασίας, αλλά με τη μέθοδο βάσει περιεχομένου να εφαρμόζεται πρώτη. Εφόσον η τεχνική αυτή δεν μπορεί να εξάγει σύσταση προϊόντος με επαρκή ακρίβεια, τότε χρησιμοποιείται η τεχνική μέσω συνεργασίας. Βέβαια εξακολουθεί να υφίσταται η αδυναμία της περίπτωσης νέου χρήστη, που χαρακτηρίζει και τις δύο τεχνικές από τις οποίες αποτελείται το σύστημα.

- **Mixed:** *Συστάσεις από διαφορετικά ΣΣ παρουσιάζονται ταυτόχρονα.* Εφαρμογή αυτής της τεχνικής έχουμε για το σύστημα “PTV” [16], όπου χρησιμοποιείται για να δημιουργηθεί σύσταση τηλεοπτικού προγράμματος. Εφαρμόζεται content-based τεχνική στις περιληπτικές περιγραφές των τηλεοπτικών εκπομπών και collaborative filtering τεχνική για τις προτιμήσεις άλλων χρηστών. Τα δύο παραπάνω συνδυάζονται προς σύσταση του τελικού προγράμματος.
- **Feature combination:** *Χαρακτηριστικά από διαφορετικές πηγές δεδομένων σύστασης εμφανίζονται μαζί σε έναν μοναδικό αλγόριθμο συστάσεων.* Πιο συγκεκριμένα, είναι ένας άλλος τρόπος να συνδυαστούν οι content-based και collaborative filtering τεχνικές, επιλέγοντας τη μία τεχνική ως το βασικό ΣΣ και την άλλη ως βοηθητικό ΣΣ, έτσι ώστε η λειτουργία του βασικού ΣΣ να εξαρτάται από δεδομένα που τροποποιούνται από το βοηθητικό. Για παράδειγμα, σε συγκεκριμένη έρευνα [17] έγιναν πειράματα σχετικά με τον κανόνα εκμάθησης Ripper, που εφαρμόζεται ώστε να προτείνονται ταινίες χρησιμοποιώντας τόσο αξιολογήσεις χρηστών όσο και χαρακτηριστικά περιεχομένου. Με αυτό τον τρόπο, πέτυχαν σημαντικές βελτιώσεις στην ακρίβεια των αποτελεσμάτων συγκριτικά με τη μέθοδο ΣΣ μέσω συνεργασίας σε περίπτωση που θα ήταν μοναδική προσέγγιση. Ο προαναφερόμενος υβριδικός συνδυασμός δίνει στο σύστημα τη δυνατότητα να χρησιμοποιεί δεδομένα από την τεχνική μέσω συνεργασίας χωρίς όμως να «βασίζεται» αποκλειστικά σε αυτήν, γεγονός που μειώνει τα αρνητικά σημεία της τεχνικής αυτής, όπως η ευαισθησία στον αριθμό των χρηστών που έχουν βαθμολογήσει ένα προϊόν.
- **Cascade:** *Ένα ΣΣ βελτιώνει τις συστάσεις που προκύπτουν από άλλο ΣΣ.* Έτσι, πρώτα εφαρμόζεται η μία τεχνική σύστασης ώστε να εξάγει μια κατάταξη «υπονήφων», και στη συνέχεια μια δεύτερη τεχνική βελτιώνει τη σύσταση μεταξύ τους. Ένα παράδειγμα τέτοιας υβριδικής μεθόδου είναι το σύστημα “EntreeC” [13], που συνδυάζει την τεχνική σύστασης βάσει γνώσης (knowledge-based) και τη σύσταση μέσω συνεργασίας. Χρησιμοποιεί τις καταχωρημένες «γνώσεις» του για τα εστιατόρια ώστε να κάνει προτάσεις βάσει ενδιαφερόντων του χρήστη, που έχει ο

ίδιος δηλώσει. Η τεχνική συνεργασίας χρησιμοποιείται για να διαμορφώσει την τελική κατάταξη των συστάσεων και, συνεπώς, το τελικό αποτέλεσμα.

- **Feature augmentation:** *Η έξοδος από μία τεχνική συστάσεων χρησιμοποιείται ως χαρακτηριστικό εισόδου σε άλλη τεχνική.* Για παράδειγμα, το σύστημα “Libra” [18] εξάγει συστάσεις βάσει περιεχομένου για βιβλία βάσει δεδομένων από το Amazon. Στα δεδομένα του συστήματος περιλαμβάνονται επίσης πληροφορίες όπως «σχετικοί συγγραφείς» ή «σχετικοί τίτλοι» που προέκυψαν από τεχνικές μέσω συνεργασίας που εφαρμόζει το Amazon.
- **Meta-level:** *Το μοντέλο εκμάθησης που παράγεται από μία τεχνική συστάσεων χρησιμοποιείται ως είσοδος σε άλλο.* Αποτελεί έναν ακόμη τρόπο για να συνδυαστούν διαφορετικές τεχνικές σύστασης προϊόντων. Η πρώτη έρευνα που επεδίωξε να εφαρμόσει την υβριδική αυτή μέθοδο ήταν το σύστημα Fab [19]. Εφαρμόζεται μια τεχνική βάσει περιεχομένου ώστε αρχικά να δημιουργήσουν για κάθε χρήστη ένα μοντέλο που περιγράφει την περιοχή ενδιαφέροντος του χρήστη. Ύστερα συλλέγονται όλα τα δεδομένα και χρησιμοποιούνται ως είσοδοι για κατασκευή κοινοτήτων ενδιαφέροντος (ΣΣ μέσω συνεργασίας). Συλλέγονται, τέλος, προϊόντα που βασίζονται στις προτιμήσεις ολόκληρης της κοινότητας, διανέμοντάς τα μετά σε κάθε χρήστη ξεχωριστά. Συνεπώς, η αρχιτεκτονική του Fab προσφέρει πρόσθετα οφέλη, τα οποία καθίστανται εφικτά από τις αλληλεπικαλύψεις στα ενδιαφέροντα των χρηστών, ταυτοποιώντας κοινότητες ενδιαφέροντος και παρέχοντας τη δυνατότητα σημαντικής εξοικονόμησης πόρων κατά την αύξηση του αριθμού των χρηστών και των προϊόντων.

Συνοψίζοντας, στην παρακάτω εικόνα φαίνονται σε διάγραμμα όλες οι κατηγορίες, και οι αντίστοιχες υποκατηγορίες τους, των ΣΣ που αναλύθηκαν στο παρόν κεφάλαιο.



Εικόνα 2.3 Κατηγορίες των Συστημάτων Συστάσεων

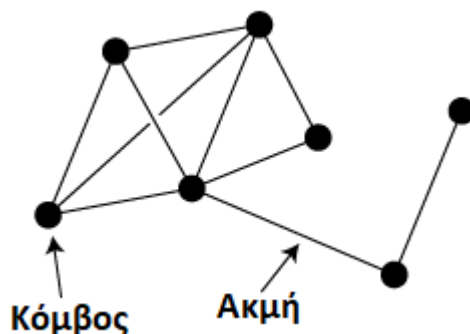
3 Κοινωνικά Συστήματα Συστάσεων (Social Recommender systems)

3.1 Ανασκόπηση Δικτύου – Κοινωνικού Δικτύου

Λόγω της αυξανόμενης χρήσης μέσων κοινωνικής δικτύωσης στις μέρες μας (όπως Facebook, Twitter και άλλα), είναι ενδιαφέρον να μελετήσουμε μια συγκεκριμένη κατηγορία των ΣΣ, τα Κοινωνικά ΣΣ (Social RS), τα οποία λαμβάνουν υπόψη τόσο τη δομή ενός κοινωνικού δικτύου όσο και τη συμπεριφορά των χρηστών μέσα σε αυτό. Τα **Κοινωνικά Δίκτυα (Social Networks)** παρέχουν σημαντικές πληροφορίες για τους χρήστες που ανήκουν στο δίκτυο και τις αλληλεπιδράσεις μεταξύ τους. Το γεγονός αυτό υποδεικνύει τη σημαντικότητα των Κοινωνικών Δικτύων για τα ΣΣ. Η βασική ιδέα είναι ότι οι συστάσεις γίνονται βάσει των προσωπικών προτιμήσεων του χρήστη, τη γενική αποδοχή ενός προϊόντος αλλά και την επιρροή από «φίλους» μέσα στο δίκτυο.

3.1.1 Ανάλυση Σύνθετων Δικτύων (Complex Network Analysis)

Ξεκινώντας από την έννοια του *Δικτύου*, ορίζεται ως ένα σύνολο από οντότητες που αλληλοεπιδρούν με ποικίλους τρόπους η μία με την άλλη [20]. Συνήθως μοντελοποιείται ως ένας γράφος που αποτελείται από *κόμβους* (nodes), οι οποίοι σχετίζονται μεταξύ τους και οι συσχετίσεις αυτές φαίνονται με τις *ακμές* (links ή edges), όπως φαίνεται απλοϊκά στην Εικόνα 3.1. Ο γράφος ορίζεται πλήρως από τους κόμβους και τη συνδεσμολογία τους.



Εικόνα 3.1 Αναπαράσταση ενός απλοποιημένου Δικτύου.

Αξίζει σε αυτό το σημείο να αναφέρουμε κάποιες σημαντικές έννοιες ενός γράφου, που θα είναι χρήσιμες σε παρακάτω ενότητες. Αν οι ακμές είναι προσανατολισμένες (ορίζονται δηλαδή από διατεταγμένα ζεύγη κόμβων) τότε ο γράφος λέγεται *κατευθυνόμενος* (directed). Ωστόσο, αν οι ακμές δεν είναι προσανατολισμένες (ορίζονται δηλαδή από μη διατεταγμένα ζεύγη κόμβων) τότε ο γράφος λέγεται *μη κατευθυνόμενος* (undirected), όπως στην περίπτωση της Εικόνας 3.1. Επίσης, αν οι ακμές είναι συνδεδεμένες με κάποια αξία (βάρος) τότε ο γράφος λέγεται *σταθμισμένος* (weighted).

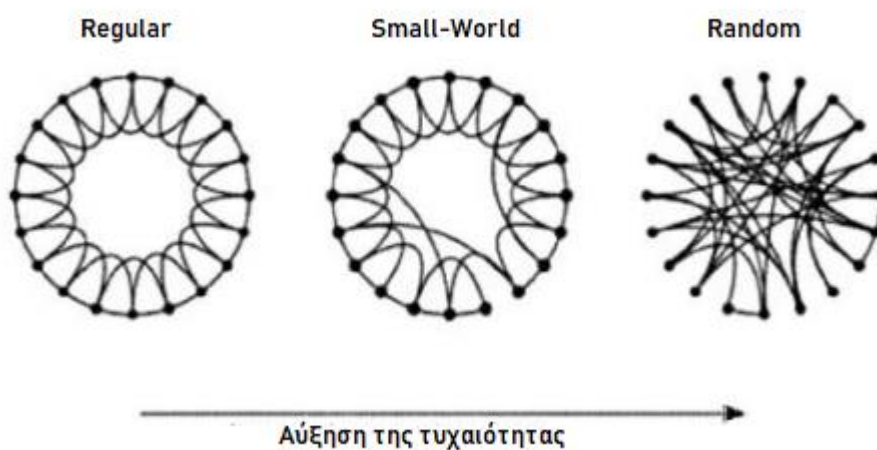
Σημαντική επίσης είναι η έννοια του *βαθμού* ενός κόμβου σε ένα γράφο. Σε μη κατευθυνόμενο γράφο, βαθμός ενός κόμβου καλείται ο αριθμός των συνδεδεμένων με αυτόν ακμών. Σε έναν κατευθυνόμενο γράφο, ορίζονται δύο διαφορετικές έννοιες:

α) *Εσωτερικός βαθμός* (in-degree) καλείται ο αριθμός των ακμών που καταλήγουν σε αυτόν.

β) *Εξωτερικός βαθμός* (out-degree) καλείται ο αριθμός των ακμών που ξεκινούν από αυτόν.

Επιπροσθέτως, για τους γράφους-δίκτυα έχουν μελετηθεί διάφορες τοπολογίες [2]. Με τον όρο τοπολογία δικτύου αναφερόμαστε στον τρόπο με τον οποίο συνδέονται οι κόμβοι ενός δικτύου και διαδραματίζουν σημαντικό ρόλο στον προσδιορισμό βασικών παραγόντων ενός δικτύου, καθώς, επίσης, και στη μέτρηση της σημασίας κάθε κόμβου στη διάδοση

πληροφοριών, στην ανίχνευση, δηλαδή, των πιο «σημαντικών» κόμβων. Δημοφιλείς τοπολογίες είναι οι *Regular*, *Random*, *Small-World*, *Scale-Free* [21]. Στην παρακάτω εικόνα (Εικόνα 3.2), φαίνεται ο τρόπος με τον οποίο ένα δίκτυο μετατρέπεται από *Regular* – Κανονικό Δίκτυο (όλοι οι κόμβοι έχουν τον ίδιο βαθμό, συνεπώς υπάρχει μηδενική τυχαιότητα στη συνδεσιμότητά τους) σε *Random* – Τυχαίο Δίκτυο (όλοι οι κόμβοι είναι συνδεδεμένοι μεταξύ τους με πλήρη τυχαιότητα). Δύο γνωστά μοντέλα των *Random* Δικτύων είναι τα *Gilbert model* και *Erdos-Renyi model*.

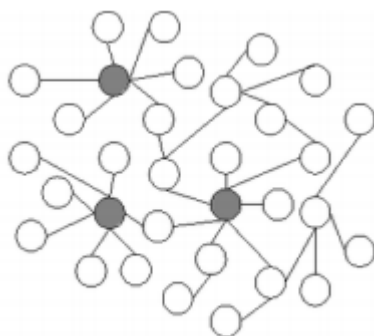


Εικόνα 3.2 Σύγκριση τοπολογιών δικτύου βάσει της τυχαιότητας [2]

Αναφορικά με τα *Small-World* Δίκτυα, πρόκειται για πολύπλοκα συστήματα όπου οι περισσότεροι κόμβοι δεν είναι γειτονικοί μεταξύ τους, αλλά είναι προσβάσιμοι ο ένας από τον άλλο με *μονοπάτια μικρού μήκους* (hops). Το επικρατέστερο μοντέλο δημιουργίας τέτοιου είδους δικτύων είναι το *Watts-Strogatz model*. Η απάντηση στην ερώτηση γιατί, αναλογικά με την τυχαιότητα, τα δίκτυα αυτά βρίσκονται μεταξύ των *Regular* και *Random* Δικτύων είναι ότι, στην περίπτωση αυτή, χρησιμοποιώντας το προαναφερθέν μοντέλο, η διαδικασία ξεκινάει από ένα *Regular* Δίκτυο, στο οποίο ύστερα επανασυνδέονται οι ακμές

με διαφορετικό τρόπο με μία σταθερή πιθανότητα $p < 1$ (δηλαδή όχι με πλήρη τυχαιότητα), εξαιρουμένων των αυτό-συνδέσεων και των διπλών συνδέσεων.

Τα Scale-Free Δίκτυα είναι ένας τύπος δικτύου που χαρακτηρίζεται από την παρουσία *κεντρικών κόμβων* (hubs), τα οποία έχουν υψηλή συνδεσιμότητα με άλλους κόμβους του δικτύου. Έτσι, υπάρχουν λίγοι κόμβοι με υψηλό βαθμό (degree), ενώ η πλειοψηφία τους έχει μικρό βαθμό. Ακολουθείται δηλαδή η λεγόμενη κατανομή νόμου δύναμης (power law). Τα δίκτυα αυτά λέγονται Scale-Free διότι τα χαρακτηριστικά τους είναι ανεξάρτητα από το μέγεθός του, δηλαδή τον αριθμό των κόμβων. Αυτό σημαίνει ότι, όταν το δίκτυο μεγαλώνει, η υποκείμενη δομή του παραμένει η ίδια, επειδή οι νόμοι δύναμης έχουν την ίδια λειτουργική δομή σε όλες τις κλίμακες (Εικόνα 3.3).



Εικόνα 3.3 Παράδειγμα Scale-Free Δικτύου. Οι σκιαγραφημένοι κόμβοι είναι τα “hubs” [2].

3.1.2 Ανάλυση Κοινωνικών Δικτύων

Το *Κοινωνικό Δίκτυο* είναι η χαρτογράφηση και η μέτρηση σχέσεων και ροών μεταξύ ανθρώπων, ομάδων και οργανισμών. Στην περίπτωση αυτή, οι άνθρωποι και οι ομάδες είναι οι κόμβοι, ενώ οι ακμές είναι οι σχέσεις-ροές μεταξύ αυτών (Εικόνα 3.4). Οι κόμβοι αυτοί μπορεί να συνδέονται μεταξύ τους με έναν ή περισσότερους τύπους αλληλεξάρτησης, όπως αξίες, οράματα, ιδέες, οικονομικές συναλλαγές, φιλία, συγγένεια, μεταφορά μολυσματικών

ασθενειών και άλλα. Ένα Μέσο Κοινωνικής Δικτύωσης αποτελείται από *i*) τους χρήστες (με δικό τους προφίλ, συνδέσεις, προτιμήσεις) και *ii*) τα «αντικείμενα» που αντιπροσωπεύουν τις δραστηριότητες του χρήστη (αλληλεπιδράσεις με άλλους χρήστες - τα λεγόμενα *likes* ή *comments* -, κοινόχρηστα βίντεο και φωτογραφίες).



Εικόνα 3.4 Παράδειγμα Κοινωνικού Δικτύου – Αλληλεπιδράσεις μεταξύ ανθρώπων σε μια μικρή κοινωνία.

Αξίζει να σημειώσουμε στο σημείο αυτό τις πιο σημαντικές μετρικές της Ανάλυσης Κοινωνικών Δικτύων που έχουν υιοθετηθεί από τα ΣΣ [2] και μπορούν να χωριστούν σε τρεις κατηγορίες. Αρχικά είναι οι **συνδέσεις** (connections) του δικτύου μέσω της μετρικής της *ομοφιλίας* (homophily), η οποία αποτελεί μέτρο της τάσης των χρηστών ενός δικτύου να συνδεθούν σε εκείνους τους χρήστες που θεωρούνται πιο «σχετικοί» με αυτούς, βασιζόμενοι, για παράδειγμα, σε δημογραφικά κριτήρια. Δεύτερον, έχουμε τις μετρικές **διανομής** (distribution) που μπορεί να είναι τοπικές μετρικές κεντρικότητας που αφορούν ξεχωριστά κάθε κόμβο του δικτύου υπολογίζοντας τη σημαντικότητά τους μέσα στο δίκτυο (degree, betweenness, closeness), ή η μέτρηση μιας μετρικής που αφορά ολόκληρο το δίκτυο, της πυκνότητας (density), η οποία χαρακτηρίζει το δίκτυο ως πυκνό, εάν ο αριθμός των ακμών είναι αρκετά μεγαλύτερος από τον αριθμό των κόμβων του δικτύου, αλλιώς θεωρείται αραιό. Τελευταία κατηγορία είναι οι μετρικές βάσει **τμηματοποίησης**

(segmentation-based), οι οποίες περιλαμβάνουν ένα σύνολο μετρήσεων που προσδιορίζουν τη δομή του δικτύου τόσο σε επίπεδο κόμβου όσο και σε επίπεδο δικτύου.

3.2 Εισαγωγή στα Κοινωνικά ΣΣ (Social RS - SRS)

Τα ΣΣ σήμερα χρησιμοποιούνται ευρέως σε πληθώρα εφαρμογών με σκοπό να προτείνουν στον χρήστη προϊόντα-υπηρεσίες βάσει ορισμένων κριτηρίων. Αξίζει, ωστόσο, να ξεχωρίσουμε τη χρήση των μετρικών της Ανάλυσης Κοινωνικών Δικτύων [21], όπως αναφέρθηκαν στην προηγούμενη ενότητα, σε ΣΣ, όπου στην περίπτωση αυτή τα ΣΣ αναφέρονται στη βιβλιογραφία ως *Social Recommender Systems* [22]. Η ραγδαία ανάπτυξη των μέσων κοινωνικής δικτύωσης έχει επιταχύνει σημαντικά την ανάπτυξη των συστημάτων αυτών. Στόχος τους είναι η διευκόλυνση των χρηστών λόγω της υπερφόρτωσης πληροφοριών, παρουσιάζοντας τα πιο ελκυστικά και σχετικά με τον χρήστη περιεχόμενα, και, κατά συνέπεια, προσφέρουν νέες ευκαιρίες για την περαιτέρω βελτίωση των ΣΣ. Επιδιώκεται, επίσης, η αύξηση της ενσωμάτωσης και της συμμετοχής νέων και υπάρχοντων χρηστών των κοινωνικών ιστοτόπων. Οι μηχανισμοί συστάσεων ετικετών, ατόμων και κοινοτήτων συχνά χρησιμοποιούν τεχνικές που προσαρμόζονται στις ανάγκες και τα ενδιαφέροντα μεμονωμένων χρηστών ή ενός συνόλου χρηστών. Παραδείγματα τέτοιων δικτύων είναι το Facebook, το Twitter και το YouTube.

Τα Μέσα Κοινωνικής Δικτύωσης και τα κλασικά ΣΣ, που αναλύσαμε στο προηγούμενο κεφάλαιο, μπορούν να ευνοηθούν το ένα από το άλλο. Πιο συγκεκριμένα, από τη μία πλευρά, τα πρώτα εισάγουν νέους τύπους δεδομένων, τα οποία είναι δημόσια σε περισσότερα του ενός άτομου, όπως, για παράδειγμα, ετικέτες, βαθμολογίες, σχόλια και συνδέσεις μεταξύ των χρηστών, και μπορούν να χρησιμοποιηθούν για την ενίσχυση των προτάσεων. Από την άλλη, οι τεχνικές που εφαρμόζονται στα ΣΣ, μπορούν να διαδραματίσουν βασικό ρόλο στην επιτυχία των εφαρμογών των μέσων κοινωνικής δικτύωσης και του κοινωνικού ιστού στο σύνολό του, διασφαλίζοντας ότι σε κάθε χρήστη παρουσιάζεται το πιο ελκυστικό περιεχόμενο, σε προσωπική βάση.

3.2.1 Βιβλιογραφική Ανασκόπηση Κοινωνικών Συστημάτων Συστάσεων

Τα Κοινωνικά ΣΣ βελτιώνουν την ακρίβεια των παραδοσιακών ΣΣ, καθώς χρησιμοποιούν σαν συμπληρωματική είσοδο στον μηχανισμό τους την αλληλεπίδραση που έχει ο ένας χρήστης στον άλλον μέσα σε ένα μέσο κοινωνικής δικτύωσης [23]. Λόγω μακροχρόνιων κοινωνικών δεσμών που έχουν χρησιμοποιηθεί, οι χρήστες είναι πιο πρόθυμοι να μοιραστούν τις προσωπικές απόψεις τους με τους διαδικτυακούς τους φίλους και να εμπιστευθούν τις προτάσεις προϊόντων και υπηρεσιών από αυτούς, παρά από αγνώστους και πωλητές. Για παράδειγμα, ένας χρήστης ίσως διαβάσει ένα συγκεκριμένο άρθρο σχετικά με ένα γεγονός απλά και μόνο επειδή το γεγονός αυτό συνέβη στην περιοχή όπου διαμένει η οικογένειά του. Επίσης, σε έναν χρήστη μπορεί να αρέσει ένα τραγούδι που προτάθηκε από κάποιον στενό του φίλο στο Facebook. Η κοινωνική εμπιστοσύνη ενός ζεύγους φίλων (u, v) μπορεί να δημιουργηθεί με βάση την άμεση ανατροφοδότηση του χρήστη u για τον χρήστη v (π.χ. με ψηφοφορία) ή από έμμεση τροφοδότηση (π.χ. με τη συχνότητα αλληλεπίδρασης και επικοινωνίας μεταξύ των u, v) [23].

