National Technical University of Athens

School of Electrical and Computer Engineering

Division of Communication, Electronic and Information Engineering

Network Management and Optimal Design Laboratory

Socio-aware Content Allocation in Complex Networks via Efficient Monitoring of Information Dissemination

Thesis for the Doctor of Philosophy (Ph. D.)

of

Margarita Vitoropoulou

Athens,

September 16, 2021

National Technical University of Athens
School of Electrical and Computer Engineering
Division of Communication, Electronic and Information Engineering
Network Management and Optimal Design Laboratory

Socio-aware Content Allocation in Complex Networks via Efficient Monitoring of Information Dissemination

Thesis for the Doctor of Philosophy (Ph. D.)

of

Margarita Vitoropoulou

Diploma in Applied Mathematical and Physical Sciences, National Technical University of Athens

Advisory Committee: Symeon Papavassiliou
Dimitrios Askounis
Ioanna Roussaki

Approved by the seven-member committee on 16 September 2021

....	....	....
S. Papavassiliou Professor NTUA	D. Askounis Professor NTUA	I. Roussaki Asst. Professor NTUA
....	....	....
T. Varvarigou Professor NTUA	V. Karyotis Assoc. Professor Ionian University	G. Matsopoulos Professor NTUA
....		
P. Stefaneas Asst. Professor NTUA		

Athens,
September 16, 2021

....

Margarita Vitoropoulou

Doctor of Philosophy in Electrical and Computer Engineering, NTUA

Abstract

This dissertation focuses on the analysis and design of content placement approaches in complex networked systems, i.e., cyber and cyber-physical networks, via efficient monitoring and inference of its constituents' interplay. Despite the fact that the types of networks under examination exhibit diverse and distinctive features, the proposed methods share a common objective, namely the utilization of minimal amount of resources in order to track, infer or predict the explicit and implicit interactions among the entities of a network, which are governed by complex constraints. Inspired by methodologies employed to solve problems of coverage in physical networks, such as Wireless Sensor Networks (WSNs), this thesis extends, through Social Network Analysis (SNA), the notion of coverage in cyber and cyber-physical networks. Problems of information tracking and inference, along with influence maximization and content allocation are mapped to covering and packing problems of combinatorial optimization, which are known to be NP-hard. To address the former, heuristic approaches as well as algorithms with provable approximation guarantees are designed and analysed.

The concept of coverage is introduced in WSNs as an important metric, which measures how well the network monitors a region of interest. In this thesis, the problem of computing a minimal amount of WSN resources to monitor an obstructed field of interest is formulated as a topology control problem. A framework based on computational geometry is introduced and two algorithms, a centralized and a distributed one, adjust the sensors' sensing ranges in order to maximize the ratio of covered area to consumed energy, while ensuring a minimum coverage percentage.

Monitoring in cyber networks, such as Online Social Networks (OSNs) is the process of tracking the users' interactions, as captured by the information flow. In networks of billions of users, it is of great significance to reduce the resources required to infer the developing diffusion dynamics. The problem of finding a minimal set of monitoring nodes is mapped to finding a maximal independent set. A greedy approach is developed for its solution, followed by a graph coloring scheme that enables information tracking and a statistical learning technique, which infers the information diffusion graph.

Approaches of monitoring or inference of information propagation can be integrated into Recommendation Systems (RSs). These RSs are known as Information Diffusion Aware

Recommendation Systems (IDARS). This dissertation treats information diffusion aware recommendations as a content allocation problem with user coverage constraints. An IDARS is designed to find a subset of users to assign them various types of content, such that eventually every user achieves a coverage goal, i.e., all users in the network receive, via information dissemination, at least a minimum number of recommendations, while the total relevance of users to the recommended items is maximized. This is translated to a generalization of the Minimum Weighted Set Cover Problem and a greedy approximation algorithm is proposed for its solution.

In Mobile Social Networks (MSNs) and platforms of streaming services the information on users' features is frequently acquired by RSs. This information along with the users' mobility patterns are exploited to derive local communities and encourage collaboration in content sharing. According to the users' physical and social ties and by acknowledging the impact of recommendations in users' content requests, the problem of content placement at edge caching networks and content sharing via Device-to-Device (D2D) communication is investigated under various objectives. At first, allocating items to a physical network, i.e., a caching network, is mapped to a problem of cache hit ratio maximization and the solution is approximated by a dynamic programming based approach. Next, content allocation is studied in a cyber-physical network formed by a caching network and a platform of streaming services. In particular, D2D-based opportunistic offloading is studied jointly with cache-aware recommendations. Multiple criteria based on user mobility patterns are proposed to determine the user equipment participating in the offloading. Expressing the user Quality of Experience (QoE) as a function of user-content relevance and its expected delivery delay, the problem of caching and recommendations is treated as a user QoE maximization problem. It is addressed by a framework that solves sequentially the problems of minimum delay content delivery, content placement in caches and cache-aware recommendations, ensuring that each user will be recommended of highly preferred content with minimum delivery delay.

Keywords: Complex Network Analysis, Information Diffusion, Monitoring, Inference, Recommender Systems, Socio-aware Content Allocation, Mobile Edge Caching.

Περίληψη

Η παρούσα διδακτορική διατριβή επικεντρώνεται στην ανάλυση και το σχεδιασμό τεχνικών ανάθεσης περιεχομένου σε σύνθετα δίκτυα, μέσω της αποδοτικής παρακολούθησης και του συμπερασμού της λειτουργίας των συστατικών τους μερών. Παρά το γεγονός ότι οι εξεταζόμενοι τύποι δικτύων παρουσιάζουν ετερογενή χαρακτηριστικά, οι προτεινόμενες μέθοδοι αναπτύσσονται σε ένα κοινό πλαίσιο που υπαγορεύει την ελαχιστοποίηση των χρησιμοποιούμενων πόρων για την παρακολούθηση, την πρόβλεψη ή το συμπερασμό των αλληλεπιδράσεων μεταξύ των οντοτήτων που τα συνιστούν. Με έναυσμα την ανάπτυξη μεθοδολογιών για την επίλυση προβλημάτων κάλυψης σε δίκτυα του φυσικού χώρου, όπως αυτά των Ασύρματων Αισθητήρων (ΑΔΑ), η έννοια της κάλυψης επανεξετάζεται μέσω της Ανάλυσης Κοινωνικών Δικτύων και επεκτείνεται σε άλλους τύπους δικτύων όπως τα δίκτυα του κυβερνοχώρου καθώς και τα κυβερνο - φυσικά δίκτυα (ΚΦΔ). Προβλήματα ιχνηλάτησης και συμπερασμού της διάδοσης πληροφορίας, μεγιστοποίησης επιρροής, ανάθεσης περιεχομένου στους χρήστες ηλεκτρονικών και κινητών μέσων κοινωνικής δικτύωσης καθώς και προσωρινής αποθήκευσης αυτού στα άκρα του δικτύου αντιστοιχίζονται σε γνωστά NP-δύσκολα (NP-hard) προβλήματα κάλυψης και συσκευασίας, για την επίλυση των οποίων, διατυπώνονται, αναλύονται και αξιολογούνται μέσω προσομοιώσεων, ευρετικοί και προσεγγιστικοί αλγόριθμοι.

Η έννοια της κάλυψης συναντάται στα Ασύρματα Δίκτυα Αισθητήρων ως σημαντική μετρική για την αξιολόγηση της απόδοσής τους στην παρακολούθηση/καταγραφή φυσικών ή περιβαλλοντικών συνθηκών σε μια περιοχή ενδιαφέροντος. Λειτουργικοί και ενεργειακοί περιορισμοί των αισθητήρων καθώς και γεωμετρικές ιδιαιτερότητες των περιοχών ενδιαφέροντος συνιστούν σημαντικές προκλήσεις στη βελτιστοποίηση της απόδοσης των ΑΔΑ. Λαμβάνοντας υπόψη τους παραπάνω περιορισμούς, διατυπώνεται ένα πλαίσιο για τη βέλτιστη κάλυψη μη κυρτών περιοχών από ΑΔΑ, τα οποία συγκροτούνται από αισθητήρες με μεταβαλλόμενες ακτίνες αίσθησης. Το προτεινόμενο πλαίσιο επιλύει ένα πρόβλημα ελέγχου τοπολογίας με στόχο τη ρύθμιση των ακτίνων αίσθησης των αισθητήρων για την ελαχιστοποίηση της ενεργειακής κατανάλωσης του δικτύου. Αξιοποιώντας μεθόδους της υπολογιστικής γεωμετρίας, δύο άπληστοι αλγόριθμοι, ένας συγκεντρωτικός και ένας κατανεμημένος, στοχεύουν στη μεγιστοποίηση του λόγου της καλυπτόμενης επιφάνειας προς το αντίστοιχο ενεργειακό κόστος, υπό τον περιορισμό της διατήρησης ενός ελάχιστου ποσοστού κάλυψης της περιοχής ενδιαφέροντος.

Η ιχνηλάτηση σε δίκτυα του κυβερνοχώρου και συγκεκριμένα στα ηλεκτρονικά μέσα κοινωνικής δικτύωσης, αναφέρεται στη διαδικασία της παρακολούθησης-παρατήρησης των αλληλεπιδράσεων των χρηστών, όπως αυτές αποτυπώνονται μέσα από την ανταλλαγή πληροφορίας. Σε δίκτυα που απαρτίζονται από δισεκατομμύρια χρήστες με ποικίλα χαρακτηριστικά, κρίνεται απαραίτητη η ανάπτυξη μεθοδολογιών προσδιορισμού των πόρων για αποδοτική παρατήρηση των αλληλεπιδράσεων και συμπερασμού των μηχανισμών διάδοσης πληροφορίας. Στην παρούσα διατριβή, ο προσδιορισμός των πόρων αντιμετωπίζεται ως πρόβλημα εύρεσης ενός μεγιστικού ανεξάρτητου συνόλου κόμβων και ένας ευρετικός αλγόριθμος διατυπώνεται για την επίλυσή του. Η παρακολούθηση της πληροφοριακής διάδοσης πραγματοποιείται με ένα σχήμα χρωματισμού ακμών σε γράφους ενώ, για το συμπερασμό της δυναμικής της, αναπτύσσεται μια τεχνική στατιστικής μάθησης.

Τεχνικές συμπερασμού ή πρόβλεψης της πληροφοριακής διάδοσης μπορούν να ενσωματωθούν στα συστήματα συστάσεων (ΣΣ) που λειτουργούν σε πλατφόρμες κοινωνικής δικτύωσης. Αυτά είναι γνωστά ως συστήματα συστάσεων με επίγνωση της διάχυσης πληροφορίας (ΣΣΕΔΠ). Η ανταλλαγή πληροφορίας μεταξύ των χρηστών αξιοποιείται από ένα ΣΣΕΔΠ ως επικουρικός μηχανισμός συστάσεων. Υπό αυτό το πρίσμα, η σύσταση περιεχομένου στους χρήστες μέσων κοινωνικής δικτύωσης διατυπώνεται ως πρόβλημα προσδιορισμού ενός συνόλου χρηστών για την ανάθεση περιεχομένου με τρόπο ώστε να μεγιστοποιείται, μέσω της πληροφοριακής διάχυσης, η συνολική σχετικότητα χρήστη-περιεχομένου στο δίκτυο, εξασφαλίζοντας παράλληλα για κάθε χρήστη, έναν ελάχιστο αριθμό προτάσεων. Αυτό αντιστοιχίζεται σε ένα γνωστό NP-hard πρόβλημα κάλυψης σε γράφους για το οποίο αναπτύσσεται και αναλύεται θεωρητικά ένας προσεγγιστικός αλγόριθμος.

Με βάση τις προτιμήσεις των χρηστών σε περιεχόμενο, όπως αυτές εξάγονται και διαμορφώνονται από τα συστήματα συστάσεων, εξετάζεται η ανάθεση περιεχομένου σε δίκτυα στο φυσικό χώρο. Συγκεκριμένα, μελετάται η προσωρινή αποθήκευση περιεχομένου στα άκρα του δικτύου, η παράδοσή του στους χρήστες κινητών μέσων κοινωνικής δικτύωσης ή πλατφορμών περιεχομένου ροής και ο διαμοιρασμός του μεταξύ των χρηστών μέσω της απευθείας επικοινωνίας των "έξυπνων" συσκευών τους. Για τον καθορισμό των σχέσεων των χρηστών εξετάζονται πέρα από τις προτιμήσεις τους, η θέση τους στο φυσικό χώρο και οι τεχνικοί περιορισμοί των συσκευών τους. Βάσει αυτών, διατυπώνονται κριτήρια επιλογής συσκευών για προσωρινή αποθήκευση περιεχομένου. Στη συνέχεια εξετάζεται η τοποθέτηση περιεχομένου στα άκρα

του δικτύου ως πρόβλημα μεγιστοποίησης του λόγου ευστοχίας της προσωρινής αποθήκευσης. Αυτό εκφράζεται ως ένα γνωστό πρόβλημα συσκευασίας και επιλύεται με μια προσεγγιστική μέθοδο δυναμικού προγραμματισμού.

Τέλος, η ανάθεση περιεχομένου μελετάται σε κυβερνο-φυσικά δίκτυα ως δισδιάστατο πρόβλημα κάλυψης, με την πρώτη διάσταση να αφορά σε δίκτυα που σχηματίζονται στο φυσικό χώρο και τη δεύτερη σε δίκτυα του κυβερνοχώρου. Αναγνωρίζοντας την επίδραση των συστάσεων στη διαμόρφωση των προτιμήσεων ενός χρήστη, η τοποθέτηση περιεχομένου στα άκρα του δικτύου αντιμετωπίζεται από κοινού με την πραγματοποίηση συστάσεων στους χρήστες κινητών μέσων κοινωνικής δικτύωσης. Στη φυσική διάσταση, ερευνάται η περίπτωση της περιστασιακής εκφόρτωσης δεδομένων σε συσκευές χρηστών. Στην περίπτωση αυτή, οι χρήστες είναι διατεθειμένοι να περιμένουν ένα εύλογο χρονικό διάστημα έως ότου συναντήσουν κάποια συσκευή που έχει αποθηκευμένο προτεινόμενο περιεχόμενο. Σκοπός είναι η μεγιστοποίηση της ποιότητας της εμπειρίας τους, η οποία εκφράζεται ως συνάρτηση της σχετικότητάς τους στο προτεινόμενο περιεχόμενο και του αναμενόμενου χρόνου παράδοσής του. Το πλαίσιο που αναπτύσσεται επιλύει διαδοχικά τα προβλήματα του ελάχιστου χρόνου παράδοσης περιεχομένου, της ανάθεσης περιεχομένου στις προσωρινές μνήμες και των συστάσεων στους χρήστες. Μέσω αυτού εξασφαλίζεται πως κάθε χρήστης θα λάβει, στον ελάχιστο δυνατό χρόνο, ένα συγκεκριμένο αριθμό συστάσεων με περιεχόμενο της προτίμησής του.

Λέξεις - Κλειδιά: Ανάλυση Σύνθετων Δικτύων, Διάδοση πληροφορίας, Παρακολούθηση, Συμπερασμός, Συστήματα Συστάσεων, Ανάθεση Περιεχομένου, Προσωρινή Αποθήκευση στα Άκρα του Δικτύου.

Acknowledgements

This dissertation bears the imprint of many people to whom I feel grateful for their help and support during my Ph.D. years.

First and foremost, I would like to express my warm appreciation and gratitude to my advisor, Professor Symeon Papavassiliou for his keen interest and constructive feedback on my research work. The completion of this dissertation could have not been possible without his expertise and valuable guidance.

I also owe a deepest thanks to Associate Professor Vasileios Karyotis for his tireless guidance and patience. His comments, suggestions and sharing of knowledge have been priceless.

I am extremely thankful to my good friend and colleague Konstantinos Tsitseklis. I feel very fortunate to have worked with him on several research works. Our collaboration has challenged me in the most fruitful way. I will never forget our long skype calls during the covid-19 lockdown.

I would also like to thank the NETMODE family, my great labmates, who lightened my mood and made my research efforts easier. They have been very motivating and caring throughout the entire time.

A very special word of thanks goes to my friends and family for their love and understanding. I am deeply grateful for their unwavering support and for always forcing me to take breaks and have fun when being overwhelmed and stressed. They kept reminding me that educational excellence cannot be fulfilling unless it meets the principles of sharing and solidarity.

Η παρούσα αφιερώνεται στον N.

και στα αποκαθηλωμένα "όταν θα".

Contents

List of Figures

List of Tables

Εκτεταμένη περίληψη στα ελληνικά

Η ταχεία ανάπτυξη νέων υποδομών κινητών δικτύων και τεχνολογιών λογισμικού έχουν οδηγήσει στο σχηματισμό πολύπλοκων, διασυνδεδεμένων συστημάτων, τα οποία προσφέρουν ποικίλες υπηρεσίες. Ο αριθμός των ατόμων που συμμετέχουν και αλληλεπιδρούν με τέτοια συστήματα αυξάνεται διαρκώς, παράγοντας δεδομένα εξαιρετικού όγκου, που ξεπερνούν την υπολογιστική δυνατότητα των σύγχρονων τεχνικών ανάλυσης και επεξεργασίας. Η παρούσα διδακτορική διατριβή επικεντρώνεται στο σχεδιασμό τεχνικών ανάθεσης περιεχομένου σε σύνθετα δίκτυα, μέσω της αποδοτικής παρακολούθησης και του συμπερασμού της λειτουργίας των συστατικών τους μερών. Παρά το γεγονός ότι οι εξεταζόμενοι τύποι δικτύων παρουσιάζουν ετερογενή χαρακτηριστικά, οι προτεινόμενες μέθοδοι αναπτύσσονται σε ένα κοινό πλαίσιο που υπαγορεύει την ελαχιστοποίηση των χρησιμοποιούμενων πόρων για την παρακολούθηση, την πρόβλεψη ή το συμπερασμό των αλληλεπιδράσεων μεταξύ των οντοτήτων που τα συνιστούν. Με έναυσμα την επίλυση προβλημάτων κάλυψης σε φυσικά δίκτυα, όπως αυτά των Ασύρματων Αισθητήρων (ΑΔΑ), η έννοια της κάλυψης επανεξετάζεται μέσω της Ανάλυσης Σύνθετων Δικτύων και επεκτείνεται σε άλλους τύπους δικτύων, όπως τα δίκτυα του κυβερνοχώρου καθώς και τα κυβερνο - φυσικά δίκτυα (ΚΦΔ). Προβλήματα ιχνηλάτησης και συμπερασμού της διάδοσης πληροφορίας, μεγιστοποίησης επιρροής, ανάθεσης περιεχομένου στους χρήστες ηλεκτρονικών και κινητών μέσων κοινωνικής δικτύωσης καθώς και προσωρινής αποθήκευσης αυτού στα άκρα του δικτύου αντιστοιχίζονται σε NP-δύσκολα (NP-hard) προβλήματα κάλυψης και συσκευασίας, για την επίλυση των οποίων, διατυπώνονται, αναλύονται και αξιολογούνται μέσω προσομοιώσεων, ευρετικοί και προσεγγιστικοί αλγόριθμοι.

Στο **Κεφάλαιο 1** παρουσιάζονται τα κίνητρα και η συμβολή της παρούσας διατριβής στην υπό μελέτη ερευνητική περιοχή. Επίσης, παρατίθεται μια σύντομη περιγραφή της δομής της.