Οι Yang et al. [23] παρουσιάζουν μια έρευνα ενός Social RS που βασίζεται σε σύσταση μέσω συνεργασίας, όπως συμβαίνει στην πλειοψηφία των υπαρχόντων συστημάτων. Ενδιαφέρον είναι το γεγονός ότι αρχικά παρουσιάζεται ο μηχανισμός του παραδοσιακού ΣΣ, χρησιμοποιώντας την τεχνική σύστασης μέσω συνεργασίας, ενώ ύστερα εξηγείται πως οι πληροφορίες του μέσου κοινωνικής δικτύωσης μπορούν να υιοθετηθούν από το ΣΣ ως συμπληρωματική είσοδος στον αλγόριθμο για περισσότερη ακρίβεια, χρησιμοποιώντας δύο διαφορετικές προσεγγίσεις, τις οποίες στο τέλος τις συγκρίνει. Συγκεκριμένα, στην μία εισάγεται ένας πιθανοτικός πίνακας εμπιστοσύνης των χρηστών με τη μέθοδο του Πίνακα Παραγοντοποίησης (Matrix Factorization based Social Recommendation Approach). Στη δεύτερη χρησιμοποιούνται αξιολογήσεις χρηστών-προϊόντων αποθηκευμένες στο σύστημα προκειμένου να προβλέψει αξιολογήσεις για νέα προϊόντα (Neighborhood based Social Recommendation Approach).

Σε μία παρεμφερή έρευνά τους, οι Yang et al. [24] προτείνουν ένα ΣΣ, στο οποίο, προκειμένου να εκμεταλλευτεί το μέσο κοινωνικής δικτύωσης, χρησιμοποιεί μία Bayesian

μέθοδο για να παρέχει ακριβείς και εξατομικευμένες προτάσεις. Σε αυτήν την τεχνική εφαρμόζεται μια πιθανοτική μέθοδος για τη μέτρηση της ομοιότητας που φέρουν οι αξιολογήσεις των χρηστών που διασυνδέονται μέσα στο δίκτυο. Είναι ιδιαίτερα χρήσιμη όταν δεν έχει κανείς αρκετά διαθέσιμα δεδομένα. Εξάγεται, έτσι, το συμπέρασμα ότι η προσέγγιση που κάνουν έχει πιο ακριβή αποτελέσματα από τα παραδοσιακά ΣΣ μέσω Συνεργασίας.

Επίσης, έχει αναπτυχθεί μια έρευνα [5] βάσει της οποίας παρατίθενται τα θετικά και τα αρνητικά σημεία των Κοινωνικών ΣΣ. Εφαρμόζοντας δύο παρεμφερείς μεθόδους, που υλοποιούνται με *memory-based Social RS* και *model-based Social RS*, εξάγεται το συμπέρασμα ότι μπορούν να βελτιώσουν σε σημαντικό βαθμό την κάλυψη του δικτύου από τα προϊόντα, δηλαδή οι προτάσεις προϊόντων που αποφασίζεται να γίνουν, ικανοποιούν περισσότερους χρήστες, σε σύγκριση με τα κλασικά ΣΣ, ιδιαίτερα για την περίπτωση που κάποια προϊόντα είναι σχετικά καινούρια και τα διαθέσιμα δεδομένα ελάχιστα. Μία άλλη προσέγγιση δόθηκε από τους Arazy et al. [25], οι οποίοι ανέπτυξαν έναν γενικευμένο αλγόριθμο που επιδιώκει να παρέχει περισσότερη ακρίβεια έναντι της Σύστασης μέσω Συνεργασίας των κλασικών ΣΣ. Η μέθοδος αυτή χρησιμοποιεί διάφορους μηχανισμούς και, συνδυάζοντας τους με το καταναλωτικό ιστορικό και τις βαθμολογίες του χρήστη, εξάγει μια τελική πρόβλεψη του συστήματος που παράγει την τελική σύσταση. Οι μηχανισμοί αυτοί αφορούν στον εντοπισμό σημαντικών δεικτών ενός μέσου κοινωνικής δικτύωσης, όπως α) της *εμπιστοσύνης* των χρηστών μεταξύ τους, β) της *δύναμης* του δεσμού μεταξύ του παραλήπτη και της πηγής (*tie strength*), η οποία διαμορφώνεται από τη διάρκεια της σύνδεσης των χρηστών και τη συχνότητα αλληλεπίδρασής τους, γ) της *υπόληψης* μιας πηγής πληροφορίας, δηλαδή η γνώμη των πολλών ή κυρίαρχη γνώμη, που διαμορφώνουν την αξιοπιστία της πηγής, και, τέλος, δ) της *ομοφιλίας*, δηλαδή της έμφυτης προτίμησης των ανθρώπων να συγχρωτίζονται με άτομα που μοιράζονται κοινά ενδιαφέροντα, κοινές αντιλήψεις και αξίες, και γενικά τους συνδέουν αρκετές «ομοιότητες».

Οι Diaz-Aviles et al. [26] προτείνουν μία προσέγγιση για την ανάλυση δεδομένων των μέσων κοινωνικής δικτύωσης, που βασίζεται στην τεχνική μέσω συνεργασίας, ώστε να

προτείνει θεματολογίες (topics) σε κάθε χρήστη ξεχωριστά. Χρησιμοποιείται Πίνακας Παραγοντοποίησης προς βελτιστοποίηση της εξατομικευμένης κατάταξης των θεματολογιών. Στο [27] παρουσιάζεται μια τεχνική (StreamRec), η οποία επίσης βασίζεται στη σύσταση μέσω συνεργασίας και δέχεται ως εισόδους α) αξιολογήσεις χρηστών σε προϊόντα (update events) και β) αιτήματα συστάσεων από κάποιον συγκεκριμένο χρήστη ή από μια εφαρμογή (recommend events).

Στην παρακάτω υποενότητα, παρατίθενται μελέτες που έχουν γίνει για τα Κοινωνικά ΣΣ αναφορικά με ένα από τα πλέον διαδεδομένα Μέσα Κοινωνικής Δικτύωσης, το Twitter.

3.2.1.1 Μελέτες Περίπτωσης Κοινωνικών Συστημάτων Συστάσεων στο Twitter

Το Twitter είναι ένα από τα πιο δημοφιλή μέσα κοινωνικής δικτύωσης. Κάθε χρήστης δημιουργεί δωρεάν το δικό του προφίλ, το οποίο αποτελείται από εικόνες και διάφορες πληροφορίες, όπως τόπος διαμονής, σπουδές, εργασία, ενδιαφέροντα. Κάθε χρήστης ακολουθεί όποιους επιθυμεί και ακολουθείται από οποιονδήποτε. Όταν ένα χρήστης ακολουθεί κάποιον άλλον, δεν είναι απαραίτητο να τον ακολουθεί και ο δεύτερος. Στην περίπτωση που ακολουθεί ο ένας τον άλλον, τότε οι πληροφορίες διαμοιράζονται και προς τις δύο κατευθύνσεις. Ο καθένας διαμορφώνει την αρχική του σελίδα με δημοσιεύσεις (tweets) και αναδημοσιεύσεις (retweets) όσων ακολουθεί. Οι δημοσιεύσεις κάθε χρήστη εμφανίζονται σε χρονολογική σειρά και οι χρήστες μπορούν να «ψάξουν» στα χρονολόγια (timelines) αυτών που ακολουθούν και να ανακαλύψουν πληροφορίες που τους ενδιαφέρουν. Ο μηχανισμός της αναδημοσίευσης επιτρέπει στους χρήστες να μοιραστούν δεδομένα με τους ακόλουθους τους και έτσι επιταχύνεται της διάδοση πληροφορίας στο μέσο κοινωνικής δικτύωσης.

Ωστόσο, λόγω του ταχέως αυξανόμενου αριθμού των δημοσιεύσεων, οι περισσότεροι χρήστες αντιμετωπίζουν σημαντικό πρόβλημα υπερφόρτωσης πληροφοριών, ειδικά εκείνοι που ακολουθούν πολλά άτομα και λαμβάνουν χιλιάδες tweets ανά ημέρα. Για αυτόν τον λόγο, έχουν δημιουργηθεί αλγόριθμοι (Συστήματα Συστάσεων), οι οποίοι προτείνουν στους

χρήστες δημοσιεύσεις που τους ενδιαφέρουν σε μεγάλο βαθμό μειώνοντας έτσι την προσπάθεια τους να αναζητήσουν χρήσιμες πληροφορίες. Συγκεκριμένα, στην έρευνα [28] χρησιμοποιείται το πλεονέκτημα της σύστασης μέσω συνεργασίας συλλέγοντας πληροφορίες για προτιμήσεις και ενδιαφέροντα από πολλούς χρήστες προκείμενου να προταθούν χρήσιμες-ενδιαφέρουσες δημοσιεύσεις σε κάθε χρήστη του Twitter. Συνδυαστικά με τη σύσταση μέσω συνεργασίας, χρησιμοποιούνται πρόσθετα χαρακτηριστικά, τα οποία μπορούν έμμεσα να επηρεάσουν την πιθανότητα αναδημοσίευσης ενός tweet. Τέτοια χαρακτηριστικά μπορεί να είναι η ποιότητα του tweet ή η αξιοπιστία του εκδότη, του χρήστη δηλαδή που έκανε τη δημοσίευση. Έχει παρατηρηθεί ότι ένας χρήστης προτιμάει δημοσιεύσεις από συγκεκριμένους χρήστες με τους οποίους έχει οικειότητα, είναι διάσημοι ή μοιράζονται κοινά ενδιαφέροντα. Επίσης, αν μία δημοσίευση δημιουργήθηκε από χρήστη που έχει την εξουσιοδότηση, δηλαδή την εξειδικευμένη γνώση, είναι πιο πιθανό να είναι μια δημοσίευση με υψηλή ποιότητα περιεχομένου. Τονίζεται ιδιαίτερα η υπεροχή της μεθόδου αυτής των Κοινωνικών ΣΣ έναντι της σύστασης βάσει περιεχομένου, καθώς στην πρώτη δεν απαιτείται πολύ καλή γνώση των χαρακτηριστικών ενός προφίλ.

Μια άλλη μελέτη [7], η οποία επίσης αφορά στο Twitter, ανέπτυξε αλγόριθμο που βασίζεται στη μέθοδο σύστασης βάσει περιεχομένου, καθώς εστιάζουν στο δίκτυο των χρηστών που ακολουθεί ένας συγκεκριμένος χρήστης. Για παράδειγμα, έχει παρατηρηθεί ότι κάποιες φορές οι χρήστες ακολουθούν κάποιους άλλους από «ευγένεια» αλλά με την πάροδο του χρόνου έχουν την τάση να σταματούν να ακολουθούν οτιδήποτε δεν τους προσφέρει πληροφορίες στο πεδίο του ενδιαφέροντος τους. Έχουν ως στόχο να δημιουργήσουν ένα δίκτυο από χρήστες, όπου η πλειοψηφία θα αποτελείται από άτομα με τα οποία επιθυμούν να παραμείνουν συνδεδεμένοι και να αλληλοεπιδρούν μέσα στο δίκτυο. Συνοπτικά, αναλύεται η ομοιότητα μεταξύ των ατόμων μέσα σε ένα δίκτυο που ακολουθεί ένας χρήστης βάσει δύο εννοιών: της δημοφιλίας (popularity) και της δραστηριότητας (activity). Η δημοφιλία ορίζεται ως η αναλογία των ατόμων που ακολουθεί ένα χρήστης και των ατόμων που τον ακολουθούν, ενώ δραστηριότητα είναι ο αριθμός των δημοσιεύσεων

ενός χρήστη από τη στιγμή που είναι εγγεγραμμένος στο συγκεκριμένο μέσο κοινωνικής δικτύωσης.

Οι Pennacchiotti et al. [29] εισάγουν το πρόβλημα της σύστασης δημοσιεύσεων για κάποιον χρήστη του Twitter με στόχο τη μεγιστοποίηση του συνολικού ενδιαφέροντος των προτεινόμενων δημοσιεύσεων, εντοπίζοντας τις δημοσιεύσεις που ταιριάζουν σε μεγάλο βαθμό στα ενδιαφέροντα του χρήστη, οι οποίες όμως δεν έχουν δημοσιευθεί από κάποιον χρήστη που ακολουθεί. Χρησιμοποιούνται δυο μοντέλα ώστε να ποσοτικοποιηθεί το ενδιαφέρον μιας δημοσίευσης σε έναν χρήστη. Το ένα αποσκοπεί στο να προτείνει τις πιο ενδιαφέρουσες δημοσιεύσεις, ενώ το δεύτερο να προτείνει ένα σύνολο δημοσιεύσεων που θα ελαχιστοποιεί τυχόν επικαλυπτόμενο περιεχόμενο. Το αθροιστικό αποτέλεσμα παρέχει μεγαλύτερη ποικιλία συστάσεων. Επιπρόσθετα, η μελέτη [30] προτείνει ένα συγκεκριμένο ΣΣ, το οποίο ονόμασε ως TeRec (Temporal Recommender System). Το τελευταίο μπορεί να μοντελοποιήσει άμεσα το ενδιαφέρον του χρήστη, χρησιμοποιώντας Πίνακα Παραγοντοποίησης, και να παρέχει συστάσεις ανάλογα με τα πιο πρόσφατα ενδιαφέροντά του. Όταν οι χρήστες δημοσιεύσουν κάποιο tweet, μπορούν να λαμβάνουν προτεινόμενα θέματα (hashtags) βάσει των ενδιαφερόντων τους σε πραγματικό χρόνο (real-time interests) και να αφήνουν τα σχόλιά τους.

Μια άλλη έρευνα [31] προτείνει μια προσέγγιση που επιλέγει τα tweets που αντιπροσωπεύουν περισσότερο τα ενδιαφέροντα του χρήστη με βάση το ιστορικό του και τις επιρροές από τον κοινωνικό του περίγυρο. Με γνώμονα το χρόνο, εφαρμόζεται ένας επαναληπτικός αλγόριθμος βελτιστοποίησης για την επιλογή δημοσιεύσεων έχοντας ως κριτήρια την καινοτομία (novelty), την κάλυψη (coverage) και την ποικιλία (diversity). Ο Shen [32] παρουσιάζει την έννοια της σύνδεσης οντοτήτων σε tweets (tweet entity linking - σύνδεση οντοτήτων, οι οποίες εντοπίζονται σε tweets, με αντίστοιχες οντότητες στον πραγματικό κόσμο), που μοντελοποιεί τα ενδιαφέροντα του χρήστη. Η προσέγγιση αυτή εντοπίζει και συνδέει οντότητες, τις οποίες οι χρήστες έχουν αναφέρει μέσα στις δημοσιεύσεις τους, και κατασκευάζει ένα γράφημα που παρουσιάζει την αλληλεπίδραση μεταξύ οντοτήτων για κάθε χρήστη. Οι Hong et al. [33] επιδιώκουν να δημιουργήσουν

προγνωστικά μοντέλα για τις αποφάσεις των χρηστών στο Twitter με τη χρήση Μηχανών Συν-Παραγοντοποίησης (Co-Factorization Machines), που είναι μια επέκταση ενός σύγχρονου ΣΣ, το οποίο μπορεί να διαχειριστεί πολλαπλές πτυχές του συνόλου των δεδομένων την ίδια στιγμή.

3.2.2 Συμπεράσματα για τα Κοινωνικά Συστήματα Συστάσεων

Παρακάτω, θα παραθέσουμε συγκεντρωτικά τα οφέλη που έχουν παρατηρηθεί από την εφαρμογή των Κοινωνικών ΣΣ, αλλά και αδύναμα σημεία που τυχόν υπάρχουν [5]. Ξεκινώντας από τα θετικά στοιχεία, οι χρήστες σε μία κοινωνική πλατφόρμα αλληλοεπιδρούν μεταξύ τους εκ φύσεως. Σπάνια παίρνουν αποφάσεις ανεξάρτητα από τους διαδικτυακούς τους φίλους και συνήθως αναζητούν συμβουλές τους για να αποφασίσουν αν θα προχωρήσουν σε κάποια συγκεκριμένη αγορά. Οι προτιμήσεις ενός χρήστη είναι πιο πιθανό να «ταιριάζουν» με άλλων χρηστών του μέσου κοινωνικής δικτύωσης τους, παρά με τυχαία επιλεγμένους χρήστες (όπως γίνεται στη σύσταση μέσω συνεργασίας). Με αυτόν τον τρόπο, έχοντας συμπληρωματικές πληροφορίες από το μέσο κοινωνικής δικτύωσης, σε συνδυασμό με την ανάλυση κοινωνικών δικτύων, δημιουργούνται οι συνθήκες για τη βελτίωση της ποιότητας των προτάσεων προϊόντων.

Επιπρόσθετα, στα κλασικά ΣΣ, προκείμενου να εξαχθούν ακριβείς προτάσεις προϊόντων, είναι απαραίτητες αρκετές αξιολογήσεις κάθε χρήστη στο παρελθόν. Ωστόσο, συνήθως ο πίνακας αξιολογήσεων που δημιουργείται από αυτή τη διαδικασία είναι συνήθως αρκετά αραιός, αφού οι περισσότεροι χρήστες βαθμολογούν μια μειοψηφία από το σύνολο των προϊόντων-υπηρεσιών που χρησιμοποιούν. Κατά συνέπεια, είναι δύσκολο ύστερα να εντοπιστούν χρήστες με παρόμοια ενδιαφέροντα και, έτσι, τα συστήματα αυτά δεν είναι ικανά να παράγουν ακριβείς συστάσεις για χρήστες που έχουν λίγες ή μηδενικές αξιολογήσεις στο ιστορικό τους (στην κατηγορία αυτή συμπεριλαμβάνονται και χρήστες που είναι νέοι στην διαδικτυακή πλατφόρμα). Αντίθετα, τα Κοινωνικά ΣΣ μπορούν να μειώσουν σημαντικά αυτό το πρόβλημα, αρκεί ο χρήστης να έχει συνδεθεί και να

αλληλοεπιδρά με επαρκές μέρος του μέσου κοινωνικής δικτύωσης. Αν τα κλασικά ΣΣ αποτυγχάνουν πλήρως σε περιπτώσεις νέων χρηστών, τα Κοινωνικά ΣΣ καταφέρνουν να εξάγουν αποτέλεσμα με πολύ μικρό σφάλμα. Τέλος, όταν υπάρχουν αρκετοί νέοι χρήστες ή και νέα προϊόντα σε μία βάση δεδομένων, δημιουργείται, επίσης, σημαντικό πρόβλημα με την κάλυψη προτάσεων προϊόντων, καθώς πολλές από τις αξιολογήσεις είναι δύσκολο να προβλεφθούν. Η έννοια της κάλυψης αναφέρεται στο σύνολο των αξιολογήσεων τις οποίες, αν τις αποκρύψουμε, τα συστήματα συστάσεων είναι σε θέση να εξάγουν μία προβλεπόμενη βαθμολογία. Ωστόσο, στα Κοινωνικά ΣΣ αυτό το πρόβλημα μπορεί να μειωθεί ικανοποιητικά μέσα από την εκμετάλλευση των αλληλεπιδράσεων των χρηστών στο μέσο κοινωνικής δικτύωσης, και να παραχθούν προτεινόμενα προϊόντα με μεγαλύτερη ακρίβεια.

Συνεχίζοντας με τα αρνητικά σημεία των Κοινωνικών ΣΣ, είναι γεγονός ότι οι συνδέσεις-φιλίες που έχει κάποιος χρήστης σε μία κοινωνική πλατφόρμα περιέχει συνήθως ένα μικρό ποσοστό από άτομα που γνωρίζει. Ως αποτέλεσμα, το κοινωνικό του δίκτυο αποτελείται από «χρήσιμες» αλλά και «θορυβώδεις» συνδέσεις, με τις δεύτερες να είναι πιο πιθανό να του προκαλέσουν σύγχυση και υπερφόρτωση πληροφοριών. Έτσι, τελικά οι κοινωνικές σχέσεις μπορεί να έχουν αρνητικές συνέπειες στα συστήματα συστάσεων. Επιπλέον, αναφέραμε προηγουμένως ότι τα Κοινωνικά ΣΣ έχουν καλύτερα αποτελέσματα από τα κλασικά ΣΣ σχετικά με χρήστες που έχουν πολύ λίγες αξιολογήσεις στο ιστορικό τους. Ωστόσο, απαραίτητη προϋπόθεση είναι να έχουν έναν ικανοποιητικό αριθμό διασυνδέσεων μέσα στο μέσο κοινωνικής δικτύωσης. Παρόλα αυτά, έχει παρατηρηθεί ότι είθισται οι χρήστες με λίγες βαθμολογίες να έχουν και λίγες συνδέσεις στο δίκτυο [5], σε αντίθεση με τους χρήστες που έχουν πολλές κοινωνικές συνδέσεις και είναι ενεργά μέλη του δικτύου έχοντας κάνει πολλές αξιολογήσεις στο παρελθόν. Συνεπώς, είναι δύσκολο τα Κοινωνικά ΣΣ να βελτιώσουν αρκετά τη διαδικασία σύστασης προϊόντων για τέτοιους χρήστες.

Τέλος, στα Κοινωνικά ΣΣ σημαντικό ρόλο διαδραματίζει η εμπιστοσύνη, η οποία στην επιστήμη υπολογιστών γενικά μπορεί να ταξινομηθεί σε δύο ευρείες κατηγορίες, του *χρήστη* και του *συστήματος* [34]. Η πρώτη μπορεί να ορισθεί τυπικά ως «η προσδοκία που έχει μία

οντότητα για τη μελλοντική συμπεριφορά κάποιου άλλου». Αυτό σημαίνει ότι η εμπιστοσύνη είναι εξατομικευμένη (personalized). Σε διαδικτυακές πλατφόρμες όπως το eBay και το Amazon, η εμπιστοσύνη βασίζεται σε δεδομένα σχετικά με προηγούμενες αλληλεπιδράσεις μεταξύ των χρηστών. Κατά αυτήν την έννοια, η εμπιστοσύνη είναι σχεσιακή (relational). Όταν δύο μέλη ενός δικτύου αλληλοεπιδρούν μεταξύ τους συχνά, ο δεσμός τους ενισχύεται και η εμπιστοσύνη εξελίσσεται με βάση την εμπειρία τους. Η εμπιστοσύνη μεταξύ των δύο μελών αυξάνεται εάν η εμπειρία είναι θετική, διαφορετικά μειώνεται. Όσον αφορά την εμπιστοσύνη του συστήματος, θεωρείται ότι αποτελείται από δύο είδη: *άμεση εμπιστοσύνη* (direct trust) και *εμπιστοσύνη σύστασης* (recommendation trust). Η πρώτη βασίζεται στην άμεση εμπειρία ενός μέλους με κάποιο άλλο, ενώ η δεύτερη στην εμπειρία άλλων μελών, εκμεταλλεύεται δηλαδή την προωθητική ιδιότητα της εμπιστοσύνης. Έτσι, γνωρίζοντας την κοινωνική εμπιστοσύνη που υπάρχει ανάμεσα στους χρήστες, παρέχονται οι απαραίτητες πληροφορίες για κάποιον χρήστη σχετικά με το περιεχόμενο που πρέπει να μοιράζεται με τους διαδικτυακούς του φίλους, αλλά και για να αξιολογεί συνετά τις πληροφορίες που δέχεται από εκείνους [5]. Ωστόσο, κάποιες φορές η εμπιστοσύνη κρύβει ρίσκο, αφού μπορεί να ερμηνευτεί με διαφορετικούς τρόπους, που μπορούν να οδηγήσουν σε διαφορετικά αποτελέσματα στη σύσταση προϊόντων. Συνήθως, υπάρχει εμπιστοσύνη μεταξύ χρηστών όταν έχουν και παρεμφερή ενδιαφέροντα και απόψεις. Αλλά υπάρχει και η περίπτωση κάποιος χρήστης να εμπιστευτεί κάποιους άλλους εξαιτίας της αξιοπιστίας τους, κι όχι επειδή συμβαδίζουν οι απόψεις τους. Επίσης, οι σχέσεις εμπιστοσύνης δεν είναι απαραίτητα ισοδύναμες σε όλους τους τύπους κοινωνικών σχέσεων, όπως, για παράδειγμα, στο Twitter το γεγονός ότι ένας χρήστης ακολουθεί κάποιον άλλον δεν συνεπάγεται ότι τον εμπιστεύεται. Συμπερασματικά, διαφορετικοί τύποι κοινωνικών σχέσεων έχουν διαφορετική επίδραση στα Κοινωνικά ΣΣ.

4

Συστήματα Συστάσεων και Διάχυση της πληροφορίας

4.1 Διάχυση πληροφορίας σε Online Social Networks

Στην προηγούμενη ενότητα αναλύσαμε διάφορα παραδείγματα εφαρμογής Κοινωνικών ΣΣ. Συμπερασματικά, μπορούμε να πούμε ότι η αλληλεπίδραση μεταξύ των μελών ενός δικτύου είναι καθοριστική για αποφάσεις που πρέπει να ληφθούν προκειμένου να δημιουργηθεί μια λίστα συστάσεων προϊόντος/υπηρεσιών ανά χρήστη. Οι παράγοντες που επηρεάζουν τη διαδικασία ποικίλουν, για αυτό και έχουν γίνει αξιολογες προσπάθειες ώστε να δημιουργηθούν μηχανισμοί με στόχο ακριβέστερες προβλέψεις.

Στην ενότητα αυτή θα ξεφύγουμε από τα γενικά χαρακτηριστικά ενός μέσου κοινωνικής δικτύωσης, τα οποία εκμεταλλεύτηκαν διάφορες έρευνες που αναφέρθηκαν στην προηγούμενη ενότητα για να βελτιώσουν τα ΣΣ. Θα εστιάσουμε, ωστόσο, στην έννοια της διάχυσης της πληροφορίας, ώστε να αναλύσουμε στην πορεία μια υποκατηγορία των Κοινωνικών ΣΣ, τα Κοινωνικά ΣΣ που λαμβάνουν υπόψιν αυτήν την έννοια (Information Diffusion-Aware Recommendation Systems - IDARS), δηλαδή τα ΣΣ που υπολογίζουν όχι μόνο τα δομικά στοιχεία του δικτύου, όπως πόσο κεντρικός είναι ένας χρήστης με βάση τις μετρικές της Ανάλυσης Κοινωνικών Δικτύων, αλλά και χαρακτηριστικά συμπεριφοράς, δηλαδή την αλληλεπίδραση χρηστών όπως μετριέται από την εμπιστοσύνη (trust), την επιρροή (influence) και άλλα.

Η διάχυση της πληροφορίας μπορεί να οριστεί τυπικά ως *η διαδικασία με την οποία η πληροφορία μεταδίδεται μέσω συγκεκριμένων καναλιών μεταξύ των χρηστών ενός μέσου κοινωνικής δικτύωσης* [2]. Οι κύριοι παράγοντες που την επηρεάζουν αφορούν στη δομή

του δικτύου (πυκνότητα, κεντρικότητα, ομαδοποίηση, συνδεσιμότητα), στη συχνότητα των αλληλεπιδράσεων μεταξύ των χρηστών και στην επιρροή που ασκεί ο ένας στον άλλον. Τα μοντέλα που χρησιμοποιούνται για τη διάχυση της πληροφορίας θα μπορούσαν να διακριθούν σε δύο κατηγορίες: *επεξηγηματικά* και *προγνωστικά*. Τα πρώτα, δεδομένων των χρονικών στιγμών που «έφτασαν» κάποιες πληροφορίες στον χρήστη, έχουν ως στόχο να συνταχθεί ένα δίκτυο που αναπαριστά τη διάδοση της πληροφορίας. Τα μοντέλα αυτά καθιστούν δυνατή την ανίχνευση της διαδρομής που ακολουθεί μια πληροφορία και είναι πολύ χρήσιμο για να κατανοήσει κανείς τον τρόπο που διαδίδεται η πληροφορία [35]. Τα μοντέλα πρόβλεψης επιδιώκουν να προβλέψουν τον τρόπο με τον οποίο μια συγκεκριμένη διαδικασία διάχυσης εφαρμόζεται σε ένα δίκτυο, βάσει χρονικών ή χωρικών χαρακτηριστικών του δικτύου. Ευρέως χρησιμοποιούμενα είναι τα παρακάτω μοντέλα, τα οποία ανήκουν στην κατηγορία των προγνωστικών μοντέλων, και συγκεκριμένα στον έναν από τους δύο υπάρχοντες άξονες ανάπτυξης, τη λεγόμενη “graph based” προσέγγιση (η άλλη είναι η “non-graph based” προσέγγιση). Αυτά τα μοντέλα είναι τα Independent Cascade (IC) και Linear Threshold (LT). Το IC είναι ένα στοχαστικό μοντέλο στο οποίο ένας κόμβος u τη χρονική στιγμή t γίνεται ενεργός (δηλαδή λαμβάνει πληροφορίες από τη διαδικασία διάχυσης) και έχει μία μόνο ευκαιρία να ενεργοποιήσει κάθε μη ενεργό γειτονικό κόμβο v τη χρονική στιγμή $t + 1$, με πιθανότητα p_{uv} . Η διαδικασία συνεχίζεται έως ότου να μην πραγματοποιείται καμία ενεργοποίηση. Το LT είναι ένα από τα πιο δημοφιλή μοντέλα κατωφλίου που χρησιμοποιούνται στη μελέτη της διάχυσης στα δίκτυα. Σε αυτό το μοντέλο, μία τιμή κατωφλίου ή ένα σύνολο τιμών κατωφλίου χρησιμοποιείται ώστε να διακριθούν εύρη τιμών στα οποία η συμπεριφορά που προβλέπεται από το μοντέλο διαφέρει σημαντικά. Κάθε ακμή (u, v) του δικτύου συνδέεται με ένα βάρος w_{uv} και κάθε κόμβος v έχει ένα κατώφλι t_v . Ο κόμβος v ενεργοποιείται εάν το σύνολο των γειτονικών κόμβων του ξεπεράσει το t_v .

Ένας εύκολος τρόπος να αντιληφθεί κανείς τον τρόπο που μπορεί να εξαπλωθεί μια πληροφορία είναι να τον παρομοιάσει με μία πανδημία και τον τρόπο εξάπλωσης ενός ιού παγκοσμίως. Πάντοτε υπάρχει μία αφετηρία – η ιστοσελίδα που δημοσιεύει το συμβάν -, ύστερα οι χρήστες που έχουν επιλέξει να ακολουθούν αυτή τη σελίδα, και, τέλος, οι

υπόλοιποι ακόλουθοι. Ο βαθμός στον οποίο ο χρήστης που ακολουθεί τον ιστότοπο, θα επηρεαστεί και θα επιλέξει να διαβάσει το άρθρο αποτελείται από πολλά κριτήρια, όπως εάν του προκαλεί ενδιαφέρον (για παράδειγμα αν έγινε κοντά στον τόπο κατοικίας του ή καταγωγής του) ή αν έχει εμπιστοσύνη στην πηγή. Φυσικά, δεν είναι πάντοτε βέβαιο ότι, όταν ένας χρήστης λάβει μια πληροφορία, θα την αναδημοσιεύσει. Εξαρτάται από την κρίση του χρήστη και το ενδιαφέρον που του δημιουργήσει.

4.2 Διάχυση πληροφορίας σε Κοινωνικά ΣΣ (Information diffusion-aware recommendation systems)

Η διάχυση πληροφορίας αποτελεί σημαντικό άξονα για την μελέτη των ΣΣ σε μέσα κοινωνικής δικτύωσης και βοηθάει σημαντικά στη μείωση του κόστους των ΣΣ. Ένας δημοφιλής τρόπος διάχυσης της πληροφορίας είναι οι *αναδημοσιεύσεις* χρηστών [36] στα Μέσα Κοινωνικής Δικτύωσης (στο Facebook αναφέρεται ως “repost”, ενώ στο Twitter ως “retweet”). Παρακάτω, μέσα από διάφορες μελέτες, θα παρατεθούν και άλλοι τρόποι.

Τα κλασικά ΣΣ δεν λαμβάνουν υπόψιν έναν σημαντικό παράγοντα ο οποίος είναι πιθανό να αλλάζει το αποτέλεσμα των συστάσεων. Αγνοούν το μηχανισμό των αναδημοσιεύσεων από τους χρήστες, οι οποίες γίνονται διαθέσιμες στους ακόλουθούς τους. Αν όμως υπολογισθεί η παράμετρος αυτή, διαφοροποιείται σημαντικά ο τρόπος με τον οποίο λαμβάνονται οι αποφάσεις και μειώνεται ο αριθμός των συστάσεων που γίνονται σε κάθε χρήστη. Αυτό είναι λογικό αφού ένας χρήστης μπορεί να λάβει την πληροφορία από μια αναδημοσίευση και είναι περιττό πλέον να γίνει σύσταση. Συνεπώς είναι πολύ σημαντικό να συνδυάζεται ένα ΣΣ με τη διάχυση πληροφοριών που δημιουργείται από την αλληλεπίδραση των χρηστών ενός μέσου κοινωνικής δικτύωσης. Δηλαδή, στο κεφάλαιο αυτό δεν λαμβάνεται απλά υπόψη το γεγονός ότι ένας χρήστης υιοθετεί τη γνώμη ενός άλλου χρήστη για την αγορά ενός προϊόντος, αλλά, υπολογίζουμε ότι πήρε, μέσα από αναδημοσίευση του άλλου χρήστη, και πληροφορίες για κάποιο νέο προϊόν, και, άρα, δεν χρειάζεται ο μηχανισμός του ΣΣ να του προτείνει ξανά το προϊόν αυτό. Στα πλαίσια της παρούσας εργασίας πραγματοποιήθηκε διεξοδική αναζήτηση στην βιβλιογραφία για εφαρμογές των

Συστημάτων Συστάσεων με κύριο παράγοντα τη διάχυση της πληροφορίας σε μέσα κοινωνικής δικτύωσης τα τελευταία χρόνια.

4.2.1 Βιβλιογραφική ανασκόπηση εφαρμογών IDARS

Σημαντικές μελέτες που έχουν διεξαχθεί στο πλαίσιο των IDARS, μπορούν να διαχωριστούν σε τρεις κατηγορίες ανάλογα με τον τρόπο με τον οποίο ο μηχανισμός της διάχυσης της πληροφορίας ενσωματώνεται με το ΣΣ [2]. Η πρώτη κατηγορία περιλαμβάνει τις μεθοδολογίες οι οποίες χειρίζονται το πρόβλημα της υπερφόρτωσης πληροφοριών. Η εκθετική αύξηση των πληροφοριών στα μέσα κοινωνικής δικτύωσης, συνοδευόμενη από την περιορισμένη ικανότητα των χρηστών να επεξεργαστούν μεγάλο όγκο δεδομένων, επηρεάζουν τη λήψη αποφάσεων και αυξάνουν την ανάγκη δημιουργίας μηχανισμών μετάδοσης πληροφοριών σχετικές με τον κάθε χρήστη. Τα κλασικά ΣΣ αντιμετωπίζουν αυτό το θέμα «φιλτράροντας» τα δεδομένα σύμφωνα με τις προτιμήσεις των χρηστών, αγνοώντας, ωστόσο, τους δεσμούς μεταξύ των χρηστών ενός μέσου κοινωνικής δικτύωσης. Παρακάτω θα αναλυθούν μερικές εφαρμογές, όπου χρησιμοποιήθηκαν διαφορετικά μοντέλα διάχυσης της πληροφορίας είτε ως εναλλακτική είτε ως επέκταση του μηχανισμού φιλτραρίσματος των ΣΣ, προσθέτοντας κοινωνικά κριτήρια (επιρροή, εμπιστοσύνη, ομοφιλία) στις συστάσεις.

Πρώτα, θα αναφερθούμε σε μία έρευνα για τη σύσταση προϊόντων με τη συμβολή της διάχυσης της πληροφορίας μέσω αναδημοσιεύσεων που διεξάχθηκε από τους Vahabi et al. [36]. Δημιούργησαν έναν αλγόριθμο, τον DifRec, βάσει του οποίου παρουσιάζεται ένα μοντέλο αναμενόμενης διάχυσης των προτεινόμενων προϊόντων σε ένα δίκτυο, όπου προβλέπονται οι αντικρουόμενες συστάσεις μεταξύ του ΣΣ και της κοινωνικής διάχυσης, προκειμένου να βελτιστοποιηθεί η σύσταση προϊόντων και να αποφευχθεί η υπερφόρτωση πληροφοριών. Μάλιστα, στα πλαίσια της συγκεκριμένης έρευνας, γίνεται σύγκριση του αλγορίθμου με άλλες μεθόδους ΣΣ, όπου η διάχυση της πληροφορίας δεν συμπεριλαμβάνεται στις μεθόδους τους. Ο αλγόριθμος αυτός θα αναλυθεί πιο διεξοδικά στο

επόμενο κεφάλαιο καθώς αποτελεί σημαντικό μέρος της έρευνας της παρούσας εργασίας προς την εξαγωγή των συμπερασμάτων. Μία αντίστοιχη έρευνα [37] συνδυάζει τη σύσταση μέσω συνεργασίας λαμβάνοντας υπόψη τη διάχυση της πληροφορίας, ώστε να δημιουργήσει ένα ΣΣ για μέσα κοινωνικής δικτύωσης που θα αντιμετωπίζει το πρόβλημα της πληροφοριακής υπερφόρτωσης. Τα πειράματα σε πραγματικά δεδομένα έδειξαν ότι η ενσωμάτωση της διάχυσης της πληροφορίας είχε θετικά αποτελέσματα σε σχέση με ένα απλό ΣΣ. Η διάχυση της πληροφορίας εδώ δε γίνεται αποκλειστικά μέσω αναδημοσιεύσεων, όπως στην προηγούμενη έρευνα, αλλά λαμβάνονται υπόψιν σημαντικές πληροφορίες που μπορούν να επηρεάσουν σε μεγάλο βαθμό τη συμπεριφορά των χρηστών, όπως χαρακτηριστικά σχετικά με τη δημοφιλία (οι χρήστες στα μέσα κοινωνικής δικτύωσης έχουν την τάση να προτιμούν τις πιο δημοφιλείς πληροφορίες), και με τους «γείτονες» ενός χρήστη (είναι οι χρήστες με τους οποίους έχει παρόμοια ενδιαφέροντα και είθισται να τον επηρεάζουν στις αποφάσεις του). Οι Ullah & Lee [38] υποθέτουν ότι η δημοφιλία των χρηστών επηρεάζει τις συστάσεις που γίνονται στις «γειτονιές» τους, που αποτελούνται από χρήστες με παρεμφερείς προτιμήσεις. Αντιμετωπίζουν το πρόβλημα της υπερφόρτωσης πληροφοριών με τον έλεγχο της διαδικασίας διάχυσης, δημιουργώντας ένα μοντέλο επιρροής των χρηστών με διάφορα κριτήρια (χωρικά, δημογραφικά, ομοιότητα απόψεων), κατατάσσοντας τους χρήστες με βάση το βαθμό της επιρροής που καταγράφηκε (influence score), που ορίζεται ως ο συνδυασμός των παραπάνω χαρακτηριστικών και της δομής του δικτύου, και, τελικά, επιλέγοντας τις πιο υψηλές βαθμολογίες.

Η δεύτερη κατηγορία ΣΣ αφορά το λεγόμενο “long-tail” πρόβλημα [2]. Παρόλο που ένα μικρό σύνολο προϊόντων είναι πολύ δημοφιλή (αναφέρονται και ως “blockbusters”), τα λιγότερο συνηθισμένα προϊόντα (αναφέρονται και ως “niche products”) έχουν μεγαλύτερο μερίδιο στις πωλήσεις. Ως εκ τούτου, είναι σημαντικό για ένα ΣΣ να παρέχει συστάσεις των μη συνηθισμένων προϊόντων. Το γεγονός αυτό εξυπηρετεί όχι μόνο τους πωλητές, οι οποίοι έχουν περισσότερα κέρδη από τις πωλήσεις αυτών των προϊόντων, αλλά και τους χρήστες-αγοραστές οι οποίοι έχουν πρόσβαση σε μεγαλύτερη ποικιλία προϊόντων και, κατά επέκταση, αυξάνεται η αφοσίωση τους στις διαδικτυακές πλατφόρμες. Τα κλασικά ΣΣ, λόγω των ανεπαρκών διαθέσιμων δεδομένων, αδυνατούν να βελτιώσουν την ποικιλία σε

επιλογές προϊόντων για τους χρήστες. Το πρόβλημα αυτό ίσως μπορεί να λυθεί συστήνοντας μη-δημοφιλή προϊόντα σε χρήστες που θεωρούνται ότι έχουν κοινά ενδιαφέροντα και προτιμήσεις με μεγάλο μέρος των χρηστών ενός δικτύου, έχοντας, ταυτόχρονα, μεγάλο εξωτερικό βαθμό (out-degree) στο μέσο κοινωνικής δικτύωσης (influential users) [39]. Το γεγονός αυτό θα οδηγούσε σε αυξημένη διάδοση αυτών των προϊόντων μέσω του μηχανισμού διάδοσης. Στο [39] παρουσιάζεται μία καινοτόμα προσέγγιση για σύσταση προϊόντων με χρήση της σύστασης μέσω συνεργασίας για τη μετάδοση της καινοτομίας (innovation diffusion). Η σύσταση της μη-συνηθισμένης (ή καινοτόμου) πληροφορίας σε έναν χρήστη που επισκέπτεται μία ιστοσελίδα, λαμβάνεται με βάση τη συμπεριφορά περιήγησης στην σελίδα, κατά τη διάρκεια μίας συγκεκριμένης χρονικής περιόδου, ενός μικρού συνόλου χρηστών που ορίζονται ως χρήστες αναφοράς (είναι εκείνοι που είχαν πρόσβαση στην ίδια σελίδα πριν από το χρήστη στον οποίο θέλουμε να γίνει η σύσταση).

Έως τώρα, παρουσιάστηκαν μελέτες στις οποίες οι τεχνικές που εφαρμόζονται ξεπερνούν τις αδυναμίες των κλασικών ΣΣ, αλλά λίγες πετυχαίνουν υψηλή αποτελεσματικότητα στην ακρίβεια των αποτελεσμάτων. Η συνεισφορά της μοντελοποίησης της διάχυσης πληροφοριών προς αύξηση της αποτελεσματικότητας σε σύσταση μέσω συνεργασίας, όσον αφορά τη βελτιωμένη ακρίβεια, παρουσιάζεται από μια έρευνα των Song et al. [40], όπου κατασκευάζονται δύο δίκτυα διάχυσης σύμφωνα με δύο παραμέτρους (πρότυπα βασιζόμενα στον χρόνο πρόσβασης του χρήστη σε έγγραφα και πρότυπα υιοθέτησης που σχετίζονται με τα θέματα (topics) των εγγράφων που ο χρήστης έχει πρόσβαση). Τότε, τρία πιθανά μοντέλα διάδοσης της πληροφορίας προτείνονται για να προβλέψουν πώς οι πληροφορίες θα εξαπλωθούν μέσα στο δίκτυο. Μια άλλη προσέγγιση παρουσιάζεται στο [41] όπου προτείνεται ένα στοχαστικό μοντέλο για τη διάχυση πληροφοριών, που λαμβάνει υπόψιν τη δύναμη των δεσμών μεταξύ των χρηστών, βάσει της οποίας μελετάται το πλεονέκτημα των χρηστών από τις κοινωνικές αλληλεπιδράσεις. Το τελευταίο ποσοτικοποιείται από τη διαφορά που παρατηρείται στις αξιολογήσεις μεταξύ χρηστών με ισχυρό δεσμό (friends).

Μια άλλη ενδιαφέρουσα προσέγγιση είναι αυτή των Wang et al. [42], οι οποίοι, προκειμένου να αυξήσουν τη διάδοση μιας δημοσίευσης αξιοποίησαν τις λεγόμενες αναφορές (ή “mentions”). Με απλά λόγια, ένας χρήστης έχει τη δυνατότητα να αφήσει ως σχόλιο σε μια δημοσίευση το όνομα ενός χρήστη, με συγκεκριμένο τρόπο ώστε ο δεύτερος να λάβει ειδοποίηση για αυτήν την αναφορά και, εφόσον του αρέσει, μπορεί να την αναδημοσιεύσει ή να αναφέρει αντίστοιχα σε σχόλιο κάποιον άλλον χρήστη. Στα πλαίσια εκείνης της έρευνας, δημιούργησαν έναν μηχανισμό, ο οποίος προτείνει στους χρήστες σε ποιους να κάνουν αναφορά (whom-to-mention). Με αυτόν τον τρόπο, ενισχύεται η διάχυση της πληροφορίας, αλλά επισημαίνονται κάποια αδύναμα σημεία, όπως η υπερφόρτωση συστάσεων, εάν οι συστάσεις επικεντρώνονται μόνο σε λίγους δημοφιλείς χρήστες. Τότε κάποιοι χρήστες θα λαμβάνουν πάρα πολλές ειδοποιήσεις, που μπορεί να προκαλέσει τα αρνητικές συνέπειες.