Στο **Κεφάλαιο 2** διατυπώνονται ορισμοί των βασικών εννοιών της Θεωρίας Γραφημάτων και μετρικών της Ανάλυσης Σύνθετων Δικτύων, όπως ο βαθμός κόμβου, το μέσο μήκος μονοπατιού, ο συντελεστής ομαδοποίησης, η κεντρικότητα βαθμού και εγγύτητας καθώς και η ενδιαμεσική κεντρικότητα. Οι μετρικές αυτές αποτελούν τα βασικά στοιχεία για την ανάλυση των εξεταζόμενων προβλημάτων καθώς και των μεθοδολογιών που αναπτύσσονται για την επίλυσή τους. Επίσης, γίνεται αναφορά στις πιο χαρακτηριστικές δομές σύνθετων δι-κτύων που συναντώνται στη διατριβή, στις ιδιότητές τους, όπως αυτές προσδιορίζονται από τις μετρικές Ανάλυσης Σύνθετων Δικτύων συμπεριλαμβανομένων της κεντρικότητας βαθμού και εγγύτητας, της ενδιαμεσικής κεντρικότητας, του συντελεστή ομαδοποίησης και του ελάχιστου μήκους μονοπατιού, καθώς και στο αντίστοιχο μαθηματικό τους μοντέλο. Συγκεκριμένα, περιγράφονται οι ιδιότητες και τα πιο αντιπροσωπευτικά μοντέλα των σχεσιακών δικτύων μικρού-κόσμου και ελεύθερης-κλίμακας, τα οποία χρησιμοποιούνται για να αναπαραστήσουν τα ηλεκτρονικά μέσα κοινωνικής δικτύωσης. Τέλος, αναλύονται τα χαρακτηριστικά του τυχαίου γεωμετρικού γράφου, ως κυρίαρχου μοντέλου των χωρικών δικτύων, το οποίο επιλέγεται στην παρούσα για την αναπαράσταση των εξεταζόμενων φυσικών δικτύων.

Στο **Κεφάλαιο 3**, εξετάζεται το πρόβλημα της αποδοτικής παρακολούθησης/καταγραφής συμβάντων από φυσικά δίκτυα και συγκεκριμένα, από Ασύρματα Δίκτυα Αισθητήρων (ΑΔΑ). Ένα ΑΔΑ αποτελείται από ένα σύνολο ενεργειακά αυτόνομων κόμβων που παρατηρούν και καταγράφουν τις μεταβολές στις φυσικές ή περιβαλλοντικές συνθήκες που επικρατούν σε μια περιοχή ενδιαφέροντος. Λαμβάνοντας υπόψη τους λειτουργικούς (επικοινωνία, αίσθηση) και ενεργειακούς (διάρκεια ζωής μπαταρίας) περιορισμούς των αισθητήρων, καθώς και την ύπαρξη κυρτών αδιαφανών εμποδίων εντός της περιοχής ενδιαφέροντος, διατυπώνεται ένα πλαίσιο για τη βέλτιστη κάλυψη μη κυρτών περιοχών από ΑΔΑ, τα οποία συγκροτούνται από αισθητήρες με μεταβαλλόμενες ακτίνες αίσθησης.

Η περιοχή ενδιαφέροντος ορίζεται να είναι μια τετραγωνική περιοχή του επιπέδου, εντός της οποίας τα εμπόδια αναπαριστώνται ως μη-αλληλοεπικαλυπτόμενα κυρτά πολυγώνα. Η τοποθέτηση των εμποδίων είναι αποτέλεσμα μιας στοχαστικής διαδικασίας που εξασφαλίζει πως τα τελευταία δεν παράγουν μια διαμέριση της περιοχής ενδιαφέροντος. Η κάλυψη από το ΑΔΑ επιχειρείται στην ελεύθερη από εμπόδια περιοχή η οποία αναφέρεται ως περιοχή υπό κάλυψη. Το σύνολο των αισθητήρων μοντελοποιείται από ένα σύνολο σημείων που ανήκουν στην υπό κάλυψη περιοχή. Η θέση τους καθορίζεται από μία διωνυμική στοχαστική διαδικασία.

Για την κάλυψη των αισθητήρων, χρησιμοποιείται το δυαδικό μοντέλο δίσκου σύμφωνα με το οποίο ένας αισθητήρας καλύπτει όλα τα σημεία της υπό κάλυψη περιοχής που βρίσκονται στην τομή του δίσκου με κέντρο τον αισθητήρα και ακτίνα την ακτίνα αίσθησης του αισθητήρα, και του πολυγώνου ορατότητας του αισθητήρα. Μετά την τοποθέτησή τους, οι αισθητήρες παραμένουν στατικοί, με μεταβαλλόμενη ακτίνα αίσθησης που λαμβάνει διακριτές τιμές και σταθερή ακτίνα επικοινωνίας, η οποία διαμορφώνεται πειραματικά στην ενότητα 3.3.5 να έχει τιμή 2.5 φορές μεγαλύτερη της μέγιστης ακτίνας αίσθησης προκειμένου να εξασφαλίζεται ασυμπτωτικά η συνεκτικότητα του δικτύου επικοινωνίας. Δεδομένης της σταθερής ακτίνας επικοινωνίας, συνεπώς και του σταθερού επικοινωνιακού κόστους, η κατανάλωση ενέργειας του δικτύου αυξάνεται με την αύξηση της ακτίνας αίσθησης των αισθητήρων. Για την κατανάλωση ενέργειας, υιοθετείται το τετραγωνικό μοντέλο που δίνεται από τη σχέση (3.3).

Το προτεινόμενο πλαίσιο επιλύει ένα πρόβλημα ελέγχου τοπολογίας με στόχο τη ρύθμιση των ακτίνων αίσθησης των αισθητήρων για την ελαχιστοποίηση της ενεργειακής κατανάλωσης του δικτύου. Αξιοποιώντας μεθόδους και έννοιες της υπολογιστικής γεωμετρίας, όπως αυτή του πεδίου ορατότητας που εκφράζεται στη σχέση (3.4), δύο άπληστοι αλγόριθμοι, ένας κεντροποιημένος και ένας κατανεμημένος, επιδιώκουν τη μεγιστοποίηση του λόγου της καλυπτόμενης επιφάνειας προς το αντίστοιχο ενεργειακό κόστος, υπό τον περιορισμό της διατήρησης ενός ελάχιστου ποσοστού κάλυψης της περιοχής ενδιαφέροντος, όπως αυτό εκφράζεται στη σχέση (3.8).

Στην ενότητα 3.3.3, περιγράφεται ο συγκεντρωτικός αλγόριθμος, ο ψευδοκώδικας του οποίου, παρατίθεται στην εικόνα 3.5. Πρόκειται για έναν επαναληπτικό αλγόριθμο που, αρχικά, θεωρεί μηδενική ακτίνα αίσθησης για κάθε αισθητήρα. Σε κάθε επανάληψή του επιλέγει να αυξήσει κατά μια σταθερή ποσότητα Δr την ακτίνα του αισθητήρα που σημειώνει το μεγαλύτερο λόγο αύξησης της κάλυψης μη καλυμμένης περιοχής προς την αύξηση της ενεργειακής κατανάλωσης που επιφέρει η αύξηση Δr της ακτίνας του αισθητήρα (3.12). Ο λόγος αυτός προσδιορίζεται από τη μετρική (3.6). Ο αλγόριθμος τερματίζει όταν επιτευχθεί ένα ελάχιστο ποσοστό κάλυψης της υπό κάλυψη περιοχής.

Στην ενότητα 3.3.4 αναλύεται ο κατανεμημένος αλγόριθμος και ο αντίστοιχος ψευδοκώδικας δίνεται στην εικόνα 3.6. Ο αλγόριθμος αυτός δέχεται ως όρισμα έναν γράφο εξαρτήσεων με κόμβους τους αισθητήρες, οι οποίοι αναπτύσσουν ακμή μεταξύ τους όταν η τομή των πεδίων ορατότητάς τους παράγει ένα μη κενό σύνολο. Στον κατανεμημένο αλγόριθμο, κάθε κόμβος μεταβάλλει ανεξάρτητα την ακτίνα αίσθησής του, εώς ότου το σύστημα συγκλίνει σε μία ανάθεση ακτίνων. Σε κάθε επανάληψη του αλγορίθμου, κάθε κόμβος υπολογίζει το ποσοστό της περιορισμένης περιμέτρου αίσθησής του που δίνεται από τον τύπο (3.13) και το οποίο καθορίζεται από το τρέχον πεδίο ορατότητάς του που δίνεται από τη σχέση (3.5), τα εμπόδια και το σύνορο της περιοχής ενδιαφέροντος. Στη συνέχεια, κάθε κόμβος συγκρίνει, για αύξηση της ακτίνας του κατά Δr, τον προκύπτοντα λόγο της αύξησης της κάλυψης μη καλυμμένης περιοχής προς την αντίστοιχη αύξηση της ενεργειακής κατανάλωσης, με τον αντίστοιχο λόγο των γειτόνων του. Ανάμεσα σε αυτούς, εκείνος με τη μεγαλύτερη τιμή επιλέγεται να αυξήσει την ακτίνα αίσθησής του κατά Δr. Όταν το ποσοστό της περιορισμένης περιμέτρου αίσθησης ξεπεράσει μια ελάχιστη προκαθορισμένη τιμή, τότε ο κόμβος παύει να μεταβάλλει την ακτίνα αίσθησής του.

Ο κεντροποιημένος και ο κατανεμημένος αλγόριθμος αναλύονται και αξιολογούνται μέσω προσομοιώσεων στην ενότητα 3.3.5. Οι μετρικές που χρησιμοποιούνται για την αξιολόγηση της επίδοσής τους είναι το ποσοστό της συνολικής κατανάλωσης ενέργειας που δίνεται από τον τύπο (3.15) και η μέση ακτίνα αίσθησης των αισθητήρων. Τα αποτελέσματα που προκύπτουν αναδεικνύουν την αποτελεσματικότητα των προτεινόμενων σχημάτων στην κάλυψη του 90% και 95% της περιοχής υπό κάλυψη και την υπεροχή τους ως προς την ενεργειακή κατανάλωση έναντι απλούστερων σχημάτων ανάθεσης ακτίνων αίσθησης.

Στο **Κεφάλαιο 4**, η έννοια της κάλυψης επεκτείνεται σε ηλεκτρονικά δίκτυα και συγκεκριμένα στα ηλεκτρονικά μέσα κοινωνικής δικτύωσης. Ο όρος της ιχνηλάτησης εισάγεται για να περιγράψει τη διαδικασία της παρακολούθησης-παρατήρησης των αλληλεπιδράσεων των χρηστών μέσων κοινωνικής δικτύωσης, όπως αυτές αποτυπώνονται από την μεταξύ τους ανταλλαγή πληροφορίας. Στο πλαίσιο αυτό, αναπτύσσεται ένα αλγοριθμικό σχήμα για την ελαχιστοποίηση των πόρων που απαιτούνται προκειμένου να επιτευχθεί η αποδοτική παρατήρηση της διάδρασης των χρηστών και κατ' επέκταση του συμπερασμού των μηχανισμών διάδοσης πληροφορίας σε ένα μέσο κοινωνικής δικτύωσης.

Αρχικά, ο προσδιορισμός των πόρων, δηλαδή των χρηστών για τους οποίους το σύστημα έχει γνώση της δραστηριότητά τους, οι οποίοι θα αναφέρονται ως κόμβοι υπό επιτήρηση, αντιμετωπίζεται ως ένα πρόβλημα εύρεσης ένος ελάχιστου μεγιστικού ανεξάρτητου συνόλου ενός γράφου και ένας ευρετικός αλγόριθμος (αλγόριθμος 1) διατυπώνεται για την επίλυσή του. Ο αλγόριθμος ταξινομεί σε φθίνουσα σειρά βαθμού κόμβου όλους τους κόμβους του δικτύου και επιλέγει διαδοχικά να εντάξει στο σύνολο των κόμβων υπό επιτήρηση τον κόμβο με το μεγαλύτερο βαθμό, του οποίου η γειτονιά δεν περιέχει κανέναν κόμβο υπό επιτήρηση.

Ακολουθεί η διατύπωση του μοντέλου της μερικής καταγραφής της πληροφοριακής διάδοσης μέσω ενός σχήματος χρωματισμού ακμών σε γράφους που βασίζεται στα εγω-δίκτυα των κόμβων υπό επιτήρηση. Συγκεκριμένα, κάθε κόμβος υπό επιτήρηση αντιστοιχίζεται σε ένα χρώμα, με το οποίο χρωματίζονται οι προσκείμενες ακμές του. Για μία ακμή που δεν είναι προσκείμενη σε κόμβο υπό επιτήρηση, αποδίδεται είτε ένα νέο χρώμα, είτε κάποιο υπάρχον χρώμα, αρκεί το τελευταίο να είναι διαφορετικό από αυτά των υπολοίπων ακμών που πρόσκεινται στους κόμβους που συνδέονται με την υπό εξέταση ακμή. Με τον τρόπο αυτό επιτυγχάνεται η αναπαράσταση διάφορων επιπέδων ιδιωτικότητας σε ένα μέσο κοινωνικής δικτύωσης.

Για τη μοντελοποίηση της πληροφοριακής διάδοσης, χρησιμοποιείται ένας αμερόληπτος τυχαίος περίπατος στο γράφο της πληροφοριακής διάδοσης, ο οποίος συνιστά ένα παράγον υπογράφημα του αρχικού γράφου των χρηστών. Οι πιθανότητες μετάβασης του τυχαίου περίπατου δίνονται από τη σχέση (4.1). Επίσης, διατυπώνονται μεροληπτικοί τυχαίοι περίπατοι με μεροληψία που βασίζεται στην Ευκλείδια απόσταση μεταξύ των κόμβων (4.2) και στην κεντρικότητα εγγύτητας (4.3), υπολογισμένες στο γράφο διάδοσης πληροφορίας. Κάθε περίπατος αναπαριστά τη διάδοση μιας κλάσης πληροφορίας που διέπεται από διαφορετική δυναμική εξάπλωσης στο δίκτυο. Κάθε φορά που μία κλάση πληροφορίας, έστω I_i φτάνει σε κάποιο κόμβο υπό επιτήρηση, έστω s_k, αφήνει ένα ίχνος στη μορφή μιας τριπλέτας (I_i, t, q), όπου t είναι η χρονική στιγμή που η πληροφορία I_i φτάνει στον κόμβο s_k και q η ακολουθία των χρωμάτων των ακμών που διήλθε η πληροφορία έως ότου φτάσει στον κόμβο υπό επιτήρηση s_k.

Για το συμπερασμό της διάχυσης των κλάσεων πληροφορίας, αναπτύσσεται μια μεθοδολογία που συνδυάζει τεχνικές οπισθοδρόμησης και στατιστικής μάθησης. Θεωρώντας το ίχνος (I_i, t, q) και ξεκινώντας από το τελευταίο χρώμα της ακολουθίας q, που αντιστοιχεί σε μια ακμή προσκείμενη στον κόμβο υπό επιτήρηση s_k, κάθε χρώμα αντιστοιχίζεται, με βάση τη δομή του αρχικού δικτύου των χρηστών, σε μια ακμή του δικτύου διάχυσης πληροφορίας. Σε πολλές περιπτώσεις, ο ακριβής συμπερασμός με οπισθοδρόμηση είναι αδύνατος, όπως στην περίπτωση που η διάδοση μιας κλάσης πληροφορίας ξεκινά από έναν κόμβο που δεν ανήκει στο σύνολο των υπό επιτήρηση κόμβων και φτάνει, σε μία μονάδα του χρόνου, σε ένα κόμβο υπό επιτήρηση (ένα παράδειγμα τέτοιας περίπτωσης παρατίθεται στην εικόνα (4.5)). Για την αντιμετώπιση αυτών των περιπτώσεων, διατυπώνεται ένα σχήμα στατιστικής μάθησης βασιζόμενο στη συχνότητα που διασχίζεται μια ακμή από την υπό εξέταση κλάση πληροφορίας. Η σχετική συχνότητα διάσχισης μιας ακμής και η αντίστοιχη κεντρικότητα που σχετίζεται με τη συχνότητα άφιξης της κλάσης πληροφορίας σε ένα κόμβο, υπολογίζονται από τους τύπους (4.4) και (4.5) α-ντίστοιχα. Βάσει αυτών των μετρικών, το προτεινόμενο σχήμα επιλέγει από το σύνολο των υποψήφιων ακμών, δηλαδή το συνόλο των ακμών του αρχικού γράφου που πρόσκεινται στον εξεταζόμενο κόμβο υπό επίβλεψη, την ακμή με τη μεγαλύτερη συχνότητα διάσχισης από την συγκεκριμένη κλάση πληροφορίας. Το σχήμα αυτό συγκρίνεται με μια πιθανοτική τεχνική όπου για το συμπερασμό της διάδοσης, επιλέγεται κάθε φορά τυχαία και ομοιόμορφα μία ακμή από το σύνολο των υποψήφιων ακμών.

Το πλαίσιο αξιολογείται μέσω προσομοιώσεων σε συνθετικά (σχεσιακά και χωρικά) και πραγματικά δίκτυα με χρήση των μετρικών της ακρίβειας και ανάκλησης που αφορούν τόσο στο αρχικό μη κατευθυνόμενο δίκτυο των χρηστών όσο και στο κατευθυνόμενο δίκτυο της διάδοσης πληροφορίας. Οι μετρικές δίνονται από τις σχέσεις (4.6), (4.7), (4.8), (4.9). Τα παραγόμενα αποτελέσματα επισημαίνουν την επίδραση των χαρακτηριστικών, όπως η κατανομή του βαθμού κόμβου και το πλήθος των μονοπατιών μήκους 2 και 3 μεταξύ κόμβων υπό επίβλεψη, που παρουσιάζουν οι εξεταζόμενοι τύποι δικτύων, καθώς και την επενέργεια του εγγενούς προβλήματος της “παγωμένης εκκίνησης” στην επίδοση του σχήματος στατιστικής μάθησης.

Στην ενότητα 4.3, αξιοποιώντας πλαίσια όπως το παραπάνω για την ιχνηλάτηση και το συμπερασμό της πληροφοριακής διάδοσης σε μέσα κοινωνικής δικτύωσης, σχεδιάζεται ένα σύστημα συστάσεων. Τα συστήματα συστάσεων που ενσωματώνουν τεχνικές συμπερασμού της διάχυσης πληροφορίας είναι γνωστά ως συστήματα συστάσεων με επίγνωση της πληροφοριακής διάδοσης (ΣΣΕΠΔ). Υπό αυτό το πρίσμα, η σύσταση περιεχομένου στους χρήστες μέσων κοι-νωνικής δικτύωσης διατυπώνεται ως πρόβλημα προσδιορισμού ενός συνόλου χρηστών για την ανάθεση περιεχομένου με τρόπο ώστε να μεγιστοποιείται, μέσω της πληροφοριακής διάχυσης, η συνολική σχετικότητα χρήστη-περιεχομένου στο δίκτυο, εξασφαλίζοντας παράλληλα, για κάθε χρήστη, έναν ελάχιστο αριθμό προτάσεων.