5 Βασικές Μέθοδοι Συστημάτων Συστάσεων Με Ενσωμάτωση Της Διάχυσης Της Πληροφορίας

Στο κεφάλαιο αυτό, θα αναλύσουμε δύο βασικές τεχνικές για συστάσεις προϊόντων οι οποίες εκμεταλλεύονται τη διάχυση της πληροφορίας ώστε να διεξάγουν συστάσεις προϊόντων σε ένα μέσο κοινωνικής δικτύωσης. Ο μηχανισμός των αναδημοσιεύσεων που διαχέονται μεταξύ των χρηστών του δικτύου είναι η κεντρική ιδέα που χρησιμοποιείται για την υλοποίηση αυτών των μεθόδων. Συγκεκριμένα, οι συστάσεις μπορούν να διακριθούν σε *άμεσες* (explicit), δηλαδή εκείνες που προκύπτουν από το ΣΣ, και σε *έμμεσες* (implicit) από τις αναδημοσιεύσεις των χρηστών, δηλαδή με ανταλλαγή πληροφορίας μεταξύ των χρηστών του μέσου κοινωνικής δικτύωσης. Αρχικά, αναλύεται το πλαίσιο συστάσεων DifRec [36], στο οποίο ο καταμερισμός των συστάσεων γίνεται με τρόπο ώστε ένα προϊόν να λαμβάνεται από ένα χρήστη είτε έμμεσα είτε άμεσα, με περιορισμό του πλήθους των άμεσων συστάσεων ανά χρήστη. Θα περιγραφεί αναλυτικά η λειτουργία του παρακάτω και θα εξαχθούν συμπεράσματα ως προς την ακρίβεια των αποτελεσμάτων και της αποδοτικότητάς του, ενώ παράλληλα θα συγκριθεί με ένα άλλο Information Diffusion-Aware Σύστημα Συστάσεων, το CoveR, στο οποίο ο καταμερισμός γίνεται με τρόπο ώστε να ελαχιστοποιηθούν οι άμεσες συστάσεις εξασφαλίζοντας πως κάθε χρήστης έχει λάβει ένα συγκεκριμένο αριθμό συστάσεων είτε άμεσα είτε έμμεσα. Η αποδοτικότητα των παραπάνω πλαισίων συστάσεων θα αξιολογηθεί μέσω προσομοιώσεων σε συνθετικά δίκτυα ώστε να αξιολογήσουμε τα υπέρ και τα κατά των δύο μεθόδων.

5.1 Ανάλυση του αλγορίθμου για το πλαίσιο συστάσεων DifRec

Πρώτο βήμα είναι να γίνει περιγραφή του συστήματος που θα χρησιμοποιηθεί. Έστω μια πλατφόρμα κοινωνικής δικτύωσης που μοντελοποιείται με έναν κατευθυνόμενο γράφο $G(V, E)$, όπου V οι χρήστες του δικτύου και E οι ακμές που συνδέουν τους χρήστες. Όταν μία κατευθυνόμενη ακμή $(i, j) \in E$ υπάρχει ανάμεσα στους χρήστες i και j , τότε λέμε ότι ο χρήστης j «ακολουθεί» τον i , δηλαδή ο j έχει πρόσβαση σε όλα τα περιεχόμενα που αναδημοσιεύονται από τον i . Επίσης, έχουμε ένα σύνολο προϊόντων I , όπου ο όρος «προϊόν» αναφέρεται σε κάθε είδους πληροφορία, όπως μία είδηση, ένα αρχείο πολυμέσων, μία ιστοσελίδα, ένα tweet και άλλα. Χρησιμοποιείται, επιπλέον, ο όρος της «αναδημοσίευσης» που υποδηλώνει τη διαδικασία κοινοποίησης συγκεκριμένων πληροφοριών από έναν χρήστη προς όλους τους ακόλουθούς του. Έτσι, έχουμε δύο διαφορετικούς τρόπους με τους οποίους ένας χρήστης i λαμβάνει ένα προϊόν:

1. Μέσω περιεχομένου από χρήστες που ακολουθεί, που χαρακτηρίζεται ως **έμμεση σύσταση**,
2. Μέσω συστάσεων που παρέχονται από ένα ΣΣ, που χαρακτηρίζεται ως **άμεση σύσταση**.

Στόχος του ΣΣ είναι να προτείνει σε κάθε χρήστη $i \in V$ ένα σύνολο L_i , υποσύνολο του συνόλου προϊόντων I , με $|L_i| \leq K, \forall i \in U$. Ως K δηλώνεται ο μέγιστος αριθμός συστάσεων σε κάθε χρήστη (πλήθος διαφορετικών προϊόντων που μπορεί να δεχθεί ένας χρήστης) και, συνήθως, είναι ένας μικρός αριθμός, ώστε να αποφεύγεται η υπερφόρτωση πληροφοριών στον χρήστη.

Επίσης, για κάθε προϊόν $t \in I$ και κάθε χρήστη $i \in U$, ορίζεται ως r_{it} η *συνάφεια* (Relevance Score) του t στον i . Ο παράγοντας r_{it} ποσοτικοποιεί το ενδιαφέρον του χρήστη i για προϊόν t , ενώ υπολογίζεται από τα προϊόντα που ο χρήστης i έχει δημοσιεύσει, παρακολουθήσει ή αναδημοσιεύσει. Στη συνέχεια, γίνεται η υπόθεση ότι σε κάθε χρήστη i αντιστοιχίζεται μία πιθανότητα αναδημοσίευσης w_i . Η πιθανότητα αυτή εξαρτάται από το

πόσο συχνά ένας χρήστης έχει την τάση να κάνει αναδημοσιεύσεις και μπορεί να υπολογιστεί από τη συνολική δραστηριότητά του μέσα στην πλατφόρμα κοινωνικής δικτύωσης.

Η κεντρική ιδέα του DifRec είναι η εξής: θεωρούμε ένα προϊόν το οποίο αναγνωρίζεται ως *συναφές* (relevant) προς κάποιον χρήστη μιας συγκεκριμένης διαδικτυακής πλατφόρμας, και, ως εκ τούτου, προτείνεται σε αυτόν. Υποθέτουμε, επίσης, ότι το προϊόν αυτό είναι σχετικό και σε έναν από τους ακολούθους του χρήστη, και, συνεπώς, προτείνεται και σε αυτόν. Αν ο πρώτος αναδημοσιεύσει το προϊόν, τότε ο ακόλουθός του θα το λάβει δύο φορές (από την αναδημοσίευση και από το ΣΣ). Καθώς ο μέγιστος αριθμός προτάσεων προϊόντων είναι περιορισμένος για συγκεκριμένη χρονική περίοδο (ώστε να αποφεύγεται η υπερφόρτωση πληροφοριών στον χρήστη, που μπορεί να οδηγήσει τελικά σε σύγχυση), το παραπάνω σενάριο οδηγεί σε μία ανεπιθύμητη κατάσταση, όπου ένας σημαντικός αριθμός προϊόντων προτείνεται στους χρήστες και δημιουργούνται διπλότυπα, αφού κάποιος μπορεί να τα λάβουν από αναδημοσιεύσεις. Για αυτόν τον λόγο, ο DifRec επιδιώκει να αποφεύγονται επαναλαμβανόμενες συστάσεις, δημιουργώντας μία συνθήκη που δεν επιτρέπει να λαμβάνουν οι χρήστες ενός δικτύου κάποιο προϊόν και έμμεσα, δηλαδή από αναδημοσίευση, και άμεσα, δηλαδή από τον μηχανισμό των ΣΣ. Παρακάτω θα γίνει ανάλυση του DifRec, ώστε να επεξηγηθούν βήμα-προς-βήμα οι βασικοί αλγόριθμοι που τον αποτελούν.

Στο συγκεκριμένο μοντέλο που χρησιμοποιείται για την υλοποίηση του DifRec, οι ακμές του γράφου, ο οποίος ορίστηκε παραπάνω, έχουν βάρη που αντιστοιχίζονται στην *πιθανότητα αναδημοσίευσης* του πηγαίου κόμβου.

5.1.1 Αλγόριθμος «Reliable Set»

Αποτελεί τον πρώτο βασικό αλγόριθμο που μοντελοποιεί τη διάχυση της πληροφορίας στο δίκτυο. Στόχος του είναι, δοθέντος του γράφου G , να υπολογίζει, με τη δειγματοληπτική τεχνική Monte Carlo, το Αξιόπιστο Σύνολο του χρήστη i (Reliable Set R_i). Πρόκειται για το

σύνολο των κόμβων που είναι προσβάσιμοι από τον i με υψηλή πιθανότητα. Με άλλα λόγια, ο αλγόριθμος αυτός δίνει το σύνολο των χρηστών που θα λάβουν με υψηλή πιθανότητα κάθε προϊόν αναδημοσιευμένο από τον χρήστη i .

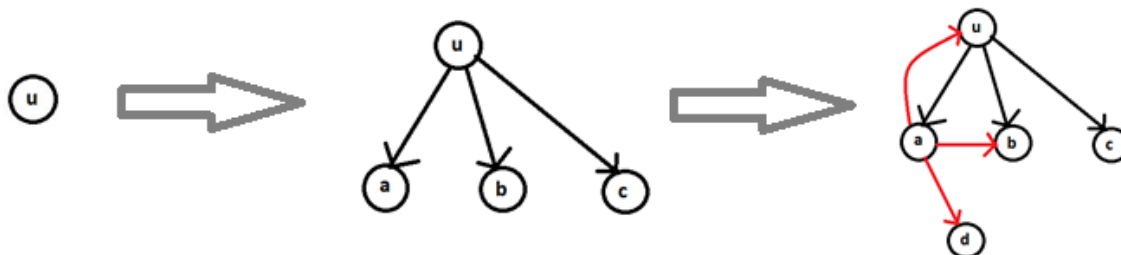
Έχοντας υπολογίσει το Reliable Set για όλους τους χρήστες του δικτύου, στο [36] ορίζεται το *Global Engagement Potential* ή *Total Relevance Score* για ένα προϊόν t που προτείνεται σε έναν χρήστη i . Συγκεκριμένα, ορίζεται ως το άθροισμα των Relevance Scores r_{jt} , για όλους τους κόμβους j που ανήκουν στο Reliable Set του i και δίνεται από τον τύπο:

$$E_{it} = \sum_{j \in R_i} r_{jt}$$

5.1.2 Αλγόριθμος «Neighborhoodness Graph»

Με αυτόν τον αλγόριθμο συγκεντρώνονται τα δεδομένα της διάχυσης της πληροφορίας στο δίκτυο όπως ορίστηκαν από τα Reliable Sets. Δημιουργείται ένας γράφος γειτνίασης (neighborhoodness graph) έτσι ώστε αν μια ακμή (i, j) υπάρχει στον γράφο αυτό, τότε όλα τα προϊόντα που αναδημοσιεύονται από τον i , θα τα λάβει και ο j (από αναδημοσίευση του i) και, άρα, δεν χρειάζεται να συσταθούν στον j . Αυτόν ακριβώς τον κανόνα ακολουθεί ο DifRec: το ίδιο προϊόν δεν προτείνεται σε δύο κόμβους-χρήστες, οι οποίοι διασυνδέονται στον γράφο γειτνίασης.

Για να δημιουργηθεί ο γράφος γειτνίασης, χρησιμοποιούνται τα Reliable Sets που υπολογίστηκαν στον προηγούμενο αλγόριθμο. Πιο συγκεκριμένα, δοθέντος ενός δικτύου $G = (V, E)$ και των reliable sets R_i για κάθε κόμβο $i \in V$, ορίζουμε το γράφο γειτνίασης $G_n = (V_n, E_n)$ ως εξής: το σύνολο των κόμβων του G_n αντιστοιχεί στο σύνολο των κόμβων G , δηλαδή $V_n = V$, ενώ η κατευθυνόμενη ακμή $(i, j) \in V_n$ υπάρχει ανάμεσα στους χρήστες $i, j \in V$ αν και μόνο αν ο κόμβος j ανήκει στο reliable set R_i του κόμβου i . Θα παραθέσουμε ένα απλό παράδειγμα δημιουργίας ενός τέτοιου γράφου. Έστω ο κόμβος u (ενός τυχαίου γράφου) με Reliable Set $\{a, b, c\}$, όπως φαίνεται στην παρακάτω εικόνα.



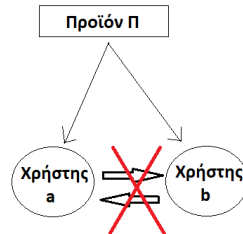
Εικόνα 5.1 Παράδειγμα δημιουργίας ενός γράφου γειτνίασης.

Τότε σχεδιάζουμε τις ακμές (u, a) , (u, b) , (u, c) , λόγω του δοθέντος Reliable Set. Έστω τώρα ο κόμβος a με Reliable Set $\{u, d, b\}$. Τότε σχεδιάζουμε τις ακμές (a, u) , (a, d) , (a, b) . Η διαδικασία αυτή ακολουθείται για όλους τους κόμβους του δοθέντος γράφου έως ότου ολοκληρωθεί ο πίνακας γειτνίασης. Αξίζει να σημειωθεί ότι και ο γράφος γειτνίασης είναι κατευθυνόμενος και μάλιστα διπλής κατευθύνσεως όπου είναι απαραίτητο.

5.1.3 Αλγόριθμος «Independent Set»

Η διαδικασία που ακολουθείται εδώ δίνει ένα Ανεξάρτητο Σύνολο (Independent Set) χρηστών στον γράφο γειτνίασης που αντιπροσωπεύει τους χρήστες στους οποίους το προϊόν μπορεί να προταθεί χωρίς να παραβιάζεται ο περιορισμός ότι το ίδιο προϊόν δεν προτείνεται σε δύο κόμβους που συνδέονται με ακμή στον γράφο γειτνίασης. Δεδομένου του γράφου γειτνίασης G_n και του προϊόντος t για το οποίο θέλουμε να «τρέξουμε» τον αλγόριθμο, ένα ανεξάρτητο σύνολο I_t στον γράφο αυτό είναι ένα υποσύνολο κόμβων, όπου κανένα ζευγάρι κόμβων στο I_t δεν συνδέεται στον γράφο γειτνίασης. Αν στους κόμβους αντιστοιχιστούν βάρη (ίσα με E_{it} σε κάθε κόμβο $i \in V_n$), προκύπτει σταθμισμένο ανεξάρτητο σύνολο I_t (Maximum-Weighted Independent Set), το οποίο δίνει μέγιστο άθροισμα βαρών των

κόμβων. Για παράδειγμα, στην παρακάτω εικόνα 5.2 το ίδιο προϊόν Π προτείνεται στους χρήστες a, b αλλά αυτοί δε συνδέονται μεταξύ τους με ακμή στον γράφο γειτνίασης.



Εικόνα 5.2 Παράδειγμα δύο χρηστών a, b που δέχονται ως εξωτερική πρόταση το ίδιο προϊόν Π , αλλά δεν συνδέονται μεταξύ τους στον γράφο γειτνίασης.

Ο αλγόριθμος που εφαρμόζεται λειτουργεί με τον εξής τρόπο για κάθε προϊόν t που πρόκειται να προταθεί (το προϊόν t εισάγεται ως είσοδος στη συνάρτηση αυτή): Οι κόμβοι i ταξινομούνται κατά φθίνουσα σειρά σύμφωνα με μια μετρική που ορίστηκε στο [36] και συμβολίζεται ως $\frac{E_{it}}{1+d_i}$, όπου η παράμετρος E_{it} ορίστηκε παραπάνω, και d_i είναι ο εξωτερικός βαθμός (out-degree) του κόμβου i .

Έτσι, αφού οι κόμβοι ταξινομηθούν, συμπεριλαμβάνονται στο ανεξάρτητο σύνολο με τη σειρά κατάταξης. Όταν ένας κόμβος i συμπεριλαμβάνεται στο σύνολο αυτό, αφαιρούμε από το γράφο αυτόν τον κόμβο και τις εξερχόμενες συνδέσεις από τον κόμβο i και τους γειτονικούς κόμβους που τον ακολουθούν (out-neighbors).

Το σκεπτικό της δημιουργίας του παραπάνω πηλίκου $\frac{E_{it}}{1+d_i}$, που συνδυάζει το E_{it} με το d_i , είναι ότι με αυτόν τον τρόπο δίνεται προτεραιότητα σε:

(α) κόμβους που παρουσιάζουν υψηλό βάρος, δηλαδή χρήστες για τους οποίους η σύσταση του προϊόντος t θα οδηγούσε σε μεγάλη διάχυση πληροφορίας, και, άρα, μεγάλο Total Relevance Score για τους χρήστες που λαμβάνουν ένα προϊόν μέσω αναδημοσίευσης,

αλλά παράλληλα σε:

(β) κόμβους με μικρό out-degree, οι οποίοι δεν συνδέονται με μεγάλο σύνολο γειτονικών κόμβων, στο οποίο το ίδιο προϊόν δεν μπορεί να προταθεί. Κόμβοι με μικρό out-degree, ως εκ τούτου, θα επέτρεπαν μεγαλύτερο ανεξάρτητο σύνολο για τα προϊόντα.

Συμπερασματικά, όσο πιο μεγάλο είναι το Relevance Score και μικρό το out-degree, τόσο καλύτερος υποψήφιος είναι ένας κόμβος για να συμπεριληφθεί στο ανεξάρτητο σύνολο, αφού για την ίδια τιμή της μετρικής, η σύσταση ενός προϊόντος σε μικρό πλήθος κόμβων με υψηλό Total Relevance Score είναι σημαντικότερη από ότι η σύσταση προϊόντος σε μεγάλο πλήθος κόμβων με χαμηλό Total Relevance Score. Επίσης, η επιλογή κόμβων με μικρό βαθμό επιτρέπει περισσότερες άμεσες (explicit) συστάσεις ενός προϊόντος, και, άρα, μεγαλύτερο έλεγχο της διάχυσης πληροφορίας από το Σύστημα Συστάσεων.

5.1.4 Αλγόριθμος «Recommend»

Είναι ο τελευταίος αλγόριθμος που δημιουργεί τις λίστες σύστασης των χρηστών από τα δεδομένα που έχει λάβει από τον αλγόριθμο IS. Για κάθε χρήστη i , η λίστα του L_i θα έχει το πολύ K προϊόντα που θα του προταθούν. Πιο συγκεκριμένα, αφού ολοκληρωθεί ο προηγούμενος αλγόριθμος, εντοπίζεται ένα ανεξάρτητο σύνολο I_t , για κάθε προϊόν $t \in I$. Ύστερα, ταξινομούνται τα προϊόντα σε φθίνουσα σειρά του βάρους τους $W(I_t)$ και εκχωρούνται τα προϊόντα στους χρήστες με αυτή τη σειρά. Όταν, για κάποιον χρήστη j τα προτεινόμενα προϊόντα γίνουν ίσα με K , δεν λαμβάνουμε πλέον υπόψιν αυτόν το χρήστη για σύσταση προϊόντος. Η διαδικασία σταματάει όταν κάθε χρήστης έχει K συστάσεις ή όταν όλα τα ανεξάρτητα σύνολα, που είναι διαθέσιμα, έχουν αναλυθεί.

5.1.5 Υπολογισμός του συνολικού Relevance Score του δικτύου

Προκειμένου να αξιολογηθεί η ποιότητα των προτάσεων που γίνονται σε ένα δίκτυο με την εφαρμογή του αλγορίθμου DifRec, υπολογίζουμε το συνολικό Relevance Score του

δικτύου, δηλαδή το Relevance Score από τα προϊόντα που «βλέπουν» οι χρήστες, είτε μέσω του $\Sigma\Sigma$, είτε μέσα από αναδημοσιεύσεις που λαμβάνουν οι χρήστες από εκείνους που ακολουθούν. Η ιδέα είναι να υπολογιστεί, για κάθε χρήστη i , το άθροισμα των Relevance Scores r_{it} για όλα τα προϊόντα t που προτάθηκαν άμεσα στον i από τη μέθοδο σύστασης, αλλά και το άθροισμα των Relevance Scores $r_{it'}$ για όλα τα προϊόντα t' που έλαβε ο χρήστης i έμμεσα από τη διάχυση της πληροφορίας.

Ορίζεται τυπικά το συνολικό Relevance Score του δικτύου ως εξής: για έναν χρήστη i , θεωρούμε ως \mathbf{R}_i το σύνολο των χρηστών των οποίων το reliable set περιέχει τον χρήστη i (εξαιρουμένου του i). Δεδομένου του συνόλου των συστάσεων από το $\Sigma\Sigma$, $\mathbf{L} = \{L_i\}_{i \in V}$, ορίζεται το *Total Relevance Score* $E(\mathbf{L}, i)$ ενός χρήστη i ως το άθροισμα δύο όρων:

1. Το Relevance Score των προϊόντων που συστάθηκαν άμεσα σε εκείνον (δηλαδή τα προϊόντα στη λίστα L_i).
2. Το Relevance Score των προϊόντων που έλαβε μέσω διάχυσης της πληροφορίας (δηλαδή τα προϊόντα που είναι στις λίστες συστάσεως όλων των χρηστών που περιέχονται στο σύνολο \mathbf{R}_i).

Συμβολικά έχουμε:

$$E(\mathbf{L}, i) = \sum_{t \in L_i} r_{it} + \sum_{t' \in \cup_{j \in \mathbf{R}_i} L_j} r_{it'}$$

Για να υπολογίσουμε το (μέσο) συνολικό Relevance Score του \mathbf{L} , και, κατά συνέπεια, όλου του δικτύου, έχουμε:

$$E(\mathbf{L}) = \frac{1}{|V|} \sum_{i \in V} E(\mathbf{L}, i).$$

5.1.6 Σημαντικά συμπεράσματα για τον DifRec

Στόχος του DifRec είναι να κατανείμει τα προϊόντα στους χρήστες λαμβάνοντας υπόψη τη μεταξύ τους ανταλλαγή πληροφορίας, ώστε να μεγιστοποιηθεί το συνολικό Relevance Score του δικτύου. Σημαντικό χαρακτηριστικό αυτής της τεχνικής είναι ότι αποκλείεται σε έναν χρήστη να προταθεί το ίδιο προϊόν έμμεσα, δηλαδή μέσω αναδημοσίευσης, και άμεσα, δηλαδή μέσω του ΣΣ.

Ωστόσο, αν και επιτυγχάνεται ένας συγκεκριμένος αριθμός K άμεσων συστάσεων, άρα K προϊόντα, σε ένα υποσύνολο χρηστών (κάποιοι θα λάβουν λιγότερα από K προϊόντα), δε συμβαίνει το ίδιο για τις έμμεσες προτάσεις, καθώς ένας χρήστης μπορεί να λάβει πολλές φορές το ίδιο προϊόν από την ανταλλαγή πληροφοριών με τους χρήστες του δικτύου που συνδέεται. Κατά συνέπεια, θα λάβει αθροιστικά περισσότερες από K προτάσεις. Επιπλέον, σε ορισμένες περιπτώσεις, μπορεί να υπάρξουν χρήστες που θα αποκλειστούν και από τα δύο είδη προτάσεων για όλα τα προϊόντα, λόγω του υπολογισμού του ανεξάρτητου συνόλου σε έναν κατευθυνόμενο γράφο. Έτσι, δεν έχουμε κάλυψη όλου του δικτύου χρηστών, ούτε με άμεσες ούτε με έμμεσες προτάσεις.

Με αφορμή τις παρατηρήσεις που παρατέθηκαν στον DifRec, δημιουργήθηκε η ανάγκη να εφαρμόσουμε μια άλλη μέθοδο, η οποία επίσης λαμβάνει υπόψη τις αλληλεπιδράσεις των χρηστών σε ένα δίκτυο, ενώ, ταυτόχρονα αναιρεί τον περιορισμό που δεν επιτρέπει σε έναν κόμβο να λάβει explicitly ούτε implicitly ένα προϊόν. Η τεχνική CoveR [43] αντιμετωπίζει το πρόβλημα της κατανομής K προϊόντων ανά χρήστη ως *πρόβλημα κάλυψης των χρηστών από K προϊόντα με τις λιγότερες δυνατές άμεσες αναθέσεις*, ώστε να μεγιστοποιείται το συνολικό Relevance Score του Δικτύου, και παράλληλα να ελαχιστοποιηθούν τα προϊόντα που προτείνονται σε έναν χρήστη πάνω από μία φορά.