Σε μια πλατφόρμα κοινωνικής δικτύωσης, που αναπαριστάται από ένα γράφο με κόμβους τους χρήστες και ακμές που υποδηλώνουν τη μεταξύ τους συσχέτιση, και χαρακτηρίζεται ως γράφος συστήματος, οι προτάσεις για νέο περιεχόμενο πραγματοποιούνται με δύο μηχανι-σμούς: τον άμεσο μηχανισμό των συστάσεων, όπου ένα σύστημα συστάσεων προτείνει ε-ξατομικευμένο περιεχόμενο σε κάθε χρήστη με βάση τις προτιμήσεις του και αναφέρεται ως ανάθεση περιεχομένου στο χρήστη, και τον έμμεσο μηχανισμό της ανταλλαγής πληροφορίας μεταξύ των χρηστών όπως προκύπτει από τις αλληλεπιδράσεις τους στην πλατφόρμα. Οι αλληλεπιδράσεις των χρηστών εξαρτώνται τόσο από την επιρροή που ασκεί ο ένας χρήστης στον άλλο όσο και από το περιεχόμενο που ανταλλάσσεται στο δίκτυο. Αυτές αποτυπώνονται στους αντίστοιχους γράφους επιρροής, οι οποίοι δημιουργούνται υπολογίζοντας, για κάθε αντικείμενο/περιεχόμενο, το αξιόπιστο σύνολο κάθε χρήστη, δηλαδή το σύνολο των χρηστών οι οποίοι θα λάβουν γνώση για το συγκεκριμένο αντικείμενο όταν αυτό ανατεθεί στον υπό εξέταση χρήστη από το ΣΣΕΠΔ. Το αξιόπιστο σύνολο ενός χρήστη για ένα αντικείμενο που ορίζεται από τη σχέση (4.11) υπολογίζεται με μια τεχνική δειγματοληψίας Monte Carlo. Η συνάφεια ενός αξιόπιστου συνόλου με ένα αντικείμενο ορίζεται να είναι το άθροισμα της αντίστοιχης συνάφειας των χρηστών που το συνιστούν και υπολογίζεται από τον τύπο (4.12).

Το πρόβλημα της ανάθεσης περιεχομένου σε χρήστες λαμβάνοντας υπόψη την δυναμική της πληροφοριακής διάδοσης αντιστοιχίζεται σε ένα NP-δύσκολο προβλήμα κάλυψης σε γράφους, γνωστό ως Διαμέριση Καλύμματος Συνόλου Ελάχιστου Βάρους (ΔΚΣΕΒ). Συγκεκριμένα, έχοντας το σύνολο όλων των δυνατών αναθέσεων (4.14), μια διαμέριση αυτού σε κλάσεις (4.15), καθεμία από τις οποίες περιλαμβάνει όλες τις δυνατές αναθέσεις αντικειμένων που αφορούν σε ένα χρήστη, την οικογένεια των αξιόπιστων συνόλων κάθε χρήστη για κάθε αντικείμενο (4.16), το κόστος της ανάθεσης κάθε αντικειμένου σε κάθε χρήστη (4.17), ο στόχος του προβλήματος είναι να βρεθεί ένα $\ell-$κάλυμμα ελάχιστου βάρους, δηλαδή, μια συλλογή από σύνολα της οικογένειας αξιόπιστων συνόλων που έχουν αθροιστικά το μικρότερο κόστος, έτσι ώστε κάθε χρήστης να λάμβανει είτε άμεσα -μέσω των συστάσεων-, είτε έμμεσα -μέσω της διάχυσης πληροφορίας- τουλάχιστον ℓ διαφορετικά αντικείμενα. Το τελευταίο εκφράζεται τυπικά από τη σχέση (4.18). Το κόστος της ανάθεσης περιεχομένου σε ένα χρήστη ορίζεται να είναι το αντίστροφο της συνάφειας του αντίστοιχου αξιόπιστου συνόλου.

Αρχικά, το πρόβλημα ΔΚΣΕΒ διατυπώνεται ως πρόβλημα μη γραμμικού ακέραιου προγραμματισμού με αντικειμενική συνάρτηση τη συνάρτηση (4.20). Ο μη γραμμικός περιορισμός (4.21), γραμμικοποιείται και εκφράζεται με τους περιορισμούς (4.28), (4.29). Το νέο πρόβλημα γραμμικού προγραμματισμού επιλύεται με τη μέθοδο βελτιστοποίησης Διαμερισμού-Φράγματος. Το πρόβλημα ΔΚΣΕΒ επιλύεται επίσης με τον άπληστο αλγόριθμο CoveR του οποίου ο ψευδοκώδικας παρέχεται στην εικόνα 4.14 και για τον οποίο αποδεικνύεται στην ενότητα 4.4.4 πως η παραγόμενη λύση προσεγγίζει τη βέλτιστη κατά $O(\frac{\Delta}{\delta} \cdot H(\Delta))$. Τα Δ, δ αφορούν στο μέγιστο και ελάχιστο βαθμό ενός γράφου αντιστοίχως, με $H(\Delta)$ να είναι ο $\Delta^{ος}$ αρμονικός μέσος. Σε κάθε επανάληψη, έχοντας το αντίστοιχο σύνολο των δυνατών αναθέσεων, ο αλγόριθμος CoveR επιλέγει από την οικογένεια αξιόπιστων συνόλων (4.16) την ανάθεση που μεγιστοποιεί το λόγο του κόστους του αξιόπιστου συνόλου προς το σχετικό μεγέθός του, όπως εκφράζεται στη σχέση (4.30). Το σχετικό μέγεθος του αξιόπιστου συνόλου, δηλαδή το πλήθος των χρηστών που δεν έχουν εκτεθεί σε προηγούμενη επανάληψη του αλγορίθμου στο υπό εξέταση περιεχόμενο, λειτουργεί ως μετρική της καλυπτικής δυνατότητας μιας ανάθεσης. Χρησιμοποιώντας αυτή τη μετρική, ο CoveR προάγει τις αναθέσεις περιεχομένου που προκαλούν μεγάλη διάχυση ως προς το πλήθος των χρηστών με αθροιστικά υψηλή συνάφεια των τελευταίων στο περιεχόμενο.

Η επίδοση του CoveR ερευνάται μέσω προσομοιώσεων σε συνθετικά και πραγματικά δεδομένα στην ενότητα 4.4.5. Ο CoveR αξιολογείται ως προς τη συνολική συνάφεια που σημειώνουν οι αναθέσεις του, το οποίο σε συνδυασμό με το μέσο αριθμό άμεσων και έμμεσων συστάσεων αντανακλά την ποιότητα της εμπειρίας του χρήστη και την αποδοτικότητα του συστήματος συστάσεων. Ο αριθμός των άμεσων συστάσεων ποσοτικοποιεί τους πόρους που δαπανώνται απο το σύστημα, ενώ το πλήθος των έμμεσων συστάσεων είναι ενδεικτικό του βαθμού αξιοποίησης της δυναμικής της διάχυσης πληροφορίας από το σύστημα συστάσεων. Μια εξίσου σημαντική μετρική είναι αυτή της κάλυψης, η οποία ορίζεται να είναι ο μέσος αριθμός άμεσων και έμμεσων συστάσεων που λαμβάνει ένας χρήστης της πλατφόρμας. Με τη μετρική αυτή εξετάζεται ο πληροφοριακός φόρτος των χρηστών. Όπως παρουσιάζεται στις εικόνες 4.22, 4.23, ο CoveR δεν υπερβαίνει τις $2.6 \cdot \ell$ συστάσεις διαφορετικών αντικειμένων ανά χρήστη. Τέλος, υιοθετούνται οι μετρικές της καινοτομίας και της ετερογένειας των συστάσεων που ορίζονται στις (4.43), (4.44).

Τα αποτελέσματα των προσομοιώσεων επιβεβαιώνουν την υπεροχή του CoveR έναντι ενός άλλου ευρετικού αλγορίθμου συστάσεων με επίγνωση της πληροφοριακής διάδοσης, γνωστού ως DifRec καθώς και ενός παραδοσιακού συστήματος συστάσεων. Οι αναθέσεις του CoveR συκρίνονται με τις αναθέσεις που προκύπτουν από την επίλυση του αντίστοιχου προβλήματος ακέραιου προγραμματισμού με τη μέθοδο Διαμερισμού-Φράγματος επιβεβαιώνοντας την ποιότητα της λύσης του και το προβάδισμά του ως προς το χρόνο εκτέλεσης.

Στο **Κεφάλαιο 5**, μελετώνται προβλήματα ανάθεσης περιεχομένου σε δίκτυα του φυσικού χώρου, όπως τα κινητά δίκτυα προσωρινής αποθήκευσης περιεχομένου και κυβερνο-φυσικά δίκτυα που συγκροτούνται από κινητά δίκτυα προσωρινής αποθήκευσης περιεχομένου και κινητά μέσα κοινωνικής δικτύωσης ή πλατφόρμες περιεχομένου ροής.

Η τεχνολογία της προσωρινής αποθήκευσης περιεχομένου στα άκρα του δικτύου χρησιμοποιεί εξυπηρετητές προκειμένου να φέρει δημοφιλές περιεχόμενο πιο κοντά στο χρήστη, μειώνοντας σημαντικά την καθυστέρηση από άκρο-σε-άκρο και τη συμφόρηση του κεντρικού δικτύου που συνεπάγεται σημαντική βελτίωση στην ποιότητα της εμπερίας του χρήστη. Το περιεχόμενο μπορεί να αποθηκευτεί στα άκρα του δικτύου (π.χ., μικροί, femto, pico σταθμοί βάσης) αλλά και σε συσκευές χρηστών, οι οποίες συγκροτούν, αθροιστικά, ένα αποθηκευτικό χώρο μεγάλης χωρητικότητας και μικρού κόστους που σε συνδυασμό με την δυνατότητα της απευθείας μεταξύ τους επικοινωνίας (χωρίς δηλαδή τη χρήση της υποδομής του δικτύου κινητής επικοινωνίας), καθιστα την προσωρινή αποθήκευση περιεχομένου πιο αποδοτική.

Στην ενότητα 5.4, εξετάζεται η ανάθεση περιεχομένου για προσωρινή αποθήκευση σε ένα δίκτυο στο φυσικό χώρο που συγκροτείται από ένα σταθμό βάσης με δυνατότητα αποθήκευσης δεδομένων και από "έξυπνες" συσκευές που ανήκουν σε χρήστες ενός μέσου κοινωνικής δικτύωσης ή πλατφόρμας με περιεχόμενο ροής, στην οποία έχουν πρόσβαση μέσω των συσκευών τους. Στη συνέχεια, μελετάται η παράδοση του περιεχομένου στους χρήστες κατόπιν σχετικού αιτήματος και ο διαμοιρασμός του μεταξύ των χρηστών μέσω της απευθείας επικοινωνίας των συσκευών τους.

Οι σχέσεις των χρηστών εκφράζονται σε δύο επίπεδα. Στο ηλεκτρονικό επίπεδο, οι σχέσεις καθορίζονται από τις προτιμήσεις των χρηστών σε περιεχόμενο που διακινείται στην πλατφόρμα, ενώ στο φυσικό επίπεδο από τη θέση τους στο φυσικό χώρο. Οι σχέσεις αυτές αναπαριστώνται με έναν γράφο ομοιότητας και έναν γράφο τοποθεσίας αντίστοιχα, όπως φαίνονται στην εικόνα 5.1. Το σύνολο των κόμβων είναι κοινό για τους δύο γράφους και μοντελοποιεί το υποσύνολο των χρηστών της πλατφόρμας που βρίσκονται στην υπό εξέταση περιοχή του χώρου.

Το σύστημα συστάσεων που λειτουργεί στην πλατφόρμα εξάγει τις προτιμήσεις των χρηστών σε αντικείμενα που ανήκουν σε έναν καταλόγο περιεχομένου, και με βάση αυτές, υπολογίζει την ομοιότητα των χρηστών. Στο γράφο ομοιότητας, δύο κόμβοι συνδέονται με ακμή όταν η ομοιότητά τους, που προσδιορίζεται με μία μετρική όπως π. χ., Jaccard, Cosine [2], ξεπερνά μία προκαθορισμένη τιμή. Για χαμηλή τιμή κατωφλιού, ο γράφος ομοιότητας γίνεται πιο πυκνός και συνδέει χρήστες που δεν έχουν μεγάλη επικάλυψη στα αντικείμενα της προτίμησής τους. Κάθε κόμβος στο γράφο τοποθεσίας χαρακτηρίζεται από μία αποτελεσματική ακτίνα που μοντελοποιεί την μέγιστη απόσταση στην οποία ένας χρήστης μπορεί να μεταδώσει ή να λάβει δεδομένα. Δύο κόμβοι συνδέονται με μια ακμή στο γράφο τοποθεσίας όταν ο ένας βρίσκεται εντός της αποτελεσματικής ακτίνας του άλλου και αντίστροφα. Οι γειτονικοί κόμβοι του γράφου τοποθεσίας μπορούν να ανταλλάξουν δεδομένα με απευθείας επικοινωνία των συσκευών τους. Η τομή του γράφου ομοιότητας με το γράφο τοποθεσίας παράγει το γράφο του συστήματος, ο οποίος αναπαριστά σχέσεις χρηστών με χωρική εγγύτητα και κοινές προτιμήσεις σε περιεχόμενο.

Στο πλαίσιο της αύξησης της αποδοτικότητας της προσωρινής αποθήκευσης περιεχομένου στα άκρα του δικτύου, αξιοποιείται ένα μέρος του εξοπλισμού των χρηστών. Η επιλογή του εξοπλισμού λαμβάνει υπόψη τόσο την ομοιότητα και την εγγύτητα των χρηστών όσο και τα τεχνικά χαρακτηριστικά των συσκευών τους. Αρχικά, ο γράφος του συστήματος διαμερίζεται σε ομάδες χρηστών μέσα από τον εντοπισμό κοινοτήτων με μια τεχνική μεγιστοποίησης της αρθρωτότητας. Στη συνέχεια, σε κάθε κοινότητα, ένα υποσύνολο των χρηστών που την συνιστούν επιλέγεται για συμμετοχή στην προσωρινή αποθήκευση, βάσει δύο σχημάτων. Το πρώτο σχήμα επιλέγει με κριτήριο την καλυπτική ικανότητα των κόμβων στην κοινότητα, όπως αυτή εκφράζεται από τη σχέση (5.2). Το διαφορετικό μέγεθος και η πυκνότητα των κοινοτήτων που προκύπτουν με την τεχνική της αρθρωτότητας, σε συνδυασμό με την περιορισμένη χωρητικότητα της μνήμης των συσκευών, λαμβάνονται υπόψη στο δεύτερο σχήμα επιλογής εξοπλισμού, στο οποίο διατυπώνεται η μετρική (5.5), η οποία υποβιβάζει τους κόμβους με μικρή χωρητικότητα και μεγάλο βαθμό, ενώ προβιβάζει τους κόμβους που χαρακτηρίζονται από μεγάλο βαθμό εντός της κοινότητας και η χωρητικότητα της μνήμης τους εξασφαλίζει την αποθήκευση ενός ελάχιστου πλήθους αντικειμένων ανά γειτονικό χρήστη στην κοινότητα.

Για την τοποθέτηση περιεχομένου στα άκρα του δικτύου (σταθμός βάσης) και στον επιλεγμένο εξοπλισμό των χρηστών, λαμβάνεται υπόψη τόσο το μέγεθος των αντικειμένων του καταλόγου, η χωρητικότητα των συσκευών μνήμης οσο και η αξία των αντικειμένων για τους χρήστες. Η αξία των αντικειμένων προσδιορίζεται από μια συνάρτηση ωφέλειας (5.14) που βασίζεται στη σχετικότητα των χρηστών με τα τελευταία. Το πρόβλημα της ανάθεσης περιεχομένου διατυπώνεται ως πρόβλημα μεγιστοποίησης του λόγου ευστοχίας της προσωρινής αποθήκευσης (5.10) και αντιστοιχίζεται σε ένα γνωστό NP-δύσκολο (NP-Hard) πρόβλημα συσκευασίας, το οποίο επιλύεται διαδοχικά, για κάθε διαθέσιμη μνήμη, με το πλαίσιο CAUSE (ο ψευδοκώδικάς του παρατίθεται στην εικόνα 5.4) που βασίζεται σε μια προσεγγιστική μέθοδο δυναμικού προγραμματισμού.

Η επίδοση του πλαισίου CAUSE αξιολογείται μέσω προσομοιώσεων σε συνθετικά χωρικά δίκτυα στην ενότητα 5.4.3.1. Μελετάται πειραματικά η επίδραση συγκεκριμένων χαρακτηριστικών του του δικτύου των χρηστών, του καταλόγου των αντικειμένων και του δικτύου προσωρινής αποθήκευσης περιεχομένου στο λόγο ευστοχίας, όπως το μέγεθος του δικτύου των χρηστών, το μέγεθος του καταλόγου του περιεχομένου, η χωρητικότητα της μνήμης του σταθμού βάσης και του εξοπλισμού των χρηστών, το πλήθος των αιτημάτων ανά χρήστη και το μοντέλο των αιτημάτων. Τα αιτήματα των χρηστών διαμορφώνονται από ένα ντετερμινιστικό και ένα πιθανοτικό σχήμα, τα οποία βασίζονται στη σχετικότητα αντικειμένου-χρήστη, όπως αυτή εκτιμάται από το σύστημα συστάσεων. Τα αποτελέσματα αναδεικνύουν τη συνεισφορά του επιλεγμένου εξοπλισμού των χρηστών στην αποδοτικότερη προσωρινή αποθήκευση περιεχομένου και την υπεροχή του CAUSE έναντι ενός βασικού σχήματος αποκλειστικής αποθήκευσης περιεχομένου στα άκρα του δικτύου.

Στην ενότητα 5.5, η ανάθεση περιεχομένου μελετάται σε κυβερνο-φυσικά δίκτυα ως δισδιάστατο πρόβλημα κάλυψης, με την πρώτη διάσταση να αφορά στο φυσικό χώρο και τη δεύτερη στον κυβερνοχώρο.

Στο φυσικό επίπεδο, η προσωρινή αποθήκευση περιεχομένου σε ένα ετερογενές δίκτυο που συγκροτείται από ένα κεντρικό σταθμό βάσης και μικρούς σταθμούς βάσης, ερευνάται ταυτόχρονα με την περιστασιακή εκφόρτωση δεδομένων σε εξοπλισμό χρηστών, οι οποίοι κινούνται με ένα συγκεκριμένο μοτίβο σε μια περιοχή ενδιαφέροντος κι έχουν πρόσβαση στην πλατφόρμα μέσω του δίκτυου κινητής επικοινωνίας. Στο κυβερνο-επίπεδο, η κατανομή περιεχομένου εξετάζεται ως πρόβλημα φιλικών-προς-το-δίκτυο συστάσεων στους χρήστες της πλατφόρμας με βάση τις προτιμήσεις τους και την απόστασή τους από την τοποθεσία που είναι αποθηκευμένο το περιεχομένο. Σκοπός είναι η επιλογή ενός υποσυνόλου του εξοπλισμού των χρηστών για περιστασιακή εκφόρτωση δεδομένων και απο κοινού προσωρινή αποθήκευση και σύσταση περιεχομένου με στόχο τη μεγιστοποίηση της ποιότητας της εμπειρίας των χρηστών.

Μέρος του περιεχομένου που παράγεται στην πλατφόρμα αποθηκεύεται στους μικρούς σταθμούς βάσης, που είναι εξοπλισμένοι με μνήμες διαφορετικής χωρητικότητας, και σε ένα περιορισμένο αριθμό από τις συσκευές των κινούμενων χρηστών. Οι χρήστες είναι διατεθειμένοι να περιμένουν ένα εύλογο χρονικό διάστημα, έως ότου συναντήσουν κάποια συσκευή που συμμετέχει στην προσωρινή αποθήκευση περιεχομένου και έχει αποθηκευμένο κάποιο προτεινόμενο σε αυτούς περιεχόμενο. Όταν ο ανεκτός χρόνος αναμονής παρέλθει, οι χρήστες λαμβάνουν το περιεχόμενο από το βασικό δίκτυο μέσω του κεντρικού σταθμού βάσης.

Το μοντέλο της κίνησης των χρηστών διαμορφώνεται στην ενότητα 5.5.2 ως εξής: Οι χρήστες κινούνται μέσα στην περιοχή ενδιαφέροντος με τρόπο ώστε το πλήθος των συναντήσεών τους να δίνεται από μια ομογενή ανέλιξη Poisson, με διαφορετική τιμή έντασης ανά ζεύγος. Για λόγους δίκαιου διαμοιρασμού του περιεχομένου, σε κάθε συνάντηση μεταξύ δύο χρηστών, μπορεί να παραδοθεί μόνο ένα αντικείμενο που είναι αποθηκευμένο στη μνήμη ενός εκ των δύο αντίστοιχων συσκευών. Λόγω της ανέλιξης Poisson, ο χρόνος μεταξύ δύο διαδοχικών συναντήσεων ενός ζεύγους χρηστών ακολουθεί εκθετική κατανομή. Συνεπώς, για μη μηδενική συνάρτηση έντασης, ο αναμενόμενος χρόνος μεταξύ δύο διαδοχικών συναντήσεων, που ερμηνεύεται ως αναμενόμενη καθυστέρηση για τη λήψη ενός περιεχομένου, θα δίνεται από το αντίστροφο της έντασης του συγκεκριμένου ζεύγους (5.15).

Με αντίστοιχο τρόπο μοντελοποιούνται οι συσχετίσεις των χρηστών με τους μικρούς σταθμούς βάσης στην ενότητα 5.5.2.4, με τη διαφορά πως η τιμή της έντασης ορίζεται να είναι μικρότερη σε σχέση με την αντίστοιχη ενός ζεύγους χρηστών, προκειμένου να αντικατοπτρίζεται το μεγαλύτερο διάστημα παραμονής ενός χρήστη εντός της περιοχής κάλυψης ενός μικρού σταθμού βάσης. Το πλήθος των αποδοτικών συναντήσεων, δηλαδή των συναντήσεων που θα έχουν ως αποτέλεσμα την επιτυχή παράδοση περιεχομένου, δίνεται από τον τύπο (5.16) για ζεύγη χρηστών και από τον τύπο (5.18) για ζεύγη από χρήστες και μικρούς σταθμούς βάσης. Και στις δύο περιπτώσεις, το πλήθος των αποδοτικών συναντήσεων καθορίζεται από τον ανεκτό χρόνο αναμονής, το πλήθος των αντικειμένων που μπορούν να ανταλλαχθούν σε μία συνάντηση, τη χωρητικότητα της υπό εξέταση μνήμης, τη διάρκεια του διαστήματος παρακολούθησης και το μέγεθος της λίστας των συστάσεων κάθε χρήστη.

Η επιλογή του εξοπλισμού των χρηστών που θα αξιοποιηθεί για την περιστασιακή εκφόρτωση δεδομένων, βασίζεται στο πλήθος των αποδοτικών συναντήσεων των χρηστών και διατυπώνεται ως πρόβλημα εύρεσης ενός συνόλου κάλυψης ελάχιστου βάρους και προκαθορισμένης πληθικότητας. Συγκεκριμένα, στην ενότητα 5.5.5.1 επιχειρείται, με έναν άπληστο αλγόριθμο, η εύρεση ενός υποσυνόλου των χρηστών το οποίο α) θα παρέχει, σε όσο το δυνατό μεγαλύτερο αριθμό χρήστών, πρόσβαση σε περιεχόμενο που βρίσκεται αποθηκευμένο σε εξοπλισμό χρηστών, β) θα εξασφαλίζει εξισορροπημένο φόρτο αιτημάτων στους χρήστες που αναλαμβάνουν την περιστασιακή εκφόρτωση περιεχομένου.

Η ποιότητα της εμπειρίας του χρήστη εκφράζεται στη σχέση (5.21) ως κυρτός συνδυασμός της ποιότητας των παρεχόμενων συστάσεων και της ποιότητας της υπηρεσίας. Η ποιότητα των συστάσεων για ένα χρήστη ορίζεται να είναι μια γραμμική συνάρτηση της σχετικότητάς του στο προτεινόμενο περιεχόμενο (5.19). Η ποιότητα της υπηρεσίας (5.20) ορίζεται ως μια συνάρτηση του αναμενόμενου χρόνου παράδοσης του προτεινόμενου περιεχομένου, ο οποίος καθορίζεται από την κινητικότητα των χρηστών.

Το πρόβλημα της από κοινού αποθήκευσης και σύστασης περιεχομένου εκφράζεται στη σχέση (5.30) ως πρόβλημα μεγιστοποίησης της εμπειρίας του χρήστη. Το πρόβλημα αυτό έχει αποδειχθεί στο [3] πως ανήκει στην κλάση των NP-δύσκολων προβλημάτων. Για την επίλυσή του αναπτύσσεται το πλαίσιο MD που επιλύει διαδοχικά τα υποπροβλήματα του ελάχιστου χρόνου παράδοσης περιεχομένου, της προσωρινής αποθήκευσης περιεχομένου στις μνήμες και των συστάσεων στους χρήστες. Μέσω αυτού εξασφαλίζεται πως κάθε χρήστης θα λάβει, στον ελάχιστο δυνατό χρόνο, ένα συγκεκριμένο αριθμό συστάσεων με περιεχόμενο της προτίμησής του.

Το πρόβλημα του ελάχιστου χρόνου παράδοσης περιεχομένου διατυπώνεται ως πρόβλημα γραμμικού προγραμματισμού το οποίο εκφράζεται από τη σχέση (5.35), το οποίο αναζητά, για κάθε χρήστη, τις συσχετίσεις του με συσκευές προσωρινής αποθήκευσης που του εξασφαλίζουν, στον ελάχιστο δυνατό χρόνο, την παράδοση τουλάχιστον ενός ελάχιστου αριθμού αντικειμένων (5.36). Στη συνέχεια, η τοποθέτηση περιεχομένου στις μνήμες γίνεται με μια επαναληπτική μέθοδο που χρησιμοποιεί τη συνάρτηση χρησιμότητας (5.38). Με αυτή, υπολογίζεται, για κάθε δυνατή ανάθεση αντικειμένου σε μνήμη, ο σταθμισμένος μέσος της σχετικότητας του αντικειμένου στους χρήστες που συσχετίζονται με τη συγκεκριμένη μνήμη. Η μέθοδος, σε κάθε επανάληψη, επιλέγει την ανάθεση με την υψηλότερη τιμή χρησιμότητας. Τέλος, η φιλική-προς-το-δίκτυο λίστα συστάσεων ενός χρήστη διαμορφώνεται με την προσθήκη περιεχομένου που βρίσκεται αποθηκευμένο στις μνήμες με τις οποίες ο τελευταίος συσχετίζεται και για το οποίο παρουσιάζει υψηλή προτίμηση.

Η επίδοση του πλαισίου MD ερευνάται μέσω προσομοιώσεων σε συνθετικά δίκτυα στην ενότητα 5.5.8. Το πλαίσιο MD αξιολογείται ως προς τις τιμές που επιτυγχάνει στις κανονικοποιημένες μετρικές της ποιότητας των συστάσεων (5.39), της ποιότητας της υπηρεσίας (5.40) και της ποιότητας της εμπειρίας (5.41). Τα παραγόμενα αποτελέσματα συγκρίνονται με αυτά της προσεγγιστικής μεθόδου JCR που διατυπώνεται στο [3], η οποία βασίζεται στην εξαντλητική αναζήτηση λύσεων στο πρόβλημα της από κοινού προσωρινής αποθήκευσης και σύστασης περιεχομένου. Το MD πλαίσιο, παρά του ότι σημειώνει λίγο χαμηλότερη ποιότητα συστάσεων, φαίνεται να αντισταθμίζει ικανοποιητικά την ποιότητα της λύσης με το χρόνο εκτέλεσης, ο οποίος, όπως φαίνεται στον πίνακα 5.1, δεν ξεπερνά τα λίγα δευτερόλεπτα, σε αντίθεση με την JCR μέθοδο που, για δίκτυα μεγάλου μεγέθους, ξεπερνά τις 8 ημέρες.

Στο **Κεφάλαιο 6**, συνοψίζονται τα συμπεράσματα της διατριβής, αναδεικνύοντας την αναγκαιότητα της αντιμετώπισης των εξεταζόμενων προβλημάτων καθώς και την αποδοτικότητα των μεθοδολογιών που αναπτύχθηκαν για την επίλυσή τους. Τέλος, γίνεται αναφορά σε ανοιχτά ερευνητικά θέματα που θα μπορούσαν να αποτελέσουν προεκτάσεις της παρούσας διατριβής. Αυτά οργανώνονται σε δύο θεματικούς άξονες που έχουν σημείο αναφοράς τις τεχνολογίες της προσωρινής αποθήκευσης περιεχομένου στα άκρα του δικτύου. Ο πρώτος άξονας αφορά στη διάδοση μη δημοφιλούς περιεχομένου μέσω φιλικών-προς-το-δίκτυο συστημάτων συστάσεων που έχουν επίγνωση του μοτίβου εναλλαγής των χρηστών μεταξύ των δικτύων σταθερής και κινητής επικοινωνίας. Ο δεύτερος άξονας αφορά στη μελέτη της μεροληψίας των μεθόδων περιστασιακής εκφόρτωσης δεδομένων και των φιλικών-προς-το-δίκτυο συστάσεων.

Τα τελευταία χρόνια ολοένα και περισσότεροι χρήστες έχουν πρόσβαση σε πλατφόρμες με περιεχόμενο ροής από τις κινητές τους συσκευές (π.χ., Youtube, Cinobo, Mubi) [4], ενώ το μεγαλύτερο μέρος της κίνησης στο κεντρικό δίκτυο οφείλεται σε αιτήματα των χρηστών για το ίδιο περιεχόμενο (δημοφιλές περιεχόμενο), η παράδοση του οποίου προκαλεί σημαντικό φόρτο στο κεντρικό δίκτυο [5]. Για την αποφόρτιση των ζεύξεων του κεντρικού δικτύου και τη μείωση του κόστους παράδοσης του περιεχομένου, το περιεχόμενο αυτό αποθηκεύεται στα άκρα του δικτύου, π.χ., σε κεντρικούς σταθμούς βάσης, μικρούς/femto/pico σταθμούς βάσης καθώς και σε εξοπλισμό χρηστών [6].

Όπως αναφέρεται στο κεφάλαιο 5, στις περισσότερες πλατφόρμες περιεχομένου ροής λειτουργούν συστήματα συστάσεων που παρέχουν εξατομικευμένες προτάσεις περιεχομένου στους χρήστες. Πρόσφατες ερευνητικές εργασίες αναδεικνύουν τη θέση του περιεχομένου στη λίστα των συστάσεων ως σημαντικό παράγοντα διαμόρφωσης των προτιμήσεων, και κατ' επέκταση, των αιτημάτων των χρηστών [7, 8]. Οι φιλικές-προς-το-δίκτυο συστάσεις προωθούν το περιεχόμενο με μικρότερο κόστος παράδοσης, τοποθετώντας το πιο ψηλά στη λίστα των συστάσεων [9]. Πρόκειται για το περιεχόμενο που βρίσκεται αποθηκευμένο κοντά στο χρήστη, το οποίο, κατά κύριο λόγο, είναι δημοφιλές [10]. Η συγκεκριμένη αναδιάταξη των λιστών, όταν γίνεται με κριτήριο το κόστος παράδοσης του περιεχομένου, έχει ως αποτέλεσμα ο χρήστης να λαμβάνει διαφορετικές συστάσεις ανάλογα με το δίκτυο μέσω του οποίου έχει πρόσβαση στην πλατφόρμα: Δεδομένου ότι το κόστος παράδοσης μη προσωρινά αποθηκευμένου περιεχομένου μέσω του δικτύου κινητής επικοινωνίας είναι σημαντικά υψηλότερο από το αντίστοιχο μέσω του δίκτυου σταθερής τηλεφωνίας, το δημοφιλές περιεχόμενο αναμένεται να είναι κυρίαρχο στις συστάσεις που γίνονται όταν ο χρήστης έχει πρόσβαση στην πλατφόρμα μέσω του δικτύου κινητής επικοινωνίας. Στην περίπτωση αυτή, οι χρήστες που ενδιαφέρονται για επίκαιρο μη δημοφιλές περιεχόμενο, είτε δε θα έχουν καθόλου πρόσβαση σε αυτό μέσω των συστάσεων ή το μη δημοφιλές περιεχόμενο θα βρίσκεται πολύ χαμηλά στη λίστα των συστάσεών τους. Αυτό θα επιφέρει αρνητικό αντίκτυπο στην ποιότητα της εμπειρίας τους.

Η γνώση του μοτίβου εναλλαγής μεταξύ των δικτύων σταθερής και κινητής επικοινωνίας, μπορεί να αξιοποιηθεί από το σύστημα συστάσεων για την διάδοση επίκαιρου μη δημοφιλούς περιεχομένου στους χρήστες με ελάχιστο κόστος παράδοσης. Το πρόβλημα της μεγιστοποίησης της διάδοσης επίκαιρου μη δημοφιλούς περιεχομένου μέσω φιλικών-προς-το-δίκτυο συστημάτων συστάσεων που έχουν επίγνωση της εναλλαγής της σύνδεσης των χρηστών μεταξύ σταθερού και κινητού δικτύου, μπορεί να αντιμετωπιστεί ως πρόβλημα βέλτιστου ελέγχου για την κατανομή περιεχομένου εντός συγκεκριμένου χρονικού ορίζοντα [11], όπου η διάδοση του μη δημοφιλούς περιεχομένου μπορεί να αναπαρασταθεί με το επιδημιολογικό μοντέλο SI [12, 13] με τους ευπαθείς κόμβους, δηλαδή τους κόμβους που δεν έχουν εκτεθεί στην πληροφορία ακόμη, να μετατρέπονται, μέσω των συστάσεων, σε μολυσμένους κόμβους, με μεγάλο κόστος, όταν αυτό συμβαίνει κατά τη διάρκεια σύνδεσης στο δίκτυο κινη-τής και με μικρότερο κόστος, όταν η μόλυνση συμβαίνει κατά τη σύνδεση στο δίκτυο σταθερής επικοινωνίας.

Το παραπάνω ζήτημα μπορεί να θεωρηθεί ως μια προσέγγιση εξασφάλισης δίκαιης/αμερόληπτης προώθησης περιεχομένου από τα φιλικά-προς-το-δίκτυο συστήματα συστάσεων, ανεξάρτητα της δημοφιλίας του τελευταίου. Η έννοια της αμεροληψίας μπορεί να οριστεί με διαφορετικούς τρόπους ανάλογα με το εξεταζόμενο σύστημα και τις εμπλεκόμενες οντότητες.

Από τη σκοπιά της συνεργατικής προσωρινής αποθήκευσης σε εξοπλισμό χρηστών με απευθείας επικοινωνία, στο [14], η αμεροληψία λογίζεται ως μια μετρική της ίσης δυνατότητας αυτών να έχουν πρόσβαση σε περιεχόμενο που παράγεται σε ένα κινητό μέσο κοινωνικής δικτύωσης. Στο [15], η αμεροληψία σχετίζεται με την εξισορρόπηση του φόρτου σε ένα ετερογενές δίκτυο προσωρινής αποθήκευσης.

Ως προς τις συστάσεις, η αμεροληψία μπορεί να οριστεί σε σχέση με τις προτιμήσεις των χρηστών σε περιεχόμενο, το οποίο στη βιβλιογραφία αναφέρεται ως κ-αμεροληψία (αμεροληψία καταναλωτή) [16] και σε σχέση με τον πάροχο-δημιουργό του περιεχομένου που είναι γνωστή ως π-αμεροληψία [17]. Στα κπ-αμερόληπτα συστήματα συστάσεων [18, 19, 20, 21], η αμεροληψία αφορά τόσο στους καταναλωτές όσο και στους παρόχους του περιεχομένου και οι συστάσεις διαμορφώνονται με τρόπο ώστε να ικανοποιούνται και τα δύο μέρη.

Συνδυάζοντας τις διαφορετικές έννοιες της αμεροληψίας, το πρόβλημα της από κοινού προσωρινής αποθήκευσης και σύστασης περιεχομένου σε ετερογενή δίκτυα που αναπτύχθηκε στην ενότητα 5.5 και διατυπώθηκε ως πρόβλημα βελτιστοποίησης της ποιότητας της εμπειρίας του χρήστη, μπορεί να μελετηθεί υπό μια σειρά περιορισμών που αφορούν στα παρακάτω:

- την εξισορροπηση του φόρτου στις διάφορες συσκευές που συμμετέχουν στην προσωρινή αποθήκευση, με βάση τα τεχνικά χαρακτηριστικά τους, όπως η χωρητικότητα σε μνήμη, η κατανάλωση ενέργειας, η διάρκεια ζωής της μπαταρίας (στην περίπτωση του εξοπλισμού των χρηστών),
- την κάλυψη των χρηστών με περιεχόμενο (κ-αμεροληψία), δηλαδή, την εξασφάλιση πως κάθε χρήστης θα λαμβάνει τουλάχιστον έναν ελάχιστο αριθμό από αντικείμενα, με μια ελάχιστη ποιότητα μετάδοσης,
- την κάλυψη των αντικειμένων (π-αμεροληψία), δηλαδή, την εξασφάλιση πως ένα υποσύνολο των αντικείμενων του καταλόγου, θα προταθεί σε έναν ελάχιστο αριθμό χρηστών, με μια ελάχιστη ποιότητα μετάδοσης.

Το πρόβλημα βελτιστοποίησης της ποιότητας της εμπειρίας του χρήστη με τους νέους περιορισμούς μπορεί να διατυπωθεί στο πλαίσιο ενός κινητού μέσου κοινωνικής δικτύωσης, στο οποίο η διάδοση πληροφορίας πραγματοποιείται είτε άμεσα, μέσα από τις συστάσεις, είτε έμμεσα, μέσω της αλληλεπίδρασης των χρηστών στο φυσικό χώρο ή στον κυβερνοχώρο, με την τελευταία να επιτρέπει την ταχύτερη διασπορά της πληροφορίας, διαμορφώνοντας αποτελεσματικά τις προτιμήσεις των χρηστών, αυξάνοντας όμως τον κίνδυνο της πληροφοριακής υπερφόρτωσής τους. Υπό αυτό το πρίσμα, τόσο η τοποθέτηση περιεχομένου στις μνήμες όσο και οι φιλικές-προς-το-δίκτυο συστάσεις θα πρέπει να σχεδιαστούν με επίγνωση της δυναμικής της αλληλεπίδρασης των χρηστών σε χωρικό και σε ηλεκτρονικό επίπεδο.

Chapter 1

Introduction

1.1 Motivation and contributions

The rapid expansion of network infrastructures, user equipment and software technologies have resulted in the development of heterogeneous, interconnected systems as the ones in Fig. 1.1, which offer various and often diverse services. In the last decade, the number of users participating and interacting with such systems has grown significantly. This interplay produces data whose volume is exceeding the computational scales offered by the state-of-the art processing algorithms [22]. Therefore, it is necessary to develop new methodologies in order to deal with the anticipated large scale of operation. This thesis focuses on the design of socio-aware content allocation approaches in complex networks via efficient monitoring in terms of utilizing only a small amount of resources to track and infer the explicit and implicit interactions between its entities. Inspired by methodologies used to address problems of coverage in physical networks, such as Wireless Sensor Networks (WSNs), this work extends the notion of coverage in cyber as well as cyber-physical networks and treats problems of information diffusion tracking and inference along with influence maximization and content allocation as combinatorial optimization problems of covering and packing, which are proved to be NP-hard. To solve these problems, heuristic approaches are designed, as well as efficient algorithms with provable approximation guarantees.