5.2 Επίλυση προβλήματος κατανομής αντικειμένων με τη μέθοδο κάλυψης δικτύου

5.2.1 Εισαγωγή

Όπως αναφέραμε και στην προηγούμενη ενότητα, στην παρούσα μελέτη, αναζητείται μία μέθοδος κατανομής προϊόντων σε ένα δίκτυο, στο οποίο οι χρήστες έχουν αλληλεπιδράσεις μεταξύ τους, αλλά δεν υφίσταται ο περιορισμός που δεν αφήνει έναν κόμβο να λάβει ένα προϊόν και από αναδημοσίευση (implicitly) και από εξωτερικό μηχανισμό ΣΣ (explicitly). Τελικός στόχος είναι η βέλτιστη διάχυση των προϊόντων στο δίκτυο, χρησιμοποιώντας τους ελάχιστους δυνατούς πόρους. Το πρόβλημα διατυπώνεται ως μια παραλλαγή του “**Weighted Set Cover**” και θα αναλυθεί εκτενέστερα παρακάτω, όπως και ο αλγόριθμος που αναπτύχθηκε για την επίλυσή του.

5.2.2 Ανάλυση προβλήματος κάλυψης δικτύου

Έστω ένα μέσο κοινωνικής δικτύωσης, το οποίο αναπαρίσταται ως ένας γράφος $G(V, E)$, όπου $V = \{u_1, \dots, u_n\}$ είναι το σύνολο των χρηστών του δικτύου και E οι ακμές που συνδέουν τους χρήστες, και ένα σύνολο προϊόντων $I = \{i_1, \dots, i_k\}$. Στόχος είναι να προταθούν τα προϊόντα αυτά στους χρήστες μέσω δύο μηχανισμών – έναν άμεσο, όπου τα προϊόντα προτείνονται απευθείας στους χρήστες μέσω Συστήματος Συστάσεων (ΣΣ), και έναν έμμεσο, όπου οι χρήστες προτείνουν προϊόντα ο ένας στον άλλον μέσω αλληλεπιδράσεων (διαδικασία διάχυσης πληροφορίας).

Για να προχωρήσουμε στα παρακάτω βήματα της ανάλυσης χρησιμοποιούμε τους δύο πρώτους αλγόριθμους του DifRec, δηλαδή εκείνους που υπολογίζουν το reliable set κάθε κόμβου του δικτύου και τον γράφο γειτνίασης G_n . Αυτό σημαίνει ότι και εδώ πρέπει να δίνονται σαν είσοδοι στο σύστημα η πιθανότητα αναδημοσίευσης κάθε κόμβου του δικτύου, καθώς επίσης και το Relevance Score κάθε προϊόντος σε κάθε χρήστη. Έτσι,

χρησιμοποιώντας τον γράφο γειτνίασης $G_n(V, E_n)$, θεωρούμε μια άμεση πρόταση του προϊόντος i στον χρήστη u . Οι χρήστες που θα λάβουν έμμεσα την πρόταση του προϊόντος i είναι το σύνολο των ακόλουθών του όπως προκύπτουν στον γράφο G_n . Το σύνολο αυτό, συμπεριλαμβανομένου και του χρήστη u , ορίζεται ως:

$$R_u = \{u\} \cup \left\{ \bigcup_{(u,v) \in E_n} u \right\}$$

Επίσης, δεδομένου του Relevance Score του χρήστη u στο προϊόν i , $r_{u,i}$, ορίζεται το Relevance Score του συνόλου R_u ως το άθροισμα των πρώτων, για κάθε χρήστη u που ανήκει στο R_u :

$$E_{u,i} = \sum_{u \in R_u} r_{u,i}$$

Στο πρόβλημα αυτό επίσης πρέπει να σημειωθεί ότι χρησιμοποιείται η έννοια της *συσχέτισης (association) των χρηστών με τα προϊόντα* ($u_j^{(i)}$ είναι ο χρήστης u_j που σχετίζεται με το προϊόν i). Με απλά λόγια, έστω ότι έχουμε ένα δίκτυο 6 χρηστών και 4 προϊόντα. Ο χρήστης 1 σχετίζεται με το προϊόν 1, το προϊόν 2, το προϊόν 3 και το προϊόν 4. Τότε θα αναφερόμαστε στον χρήστη 1 ως εξής: $1^{(1)}, 1^{(2)}, 1^{(3)}, 1^{(4)}$. Αντίστοιχα ισχύει και για τους υπόλοιπους χρήστες. Δηλαδή, εδώ η αναφορά, για παράδειγμα, στον χρήστη 1 είναι απροσδιόριστη, καθώς πρέπει να αναφέρουμε και το προϊόν στο οποίο αναφέρεται.

Έτσι πλέον μπορούμε να ορίσουμε το σύνολο όλων των στοιχείων του δικτύου, το πλήθος των οποίων είναι $n \times k$, λόγω της συσχέτισης μεταξύ των στοιχείων και των προϊόντων. Συμβολίζεται ως εξής:

$$S = \{u_1^{(1)}, u_1^{(2)}, \dots, u_1^{(k)}, \dots, u_n^{(1)}, u_n^{(2)}, \dots, u_n^{(k)}\}$$

Ορίζοντας, επίσης, τα σύνολα $U_j = \bigcup_{m=1}^k u_j^{(m)}$ τότε θα ισχύει:

$$S = \bigcup_{j=1}^n U_j$$

Τα σύνολα U_j περιέχουν όλες τις συσχετίσεις των u_j με το σύνολο των προϊόντων. Ομοίως, ορίζουμε μια οικογένεια από $n \times k$ υποσύνολα του S , που ορίζεται από τα Reliable Sets των χρηστών συσχετιζόμενα με τα προϊόντα:

$$F = \{R_{u_1}^{(1)}, R_{u_1}^{(2)}, \dots, R_{u_1}^{(k)}, \dots, R_{u_n}^{(1)}, R_{u_n}^{(2)}, \dots, R_{u_n}^{(k)}\}$$

Για την ανάπτυξη του αλγορίθμου, θα ορίσουμε μια **συνάρτηση κόστους** που σχετίζεται με την F ως εξής:

$$c(R_u^{(i)}) = \frac{1}{E_{u,i}+1}, \forall i \in I, u \in V.$$

Στόχος είναι να εντοπιστεί ένα υποσύνολο του F τα στοιχεία του οποίου ελαχιστοποιούν το κόστος, έτσι ώστε κάθε χρήστης να λάβει τουλάχιστον K προτάσεις προϊόντων (K είναι το πλήθος των διαφορετικών προϊόντων που μπορεί να δεχθεί ένας χρήστης και, στο πρόβλημα αυτό, ορίζεται ως το ελάχιστο πλήθος προτάσεων που πρέπει να λάβει ένας χρήστης είτε άμεσα είτε έμμεσα). Σε αυτήν την περίπτωση το δίκτυο είναι K – covered (αφαιρώντας μία πρόταση – recommendation, θα έχει ως αποτέλεσμα κάποιοι χρήστες να λάβουν λιγότερα από K προϊόντα). Παρακάτω παρουσιάζεται ο αλγόριθμος ο οποίος χρησιμοποιείται για την επίλυση αυτού του προβλήματος.

5.2.3 Ανάλυση του αλγορίθμου για το πλαίσιο συστάσεων Cover

Θεωρούμε ότι X_t είναι το σύνολο των στοιχείων του δικτύου που δεν έχουν καλυφθεί (uncovered elements) τη χρονική στιγμή t , με $X_1 = S$, το σύνολο δηλαδή όλων των

στοιχείων του δικτύου, που ορίστηκε παραπάνω. Θα επιλέξουμε το υποσύνολο του F που ελαχιστοποιεί το λόγο:

$$\frac{c(R_u^{(i)}) * |R_u^{(i)}|}{|R_u^{(i)} \cap X_t|}$$

Ο λόγος αυτός αναδεικνύει τις προτάσεις προϊόντων, οι οποίες ελαχιστοποιούν το κόστος και μεγιστοποιούν τους κόμβους που καλύπτονται.

Σε κάθε επανάληψη ακολουθείται η εξής διαδικασία: Δηλώνουμε το σύνολο των στοιχείων που δικτύου που έχουν «καλυφθεί» (covered elements), δηλαδή έχουν λάβει προτάσεις προϊόντων, στην αρχή της επανάληψης t ως C_t . Υπολογίζονται οι λόγοι για όλα τα στοιχεία του X_t και, μόλις επιλεγθεί το $R_u^{(i)}$ με τον μικρότερο λόγο, τότε διακρίνουμε δύο συνθήκες ώστε να προχωρήσουμε στην επόμενη επανάληψη:

- Αν $|\bigcup_{j=1}^k R_u^{(j)} \cap C_t| < l$, τότε $X_{t+1} = X_t \setminus R_u^{(i)}$
- Αν $|\bigcup_{j=1}^k R_u^{(j)} \cap C_t| \geq l$, τότε $X_{t+1} = X_t \setminus \bigcup_{i=1}^k R_u^{(i)}$

Οι συνθήκες αυτές εξασφαλίζουν ότι οι χρήστες θα λάβουν τουλάχιστον K συστάσεις, (K - κάλυψη δικτύου) όπως ορίζει το πλαίσιο συστάσεων CoveR, διαφορετικά θα πρέπει να συνεχίσουν οι συστάσεις για αυτούς μέχρι να «φτάσουν» το απαραίτητο κατώφλι.

Η διαδικασία συνεχίζεται μέχρι την επανάληψη t' , όπου $X_{t'} = \emptyset$. Τελικά, βάσει του συνόλου $C_{t'}$, που προκύπτει στην τελευταία επανάληψη, προκύπτει ένας πίνακας M , που περιέχει τις τελικές συστάσεις στο δίκτυο.

5.3 Αποτελέσματα των μεθόδων σε Scale-Free και Small-World δίκτυα

Η χρησιμότητα των παραπάνω αλγορίθμων DifRec και CoveR, αλλά και τα αδύναμα σημεία τους μπορούν να αναδειχθούν με μεγαλύτερη ακρίβεια σε μεγάλα δίκτυα, καθώς είναι πιο ρεαλιστικά και μπορούν να απεικονίσουν ένα πραγματικό μέσο κοινωνικής δικτύωσης με μεγάλο αριθμό χρηστών και πληθώρα αλληλεπιδράσεων μεταξύ τους. Παραδείγματα τέτοιων συνθετικών δικτύων είναι τα Scale-Free (S-F) και Small-World (S-W). Παρακάτω παρουσιάζεται η μελέτη και σύγκριση που πραγματοποιήθηκε σε τέτοια δίκτυα στο [43], προκειμένου να εξαχθούν συμπεράσματα για τους δύο αλγορίθμους.

Πιο συγκεκριμένα, εξετάζονται τα αποτελέσματα των προσομοιώσεων σε 25 δίκτυα Scale-Free και 25 δίκτυα Small-World για τις παρακάτω περιπτώσεις συνδυαστικά:

- Πλήθος κόμβων: 100, 200 και 300
- Πλήθος προϊόντων: 20, 40, 100
- Πλήθος διαφορετικών προϊόντων που μπορεί να δεχθεί ένας χρήστης: 2 και 5

Χρησιμοποιήθηκαν οι παρακάτω μετρικές, οι οποίες βοηθούν στην σύγκριση των αποτελεσματικότητας των δύο μεθόδων:

1. $|V|$: το πλήθος των χρηστών
2. $|I|$: το πλήθος των προϊόντων
3. K : πλήθος διαφορετικών προϊόντων που μπορεί να δεχθεί ένας χρήστης
4. $|R_e|$: ο μέσος αριθμός των άμεσων προτάσεων
5. $|R_i|$: ο μέσος αριθμός των έμμεσων προτάσεων
6. S_c : το μέσο Relevance Score

7. $|I_r|$: ο μέσος αριθμός των προϊόντων που προτάθηκαν από το Σύστημα Συστάσεων
8. $|V_c|$: ο μέσος αριθμός των χρηστών που παρουσιάζουν τουλάχιστον μία από τις παρακάτω ιδιότητες:
 - α) Έχουν λάβει πρόταση για περισσότερα από K προϊόντα.
 - β) Έχουν λάβει ως πρόταση κάποιο προϊόν για περισσότερες από μία φορές.

Τότε για τους χρήστες αυτούς λέμε ότι συνέβη «παραβίαση χωρητικότητας» (capacity violation).
9. $|V_0|$: ο μέσος αριθμός των χρηστών που δεν έλαβαν καμία πρόταση
10. $|IV_c|$: ο αριθμός των προϊόντων ανά χρήστη για χρήστες στους οποίους έγινε παραβίαση χωρητικότητας

Παρακάτω φαίνονται οι πίνακες των αποτελεσμάτων των δύο αλγορίθμων ανά περίπτωση δικτύου και παραμέτρων.

Πίνακας 5.1 DifRec σταθμισμένα αποτελέσματα σε 25 Scale-Free δίκτυα ($|V|=\{100,200,300\}$, $|I|=\{20,40,100\}$, $K=\{2,5\}$)

$ U $	$ I $	K	$ R_e $	$ R_i $	Sc	$ I_r $	$ V_c $	$ V_0 $	$ IV_c $
100	20	2	173.04	33.32	108.91	9	13.3	12.36	4.08
100	20	5	427.48	70.4	260.533	16.8	13.16	12.36	9.39
100	40	2	174.48	38.4	119.902	10.76	14.6	11.8	4.23
100	40	5	433.36	85.56	284.835	20.84	14.64	11.8	9.91
100	100	2	175.28	42.36	117.663	11.52	15.5	11.36	4.26
100	100	5	436.56	96.92	288.653	24.6	15.52	11.36	10.3
200	20	2	354.68	63.4	218.687	11.16	25.4	20.12	4.1
200	20	5	875.84	129.68	515.822	18.52	25.16	20.12	9.21
200	40	2	357.6	73.92	230.474	13.84	28.1	18.92	4.22
200	40	5	887.16	158.76	557.537	25.2	28	18.92	9.75
200	100	2	360.16	85.52	236.394	15.64	30.9	17.84	4.35
200	100	5	895.88	191.56	575.681	31.6	30.88	17.84	10.2
300	20	2	536.56	91.8	326.727	13	37.7	29.04	4.11
300	20	5	1327.04	193	782.167	19.32	37.36	29.04	9.35
300	40	2	539.64	101.8	336.763	16.08	40.5	27.96	4.2
300	40	5	1340.48	225.32	820.825	28.4	40.36	27.96	9.83
300	100	2	542.04	113.52	341.814	17.32	43	26.8	4.3
300	100	5	1350.72	261.84	842.106	36	42.88	26.8	10.3

Πίνακας 5.2 *DifRec* σταθμισμένα αποτελέσματα σε 25 *Small-World* δίκτυα ($|V|=\{100,200,300\}$, $|I|=\{20,40,100\}$, $K=\{2,5\}$)

$ U $	$ I $	K	$ R_c $	$ R_i $	Sc	$ I_r $	$ V_c $	$ V_0 $	$ IV_c $
100	20	2	108.16	394.96	212.74	17.28	62	16.92	6.05
100	20	5	216.56	624.16	352.02	19.88	55.48	16.92	10.6
100	40	2	123.88	525.84	282.79	25.12	68.9	13.52	7.36
100	40	5	266.6	929.52	510.99	36.48	65.76	13.52	13.7
100	100	2	140.96	716.64	372.11	35.2	75.6	10.28	8.93
100	100	5	320.8	1413.52	742.77	61.72	74.12	10.28	18.2
200	20	2	252.72	886.4	479.98	19.32	147	14.68	5.94
200	20	5	515.88	1429.52	818.19	20	135.4	14.68	10.5
200	40	2	287.92	1174.96	633.70	33.12	158	10.36	7.31
200	40	5	625.80	2108.44	1167.6	39.56	152.48	10.36	13.7
200	100	2	321.56	1543.28	814.17	51.60	168	6.44	9
200	100	5	736.12	3068.8	1630.84	80.48	164.64	6.44	18.3
300	20	2	388.36	1312.44	730.21	19.8	216	23.72	6.03
300	20	5	793.08	2106.16	1236.24	20	199.68	23.72	10.6
300	40	2	439.28	1729.8	948.74	36.36	235	16.56	7.31
300	40	5	954.36	3108.12	1752.29	39.84	225.04	16.56	13.8
300	100	2	491.28	2292.84	1214.36	61.76	250	10.72	9.01
300	100	5	1125.76	4558.88	2438.05	88.88	245.60	10.72	18.3

Πίνακας 5.3 *CoveR* σταθμισμένα αποτελέσματα σε 25 *Scale-Free* δίκτυα ($|V|=\{100,200,300\}$, $|I|=\{20,40,100\}$, $K=\{2,5\}$)

$ U $	$ I $	K	$ R_c $	$ R_i $	Sc	$ I_r $	$ V_c $	$ IV_c $
100	20	2	32.16	645.08	189.95	14.68	58.4	4.27
100	20	5	81.04	1614.96	387.75	19.08	69.4	8.07
100	40	2	31.64	651.6	205.33	18.84	61.56	4.34
100	40	5	79.4	1605.88	450.50	31.12	65.64	9.30
100	100	2	31.76	653.08	246.35	23.96	64.4	5.14
100	100	5	79	1629.32	547.00	47.56	69.04	11.13
200	20	2	65.6	2014.88	403.58	18.48	139	4.31
200	20	5	164.6	5033.4	821.24	20	151.92	8.37
200	40	2	63.92	2049.8	464.45	29.12	151.84	4.65
200	40	5	160.04	5107.56	967.85	37.92	154.52	9.62
200	100	2	63.36	2016.32	533.45	40.76	150.2	5.44
200	100	5	159.64	4993.0	1114.24	72	156.16	10.99
300	20	2	94.96	4127.6	657.26	19.52	210.52	4.74
300	20	5	238.8	10335.84	1275.59	20	230.64	8.74
300	40	2	97.24	4010.88	753.85	34.52	216.72	5.30
300	40	5	242.72	10056.04	1563.73	39.68	231.32	10.64
300	100	2	95.76	4184.12	933.11	56.6	227.6	6.57
300	100	5	240.52	10509.2	1929.91	85.32	232.48	13.31

Πίνακας 5.4 *CoveR* σταθμισμένα αποτελέσματα σε 25 *Small-World* δίκτυα ($|V|=\{100,200,300\}$, $|I|=\{20,40,100\}$, $K=\{2,5\}$)

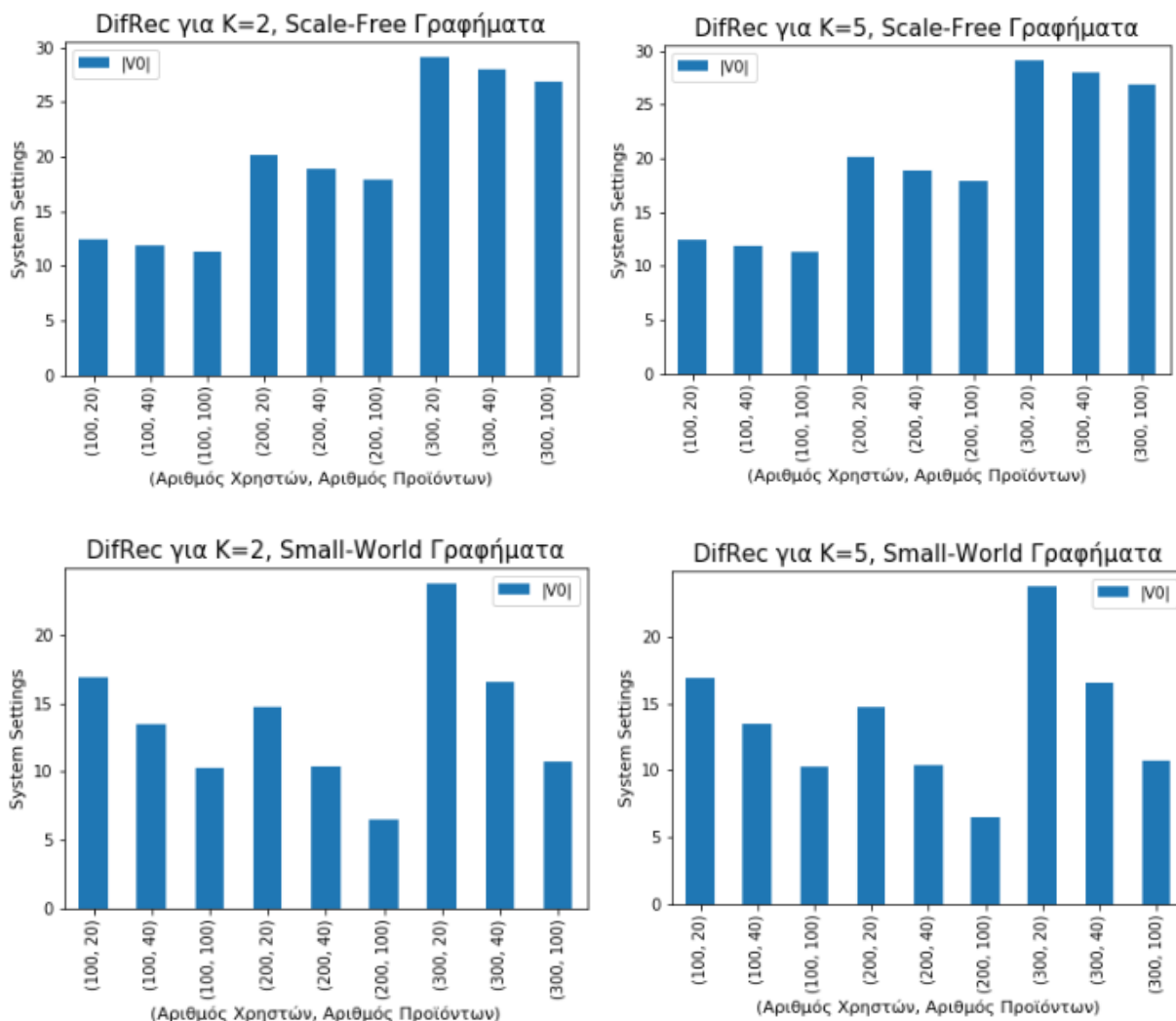
$ U $	$ I $	K	$ R_c $	$ R_i $	S_c	$ I_r $	$ V_c $	$ IV_c $
100	20	2	22.2	519.72	203.63	11.96	66.96	4.38
100	20	5	55.24	1286.92	422.49	17.92	71.64	8.95
100	40	2	21.76	500.16	243.68	14.72	67.84	5.07
100	40	5	54.36	1240.12	510.64	25.68	72.04	10.42
100	100	2	21.12	492	267.60	17.8	67.52	5.45
100	100	5	53.32	1246.68	598.52	32.64	74.24	11.73
200	20	2	68.48	874.24	422.67	18.72	113.24	4.65
200	20	5	172.88	2173.16	860.87	20	124.76	9.07
200	40	2	67.08	859.92	462.04	30.32	113.60	4.87
200	40	5	169.12	2154.4	1011.58	38.52	125.8	10.25
200	100	2	67.72	864.04	520.54	45.56	118	5.28
200	100	5	169.04	2171.44	1181.20	76.28	126.44	11.70
300	20	2	104.96	1290.76	641.68	19.68	170.96	4.66
300	20	5	262.08	3240.6	1325.19	20	185.08	9.31
300	40	2	102	1268.92	706.40	35.36	168.76	5.01
300	40	5	255.24	3199.76	1549.02	39.76	186.44	10.58
300	100	2	103.64	1288.36	784.50	63.12	171.76	5.41
300	100	5	260.96	3245.76	1809.66	91.32	185.2	12.18

Με την βοήθεια των πινάκων αυτών και με χρήση διαγραμμάτων στις παρακάτω υποενότητες, θα γίνει σύγκριση των δύο μεθόδων ώστε να καταλήξουμε σε ένα ασφαλές συμπέρασμα σχετικά με την αποτελεσματικότητά τους.