The key contributions of this dissertation are highlighted in Fig 1.2 and can be summarized as follows:

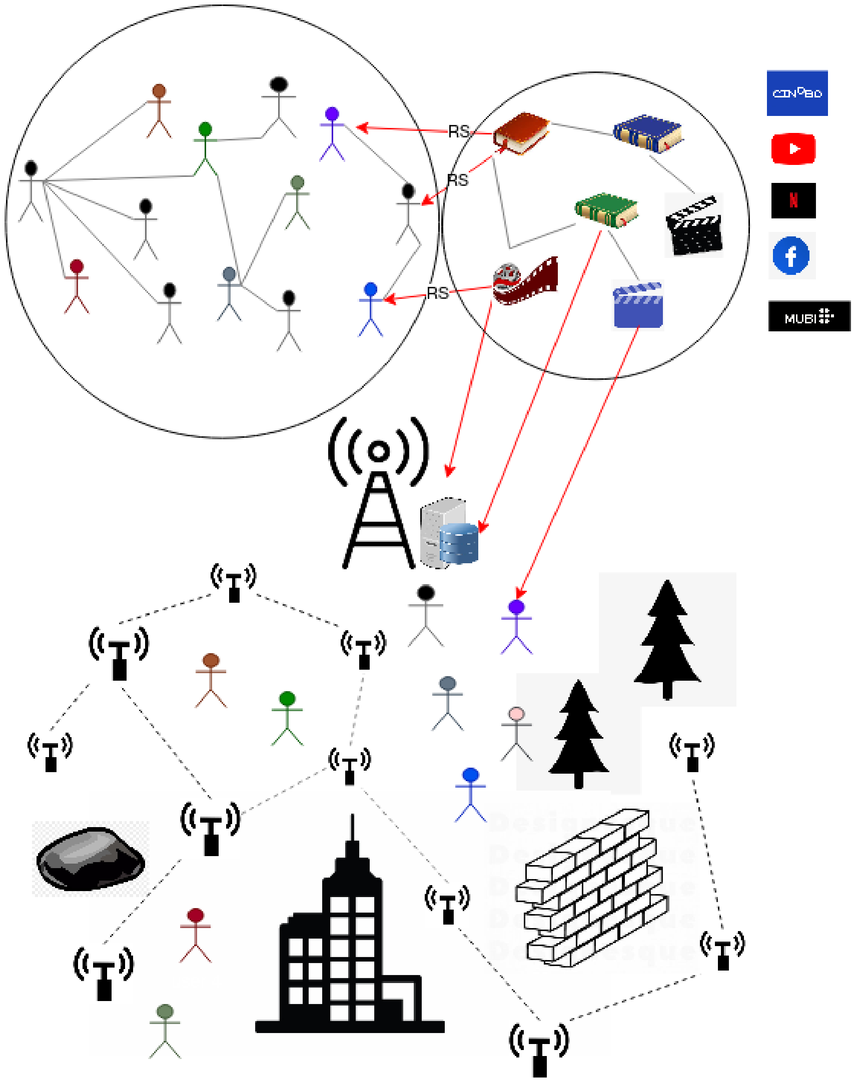


Figure 1.1: An example of a heterogeneous cyber-physical network.

- Coverage is introduced in Wireless Sensor Networks as an important metric, which measures how well the network monitors a region of interest. The problem of utilizing a minimal amount of resources in a WSN in order to monitor an obstructed field of interest is formulated as a topology control problem. Assuming that the devices have the ability to modify their sensing ranges dynamically, the objective is to maximize the area covered by randomly dispersed sensors, while reducing their sensing range as much as possible, resulting in low energy consumption, in the presence of convex opaque obstacles. A framework capitalizing on the notion of the visibility polygon

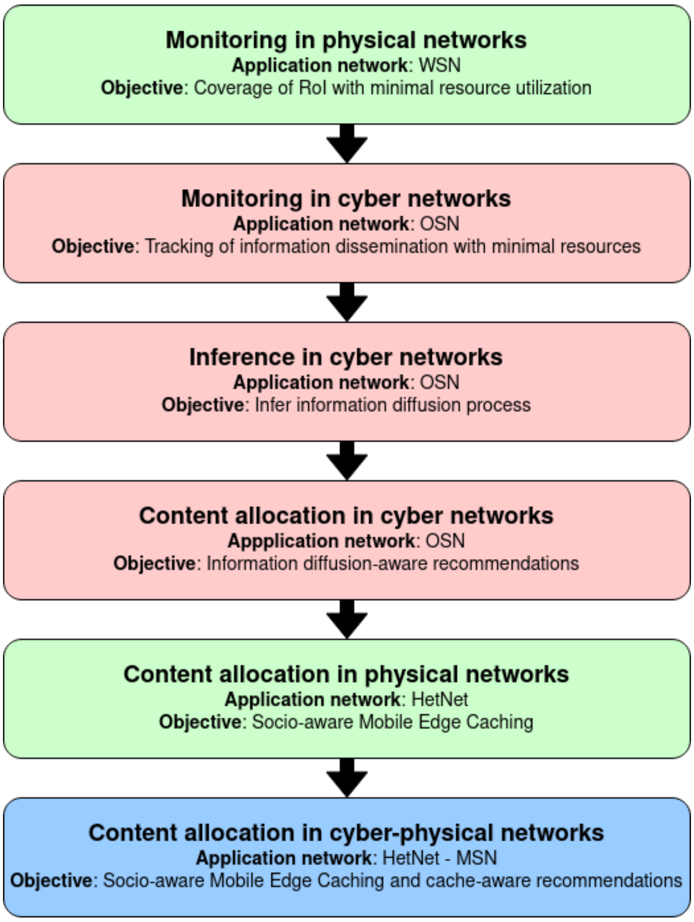


Figure 1.2: Considered systems, addressed problems and key contributions.

is introduced and two algorithms, a centralized and a distributed one are proposed. The algorithms aim to maximize the ratio of covered area to consumed energy, while ensuring a minimum coverage percentage. Analysis and simulation results indicate that the proposed schemes achieve energy-efficient coverage, outperforming the plain assignment of maximum sensing range across the network.

- Monitoring in cyber networks, such as Online Social Networks (OSNs), is defined as the process of tracking the interactions among network entities, expressed by the information flow. It is thus of great significance to determine a minimum set of nodes in an social graph that have to be monitored in order to infer its diffusion dynamics. The set of monitoring nodes chosen to recover the information propagation graph, in terms of who influences whom in the OSN, is referred to as the monitoring cover of a social graph. Finding a monitoring cover is treated as a problem of finding a maximal independent set. A greedy methodology is introduced for its solution, followed by a graph coloring scheme and a statistical learning technique for the inference of the information diffusion graph.

- The knowledge on the spreading dynamics of an OSN can be incorporated to a Recommender System (RS) through monitoring or inference of information propagation. These RSs are known as Information Diffusion Aware Recommendation Systems (IDARS). Information diffusion awareness increases the diversity of recommended items by avoiding redundant recommendations, which users are expected to attain via their connections in the OSN. The problem of information diffusion aware recommendations is studied from the viewpoint of user coverage. An IDARS is designed to utilize minimal amounts of resources in order to allocate content that best matches user preferences and respect their tolerance to information, in terms of one's capacity for distinct items. This recommendation problem is mapped to the Minimum Weighted ℓ-Cover problem, which is a generalization of the well-studied Minimum Weighted Set Cover Problem. An ℓ-cover is defined as a set of assignments to users who maximize both the spread of recommendations and the total user-to-item relevance, so that each user in the network is covered by at least ℓ items. In order to solve the ℓ-Coverage problem (find an ℓ-cover for the OSN), a greedy algorithm is proposed, which is proved to be an $O(\frac{\Delta}{\delta}H(\Delta))$-approximation for the ℓ-Coverage problem, where Δ, δ are the maximum and minimum degree of the network, respectively, and $H(\Delta)$ is the Δ^{th} harmonic number.

- In Mobile Social Networks (MSNs) and platforms of streaming services, information on users' features is acquired by Recommendation Systems. This information is exploited to derive communities and encourage user collaboration in local content sharing. By considering the case where content and network providers form one entity, two types of relationships between the users of a MSN are assumed, physical and social ones. Based on the former, Mobile Edge Caching (MEC) is treated as a content allocation problem in physical networks. In particular, the problem under examination is the one of limited utilization of User Equipment (UE) for Base Station (BS) assisted caching and content sharing through Device-to-Device (D2D) communication. By combining the concepts of monitoring and coverage as presented in physical and cyber networks respectively, content allocation in UE caches is perceived as a two-dimensional problem of content coverage. The first dimension is the physical distance that measures the proximity of users' devices, leading to physical ties. The second dimension refers to the users' social ties, determined by their similarity in content preference, which is predicted by a Recommender System that operates in the network. The problem of UE assisted caching and content sharing via D2D communication is divided into two subproblems. Initially, the selection of the assisting caches (i.e., UE caches) is translated to a problem of community detection and influence maximization based on user similarity and proximity. Then, the problem of allocating items to caches, with respect to user preferences and UE limited capacity, is mapped to the multiple knapsack problem. A dynamic programming based approach is employed to acquire an approximation of the solution that maximizes the cache hit ratio.

- Mobile Edge Caching and recommendations in MSNs or platforms of streaming services are treated as content allocation procedures in the physical and the cyber level respectively. The objective of increasing user engagement to the streaming platforms or MSNs via recommendations, while minimizing the cost associated to the content delivery is examined in the case of opportunistic offloading via D2D communication in a heterogeneous network of small cells and mobile users. Users may wait for a tolerable amount of time (the "hand-off delay") in order to consume recommended content cached in an encountering device. Based on the users' mobility pattern, a coverage-inspired approach is developed for the selection of UEs that will serve as helper caches. This approach balances the distribution of the recommended content's requests to UEs who can provide in aggregate cached content to most of the network users. Expressing the user QoE as a function of user-content relevance and its expected delivery delay, the problem of joint caching and recommendations is expressed as a user QoE maximization problem. It is addressed by a framework that solves sequentially the subproblems of minimum delay content delivery, content caching and cache-aware recommendations, ensuring that each user will be recommended of highly preferred content with minimum delivery delay.

1.2 Outline

The dissertation is organized as follows:

Chapter 2 reviews the fundamental concepts and structures that emerge in the study and analysis of Complex Networks (CNs) and will be useful in the context of this thesis.

Chapter 3 describes the notion of monitoring in physical Networks and presents a topology control approach to address the problem of Minimum Variable Radii Sensor Cover in planar obstructed regions.

Chapter 4 introduces techniques of monitoring, information tracking and content allocation in cyber networks. In particular, an approach of Information Diffusion inference by monitoring the users' interactions in Online Social Networks and a scheme of Information-Diffusion-Aware Recommendations to the users of an OSN are analytically presented.

Chapter 5 focuses on methodologies of content placement in physical and cyber-physical Networks. By acknowledging the impact of recommendations in users' content requests, the problem of content placement at heterogeneous caching networks and content sharing via Device-to-Device (D2D) communication is investigated under various objectives.

Chapter 6 concludes this dissertation with an overview of the study and a summary of the research results. It also discusses perspectives of information dissemination and content allocation in physical, cyber and cyber-physical networks for future research.

Chapter 2

Theoretical background on complex networks

This Chapter initially provides a background on Graph Theory, which is the main mathematical tool to describe the fundamentals of network structure and dynamics. It then reviews the most prominent evaluation metrics of Complex and Social Network Analysis as well as the most characteristic types of emerging network structures along with the models of their development and formation, which will be used to represent the complex networks under examination.

Definition 1. *(Complex networks) A Complex network is one that exhibits emergent behaviors that cannot be predicted a priori from known properties of the network's constituents.*

The above definition found in [23], focuses on the characterization of networks as complex in terms of the observed behaviors of its constituents, which may be diverse within the same domain (e.g., social networks), or surprisingly similar across diverse domains (e.g., virus propagation in human networks). The field of complex networks covers a wide range of network types that vary in many aspects of their structure, operation and application scope, due to which several classifications emerge [23]. In this thesis, the complex networks are segregated into physical, cyber and cyber-physical networks. The most prominent metrics of SNA will be employed to develop and evaluate methodologies of topology control, inference of diffusion processes and content allocation to Wireless Sensor Networks, Online Social Networks and Mobile Social Networks, as typical examples of the aforementioned classes.

2.1 Graph theory fundamentals

An **undirected graph** $G(V, E)$ consists of a pair of finite and nonempty set $V = V(G)$ of $|V| = n$ points, referred to as vertices, with a set of $|E|$ unordered pairs (i, j), $i, j \in V$ of distinct points of V, referred to as edges. If an edge $e = (i, j)$ exists, we say that i and j are adjacent vertices and vertex i and edge e are incident with each other. Two distinct edges e, f incident to a common vertex are called adjacent edges.

A **directed** graph or digraph consists of directed edges, therefore the pair (i, j) is ordered, visualized by an arrow beginning from node i and pointing to node j. Contrary to the undirected graph, the edges (i, j) and (j, i) are different and the existence of one of these does not imply the existence of the other.

A graph $G' = (V', E')$ is a **subgraph** of $G = (V, E)$ and it is denoted by $G' \subset G$, if $V' \subset V$ and $E' \subset E$. If G' contains all the edges of G that join the vertices in V', then G' is called the induced subgraph and it is denoted by $G[V']$. If $V' = V$, then G' is called the spanning subgraph of G.

A graph $G = (V, E)$ (directed or indirected) is **weighted**, if a measurable quantity, denoted by w and defined as weight, is assigned to each edge in E, $w : E \to \mathbb{R}$.

The order $|G|$ of G is the number of vertices $|G| = |V(G)|$, and the size of G, denoted by $e(G)$ is the number of edges in G, i.e., $e(G) = |E(G)|$. The open **neighborhood** of a node $i \in V(G)$ denoted by $N(i)$ is the set of all vertices of G that are adjacent to i, whereas the closed neighborhood of vertex i, denoted by $N[i]$, is the set of vertices adjacent to i, including i itself. The **degree** of a vertex i, $d(i)$, in an undirected graph, is the number of edges that have as one of their endpoints the vertex i. In directed graphs, each node is characterized by two degrees, the in-degree $d^{in}(i)$, which is equal to the number of edges pointing to node i, and the out-degree $d^{out}(i)$, which is equal to the number of edges starting from node i.

The **adjacency matrix** $A = [a_{ij}]$ of a graph is defined as a $0 - 1$ element matrix, where $a_{ij} = 1$ if the edge (i, j) exists, otherwise $a_{ij} = 0$. Thus, the adjacency matrix of a graph contains all the information about its connectivity. If the graph is undirected, the adjacency matrix is symmetric. Similarly to the adjacency matrix, the weight matrix $W = [w_{ij}]$ is defined, where w_{ij} is the weight of the edge (i, j).

A **walk** of a graph G is an alternating sequence of vertices and edges, starting and ending with vertices in which each edge is incident with the two vertices preceding and following it in the given sequence. The walk $u_0, e_1, u_1, ..., u_{n-1}, e_n, u_n$ is referred to as a $u_0 - u_n$ walk and it is closed if $u_0 = u_n$, otherwise it is open. If the edges of a walk are distinct, then it is called a trail. If the vertices of a walk are distinct, therefore, its edges are also distinct, the walk is called a path. A cycle is a closed walk with distinct vertices and it is denoted by C_n. The C_3 graph is called a triangle. The length of a walk is equal to the number of the traversed edges in the network.

Covering, also referred to as **coloring**, is an important concept in Graph Theory. In the problem of inference of diffusion processes presented in section 4.2, it is required to distinguish the vertices of a graph in disjoint sets where the nodes that belong to the same set are non-neighboring. These disjoint sets may be represented by different color classes so that the color of each class is assigned to the vertices/edges of a network so that adjacent vertices/edges have different colors. The maximal number of colors in a vertex coloring is defined as chromatic number $\chi(G)$ and the minimal number of colors in an edge coloring is defined as edge-chromatic number $\chi'(G)$. An $n-$coloring of a graph uses n colors, therefore, it partitions the vertex set V into n color classes. A graph is $n-$colorable if $\chi(G) \leq n$ and $n-$chromatic if $\chi(G) = n$. For the edge-chromatic number it holds that $\chi'(G) \geq \Delta(G)$, where $\Delta(G)$ denotes the maximal network degree.

In order to study the dynamic or the stochastic behavior of networks, **random graphs** are employed. Two basic and closely related models of Random Graphs are encountered in the literature, the Erdõs-Rényi graph denoted by $G(n, M)$ or G_M and the Gilbert graph denoted by $G(n, p)$, where n is the cardinality of the set of vertices V, [23]. The space $G(n, M)$ consists of all $\binom{N}{M}$ graphs of n nodes and M edges, where $N = \binom{n}{2}$. All the graphs of this space have equal probability to be selected. Due to the assignment of a probability measure to the graphs, the graph space becomes a probability space. The probability that G_M is precisely a fixed graph H with n vertices and M edges is:

$$\mathbf{P}_M(G_M) = \binom{N}{M}^{-1}, \tag{2.1}$$

where each of the m edges of H have to be selected and none of the $n - M$ edges are allowed to be selected. The space $G(n,p)$ is defined for probability $0 \leq p \leq 1$. An element of this space is a graph of n nodes and edges that are selected independently with probability p, for all possible edges. Similarly to the $G(n,M)$ model, the probability of a fixed graph with n nodes and m edges is

$$p^m(1-p)^{N-m}. \tag{2.2}$$

The expected number of edges in $G(n,p)$ is $\binom{n}{2}p$ and by the law of large numbers, any graph in $G(n,p)$ will almost surely have approximately that many edges, provided that the expected number of edges tends to infinity. In this case, it has been proven that $G(n,M)$ and $G(n,p)$ are practically, in many cases, interchangeable with $M = \binom{n}{2}p$.

The distribution of the node degree in $G(n,p)$ is binomial (Fig. 2.2):

$$P(deg(u)=k) = \binom{n-1}{k}p^k(1-p)^{n-1-k}. \tag{2.3}$$

Since for $n \to \infty$ and np =constant it holds that

$$P(deg(u)=k) \to \frac{(np)^k e^{-np}}{k!}, \tag{2.4}$$

which is a Poisson distribution with parameter np.

A Random Geometric Graph (RGG), denoted as $G(N,r)$, is a spatial network constructed by randomly placing N nodes in a metric space, according to a specified probability distribution. Two nodes of G are connected by an edge if and only if their distance is smaller than a certain neighborhood radius, r. The Wireless Sensor Networks in Section 3.3 as well as the Mobile Social Networks in Section 5.4 are modeled as RGGs.

2.2 Complex and Social Network Analysis metrics and features

2.2.1 Degree distribution

The degree of a node represents the neighboring relations between the node itself and the nodes interacting directly with it. The node degree distribution describes cumulatively measures of direct neighboring relations among the nodes of a network. The node degree may depend on the Euclidean distance of nodes, as in the previously mentioned RGGs. In general, the neighboring relations are determined by different factors, therefore the degree distribution is derived by non-distance based metrics. The degree distribution may have a deterministic or probabilistic form according to the application domain of the networks examined. In networks where neighborhood relations do not vary, the degree distribution is the full spectrum of node degree values, whereas in networks of varying neighborhood relations where node connectivity is stochastically defined, the degree distribution $P(k)$ is the probability that a node has k neighbors. In undirected networks, a single degree distribution characterizes the whole network, whereas in directed networks, two distributions are required, one for the in-degree and one for the out-degree of the nodes, so that neighboring information is provided for both directions of the flow in the network. The degree distribution is characteristic of a network topology. For example, the degree distributions of the Online Social Networks examined in Chapter 4, are characterized by heavy tails in the plot of node degree-number of nodes, which highlights the fact that a large number of nodes have small degree.

2.2.2 Average path length

The average path length is a network-wide defined metric, as opposed to node degree, which is a node-specific metric. It is defined as the average of the shortest path lengths between all pairs of nodes in the network. The computation of the average path length is a centralized operation, since one has to obtain all possible node pairs in the network, then compute the length of the shortest path for each pair and average over all pairs. Most frequently, the considered metric space is the discrete graph space, where distance between nodes is measured in hops, thus the length of the shortest path for a pair of nodes equals to the least possible number of the edges separating them.

2.2.3 Clustering coefficient

The clustering coefficient characterizes the structure of a network both locally and globally by expressing the extent of the triadic closure process in the network, which occurs when two neighbors of a node become themselves neighbors. A high clustering coefficient implies significant participation of the triadic closure in the evolution of the network. The local clustering coefficient C_i of node i is a measure of direct connectivity between the neighbors of i:

$$C_i = \frac{\text{number of edges between the neighbors of } i}{\text{number of all possible edges between the neighbors of } i}. \tag{2.5}$$

The network local clustering coefficient is the average over all network nodes N,

$$C_{net} = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{2.6}$$

The network global clustering coefficient is

$$C_G = \frac{\sum_i \text{number of edges between the neighbors of } i}{\sum_i \text{number of all possible edges between the neighbors of } i}. \tag{2.7}$$

2.2.4 Centrality measures

The centrality metric of a node is a measure of its importance in terms of network structure, operation or applications, based on which various centrality definitions have been employed in social and communication networks. A network-wide version of centrality metric can also be defined. It characterizes the expected significance of each node in the network on average. The metrics foremost employed in this thesis are the degree, closeness and betweenness centrality.

Degree centrality quantifies the potential of a node to manage the information flow in the network through popularity and it is commonly a linear function of the value of node degree. The network degree centrality is a linear function of the average node degree. Assuming that $A = [a_{ij}]$ is the adjacency matrix of a network topology, the degree centrality of node k is

$$C_D(k) = \sum_{i=1}^{n} a_{ik}. \tag{2.8}$$

To obtain a measure that is independent of network size, the relative degree centrality is introduced as

$$C'_D(k) = \frac{\sum_{i=1}^{n} a_{ik}}{n-1}. \tag{2.9}$$

Closeness centrality, otherwise referred to as proximity or path based centrality, identifies the most spatially important nodes of a network, hence it is dependent on the considered distance metric which, in this thesis, is that of hop count. The definition of closeness centrality is based on the distance of each node from the rest of the network nodes. A node has high closeness centrality, thus it is central in terms of proximity, if it is relatively close to most of the network nodes. The distance between two nodes i, j is denoted by $d(i, j)$ and it is defined to be the length of the shortest path from i to j in a given network topology. The closeness centrality of node k is

$$C_P(k) = \frac{1}{\sum_{i=1}^{n} d(i,k)}, \tag{2.10}$$

and the relative closeness centrality of node k is

$$C'_P(k) = \frac{n-1}{\sum_{i=1}^{n} d(i,k)}. \tag{2.11}$$

Closeness centrality exhibits quantitative problems when a network is disconnected (shortest paths have infinite values) or its connectivity changes dynamically.

Betweenness centrality is based, similarly to closeness centrality, to the notion of geodesics (shortest paths) and quantifies the frequency with which a node participates in the geodesics connecting other pairs of vertices in the underlying network topology. Additionally, it does not suffer from the previously mentioned computational problem in intermittently connected networks. In order to compute the betweenness centrality of a node k, one has to compute at first its partial betweenness. Given an unordered pair of vertices $\{i, j\}$ where $i \neq j \neq k$, the partial betweenness $b_{ij}(k)$ of node k with respect to the pair (i, j) is

$$b_{ij}(k) = \frac{g_{ij}(k)}{g_{ij}}. \tag{2.12}$$

The overall betweenness centrality of vertex k is then computed by the sum of its partial betweennesses for all the unordered pairs of vertices $i \neq j \neq k$:

$$C_B(k) = \sum_{i \neq j \neq k}^{n} \sum_{i<j}^{n} b_{ij}(k), \tag{2.13}$$

where n is the number of vertices in the graph. A major drawback of this metric is the requirement of computing all the shortest paths in a given network topology for each node which is computationally demanding, especially in large scale networks. To address this issue, approximation methods are employed, some of which can be found in [23].

2.3 Distinctive structure of complex networks

The term network structure reflects all the properties of the network related to the degree of the nodes, the distance between node pairs, the connectivity, the clustering coefficient, etc. The structures of complex networks employed in this dissertation are the small-world (SW) and the scale-free (SF) networks, which belong to the class of relational graphs and are considered the most appropriate for representing the structure and evolution of Online Social Networks. Contrary to spatial graphs, where the network is embedded in a metric space and the connected node pairs are selected according to a predefined distance threshold, relational graphs represent systems of interactions where any two nodes may possibly become neighbors.

2.3.1 Small-world networks

A small-world network refers to a growing graph whose average path length increases proportionally to the logarithm of the number of network nodes. It lies between a regular graph (i.e., a graph where each node connects only to close neighbors in a specific manner) and a random graph which results in exhibiting high clustering coefficient and short average path length. In terms of its corresponding random graph, a small-world graph can be defined as follows:

Definition 2. *(small-world graphs) A small-world graph is a graph with n-vertices and average degree k that exhibits $L \approx L_{random}(n,k)$, but $\gamma \gg \gamma_{random} \approx \frac{k}{n}$, where L_{random} and γ_{random} are the average path length and the clustering coefficient correspondingly of the random graph with n vertices and average degree k.*

The mathematical model employed for the formation of small-world graphs in this thesis is the one of Watts and Strogatz, otherwise referred to as β-model, which starts from the ordered structure of a ring lattice with each node having k neighbors and rewires edges randomly, denoted as shortcuts, with increasing probability β up to the point that a random graph's topology is achieved. Parameter β controls the degree of randomness of the graph since it dictates the transformation of the initial regular lattice to a random graph with asymptotically known properties. Each value of β corresponds to a different type of graph structure, from totally ordered graphs to small-world graphs and finally to totally random graphs, providing a graph structure continuum from regularity to randomness.

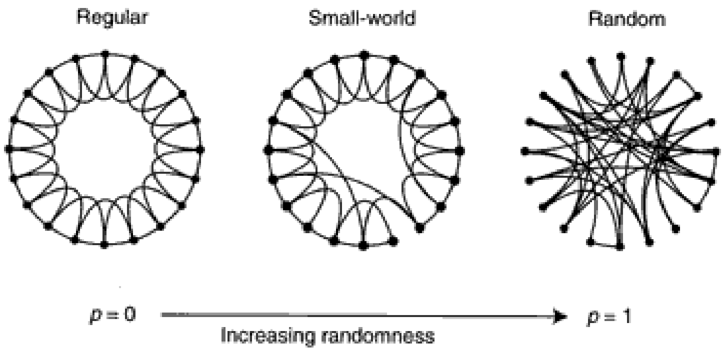


Figure 2.1: The β-model of Watts-Strogatz [1].

2.3.2 Scale-free networks

The concept of "scale-free" captures the lack of scale in the degree distribution of complex networks with different node groups exhibiting differences in scaling of their node degree. The degree distribution of random networks follows a Poisson distribution, as presented in Fig. 2.2, which means that the probability of a node to have k connections decreases exponentially for large k, therefore it is extremely rare to find nodes having significantly more or fewer links than the average. However, as confirmed by small-world networks, complex networks present more complicated features regarding structure due to the way their nodes are interconnected as well as their interactions as the network evolves. The formation of new connections is dependent to the existing ones which results in imbalanced degree of the network nodes.

The typical evolution of complex networks is based on two mechanisms not considered by the random graph model. The first one is growth which indicates that the newly added network nodes tend to make links with the existing ones. The second mechanism is known as preferential attachment and it describes the tendency of newly added nodes to connect to existing ones with probability proportional to the popularity of the latter. This implies that the probability of a new node linking to existing ones is not uniform but it is higher for the nodes displaying larger connectivity or degree, creating two extreme groups, one of few high degree nodes referred to as hubs, and one including the rest of the nodes which exhibit low degree (Fig. 2.2). Preferential attachment is mostly considered to be linear with respect to node degree. In scale-free networks, the mechanisms of growth and preferential attachment are reflected by the power-law distribution, namely, the probability that a vertex connects with k other vertices is following a model such as $P(k) = k^{\gamma}$ with the exponent $\gamma \in [2.1, 4]$ as shown in real world experiments [24].

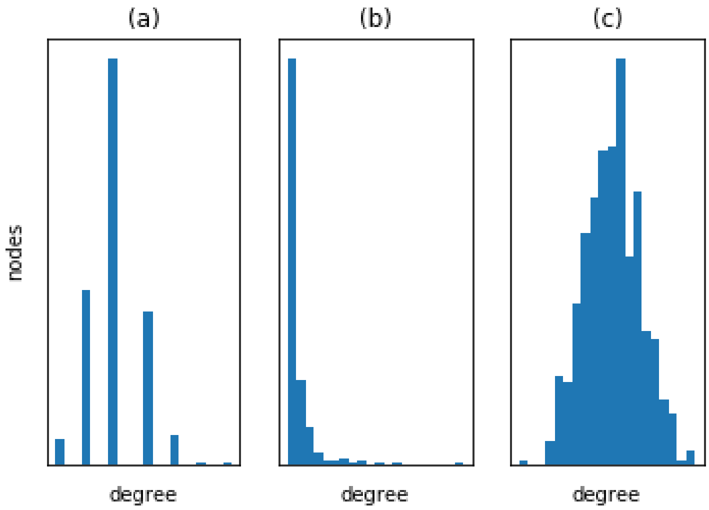


Figure 2.2: Examples of degree distributions for (a) small-world (heavy tailed), (b) scale-free (power-law) and (c) Random networks (Poisson). Topologies of 100 nodes are generated using the models of (a) Watts-Strogatz with $\beta = 0.2$, (b) Barabasi-Albert with $k = 4$ and (c) Erdos-Renyi with $M = 700$.

The most popular model of networks' evolution dynamics in terms of network elements addition/deletion is the one of Barabasi and Albert which, contrary to the small-world model of Watts and Strogatz, does not enhance an existing topology with particular features but constructs one by combining the mechanisms of growth and preferential attachment. These two principles are shown by continuum theory to lead to power-law degree distributions [23].

2.3.2.1 Preferential attachment

The linear preferential attachment is defined with respect to parameter x_i, which in this thesis represents the node degree. It is the procedure of choosing a node with probability proportional to the value of x_i expressed, in the general case, as follows:

$$\Pi(x_i) = \frac{f(x_i)}{\sum_j f(x_j)}, \tag{2.14}$$

where $f(x_i)$ is an increasing function of the quantity x_i. The sum at the denominator spans all network nodes.

2.3.2.2 Barabasi-Albert algorithm (BA)

In the BA model, the time is considered slotted. The initiating network consists of m_0 nodes. At time slot t, each newcomer node is linked to the network through m connections with m existing nodes (injective link-node mapping) selected according to the preferential attachment rule. At the end of time slot t, the network size is equal to $N_t = m_0 + t$.

Chapter 3

Monitoring in physical networks

3.1 Wireless Sensor Networks

The growth in micro-electro-mechanical systems technology and wireless communications has facilitated the design of low-cost, low-power, small-sized inter-connected devices, which form a Wireless Sensor Network. A large number of sensor nodes are often densely deployed in a deterministic or random manner and are organized into a multi-hop wireless network. Each sensor may consist of sensing, processing, storage and communication units. The sensing unit of each sensor node acquires information from the physical surroundings such as temperature, pressure, etc., which is stored in its storage component. The sensor's processing unit realizes simple computations on the sensed data, while the communication unit transmits the processed data to intermediate nodes and controls the packets across the network. Wireless Sensor Networks are developed in a wide range of application fields such as healthcare, home automation, environment, security and transportation [25]. The basic functionality of a WSN is to monitor the region in which its sensors are deployed, referred to as Region of Interest (RoI), generate the data corresponding to the monitored events and find a path for the data to reach the designated data sink. Coverage and Connectivity are considered important evaluation metrics of Quality-of-Service (QoS) in the WSNs, which reflect how well the network monitors a Region of Interest and whether the detected/aggregated information can be forwarded successfully.

Due to the sensors' resource constraints (limited battery capacity), nodes are prone to failures, which in turn result in frequent changes in the network topology. The design and analysis of efficient coverage schemes, where connectivity is guaranteed, energy consumption is minimized and network lifetime is prolonged, are currently very important research directions [26].

3.1.1 Fundamental concepts of WSNs

Deployment strategy. Sensor deployment can be deterministic or random. In deterministic deployment, each sensor is placed at predetermined coordinates in order for the network to achieve a high level of target monitoring, low energy consumption and prolonged lifetime. In harsh environments, such as disaster regions, where deterministic deployment may be infeasible, sensor nodes are distributed within the field stochastically (e.g., air-dropped). This may result in a partition of the RoI to densely and sparsely monitored regions, which in turn leads to lower performance compared to the deterministically deployed network. The RoI may be convex or non-convex (corresponding to the existence of obstacles).

WSN types. A WSN is characterized as homogeneous or heterogeneous based on its sensor types. In a homogeneous WSN, all sensor nodes are identical in terms of battery power and hardware complexity (sensing range, communication range, processing capability). A heterogeneous WSN consists of two or more different types of nodes with different battery power and functionality, where usually, the set of more powerful nodes, known as cluster heads, are employed to receive and process data from less powerful nodes.

Sensing model. The simplest model is the **binary disc sensing model** [27] according to which a sensor is able to monitor events or sense the environment only within its sensing range. In this model, the sensing range of a node is determined within a circular disc of radius r_s, referred to as sensing radius. Assuming that sensor s_i is deployed at point (x_i, y_i), the probability that point $p = (x, y)$ is covered by s_i equals to 1, if $d(s_i, p) < r_s$, where $d(s_i, p)$ is the Euclidean distance between sensor s_i and point p,

$$\mathbb{P}_p(s_i) = \begin{cases} 1, & d(s_i, p) < r_s, \\ 0, & otherwise. \end{cases} \tag{3.1}$$

In the **probabilistic sensing model** [27], the probability of an event at point p to be detected by sensor s_i decreases as the Euclidean distance between them increases. This is formally expressed as follows:

$$\mathbb{P}_p(s_i) = \begin{cases} 0, & r_s + r_e \leq d(s_i, p), \\ e^{-\lambda a^{\beta}}, & r_s - r_e < d(s_i, p) < r_s + r_e, \\ 1, & r_s - r_e \geq d(s_i, p). \end{cases} \tag{3.2}$$

where $a = d(s_i, p) - (r_s - r_e)$, $r_e < r_s$ is a measure of the uncertainty in sensor monitoring, while λ and β are parameters that measure the monitoring probability when a target is at location $[r_s - r_e, r_s + r_e]$.

Communication model. In the **binary disk communication model** [28], which is the one employed in this thesis, a node s_i is able to communicate only up to a certain threshold distance from itself, which is referred to as communication radius and is denoted by $r_c(s_i)$. Sensors can have different communication ranges depending on their transmission power levels. Sensors s_1, s_2 are able to communicate with each other if the Euclidean distance between them is less than or equal to the minimum of their communication radii, thus $d(s_i, s_j) \leq \min\{r_c(s_i), r_c(s_j)\}$. More complex communication models that capture physical characteristics of a wireless channel (e.g., multi-path fading, interference, etc.) or the deployment environment are presented in [29]. The connectivity of a WSN requires that any active node is within the communication range of one or more active nodes, such that all active nodes form a communication backbone.

WSN model. A graph theoretic model is typically employed to represent WSNs. In particular, the structure of Random Geometric Graphs resembles the topological structure of randomly deployed sensor networks [30]. A WSN is modeled by a Random Geometric Graph $G(V, E, r)$, where V is the set of sensors and r is a predetermined threshold distance between the nodes that determines the set of network edges E. The induced communication graph of G is denoted by $G_c = (V, E_c, \mathbf{r_c})$, where $\mathbf{r_c}$ is the vector of sensors' communication radii for which it holds that $\mathbf{r_c}(i) \leq r$, $\forall i \in V$ and $E_c \subseteq E$ is the set of edges such that an edge exists between any two nodes if their Euclidean distance is less than the communication radius r_c. The network formed by the induced communication graph is said to be connected if every pair of nodes is connected by a path in G_c.

Sensor mobility. In a stationary sensor network the location of a node is determined at the initial configuration and remains fixed over time after deployment. Mobile nodes may change their position depending on the objective of the WSN. Mobile nodes have all the features of fixed nodes (e.g., collect and process data, transmit and receive messages), but it is the mobility feature that enhances network performance by addressing coverage issues (e.g., densify the network in sparsely monitored regions) and connectivity maintenance that may arise after deployment.

Coverage schemes. According to the type of targets monitored by the WSN, coverage can be classified into *area*, *target* and *barrier* coverage. Area or blanket coverage refers to the monitoring of an entire region. Either full or partial coverage of the region is required according to the nature of the application. In target coverage, the objective is to surveil a fixed number of targets, referred to as Points of Interest (PoI), within a given region. The notion of barrier coverage concerns the detection of penetration through the region of interest, also referred to as intruder detection. Contrary to blanket coverage, target and barrier coverage do not require the whole RoI to be covered by sensors, thus, the cost of sensor deployment is significantly reduced.

Classification of coverage algorithms. Most of the available works on coverage in WSNs can be segregated in centralized and distributed. Centralized approaches require that a single node collects information transmitted by all nodes of the network, which results in high communication cost. Distributed algorithms allow each node to make independent decisions based on the information exchanged with its one-hop neighborhood. Even though distributed approaches exhibit an increase in the overall processing cost, communication cost is reduced compared to the centralized approaches.

3.2 Coverage problems in Wireless Sensor Networks

Recently, research works aim to optimize either individually or jointly the number of active nodes, coverage, energy efficiency, communication overhead, fault tolerance, scalability and lifetime of WSNs. Challenges in area coverage and connectivity are presented from the perspective of energy consumption and focus on optimal deployment and topology control. Coverage is investigated in conjunction to connectivity, since coverage without connectivity provides no guarantee that the data will arrive at the designated sink, while connectivity without coverage results in unmonitored regions, referred to as coverage holes, in the Region of Interest. In [31] a sufficient condition for a covered network is provided to imply connectivity in convex regions. It is proved that if a convex region is completely covered by a set of nodes with sensing radius r_s and communication radius r_c, the induced communication graph is connected when $r_c \geq 2 \cdot r_s$. In other words, if $r_c \geq 2 \cdot r_s$, the configuration that guarantees coverage will also ensure connectivity for the network.

3.2.1 Deployment approaches

Energy conservation in WSNs is closely related to the design of deployment schemes, which compute the minimum number of nodes to be placed in the RoI while ensuring coverage and connectivity.