5.3.1 Αποτελέσματα για τον DifRec σε Scale-Free και Small-World Δίκτυα

Ξεκινώντας από το πλαίσιο συστάσεων DifRec, θέλουμε αρχικά να εντοπίσουμε και να επιβεβαιώσουμε τα βασικά χαρακτηριστικά του που εξηγήσαμε στην ενότητα 5.1. αρχικά, φαίνεται ότι σε όλα τα γραφήματα υπάρχει ένας μικρός αριθμός χρηστών οι οποίοι δεν έλαβαν καμία σύσταση προϊόντων (μετρική $|V_0|$), ούτε άμεσα ούτε έμμεσα. Αυτό είναι ένα βασικό στοιχείο του DifRec και ένα σημαντικό αρνητικό σημείο του, καθώς δεν καλύπτει με συστάσεις όλο το δίκτυο των χρηστών. Παρακάτω, φαίνονται τα διαγράμματα για αυτή τη μετρική όσον αφορά τις προσομοιώσεις που έγιναν σε S-F και S-W δίκτυα, με τους συνδυασμούς πλήθους κόμβων και προϊόντων που αναφέραμε παραπάνω και φαίνονται στον οριζόντιο άξονα, ενώ στον κατακόρυφο παρουσιάζονται οι παράμετροι που

εξετάζονται. Παρατηρούμε από αυτά τα διαγράμματα ότι ο αριθμός των χρηστών που δε θα λάβουν καμία σύσταση προϊόντος, δεν εξαρτάται από τον αριθμό K , αλλά επηρεάζεται από τον αριθμό των κόμβων του δικτύου και τα προϊόντα που εισάγονται στο ΣΣ. Συγκρίνοντας τα S-F και S-W, οι τιμές της μετρικής είναι στην ίδια τάξη μεγέθους και το πλήθος των χρηστών με αυτήν την ιδιότητα είναι προσεγγιστικά το 10% του συνόλου των χρηστών.



Εικόνα 5.3 Χρήστες που δεν έλαβαν καμία σύσταση προϊόντος – μετρική $|V_0|$, με εφαρμογή του DifRec.

Επιπρόσθετα, παρατηρούμε ότι έχει γίνει παραβίαση του πλήθους των διαφορετικών προϊόντων που μπορεί να δεχθεί ένας χρήστης (capacity violation - μετρική $|V_c|$), όπως

αναμέναμε να συμβαίνει στην εφαρμογή του DifRec (Εικόνα 5.4). Δεν πρέπει να ξεχνάμε ότι ο αλγόριθμος αυτός δεν επιτρέπει να γίνει άμεση και έμμεση σύσταση κάποιου προϊόντος στον ίδιο χρήστη. Έτσι, η παραβίαση, από τη μία πλευρά, έχει γίνει λόγω των έμμεσων συστάσεων που προτείνονται στον ίδιο χρήστη για περισσότερες από μία φορές (duplicate items), το οποίο είναι επίσης συναντάται στο πλαίσιο συστάσεων DifRec. Από την άλλη πλευρά, μπορεί να συμβαίνει όταν ένας χρήστης λαμβάνει K συστάσεις προϊόντων μέσω του ΣΣ, όσα είναι δηλαδή και τα διαφορετικά προϊόντα που μπορεί να λάβει ο χρήστης (υπενθυμίζουμε ότι ο DifRec ορίζει ως K τον μέγιστο αριθμό συστάσεων που μπορεί να λάβει κάθε χρήστης από το ΣΣ). Άρα, αν δεχθεί και έμμεσες συστάσεις από αναδημοσίευση «γειτονικού» χρήστη, αθροιστικά θα ξεπεραστεί το «όριο» των K προϊόντων. Η παράμετρος K είναι ένας τυπικός αριθμός που ορίζεται από το εκάστοτε σύστημα ανάλογα με τις συνθήκες που επικρατούν. Η χρησιμότητά του έγκειται στο γεγονός ότι από τα ΣΣ επιδιώκεται ο χρήστης να ικανοποιείται από τις συστάσεις που δέχεται, και όχι να του προκαλείται σύγχυση, καθώς τότε τα αποτελέσματα πιθανόν να απέχουν από τα επιθυμητά. Για την μετρική $|V_c|$ μπορούμε επίσης να παρατηρήσουμε ότι δεν μεταβάλλεται με την αύξηση του K , αλλά αυξάνεται ελαφρώς με την αύξηση των προϊόντων, με μεγαλύτερη αύξηση να υφίσταται κατά την αύξηση του πλήθους των κόμβων. Σημαντικό είναι το γεγονός ότι συγκρίνοντας τα S-F και S-W, το capacity violation είναι αξιοσημείωτα μεγαλύτερο στα S-W (σε ποσοστό 50–80% προσεγγιστικά), σε αντίθεση με τα S-F (12–15% προσεγγιστικά). Η απόκλιση αυτή οφείλεται στα χαρακτηριστικά των δύο τοπολογιών όσον αφορά τη συνδεσμολογία τους και την πυκνότητα.

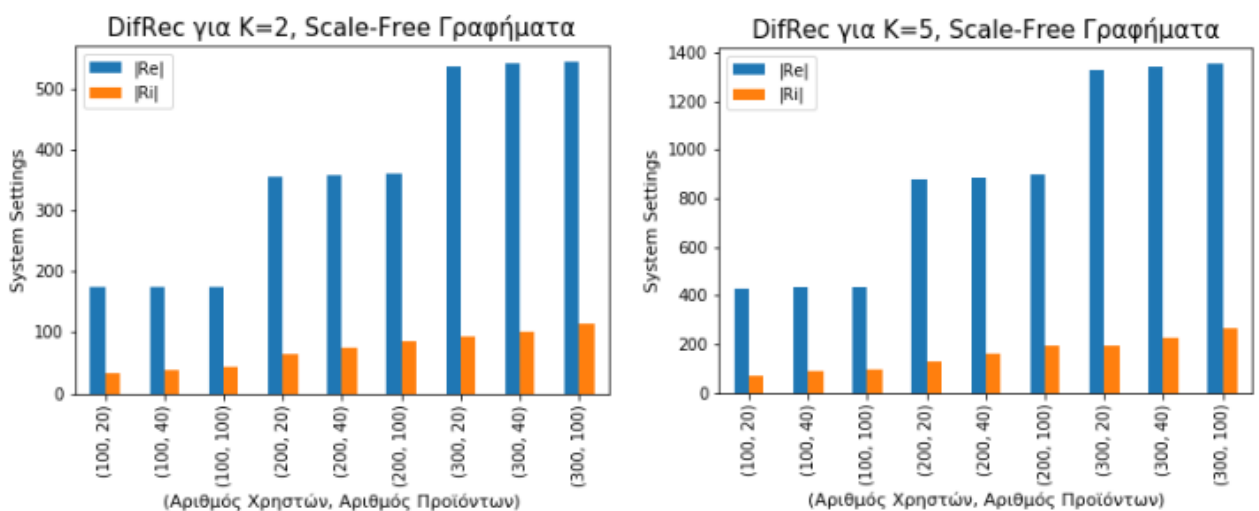


Εικόνα 5.4 Χρήστες στους οποίους παρατηρήθηκε capacity violation – μετρική $|V_c|$, με εφαρμογή του DifRec.

Στα διαγράμματα των Εικόνων 5.5 και 5.6 φαίνονται οι μέσες τιμές των άμεσων και έμμεσων συστάσεων για S-F και S-W δίκτυα αντίστοιχα. Το σημαντικό που παρατηρούμε στα διαγράμματα αυτά είναι η διαφορά τιμής με την οποία υπερτερούν οι άμεσες συστάσεις έναντι των έμμεσων συστάσεων για τα S-F, ενώ αντίθετο αποτέλεσμα έχουμε για τα S-W. Οι άμεσες συστάσεις είναι ίδιες κλίμακας για τα δύο είδη δικτύων και μάλιστα ανταποκρίνονται στον περιορισμό του πλήθους των άμεσων συστάσεων ανά χρήστη, αφού αναλογικά με το πλήθος των κόμβων και την παράμετρο K , δεν υπερβαίνεται ο μέγιστος

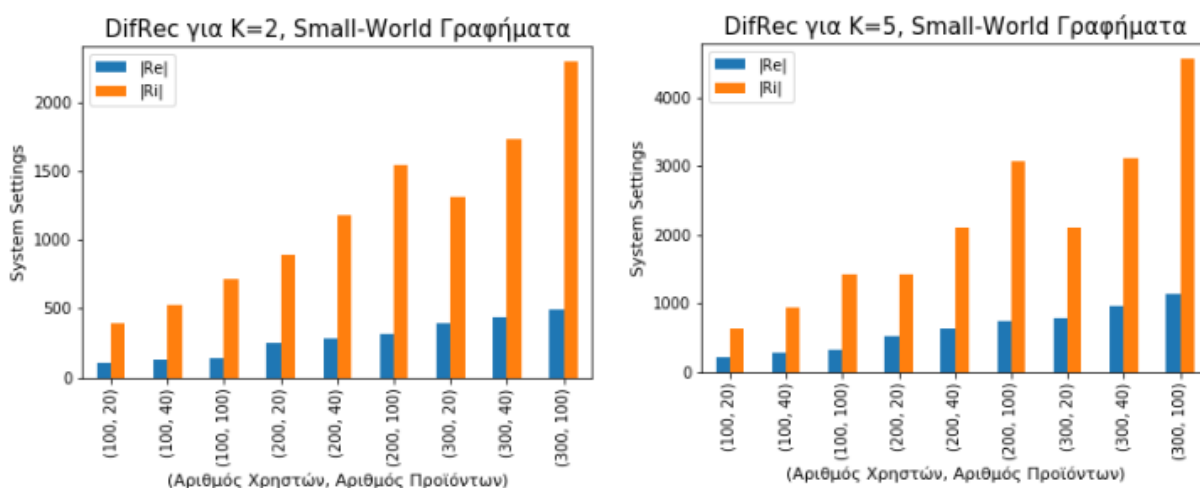
αριθμός άμεσων προτάσεων που επιτρέπεται. Για παράδειγμα, για $|U| = 200$ και $K = 5$, στο S-F και το S-W δίκτυο έχουμε αριθμό άμεσων προτάσεων μικρότερο των 1000. Όσον αφορά τις έμμεσες συστάσεις, για κάθε άμεση σύσταση προϊόντος i που γίνεται σε έναν χρήστη u , γίνεται έμμεση σύσταση του ίδιου προϊόντος στους γειτονικούς κόμβους του που προκύπτουν από τον γράφο γειτνίασης. Ωστόσο, εφόσον ο αλγόριθμος αυτός έχει τον περιορισμό ότι δεν επιτρέπεται άμεση και έμμεση σύσταση ενός συγκεκριμένου προϊόντος στον ίδιο χρήστη, μπορούμε να συμπεράνουμε τα ακόλουθα.

- Για τα S-F δίκτυα, έχουμε αναφέρει σε προηγούμενη ενότητα ότι χαρακτηρίζονται από κεντρικά σημεία «hubs» που έχουν πολύ υψηλή συνδεσιμότητα. Υπάρχουν λίγοι τέτοιοι κόμβοι, ενώ η πλειοψηφία των κόμβων έχει μικρό βαθμό. Έτσι, μπορούμε να υποθέσουμε ότι αυτό έχει σαν αποτέλεσμα να μην είναι μεγάλος ο τελικός αριθμός των αναδημοσιεύσεων, δηλαδή των έμμεσων συστάσεων, το οποίο προκύπτει βεβαίως σε συνδυασμό με τον περιορισμό του DifRec για μη σύσταση του ίδιου προϊόντος σε ένα χρήστη άμεσα και έμμεσα. Οι αναδημοσιεύσεις κατά κύριο λόγο γίνονται από τα hubs που είναι η μειοψηφία κόμβων μέσα στο δίκτυο.



Εικόνα 5.5 Μέσος όρος άμεσων και έμμεσων συστάσεων για Scale-Free Δίκτυα, με εφαρμογή του DifRec.

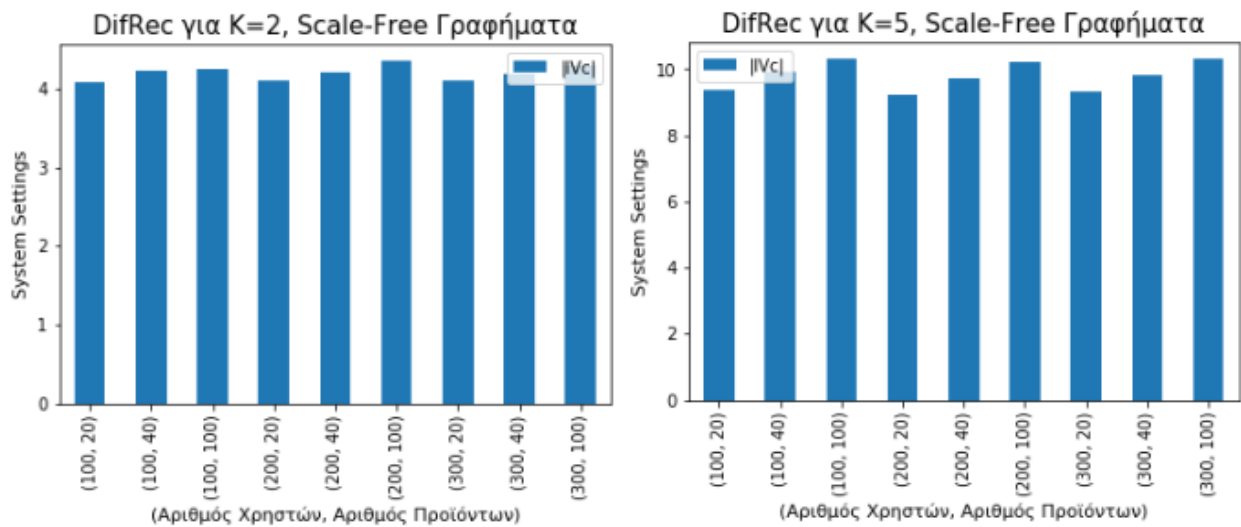
- Για τα S-W δίκτυα, παρατηρούμε ένα αντίθετο αποτέλεσμα, με τις έμμεσες συστάσεις δηλαδή να είναι αρκετά περισσότερες από τις άμεσες συστάσεις. Η διαφορά αυτή οφείλεται στη δομή των δικτύων αυτών, καθώς υπάρχει αρκετά μεγάλη συνδεσιμότητα μεταξύ των κόμβων του δικτύου, είναι, συνήθως, δίκτυα με μεγαλύτερη πυκνότητα. Κατά συνέπεια, η διάχυση της πληροφορίας μέσω αναδημοσιεύσεων των χρηστών υφίσταται σε σημαντικά πιο έντονο βαθμό από την περίπτωση των S-F δικτύων. Με τον τρόπο αυτό εξηγείται, επίσης, ο λόγος για τον οποίο παρατηρούμε capacity violation για περισσότερους χρήστες. Έχουμε έναν πολύ μεγάλο αριθμό από έμμεσες συστάσεις, που συνεπάγεται ότι αρκετά προϊόντα έχουν προταθεί στον ίδιο χρήστη πολλές φορές και χρήστες που έχουν λάβει περισσότερες από K συστάσεις.



Εικόνα 5.6 Μέσος όρος άμεσων και έμμεσων συστάσεων για Small-World Δίκτυα, με εφαρμογή του DifRec.

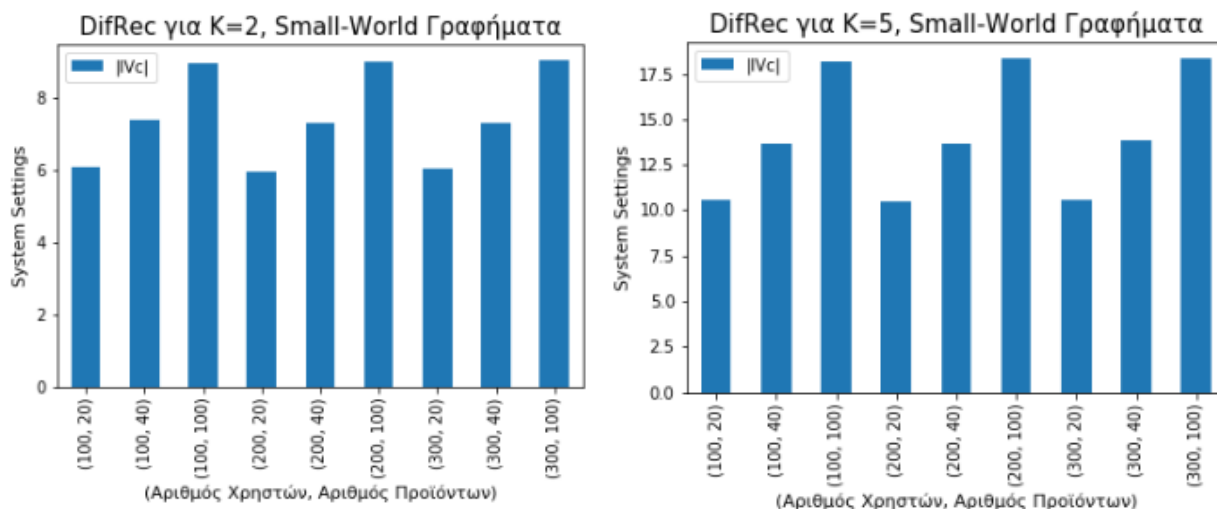
Εξετάζοντας τον αριθμό των προϊόντων που δέχθηκε κάθε χρήστης, για τον οποίο σημειώθηκε capacity violation, δηλαδή την μετρική $|IVc|$, μπορούμε να δούμε, για την περίπτωση των S-F, ότι είναι σχεδόν ανεξάρτητη από τον αριθμό των χρηστών του δικτύου και από το πλήθος των προϊόντων. Αυξάνεται, όμως, με την αύξηση του K . Για παράδειγμα, παρατηρώντας τα παρακάτω διαγράμματα, για $K = 2$, οι χρήστες έχουν δεχθεί

προσεγγιστικά τέσσερα προϊόντα αντί για δύο, ενώ για $K = 5$, δέκα προϊόντα έχουν προταθεί αντί για πέντε.



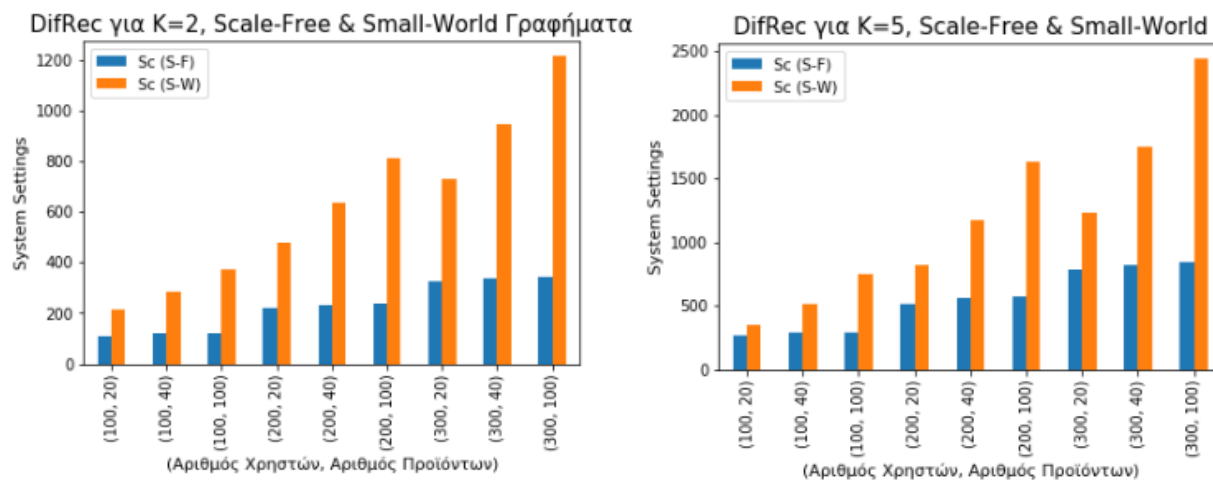
Εικόνα 5.7 Αριθμός των προϊόντων που δέχθηκε ένας χρήστης με capacity violation – μετρική $|V_c|$, για Scale-Free δίκτυα, με εφαρμογή του DifRec.

Για τα S-W δίκτυα, η μετρική επηρεάζεται και από την αύξηση των προϊόντων. Επίσης, στα διαγράμματα είναι πιο εμφανές το capacity violation. Για παράδειγμα, για $K = 5$ ένας χρήστης μπορεί να φτάσει στην ακραία περίπτωση να λάβει 18 προϊόντα, ενώ έχει capacity 5 προϊόντα.



Εικόνα 5.8 Αριθμός των προϊόντων που δέχθηκε ένας χρήστης με capacity violation – μετρική $|IV_c|$, για Small-World δίκτυα, με εφαρμογή του DifRec.

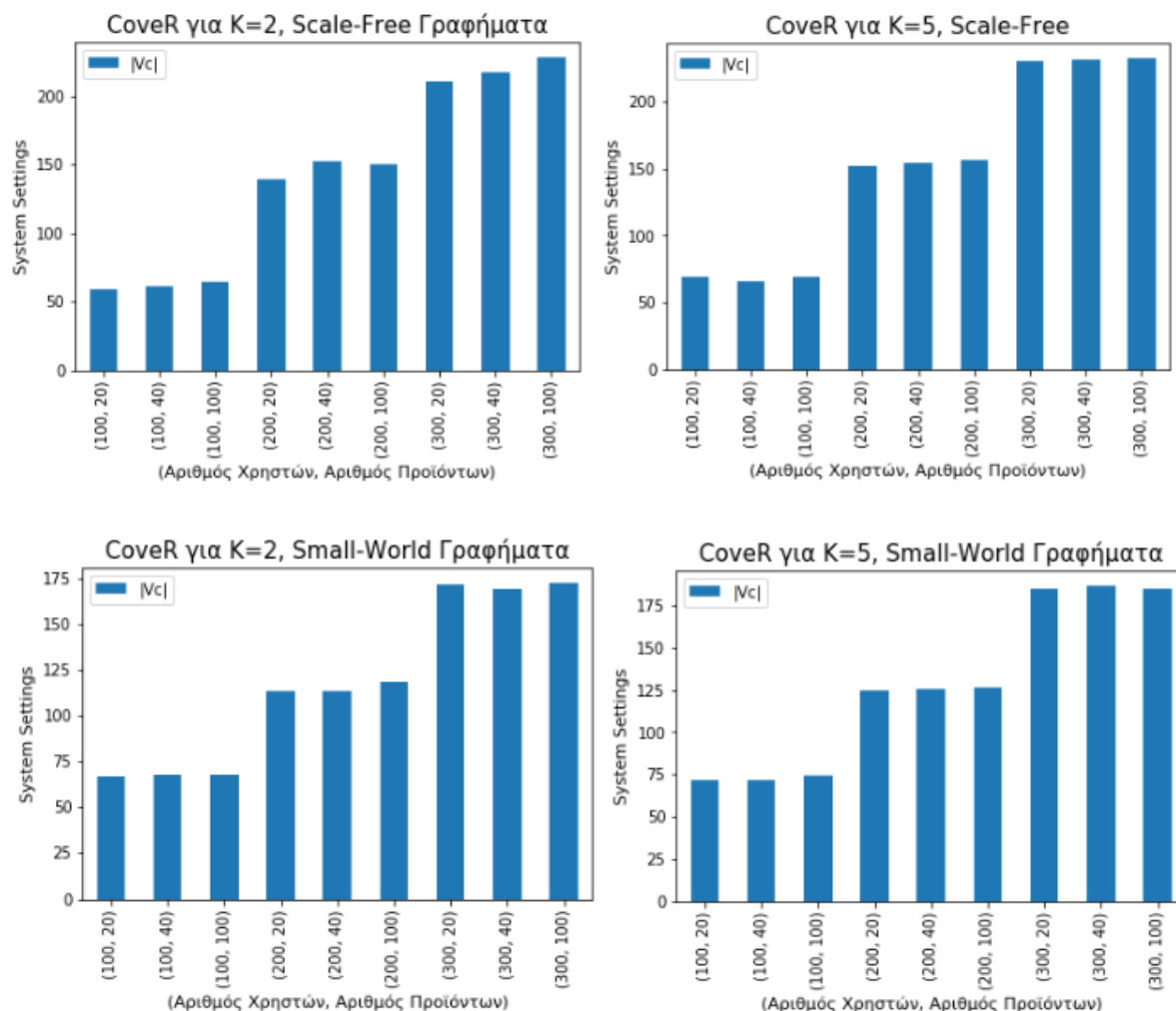
Συνοψίζοντας, βάσει των προσομοιώσεων σε S-F και S-W δίκτυα και τα διαγράμματα που προέκυψαν, ήταν εφικτό να εντοπίσουμε βασικά χαρακτηριστικά του DifRec. Πιο συγκεκριμένα, εντοπίστηκαν χρήστες που δεν έχουν λάβει σύσταση προϊόντων ούτε άμεσα ούτε έμμεσα, αλλά και χρήστες που έχουν λάβει έχουν λάβει περισσότερα διαφορετικά προϊόντα από όσα μπορούν να δεχθούν. Επιπροσθέτως, θα μπορούσαμε να συμπεράνουμε ότι ο αλγόριθμος αυτός έχει καλύτερα αποτελέσματα σε δίκτυα που έχουν τη μορφή Scale-Free δικτύων, αν ληφθεί υπόψη το capacity violation στις δύο τοπολογίες. Συγκρίνοντας, ωστόσο, το μέσο Relevance Score, μπορούμε να πούμε ότι τα Small-World δίκτυα έχουν καλύτερη απόδοση (Εικόνα 5.9).



Εικόνα 5.9 Μέσο Relevance Score για Scale-Free και Small-World στο πλαίσιο συστάσεων DifRec.

5.3.2 Αποτελέσματα για τον CoveR σε Scale-Free και Small-World Δίκτυα

Εφαρμόζοντας τον αλγόριθμο CoveR σε S-F και S-W δίκτυα, το πρώτο που αξίζει να παρατηρήσουμε είναι ότι δεν υπάρχει κάποιος χρήστης στον οποίο δεν έγινε κάποια σύσταση προϊόντος (μηδενική μετρική $|V_0|$). Το αποτέλεσμα αυτό ήταν αναμενόμενο καθώς είναι ένα από τα κύρια χαρακτηριστικά του αλγορίθμου. Υπάρχει πλήρης κάλυψη του δικτύου, δηλαδή κάθε χρήστης λαμβάνει προτάσεις προϊόντων. Μάλιστα, λαμβάνει τουλάχιστον K προϊόντα είτε άμεσα είτε έμμεσα, αφού η παράμετρος K στο CoveR ορίζεται ως ο ελάχιστος αριθμός προϊόντων που πρέπει να λάβει κάθε χρήστης (K -κάλυψη δικτύου). Επιπρόσθετα, σε αυτήν την περίπτωση δεν ισχύει ο περιορισμός ότι ένα προϊόν δεν μπορεί να προταθεί άμεσα και έμμεσα σε κάποιον χρήστη (περιορισμός του DifRec). Για τους παραπάνω λόγους και όπως μπορεί να παρατηρήσει κανείς στα παρακάτω διαγράμματα, το capacity violation είναι αρκετά υψηλό και για τα δύο είδη δικτύων (55–75% προσεγγιστικά). Παρατηρούμε επίσης ότι η μετρική $|Vc|$ δεν επηρεάζεται από την παράμετρο K και από τον αριθμό των προϊόντων $|I|$, αλλά αυξάνεται με την αύξηση του πλήθους των χρηστών του δικτύου.



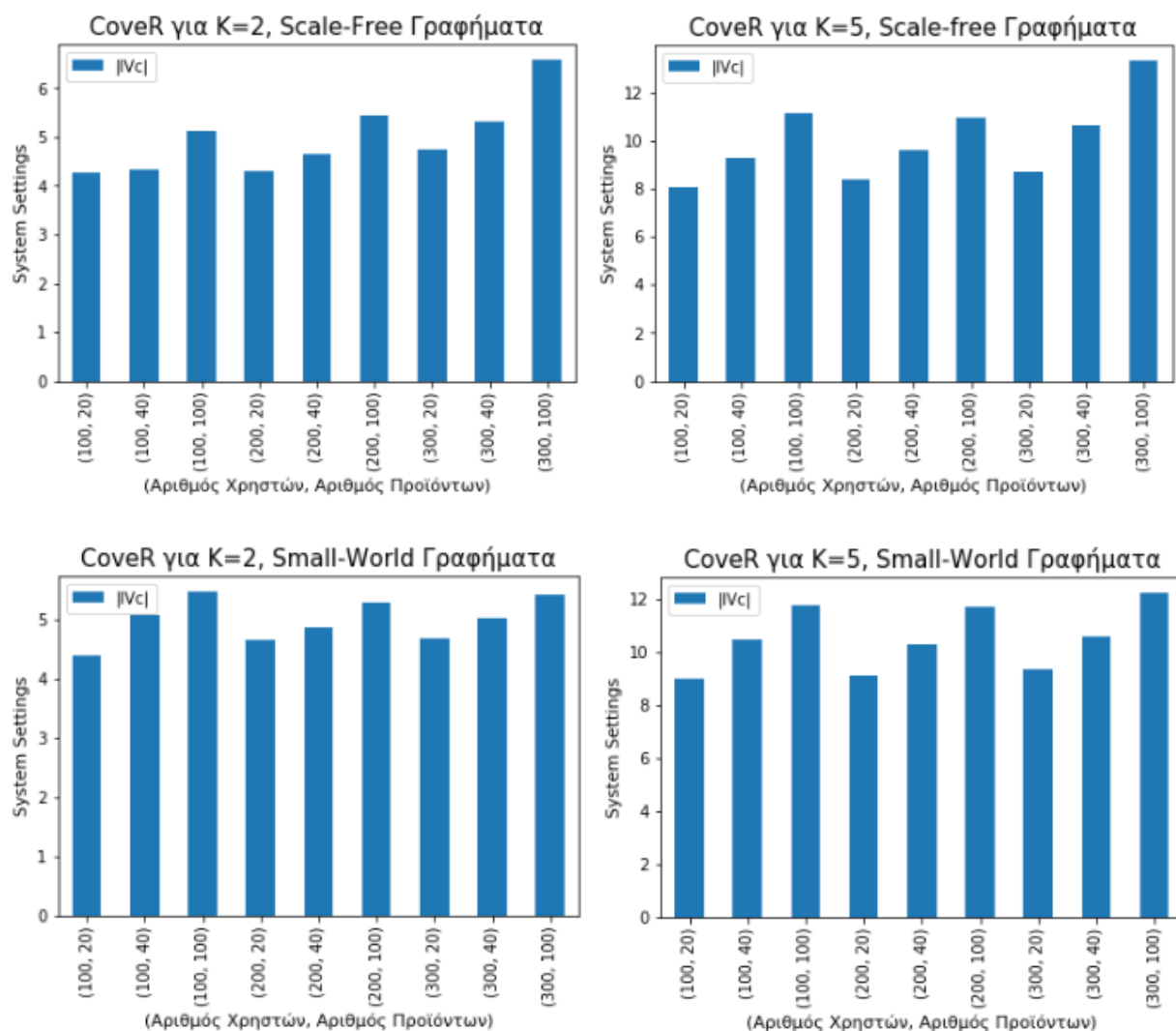
Εικόνα 5.10 Χρήστες στους οποίους παρατηρήθηκε capacity violation – μετρική $|V_c|$, με εφαρμογή του CoveR.

Συγκρίνοντας τις έμμεσες και άμεσες συστάσεις προϊόντων, είναι εμφανής η διαφορά μεταξύ τους και για τα δύο είδη δικτύων με τις πρώτες να υπερτερούν κατά μεγάλο βαθμό. Αυτό οφείλεται κατά κύριο λόγο στην άρση του περιορισμού που ισχύει στον DifRec. Και τα δύο είδη συστάσεων, παρόλα αυτά, αυξάνονται με την αύξηση της παραμέτρου K και το πλήθος των χρηστών, αλλά ελάχιστα επηρεάζονται από τον αριθμό των προϊόντων (Εικόνα 5.11).



Εικόνα 5.11 Μέσος όρος άμεσων και έμμεσων συστάσεων για Scale-Free Δίκτυα, με εφαρμογή του CoveR.

Όσον αφορά τον αριθμό των προϊόντων που δέχθηκε κάθε χρήστης, για τον οποίο σημειώθηκε capacity violation, δηλαδή την μετρική $|IVc|$, παρατηρούμε μια παρόμοια κατάσταση για τις δυο τοπολογίες, όπως φαίνεται στα παρακάτω διαγράμματα. Υπάρχει ελάχιστη ή μηδενική επιρροή από τον αριθμό των χρηστών στο μέσο κοινωνικής δικτύωσης, ενώ μεγαλύτερη επίδραση υφίσταται σε μεταβολή της μεταβλητής K και του πλήθους των προϊόντων.



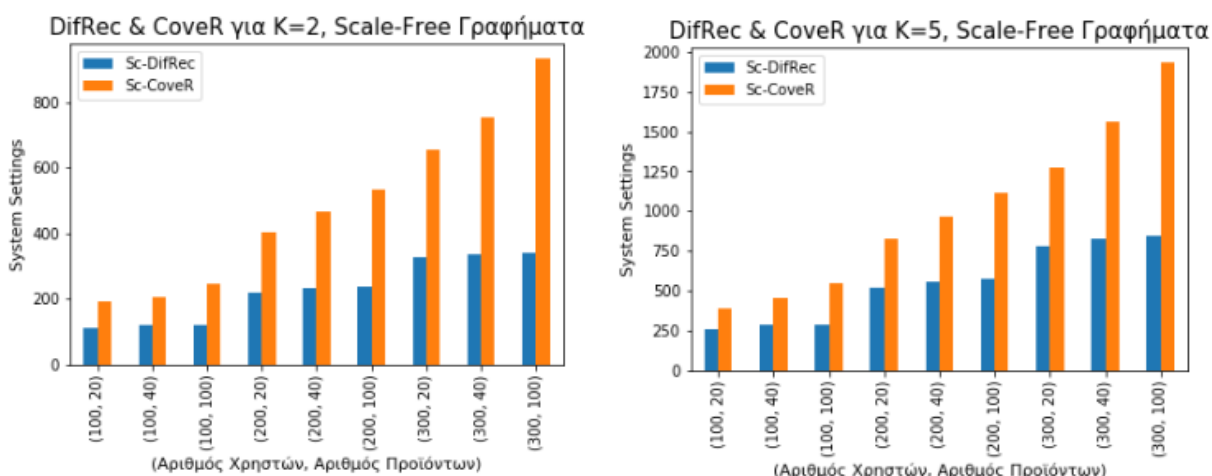
Εικόνα 5.12 Αριθμός των προϊόντων που δέχθηκε ένας χρήστης με capacity violation – μετρική $|V_c|$, με εφαρμογή του CoveR.

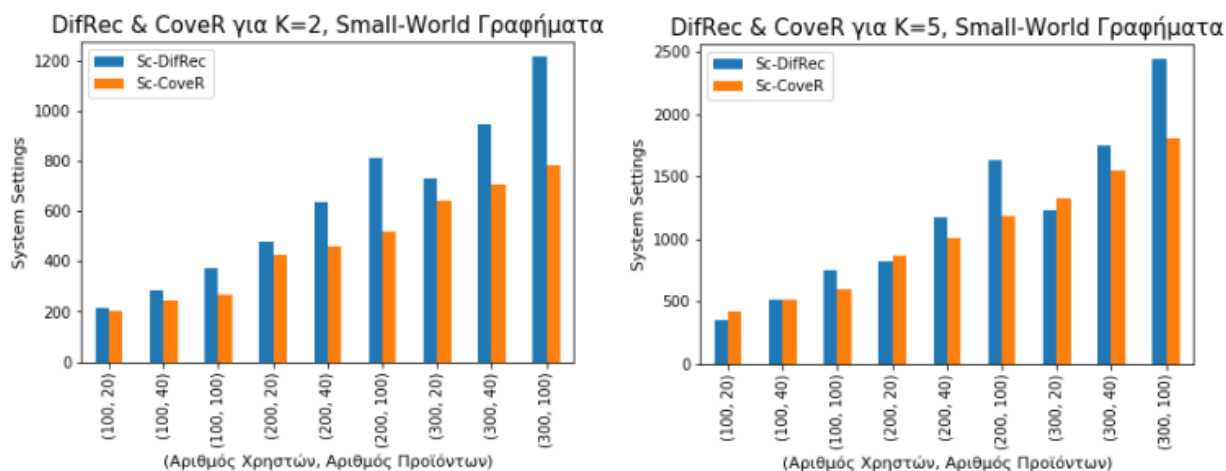
Συνοψίζοντας, με την εφαρμογή του αλγορίθμου CoveR, παρατηρούμε παρεμφερή συμπεριφορά στις δύο τοπολογίες αυτή τη φορά, γεγονός που μπορεί να εξηγηθεί από δύο παράγοντες. Αρχικά, ο χαμηλός αριθμός από άμεσες συστάσεις εξηγείται από το κριτήριο που χρησιμοποιείται για τη λήψη αποφάσεων στον CoveR που ελαχιστοποιεί τον αριθμό των χρηστών που θα λάβουν σύσταση, θέτοντας όμως ένα κατώτατο όριο προϊόντων που πρέπει τους προταθούν είτε άμεσα είτε έμμεσα. Επίσης, η άρση της απαγόρευσης που

ισχύει στον DifRec επηρεάζει το αποτέλεσμα εδώ δίνοντας περισσότερες έμμεσες συστάσεις.

5.3.3 Σύγκριση αποτελεσμάτων των αλγορίθμων DifRec και Cover

Βασικός στόχος στην παρούσα εργασία είναι, πέρα από την ανάλυση των πλαισίων συστάσεων DifRec και Cover, να εξαχθούν συμπεράσματα για το ποιο από τα δύο δίνει καλύτερα αποτελέσματα. Το πιο βασικό κριτήριο επιλογής είναι το συνολικό Relevance Score του δικτύου, δηλαδή των προϊόντων που «βλέπουν» οι χρήστες του δικτύου (μετρική $|Sc|$) είτε μέσω του ΣΣ είναι από αναδημοσιεύσεις των χρηστών που ακολουθεί. Στα παρακάτω διαγράμματα μπορεί να δει κανείς συγκεντρωτικά τα αποτελέσματα των προσομοιώσεων των δύο τοπολογιών για το Relevance Score που υπολογίστηκε. Παρατηρούμε μια διαφορετική εικόνα για S-F και S-W δίκτυα. Στα πρώτα είναι εμφανές ότι ο Cover έχει μεγαλύτερο Relevance Score, με τη διαφορά μεταξύ των δύο μεθόδων να αυξάνεται με την αύξηση του πλήθους των κόμβων. Το Sc για τον Cover επηρεάζεται από τη μεταβολή και των τριών παραμέτρων (αριθμός χρηστών, αριθμός προϊόντων, K), ενώ ο DifRec δεν έχει σημαντική επίδραση από την αύξηση του πλήθους των προϊόντων, αλλά από τον αριθμό των χρηστών και την παράμετρο K .

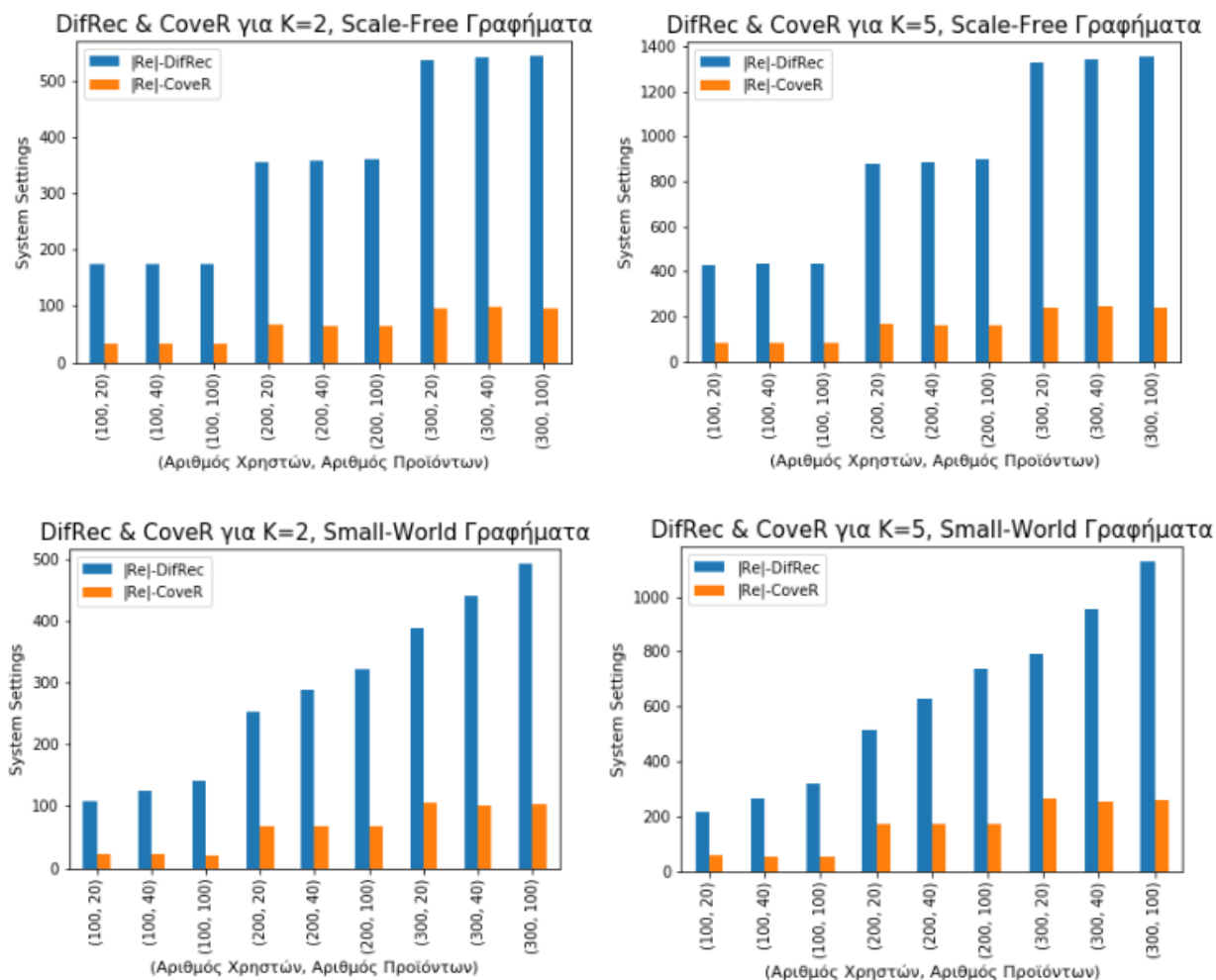




Εικόνα 5.13 Μέσο Relevance δικτύου – μετρική S_C , για DifRec και CoveR.

Αντιθέτως, στην περίπτωση των S-W δικτύων, υπάρχει ένα παρεμφερές αποτέλεσμα και για τις δύο τεχνικές, με τον DifRec να έχει μεγαλύτερο συνολικό Relevance Score σε κάποιες περιπτώσεις, αλλά η διαφορά δεν είναι σημαντική. Σε γενικές γραμμές, εδώ παρατηρούμε αύξηση του S_C με την αύξηση των αριθμών των χρηστών, αλλά και με την αύξηση των προϊόντων και της μεταβλητής K .

Σημαντική διαφορά παρατηρείται, ωστόσο, στην παράμετρο $|R_e|$ για τα δύο είδη δικτύου, δηλαδή στον μέσο όρο του πλήθους των συστάσεων που έχουν δεχθεί οι χρήστες από το ΣΣ, όπως φαίνεται στα παρακάτω διαγράμματα.



Εικόνα 5.14 Μέσος όρος άμεσων συστάσεων για DifRec και CoverR.