Virtual Force based approaches. When random deployment is applied, sensors may be concentrated to certain parts of the RoI resulting in coverage holes. Virtual Force Algorithm (VFA) is a mobile sensor redeployment strategy, which enhances coverage after the initial configuration. The VFA method was inspired by the potential field theory used to avoid obstacles in mobile robot movement [32]. Sensors, a priori known obstacles and preferential areas in the RoI are modeled as attractive or repulsive virtual points. According to the desired distances among sensors, a threshold value determines the repulsive and attractive forces to which each sensor is subjected, e.g., when two nodes are closer than the predetermined threshold, the force is in repulsive pattern, intending to separate them. Obstacles and preferential areas also exert forces to sensors. The movement of each sensor is determined by the summation of the force vectors. VFA is a centralized approach in which the sink node calculates the new positions and the direction of movement for each sensor node. A distributed VFA-based algorithm is presented in [33] whose objective is, based on the coordinate's information of the nodes and their neighbors, to redeploy the sensors to ensure coverage and connectivity, while minimizing the redeployment energy cost per node in terms of travelled distance. The VFA concept is also extended in [34] to consider the connectivity maintenance of a network with arbitrary sensor communication/sensing ranges or node densities where there is no knowledge of the field layout. The goal is to maximize sensing coverage and guarantee connectivity, in a distributed manner, at the cost of a small moving distance. Even though centralized schemes are computationally demanding for the sink node, distributed schemes consume more energy per node, which results in shorter network lifetime.

Computational geometry based approaches. Computational geometry is a research area focused on the systematic study of algorithms and data structures for geometric objects, such as points, line segments and polygons, aiming to exact algorithms that are asymptotically fast, [35]. Voronoi diagram and its dual, Delaunay triangulation, are the most popular computational geometry methods used in WSNs. The Voronoi diagram is a method of partitioning the Region of Interest into a number of polygonal regions based on the distance between sensor nodes. Each node occupies only one polygon, where all the interior points are closer to it than to any other sensor. Delaunay triangulation is constructed by connecting every two adjacent points of the Voronoi diagram whose polygons share a common edge. The vector-based (VEC), the Voronoi-based (VOR) and the Minimax algorithm address coverage hole problems in a distributed manner [36]. VEC is inspired by electromagnetic particles and it removes sensor nodes from a densely covered area by exerting a repulsive force between them, namely, a virtual force pushes the sensors away from each other if there is a coverage hole in either of their Voronoi Polygons, while VOR pulls sensor nodes to the sparsely monitored area. Similarly to VOR, but with a lower limit on the maximum movement of nodes, the Minimax algorithm reduces coverage holes by pulling sensor nodes closer to coverage gaps.

Grid based approaches. Deterministic placement of sensors is provided by grid-based approaches [37]. Nodes' positions are fixed according to a regular grid pattern such as a triangular, square or hexagonal grid. The monitoring area is divided into small grids and the location of sensor nodes is specified by the employed algorithm to be in the grid center or grid vertices. Within the scope of fire detection, a grid-based approach that aims to prolong network lifetime and achieve full area coverage by minimizing the utilization of sensors, is discussed in [38]. Grid deployment is compared to triangular and strip deployment and it is shown to outperform the former when the number of nodes is considered a critical issue. A virtual square grid-based coverage algorithm for WSN is introduced in [39], where each sensor node divides its sensing range into virtual square grids. If all the virtual grids of a sensor are covered by neighboring nodes, the sensor under examination is considered redundant and sleep scheduling is applied, leading to low energy consumption while at the same time the active nodes guarantee coverage and connectivity in the whole network.

3.2.2 Topology Control approaches

In randomly deployed static WSNs, sensors are usually placed densely, thus, significant energy savings may be achieved by exploiting node redundancy in the network. This is realized by either adjusting the sensing range of the wireless nodes in order to reduce coverage overlaps or by dynamically managing the node duty cycles, with some nodes being scheduled to enter a power saving mode, while the remaining active nodes provide continuous monitoring of the RoI. Both approaches fall under the scope of Topology Control [40]. Therefore, one has to compute the minimum number of active nodes and also, in the case of varying sensing radii, determine the corresponding sensing ranges that ensure sufficient coverage and connectivity with a minimum total energy cost. This problem, which is referred to as **variable radii connected sensor cover problem (VRCSC)**, is studied for WSNs deployed in regular regions, that is, regions that can be described by a bottom left corner and a top right corner, in [41, 42] with methods based on Computational Geometry and in [43] with a multi-objective genetic algorithm approach. A greedy centralized and a distributed scheme, using the concepts of Voronoi diagram and Relative Neighbor Graph, are proposed in [41] for the assignment of sensing and transmission radii and for the sleep scheduling of sensors such that coverage and connectivity are ensured. The presented algorithms are designed to take into consideration non-circular sensing and transmission ranges caused by noise properties and irregularity in radio propagation respectively. Furthermore, in [42], a distributed algorithm with one hop approximation of Delaunay triangulation is developed for sensing radii assignments with minimal energy consumption and energy balancing. A common assumption is that the deployment area is free of obstacles. However, in practical scenarios, obstacles can prevent transmission and/or coverage. Such physical limitations, if not properly considered, may introduce several burdens in the development, provisioning and effectiveness of various services required for the realization of a monitored environment [44, 45, 46, 47]. In the following section, the problem of the minimum variable radii connected sensor cover is analyzed in a non convex region where the lack of convexity is due to the presence of arbitrary shaped, convex opaque obstacles. The impact of obstacles on the coverage and the connectivity of a WSN is studied with heuristic methods, which avoid the high computational cost of constrained Voronoi tessellation and Delaunay triangulation.

3.3 The problem of Minimum Variable Radii Sensor Cover in obstructed network regions

In this thesis, the problem of minimum variable radii connected sensor cover, which is described in section 3.2.2, becomes more complex as it is studied in irregular (i.e., obstructed) regions. Due to the presence of obstacles, connectivity cannot be ensured as in [31]. These environments represent typical cases of realistic WSNs and Internet of Things (IoT) deployments. Assuming the devices have the capability to modify their sensing ranges, the area covered by randomly dispersed sensors has to be maximized, while reducing the sensing energy consumption as much as possible despite the presence of convex obstacles. To address this problem, a framework capitalizing on the notion of the visibility polygon is introduced and two algorithms are proposed, a centralized (and a randomized version thereof) and a distributed one. The algorithms aim to maximize the ratio of covered area to consumed energy, while ensuring a minimum coverage percentage. The presented schemes solve a variation of the original problem of VRCSC, which is referred to as the minimum variable radii partial sensor cover, outperforming the plain assignment of maximum sensing range across the network.

3.3.1 WSN model

A WSN is deployed over the region of interest, denoted as F, which without loss of generality is assumed to be a square of side L_0, i.e., $F = [0, L_0]^2 \subset \mathbb{R}^2$. F contains arbitrary convex opaque obstacles and sensor nodes. The set of obstacles $\mathbf{O} = \{O_i\}$ is modeled as a set of non-overlapping, convex polygons in region F, as shown in Fig. 3.1. Both the size and the number of obstacles are predetermined so that $\mathbf{O}$ cannot be a partition of F. The term *coverable region* F_c will be used for the region F without the obstacles, which has an irregular shape. The objective is to minimize the sensing energy while covering F_c up to a given percentage of the maximally covered area of the sensor network, i.e., the area covered when all deployed sensors employ the maximum sensing range, which is referred to as *feasible coverable region* F'_c. F'_c is determined by the surrounding obstacles, the node locations and the boundary of the coverable region. It should be noted that in principle, covering all F_c might not be possible, i.e., the coverage of F_c' may not correspond to the coverage of F_c, due to the random deployment of sensor positions.

The set of sensors $S = \{s_1, ..., s_n\}$ is modeled by a set of points $\mathbf{P} = \{p_1, p_2, ..., p_n\}$ with $P \subset F_c$. Since the random deployment scheme is a commonly used assumption for WSNs [48], sensors are distributed within the coverable region stochastically and independently according to a Binomial Point Process [49]. The binary disk model is used for the calculation of coverage [48], in which a point in F_c' is covered by a sensor s_i, if and only if it lies within the disk of center s_i and radius r_i and there is line-of-sight from s_i. The presence of opaque obstacles blocks line-of-sight in certain subsets of the region, leading to non-convex sensing patterns.

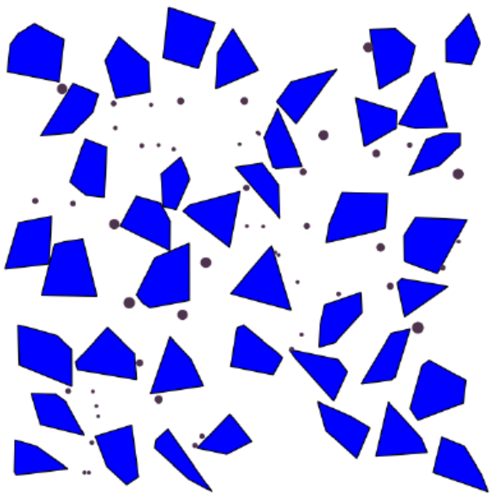


Figure 3.1: Illustration of the obstructed region of interest with the randomly deployed sensors. The sensors are represented by the black dots of different size, which model the inhomogeneity in the network. The obstacles are modeled by the blue convex shapes.

Once dispersed, sensors are assumed to remain static with adjustable homogeneous sensing range $r \in [0, r_{max}]$ (topology control enabled). Their transmission range r_t remains fixed and it should be properly determined to ensure network connectivity. Furthermore, r_t is associated with the maximum sensing range. In this problem, since the presence of obstacles results in a partial coverage of a non-convex area, the condition of $r_t \geq 2r_{\max}$ proposed in [50] is adopted, where r_{max} is the maximum sensing range. The evaluation results in section 3.3.5 indicate that setting the transmission range to $r_t = 2.5r_{\max}$ in highly obstructed networks with more than 150 nodes results in the majority of sensors ($> 95\%$) being connected.

Considering constant transmission range, thus constant communication cost, it is assumed that energy consumption increases with higher sensing radius r and it is proportional to a power of r [41]. The quadratic model that has been extensively used in the literature, is adopted [51]. For sensor s_i, the energy consumption is defined as $c_i \cdot r_i^2$, where c_i is a sensing constant, and **the total energy consumption** is the sum of individual consumptions [43], defined as

$$Energy(\mathbf{r}) = \sum_i c_i r_i^2. \tag{3.3}$$

A power saving technique is proposed through the adjustment of sensing ranges

$$\mathbf{r} = (r_1, r_2, .., .r_n) \in \mathcal{D},$$

where $\mathcal{D} = [0, r_{\max}]^n \subset \mathbb{R}^n$ is the space of valid sensing radii assignments. The proposed technique reduces coverage overlaps, while maintaining a predefined coverage goal given by a threshold value.

3.3.2 Analysis of the Minimum Variable Radii Partial Sensor Cover problem

3.3.2.1 Fundamental concepts

The *Visibility Polygon* [52] of a node i, depicted in an example in Fig. 3.2a, is denoted as $\mathcal{VP}_i$ and corresponds to the set of points $v \in F$ such that the line (v, p_i) does not intersect any obstacle. It contains all the points that are in line-of-sight from node i.

The *Field of Vision* of node i, shown in Fig. 3.2b, is defined as

$$\mathcal{FOV}_i := \mathcal{VP}_i \cap DISK(p_i, r_{max}), \tag{3.4}$$

and describes the maximal set of points a node covers when it is set to its maximum sensing radius $r_{\max}$.

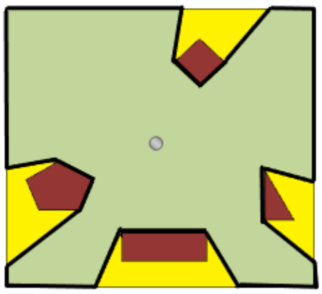
The *Current Field of Vision* of sensor i with sensing radius r_i (recall that nodes have variable radius) is

$$\mathcal{CFOV}_i(r_i) := \mathcal{FOV}_i \cap DISK(p_i, r_i). \tag{3.5}$$

The *Reduced Field of Vision* of a node i, which depends on the current assignment $\mathbf{r}$, describes the field of vision covered solely by node i, and is defined as

$$\mathcal{RFOV}_i(\mathbf{r}) := \mathcal{FOV}_i \setminus \bigcup_{i \neq j} \mathcal{CFOV}_j(r_j). \tag{3.6}$$

The *Total Coverage* $\mathcal{TC}(\mathbf{r}) := \bigcup_i \mathcal{CFOV}_i(r_i)$ represents the set of all the points covered by at least one node for a particular radii assignment. The *area covered in assignment* $\mathbf{r}$ is defined as $Covered(\mathbf{r}) := m(\mathcal{TC}(\mathbf{r}))$, where $m(\cdot)$ is the Lebesque measure of a set [53].



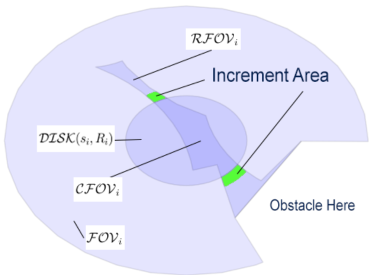


Figure 3.2: Illustration of the basic structures. (a) Visibility Polygon of a sensor (bold boundary outline), (b) Current Field of Vision, Reduced Field of Vision and obstacle area (the latter depicted as white space).

3.3.2.2 Maximizing the coverage-to-energy ratio

To solve the problem of finding an assignment of sensing radii, which has high area coverage and low energy consumption, formulations and solutions that maximize the ratio of *covered area* to *consumed energy* are formulated. Specifically, the initial objective is defined as finding the assignment, which yields the maximum ratio of the *Covered* area to the *Energy* consumption in a particular assignment $\mathbf{r}$, requiring that a specified coverage threshold T is achieved. Formally:

$$\arg\max_{\mathbf{r} \in \mathcal{S} \subseteq \mathcal{D}} \left\{ \frac{Covered(\mathbf{r})}{Energy(\mathbf{r})} \right\}, \tag{3.7}$$

where the feasible set $\mathcal{S}$ is defined as follows:

$$\mathcal{S} = \left\{ \mathbf{r} \in \mathcal{D} \;\middle|\; \frac{Covered(\mathbf{r})}{Covered(\mathbf{r}_{\max})} \geq T \right\}. \tag{3.8}$$

Maximizing this ratio is not trivial. The optimization problem (3.7) is non-convex and *Covered* is a function determined by the obstacles and the sensors. Therefore, in the following, a modified formulation is presented, which subsequently is shown to be equivalent with the original one within a specific area of interest, i.e., set of assignments.

3.3.2.3 Minimizing the energy

The alternative approach focuses on the minimization of the energy of the radii assignment with the coverage requirements of the Expression (3.8), considering the feasible set $\mathcal{S}$. This leads to the following formulation:

$$\underset{\mathbf{r} \in \mathcal{S} \subseteq \mathcal{D}}{\arg\min} \left\{ Energy(\mathbf{r}) \right\}. \tag{3.9}$$

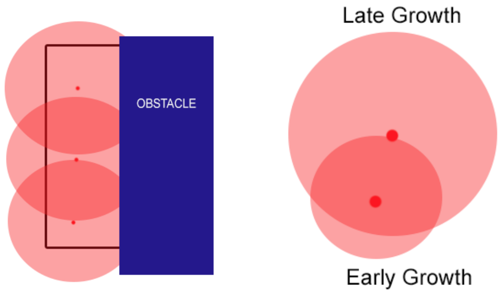
The energy consumption function, *Energy*, is monotonically increasing. That is, for the component-wise inequality of $\mathbf{r}_1, \mathbf{r}_2$ denoted as $\mathbf{r}_1 \preceq \mathbf{r}_2$, $\mathbf{r}_1 \neq \mathbf{r}_2$, $Energy(\mathbf{r}_1) < Energy(\mathbf{r}_2)$. Specifically, consider a point $\mathbf{r}$ in the interior $int\mathcal{S}$. Such point has the property that there is an open n-ball, centered at $\mathbf{r}$ that is fully contained in the set $\mathcal{S}$. It suffices to take a point $\mathbf{r}' = \mathbf{r} - \epsilon$, with $\epsilon \succeq \mathbf{0}$, thus $\mathbf{r}' \preceq \mathbf{r} \Rightarrow Energy(\mathbf{r}') < Energy(\mathbf{r})$.

Therefore, the minimum cannot be achieved in the interior $int\mathcal{S}$, and the search can be restricted in a subset of the boundary $\partial\mathcal{S}$, denoted by $\hat{S}$, which is called the frontier and it is defined as $\hat{S} := cl(\partial\mathcal{S} \setminus \partial\mathcal{D}) \subseteq \partial\mathcal{S}$, where $cl(\cdot)$ denotes the closure of a set [53].

The set $\hat{S}$ is related to the set $\tilde{\mathcal{S}}$ of the assignments that achieve T% of the maximum possible coverage (T-assignments):

$$\tilde{\mathcal{S}} = \left\{ \mathbf{r} \in \mathcal{D} \;\middle|\; \frac{Covered(\mathbf{r})}{Covered(\mathbf{r}_{\max})} = T \right\}. \tag{3.10}$$

In the case of partial coverage ($T < 1$), $\hat{S} = \tilde{\mathcal{S}}$ since for every T-assignment there is always at least one sensor that will result in greater coverage by increasing its sensing radius. In the case of maximum coverage ($T = 1$), it holds that $\hat{S} \subseteq \tilde{\mathcal{S}}$. This is because the maximum coverage can be achieved with $\mathbf{r} < \mathbf{r}_{\max}$, i.e., $\mathbf{r}_{\max} \in \tilde{\mathcal{S}}$ and $\mathbf{r}_{\max} \notin \hat{S}$. This is clarified in Fig. 3.3a, where the overlaps in sensors' maximum sensing range indicate that complete coverage of the rectangle area (adjacent to the blue obstacle) can be obtained by sensing radii smaller than the one illustrated.



(a) (b)

Figure 3.3: (a) An example of maximum coverage achieved by $\mathbf{r} < \mathbf{r}_{max}$, (b) An example of the overlap of two sensing disks, where radius decrease is performed by the Greedy Centralized Algorithm.

3.3.2.4 Relationship between the two formulations

For a point $\mathbf{r}$ in the constrained set $\tilde{\mathcal{S}}$, it holds that $Covered(\mathbf{r}) = T \cdot Covered(\mathbf{r}_{\max})$. The right hand side expression is constant, and as a result, $Covered$ is constant in that set. This leads to:

$$\begin{aligned} \arg\max_{\mathbf{r}\in\tilde{\mathcal{S}}} \left\{ \frac{Covered(\mathbf{r})}{Energy(\mathbf{r})} \right\} &= \arg\max_{\mathbf{r}\in\tilde{\mathcal{S}}} \left\{ \frac{(Constant)}{Energy(\mathbf{r})} \right\} \\ &= \arg\min_{\mathbf{r}\in\tilde{\mathcal{S}}} \left\{ Energy(\mathbf{r}) \right\}, \end{aligned} \tag{3.11}$$

which shows that the two problems are equivalent in the set $\tilde{\mathcal{S}}$. In the following sections, two heuristic methods are proposed to address both (3.7) and (3.9).

3.3.3 Greedy centralized algorithm

To solve the above problem, a greedy centralized algorithm is presented (Fig.3.5) that iterates until the required percentage of coverage is achieved, or once exceeded. At iteration t, sensor s_i is chosen to increase its radius by Δr, if the latter has the largest ratio of increased area per increased energy defined as:

$$R(i,t) = \frac{m(\mathcal{RFOV}_i(\mathbf{r}') \setminus \mathcal{RFOV}_i(\mathbf{r}))}{Energy(\mathbf{r}') - Energy(\mathbf{r})}, \tag{3.12}$$

where $\mathbf{r}' = \mathbf{r} + \mathbf{e}_i \min\{r_i + \Delta r, r_{\max}\}$ considering nodes for which $\mathbf{r}' \neq \mathbf{r}$. The element e_i is the i^{th} vector of the standard basis of the n-dimensional Euclidean space. If nodes s_i, s_j at iteration t have equal ratios $R(i,t) = R(j,t) = R_{\max}$, the algorithm selects one of them arbitrarily ($R(i,t)$ corresponds to $incrementRatio(i)$).