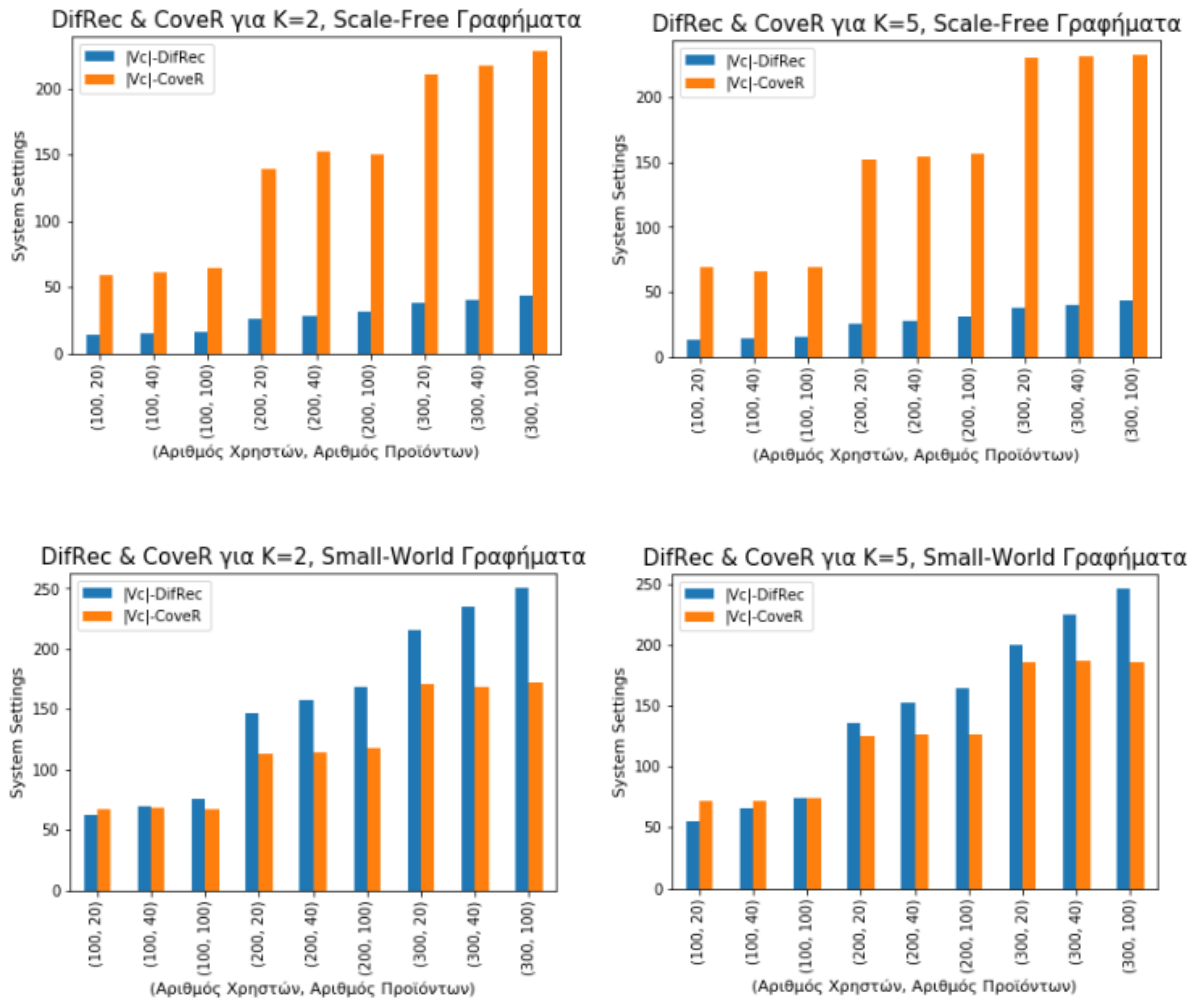
Η διαφορά αυτή είναι αναμενόμενη αν αναλογιστούμε ότι ο CoverR υλοποιείται έτσι ώστε όχι μόνο να αποδώσει μέγιστο συνολικό Relevance Score, αλλά και να μεγιστοποιηθούν οι χρήστες που καλύπτονται με τέτοιο τρόπο ώστε να μειώνονται οι άμεσες συστάσεις, σύμφωνα με το κριτήριο επιλογής που ορίστηκε στην αντίστοιχη ενότητα, όπου γίνονται οι επιλογές συστάσεων προκειμένου να το ελαχιστοποιούν. Τέλος, όπως έχουμε ήδη αναφέρει, σημαντική διαφορά μεταξύ του CoverR και του DifRec είναι ότι στον πρώτο δεν υπάρχουν χρήστες που δεν έλαβαν σύσταση με κανέναν από τους δύο τρόπους.

Συμπερασματικά, είδαμε ότι ως προς το συνολικό Relevance Score, δεν υπάρχουν σημαντικές διαφορές μεταξύ των δύο μεθόδων, με τον CoveR να υπερτερεί στην περίπτωση των S-F δικτύων. Έτσι, για να φτάσουμε σε ένα τελικό συμπέρασμα, θα πρέπει να λάβουμε και άλλες παραμέτρους υπόψη. Για παράδειγμα, είναι λογικό στην περίπτωση που η τιμή του Sc είναι ίδια για τις δύο τεχνικές, να επιλέξουμε εκείνη που δίνει ίδια απόδοση αλλά με συστάσεις σε αρκετά λιγότερους χρήστες, δηλαδή τον CoveR σε αυτήν την περίπτωση, ο οποίος, παράλληλα, καλύπτει πλήρως το δίκτυο με τις άμεσες συστάσεις του. Ωστόσο, δεν πρέπει να αγνοούνται τα αδύναμα σημεία του CoveR, όπως το σημαντικά μεγάλο πλήθος έμμεσων συστάσεων που παρουσιάζει και ο μεγάλος αριθμός χρηστών που έχουν capacity violation και για τις δύο τοπολογίες.

Αν συγκρίνουμε ξεχωριστά τις δύο μεθόδους ανά περίπτωση τοπολογίας μπορούμε να συμπεράνουμε τα εξής:

- Σε S-F δίκτυα, τα υπέρ του DifRec είναι ότι παρουσιάζει σημαντικά μικρότερο αριθμό χρηστών με capacity violation με τη διαφορά τους να αυξάνεται όσο αυξάνεται ο αριθμός των χρηστών του δικτύου, όπως φαίνεται στα παρακάτω διαγράμματα, και αρκετά μικρότερο αριθμό έμμεσων συστάσεων. Ωστόσο, ο CoveR έχει μεγαλύτερο συνολικό Relevance Score, λιγότερες άμεσες συστάσεις και πλήρη κάλυψη του δικτύου. Συνεπώς, είναι αρκετά σύνθετο στην περίπτωση αυτή να αποφανθεί κανείς. Θα μπορούσαμε να πούμε ότι εξαρτάται από την εφαρμογή του εκάστοτε δικτύου και σε ποιες παραμέτρους ο ερευνητής θέλει να αποδώσει βαρύτητα.
- Σε S-W δίκτυα, φαίνεται ότι το συμπέρασμα μπορεί να εξαχθεί πιο εύκολα, καθώς το capacity violation είναι παρόμοιο και στα δύο δίκτυα, με τη διαφορά τους να αυξάνεται ελάχιστα με την αύξηση του πλήθους των κόμβων (με τον DifRec να αποκτά περισσότερους χρήστες με capacity violation), ενώ και οι έμμεσες συστάσεις είναι αυξημένες σημαντικά και για τον DifRec, όπως συμβαίνει έτσι κι αλλιώς για τον CoveR. Επίσης, ο τελευταίος έχει μικρότερο μέσο όρο άμεσων συστάσεων. Αλλά, όσον αφορά το Sc , είδαμε ότι ο DifRec υπερिशχύει στις περισσότερες

περιπτώσεις, όμως με πολύ μικρή διαφορά. Συμπερασματικά, θα μπορούσαμε να πούμε ότι ο CoveR δίνει καλύτερα αποτελέσματα για αυτήν την τοπολογία.



Εικόνα 5.15 Χρήστες με capacity violation σε DifRec και CoveR.

6

Σύνοψη & Συμπεράσματα

Τις τελευταίες δεκαετίες, τα Συστήματα Συστάσεων χρησιμοποιούνται ευρέως στις διαδικτυακές πλατφόρμες, καθώς συμβάλλουν σημαντικά στη μετρίαση της υπερφόρτωσης πληροφορίας των χρηστών. Διευκολύνεται η αναζήτηση προϊόντων του χρήστη προτείνοντας του προϊόντα τα οποία, σύμφωνα με διάφορα κριτήρια που εφαρμόζονται, είναι πολύ πιθανό να ανήκουν στο πεδίο ενδιαφέροντός του. Δύο βασικά αποτελέσματα θα μπορούσαμε να διακρίνουμε: α) θετική εμπειρία του χρήστη για τον ιστότοπο και β) αύξηση των εσόδων για τον έμπορο. Συνεπώς, από τη μία πλευρά, επιτυγχάνεται η ικανοποίηση του καταναλωτή και αυξάνεται η πιθανότητα να επιστρέψει στο μέλλον και, από την άλλη πλευρά, το ηλεκτρονικό κατάστημα αυξάνει τις πωλήσεις του και, ως εκ τούτου, τα κέρδη του, καθώς, επίσης, αποκτά «πιστούς» πελάτες. Η έννοια του προϊόντος αφορά από αντικείμενα προς πώληση (Amazon), έως και προσφορά υπηρεσιών και ψυχαγωγίας, όπως οι ταινίες στο Netflix, στο οποίο τα έσοδα προκύπτουν από μηνιαία συνδρομή του χρήστη, οπότε και είναι επιτακτική η ανάγκη να παραμένει αμείωτο το ενδιαφέρον του κατά τη διάρκεια που είναι συνδεδεμένος στην πλατφόρμα.

Υπάρχουν διάφορες τεχνικές που μπορούν να χρησιμοποιηθούν για να υλοποιηθούν τα ΣΣ, για αυτό και έχει γίνει διαχωρισμός σε τρεις βασικές κατηγορίες, οι οποίες αναλύονται στην παρούσα εργασία. Η πρώτη κατηγορία αφορά τα ΣΣ **βάσει περιεχομένου**, όπου ο μηχανισμός επικεντρώνεται στα ενδιαφέροντα ενός χρήστη μεμονωμένα και σε χαρακτηριστικά προϊόντων που έχει βαθμολογήσει ήδη σε παρελθοντικές αγορές του, προκειμένου να εντοπιστούν εκείνα τα προϊόντα που θα πλησιάζουν τις προτιμήσεις του. Ακολουθούν τα **συνεργατικά** ΣΣ, τα οποία επιδιώκουν να εντοπίσουν ομοιότητες μεταξύ

χρηστών σε ένα δίκτυο, ή μεταξύ των διαθέσιμων προϊόντων, ανάλογα με την προσέγγιση που επιλέγεται. Έτσι, για παράδειγμα, για την πρώτη περίπτωση, δημιουργούνται «κοινότητες» χρηστών με παρεμφερή ενδιαφέροντα και, κατά συνέπεια, είναι ασφαλές να προταθεί κάποιο προϊόν σε κάποιον χρήστη, αν το προϊόν αυτό έχει ήδη επιλεγθεί και βαθμολογηθεί θετικά από άλλον χρήστη μέσα στην ίδια κοινότητα. Και οι δύο προαναφερθείσες κατηγορίες φέρουν δυνατά και αδύναμα σημεία. όπως έχουν ήδη αναλύσει σε προηγούμενο κεφάλαιο. Η τρίτη κατηγορία, τα ΣΣ με **υβριδικό φιλτράρισμα**, δημιουργήθηκαν ώστε να συνδυάζουν περισσότερες από μία μεθόδους, περιορίζοντας έτσι τα μειονεκτήματα των προηγούμενων, όπως αυτά που προκύπτουν όταν υπάρχουν λίγα διαθέσιμα δεδομένα και ένας χρήστης ή προϊόν είναι νέα στο σύστημα.

Στη συνέχεια, παρουσιάστηκε η έννοια του δικτύου και βασικών μετρικών της Ανάλυσης Κοινωνικών Δικτύων, που εφαρμόζονται στα Κοινωνικά ΣΣ, τα οποία εκμεταλλεύονται τους δεσμούς και τις αλληλεπιδράσεις μεταξύ των χρηστών μέσα στο δίκτυο ώστε να προσφέρουν συμπληρωματικές πληροφορίες στα ΣΣ προς διευκόλυνση της εξαγωγής συστάσεων και μεγαλύτερη ακρίβεια και αξιοπιστία των αποτελεσμάτων. Ιδιαίτερη έμφαση δόθηκε, ύστερα, σε μία υποκατηγορία στον συστημάτων αυτών, τα Information Diffusion-Aware Συστήματα Συστάσεων. Τα τελευταία επικεντρώνονται στη διάχυση και ανταλλαγή πληροφοριών που υφίστανται μέσα σε ένα μέσο κοινωνικής δικτύωσης. Η διάχυση αυτή μπορεί να επιτευχθεί με διάφορα μοντέλα (επεξηγηματικά ή προγνωστικά) και η συμβολή της είναι σημαντική για την επίλυση σημαντικών ζητημάτων, όπως η υπερφόρτωση πληροφοριών και το “long-tail” πρόβλημα (υπερίσχυση κυρίως των δημοφιλών προϊόντων σε συστάσεις). Τα συστήματα αυτά έρχονται να υπερνικήσουν αυτά τα θέματα. Στην παρούσα εργασία εστίασαμε σε δύο IDARS, που εκμεταλλεύονται τη διάχυση της πληροφορίας προκειμένου να παράγουν συστάσεις σε ένα μέσο κοινωνικής δικτύωσης. Οι συνολικές συστάσεις που δέχονται οι χρήστες, μπορούν να διακριθούν σε *άμεσες*, δηλαδή εκείνες που προκύπτουν από το ΣΣ, και σε *έμμεσες*, από τις αναδημοσιεύσεις των χρηστών, δηλαδή με ανταλλαγή δεδομένων μεταξύ των χρηστών του δικτύου. Αρχικά, αναλύθηκε το πλαίσιο συστάσεων DifRec, στο οποίο ο καταμερισμός τους γίνεται έτσι ώστε ένα προϊόν να προτείνεται σε έναν χρήστη είτε έμμεσα είτε άμεσα, με περιορισμό του πλήθους των

άμεσων συστάσεων ανά χρήστη. Το δεύτερο σύστημα που αναλύθηκε είναι το CoveR, στο οποίο ο καταμερισμός γίνεται με τρόπο ώστε να ελαχιστοποιηθούν οι άμεσες συστάσεις εξασφαλίζοντας πως κάθε χρήστης έχει λάβει ένα συγκεκριμένο αριθμό συστάσεων είτε άμεσα είτε έμμεσα. Η αποδοτικότητα των παραπάνω πλαισίων αξιολογήθηκε στη συνέχεια μέσω προσομοιώσεων σε συνθετικά δίκτυα ώστε να αποφανθούμε ποιο από τα δύο είναι καλύτερη επιλογή για σύσταση προϊόντων.

Τα αποτελέσματα των προσομοιώσεων αρχικά συνέβαλαν στον εντοπισμό και τη διαπίστωση χαρακτηριστικών των δύο πλαισίων συστάσεων, DifRec και CoveR, που αναφέρθηκαν κατά την ανάλυση της μεθοδολογίας τους. Εφαρμόζοντας τη μεθοδολογία του DifRec, υπάρχουν χρήστες που δεν λαμβάνουν συστάσεις προϊόντων ούτε άμεσα ούτε έμμεσα, με αποτέλεσμα να μην υπάρχει κάλυψη του δικτύου. Επίσης, κάποιοι χρήστες έλαβαν περισσότερα προϊόντα συγκριτικά με το πλήθος των διαφορετικών προϊόντων που μπορούν να δεχθούν ή έλαβαν έμμεσα το ίδιο προϊόν για περισσότερες από μία φορές, γεγονός που παρατηρήθηκε σε μεγαλύτερο βαθμό για Small-World τοπολογίες. Η μεθοδολογία του DifRec δεν επιτρέπει σε κάποιον χρήστη να λάβει το ίδιο προϊόν και από το ΣΣ και από αναδημοσιεύσεις. Τέλος, αναδείχθηκε ο περιορισμός ότι κάθε χρήστης λαμβάνει έως ένα συγκεκριμένο αριθμό άμεσων συστάσεων. Όσον αφορά τον CoveR, διαπιστώθηκε ότι κανένας χρήστης του δικτύου δεν έμεινε χωρίς σύσταση, όπως ήταν αναμενόμενο καθώς η μεθοδολογία του δημιουργεί συνθήκες για κάλυψη όλων των χρηστών του δικτύου. Συγκρίνοντας τις δύο μεθόδους, δεν ήταν εύκολο να αποφανθούμε ποια είναι πιο αποδοτική, λόγω των διαφορών κάποιων παραμέτρων που σε κάποιες περιπτώσεις υποδεικνύουν τον CoveR, σε κάποιες άλλες τον DifRec. Ωστόσο, στην πλειοψηφία των περιπτώσεων, ο CoveR δίνει καλύτερα αποτελέσματα αφού δίνει καλύτερο Relevance Score και οι τελικές συστάσεις γίνονται σε λιγότερους χρήστες. Σημαντικό είναι και το γεγονός ότι εξάγονται συστάσεις για όλους τους χρήστες του μέσου κοινωνικής δικτύωσης.

Συνοψίζοντας, μέσα από την παρούσα διπλωματική εργασία μπορεί επίσης κανείς να διαπιστώσει τη σημασία των δεσμών ανάμεσα σε χρήστες ενός μέσου κοινωνικής

δικτύωσης και πώς η εκμετάλλευσή τους δύναται να συμβάλλει στη βελτιστοποίηση των κλασικών Συστημάτων Συστάσεων. Επίσης, αναδεικνύεται η αποτελεσματικότητα των Information Diffusion Aware Συστημάτων Συστάσεων, τα οποία συνεισφέρουν στην αντιμετώπιση της πληροφοριακής συμφόρησης και εξάγουν πιο ακριβείς συστάσεις χρησιμοποιώντας τη διάχυση των πληροφοριών μεταξύ των χρηστών ενός μέσου κοινωνική δικτύωσης,

7

Βιβλιογραφία

- [1] Symeonidis, P., Nanopoulos, A., & Manolopoulos, Y. (2008). Providing justifications in recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(6), 1262-1272.
- [2] Karyotis, V., Vitoropoulou, M., Kalatzis, N., Roussaki, I., & Papavassiliou, S. (2019). Efficient and socio-aware recommendation approaches for bigdata networked systems. *Big Data Recommender Systems: Algorithms, Architectures, Big Data, Security and Trust*, 1, 41.
- [3] Aggarwal, C. C. (2016). An introduction to recommender systems. In *Recommender systems* (pp. 1-28). Springer, Cham.
- [4] Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1-35). Springer, Boston, MA.
- [5] Tang, J., Hu, X., & Liu, H. (2013). Social recommendation: a review. *Social Network Analysis and Mining*, 3(4), 1113-1133.

- [6] Thorat, P. B., Goudar, R. M., & Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 110(4), 31-36.
- [7] Garcia-Gavilanes, R., & Amatriain, X. (2010). Weighted content based methods for recommending connections in online social networks.
- [8] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325-341). Springer, Berlin, Heidelberg.
- [9] Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web* (pp. 291-324). Springer, Berlin, Heidelberg.
- [10] Breese, J. S., Heckerman, D., & Kadie, C. (2013). Empirical analysis of predictive algorithms for collaborative filtering. *arXiv preprint arXiv:1301.7363*.
- [11] Do, M. P. T., Nguyen, D. V., & Nguyen, L. (2010). Model-based approach for collaborative filtering. In *6th International Conference on Information Technology for Education*.
- [12] Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261-273.
- [13] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4), 331-370.

- [14] Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5-6), 393-408.
- [15] Billsus, D., & Pazzani, M. J. (2000). User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10(2-3), 147-180.
- [16] Smyth, B., & Cotter, P. (2000). A personalised TV listings service for the digital TV age. *Knowledge-Based Systems*, 13(2-3), 53-59.
- [17] Basu, C., Hirsh, H., & Cohen, W. (1998, July). Recommendation as classification: Using social and content-based information in recommendation. In *AAAI/IAAI* (pp. 714-720).
- [18] Mooney, R. J., & Roy, L. (2000, June). Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries* (pp. 195-204).
- [19] Balabanović, M., & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66-72.
- [20] Apt, K. R., Markakis, E., & Simon, S. (2016). Paradoxes in social networks with multiple products. *Synthese*, 193(3), 663-687.
- [21] Karyotis, V., Stai, E., & Papavassiliou, S. (2013). *Evolutionary dynamics of complex communications networks*. CRC Press.

- [22] Guy, I. (2015). Social recommender systems. In *Recommender systems handbook* (pp. 511-543). Springer, Boston, MA.
- [23] Yang, X., Guo, Y., Liu, Y., & Steck, H. (2014). A survey of collaborative filtering based social recommender systems. *Computer communications*, 41, 1-10.
- [24] Yang, X., Guo, Y., & Liu, Y. (2012). Bayesian-inference-based recommendation in online social networks. *IEEE Transactions on Parallel and Distributed Systems*, 24(4), 642-651.
- [25] Arazy, O., Kumar, N., & Shapira, B. (2009). Improving social recommender systems. *IT professional*, 11(4), 38-44.
- [26] Diaz-Aviles, E., Drumond, L., Schmidt-Thieme, L., & Nejdl, W. (2012, September). Real-time top-n recommendation in social streams. In *Proceedings of the sixth ACM conference on Recommender systems* (pp. 59-66).
- [27] Chandramouli, B., Levandoski, J. J., Eldawy, A., & Mokbel, M. F. (2011, June). StreamRec: a real-time recommender system. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data* (pp. 1243-1246).
- [28] Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., & Yu, Y. (2012, August). Collaborative personalized tweet recommendation. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval* (pp. 661-670).

- [29] Pennacchiotti, M., Silvestri, F., Vahabi, H., & Venturini, R. (2012, October). Making your interests follow you on twitter. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 165-174).
- [30] Chen, C., Yin, H., Yao, J., & Cui, B. (2013). Terec: A temporal recommender system over tweet stream. *Proceedings of the VLDB Endowment*, 6(12), 1254-1257.
- [31] Ren, Z., Liang, S., Meij, E., & de Rijke, M. (2013, July). Personalized time-aware tweets summarization. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 513-522).
- [32] Shen, W., Wang, J., Luo, P., & Wang, M. (2013, August). Linking named entities in tweets with knowledge base via user interest modeling. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 68-76).
- [33] Hong, L., Doumith, A. S., & Davison, B. D. (2013, February). Co-factorization machines: modeling user interests and predicting individual decisions in twitter. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 557-566).
- [34] Sherchan, W., Nepal, S., & Paris, C. (2013). A survey of trust in social networks. *ACM Computing Surveys (CSUR)*, 45(4), 1-33.
- [35] Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *ACM Sigmod Record*, 42(2), 17-28.

- [36] Vahabi, H., Koutsopoulos, I., Gullo, F., & Halkidi, M. (2015, October). Difrec: A social-diffusion-aware recommender system. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (pp. 1481-1490).
- [37] Pan, Y., Cong, F., Chen, K., & Yu, Y. (2013, October). Diffusion-aware personalized social update recommendation. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 69-76).
- [38] Ullah, F., & Lee, S. (2016). Social content recommendation based on spatial-temporal aware diffusion modeling in social networks. *Symmetry*, 8(9), 89.
- [39] Ishikawa, M., Geczy, P., Izumi, N., & Yamaguchi, T. (2008, December). Long tail recommender utilizing information diffusion theory. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 785-788). IEEE.
- [40] Song, X., Tseng, B. L., Lin, C. Y., & Sun, M. T. (2006, August). Personalized recommendation driven by information flow. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 509-516).
- [41] Wong, F. M. F., Liu, Z., & Chiang, M. (2015). On the efficiency of social recommender networks. *IEEE/ACM Transactions on Networking*, 24(4), 2512-2524.
- [42] Wang, B., Wang, C., Bu, J., Chen, C., Zhang, W. V., Cai, D., & He, X. (2013, May).

Whom to mention: expand the diffusion of tweets by@ recommendation on micro-blogging systems. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1331-1340).

- [43] Vitoropoulou, M., Tsitseklis K., Karyotis, V. & Papavassiliou, S. (2020). «Technical Report on IDARS»
http://www.netmode.ntua.gr/main/index.php?option=com_content&view=article&id=122&Itemid=87, Zografou, Greece.
- [44] Garcia-Gavilanes, R., & Amatriain, X. (2010). Weighted content based methods for recommending connections in online social networks.