It is noted that nodes are selected based on the value of the uncovered area per additional energy requirements ratio in the ring defined by r_i and $r_i + \Delta r$. If the $\mathcal{FOV}$ of a node is partitioned in several such rings, each ring can be assigned a quality measure based on the ratio in the corresponding region. This is visualized in Fig. 3.4. The quality of each ring in $\mathcal{FOV}_i$ is constant, whereas the quality of the rings in $\mathcal{RFOV}_i$ varies with the development of the algorithm's assignment. Sensors with rings that initially delivered a good ratio of area per energy, can deteriorate over iterations that include redundant nodes or, more frequently, nodes with already unfavorable ratios. To mitigate this, the radius of the node is stochastically reduced with the least ratio of decreased area per decreased energy over a change of Δr at each iteration (see Fig. 3.3b).

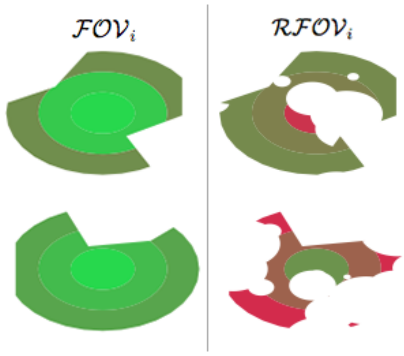


Figure 3.4: Two instances of $\mathcal{FOV}_i$ (left) and $\mathcal{RFOV}_i$ (right) partitioned in 3 rings. Green color indicates a large ratio of area per energy in the region defined by the corresponding ring, whereas Red color indicates a small ratio. White spaces are due to obstacles and the operation of neighboring sensors.

Greedy Centralized Algorithm
Input: Set of obstacles $\mathbf{O}$, set of sensor points $\mathbf{P}$, Coverage threshold T.
Output: Radii assignment $\mathbf{r}$.

Initialization: $\mathbf{r} = \mathbf{0}$.
function GreedyAssignment $(\mathbf{O}, \mathbf{P}, T)$
while $CoverageFraction(\mathbf{r}) < T$ **do**
  S $= \arg\max_{i \in \mathbb{N}_n} \{incrementRatio(i)\}$
  i $= selectRandom(S, 1)$
  $r_i = \min\{r_i + \Delta r, r_{max}\}$
  if BernoulliTrial(p) **is** successful **then**
    S $= \arg\min_{i \in \mathbb{N}_n} \{decrementRatio(i)\}$
    i $= selectRandom(S, 1)$
    $r_i = \max\{r_i - \Delta r, 0\}$
  end if
end while

Figure 3.5: Greedy centralized algorithm for the variable radii connected sensor cover problem.

3.3.4 Greedy distributed algorithm

Furthermore, a distributed scheme is proposed, where each node acts in an independent manner, modifying its radius, until the whole system converges to a sensing radii assignment.

3.3.4.1 Dependency graph

The graph of dependencies encapsulates the inter-dependence of nodes for the cooperative coverage of the coverable region F'_c. The dependency graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, is the undirected intersection graph of the sets $\{\mathcal{FOV}_i\}$, that is, a graph in which

$$\{v_i, v_j\} \in \mathcal{E} \Leftrightarrow FOV_i \cap \mathcal{FOV}_j \neq \emptyset.$$

Thus, in this graph, an edge indicates that the two adjacent vertices cover common parts of the region of interest, and as such, they will have to communicate with each other to decide locally for their sensing radius increase.

3.3.4.2 Local convergence criterion

An emerging issue with the distributed algorithm is to determine a criterion for a node to stop modifying its radius so that a sufficient overall coverage is achieved. For this, an extension of the general result for full coverage found in [54] is implemented. At each iteration, a node will compute the percentage of its *restricted sensing perimeter*, defined as the length of its sensing circumference $\partial(\mathcal{CFOV}_i)$, which is restricted by each neighboring sensor's $\mathcal{CFOV}_j$, its k surrounding obstacles $\mathcal{O}_k$ and the boundary of the region of interest:

$$\mathcal{PCOV}_i := \frac{length(\mathcal{CCIRC}_i)}{length\Big(\partial(\mathcal{CFOV}_i)\Big)}, \tag{3.13}$$

where

$$\mathcal{CCIRC}_i = \partial(\mathcal{CFOV}_i) \cap \left(\Big(\bigcup_{j \neq i} \mathcal{CFOV}_j \Big) \cup \Big(\bigcup_{k} O_k \Big) \cup \mathcal{F}^c \right). \tag{3.14}$$

If the percentage of its *restricted sensing perimeter* is larger than or equal to a predetermined limit ℓ, the node will stop altering its radius. This criterion results in varying percentages of coverage for a constant ℓ, but according to the results in [54] the variance of the finally achieved coverage percentage is expected to decrease as the network size increases.

The algorithm iterates until all nodes satisfy the local convergence criterion. At each iteration, nodes that fail the local convergence criterion, evaluate their neighborhood in the dependency graph. Each of these nodes compares its ratio to those of its neighbors, and if it is among the nodes with the greatest increment Ratio, it is included in the set of sensors that will increase their radii concurrently by Δr. The pseudocode for this scheme is provided in Fig. 3.6.

3.3.5 Evaluation of the centralized and distributed algorithms

The performance evaluation of the proposed schemes is achieved via modeling and simulation. For this purpose, typical metrics that have been used in similar evaluations are employed, namely the average sensing radius (also used in [42]), and the energy consumption ratio defined as:

$$ConsumRatio(\mathbf{R}) := \frac{Energy(\mathbf{R})}{Energy(\mathbf{R}_{max})}. \tag{3.15}$$

The performance of a randomized version of the Greedy Centralized Algorithm denoted by Check-c-R, is also examined. This version may address scalability issues of the Greedy Centralized Algorithm. At each iteration of Check-c-R, only a subset S' of the set of sensors S are candidates for radius increase, with $|S'| = c$. The set S' is chosen uniformly at random among all the possible subsets of cardinality c. Consequently, the Greedy Centralized

Greedy Distributed Algorithm
Input: Set of obstacles $\mathbf{O}$, set of sensor points $\mathbf{P}$, sensing perimeter limit l.
Output: Radii assignment $\mathbf{R}$.

Initialization: $\mathbf{R} = \mathbf{0}$.
function GreedyDistributedAssignment $(\mathbf{O}, \mathbf{P}, l)$
while Change of $\mathbf{R}$ in the last iteration **do**
  $SetToIncrease = \emptyset$
  for i **from** 1 **to** N **do**
    if $\mathcal{PCOV}_i \geq l$ **then**
      Continue
    end if
    if $i \in \arg\max_{j \in (nei) \cup \{i\}} \{incrementRatio(j)\}$ **then**
      $SetToIncrease = SetToIncrease \cup \{i\}$
    end if
    nei = Neighborhood($\mathcal{G}$,i)
  end for
  SetToZero = FindRedundantNodes(Sensors)
  IncreaseEachInSetBy(SetToIncrease,Δr)
  SetZeroEachInSet(SetToZero)
end while

Figure 3.6: Greedy distributed algorithm for the variable radii connected sensor cover problem.

Algorithm can be derived by Check-c-R, for $c = |S|$.

The simulation parameters are: $L_0 = 10m, r_{\max} = 1m, \Delta r = \frac{r_{\max}}{5} = 0.2m$. The sensing constant is set $c_i = 1$. Unless otherwise stated, the number of obstacles considered is 40 and the number of nodes is assumed 200. The obstacles are generated from a convex hull of 7 points dropped uniformly at random inside a unit square.

3.3.5.1 Energy consumption ratio

In Fig. 3.7, a comparison among the three aforementioned alternatives (Greedy Centralized: Check-all-R, Check-c-R, and Distributed) is presented with respect to the metric *ConsumRatio* for two different scenarios of targeted coverage, i.e., 90% and 95%. At first the effect of network size on energy consumption is examined. It should be noted that the coverage percentage in the distributed algorithm is a random variable. To enable a fair comparison with the centralized approaches, the Restricted Sensing Perimeter limit is chosen to induce a percentage of area coverage of an average value close to the targeted coverage ($\approx 91\%$ and $\approx 94\%$), as presented in Fig. 3.8.

It can be observed that for all algorithms, as the number of sensors increases, the con-

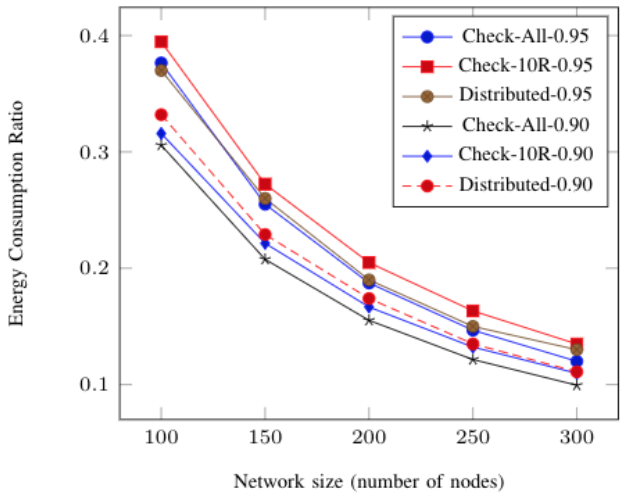


Figure 3.7: Comparison of Average Energy Consumption Ratio for the three algorithms.

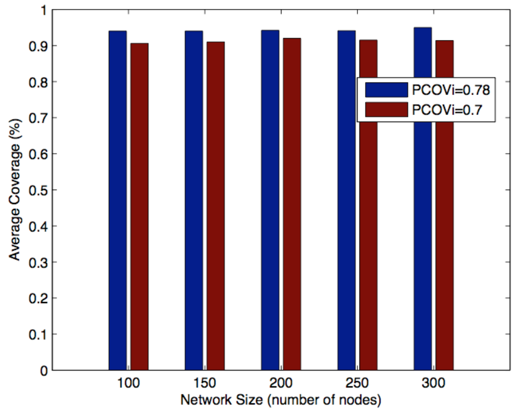


Figure 3.8: Average Coverage for different values of the Restricted Sensing Perimeter ($PCOV$) in the Greedy Distributed Algorithm.

sumption ratio decreases, leading to more efficient radii assignments. This is reasonable, since when more sensors exist, a smaller sensing range will be required for each sensor, in order for the same percentage coverage to be achieved. This leads to greater energy efficiency. A smaller required coverage percentage, i.e. 90% as opposed to 95%, leads to similar results, as it is not required for some nodes to increase their sensing radii more than the rest in order to cover stringently obscured areas.

Fig. 3.9 compares the average consumption ratio for varying number of obstacles. The distributed approach behaves similarly to the randomized centralized for a small number of obstacles and similarly to the greedy centralized for more obstacles. This is due to their fundamental operation. In the greedy centralized, the sensing radius assignment depends on the area between the increment rings. As the number of obstacles increases, less options are left in this area. On the contrary, in the distributed algorithm, the decision depends on the restricted sensing perimeter, which is a function of the obstacles, the boundary of the region and other sensors. The boundary and other sensors have a fixed contribution to the restricted sensing perimeter, due to their specific shape. In this case, the restricted sensing perimeter can obtain a value greater than the current sensing circle. The greater the number of obstacles around a sensor, the greater the value of the restricted sensing radius will be (due to the shape of the obstacles), and each sensor will stop faster increasing its sensing range, leading to an assignment with smaller radii overall.

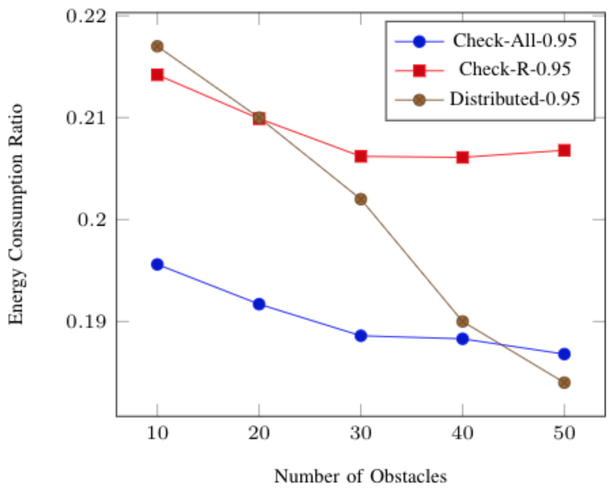


Figure 3.9: Comparison of Average Energy Consumption Ratio for different number of obstacles (results averaged over 50 topologies).

3.3.5.2 Average sensing radius

Additional characteristics of the behavior of the distributed algorithm are presented in Table 3.1 and Fig. 3.10. Specifically, Table 3.1 lists the Average Coverage (AvgCov), the standard deviation of the Coverage (CovStd), the Average Sensing Radius (AvgRad) and the number of iterations required for convergence (Iters).

The results are in accordance with the previous ones. Furthermore, it can be verified that for the distributed algorithm the standard deviation of the coverage percentage is small, indicating that despite its stochastic nature, the outcome of the scheme has a good eventual

Table 3.1: Performance indices for the Distributed Algorithm

Nodes	AvgCov	CovStd	AvgRad	Iters
Perimeter Coverage 78%				
Nodes	AvgCov	CovStd	AvgRad	Iters
100	0.94	0.0145	0.455	32.8
150	0.94	0.0129	0.353	35.9
200	0.942	0.0277	0.292	39.1
250	0.941	0.0387	0.253	41.3
300	0.950	0.0241	0.229	43.8
Perimeter Coverage 70%				
Nodes	AvgCov	CovStd	AvgRad	Iters
100	0.906	0.0314	0.426	30.1
150	0.910	0.0134	0.333	33.4
200	0.920	0.0147	0.280	36.5
250	0.915	0.0227	0.244	38.6
300	0.914	0.0353	0.219	41.2

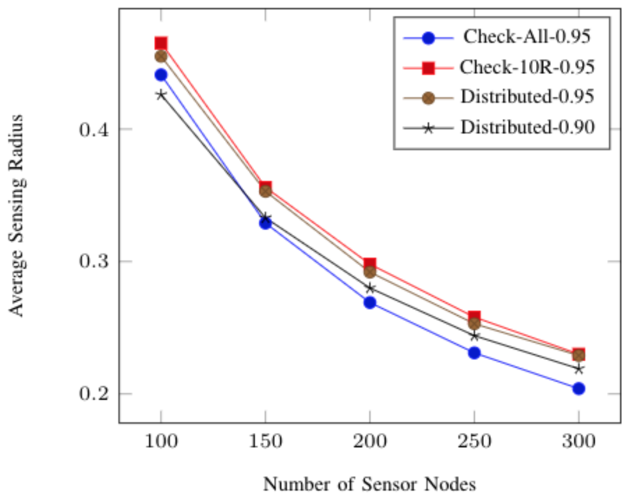


Figure 3.10: Comparison of Average Sensing Radius for different network sizes.

behavior. Fig. 3.11 presents the average sensing radius for varying number of obstacles, which are in accord with the results in Fig. 3.9.

3.3.5.3 Network connectivity

The connectivity of various deployments is investigated to verify that information exchange is possible. Connectivity depends on the number of nodes per unit area and the number of obstacles, which are considered of constant size. In Fig. 3.12, the connectivity of the simulated topologies is shown, varying from a moderately obstructed environment with

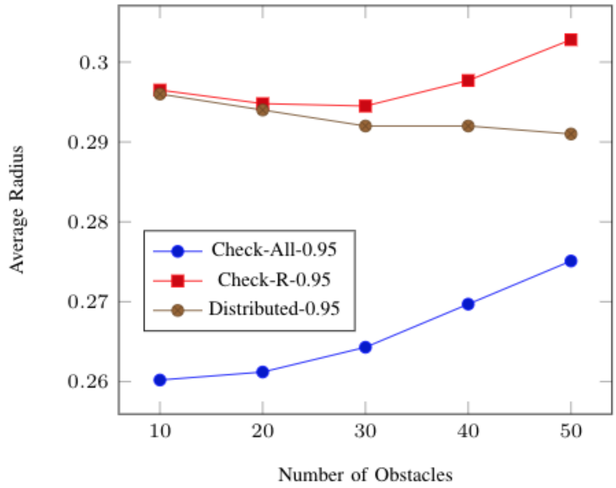


Figure 3.11: Comparison of Average Sensing Radius for different number of obstacles (results averaged over 50 topologies).

10 obstacles, to a heavily obstructed environment with 50 obstacles. The comparison is between a communication radius $r_t = 1$, equal to the maximum sensing range $r_{\max}$, and $r_t = 2.5 \cdot r_{\max} = 2.5$.

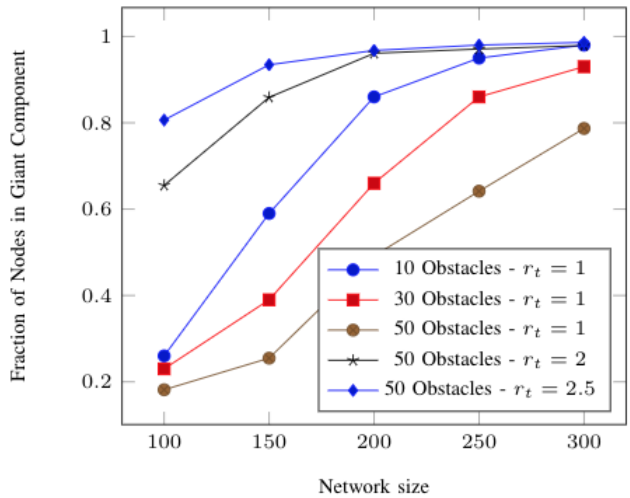


Figure 3.12: Average Size of Giant Connected Component (results averaged over 50 topologies).

In the vertical axis, the size of the largest component (Giant Connected Component) as a fraction of the total nodes is presented. The results indicate that a communication range of $r_t = 1$ is insufficient to provide a giant component that contains the majority of nodes.

Contrary to the previous case, for $r_t \geq 2$, even the most obstructed networks are very well connected for sufficiently dense deployments – one example is the case of 200 nodes, in which an average of 96.1% of the nodes are included in the giant component.

The approaches presented in 3.3 monitor efficiently an irregular region of interest by solving a problem of coverage under complex constraints in a Wireless Sensor Network. In Chapter 4, the notion of monitoring and coverage is extended, via Social Network Analysis, in cyber networks.

Chapter 4

Monitoring of information diffusion and socio-aware content allocation in cyber networks

4.1 Information diffusion in Online Social Networks

A definition of information diffusion in social networks is given in [55] as follows:

Definition 3. *Information diffusion is defined as the process by which information is communicated through certain channels among the users of a social network.*

The main factors influencing information diffusion are the network structure (node density, centrality, clustering, connectivity), the strength of ties (frequency of interactions, strength of influence) [56, 57] and its temporal dynamics, namely, how the diffusion rate evolves over time. Research works in information diffusion address three main categories of problems: detection of popular topics, inference/prediction of the diffusion network and identifying the most important nodes in a diffusion process, which is closely related to the problem of influence estimation and maximization [58].

4.1.1 Detection and prediction of popular topics

For the detection of popular topics in OSNs, most methods in the literature are based on finding unusual term frequencies in information sharing [59]. In [60] a state machine is used to model the arrival times of documents in a stream in order to identify bursts, that is, topics that grow in intensity for a period of time and then fade away, under the assumption that all documents belong to the same thematic category. In [61] a method of tracking units of information in terms of short distinctive phrases, referred to as memes, is developed through clustering textual variants of memes. The method is applied to real data and the results show that the news cycle in social media exhibits a persistent temporal pattern of growing and decaying popularity around the point of peak intensity. An approach that predicts the topics of great popularity in the near future is developed in [62]. A tendency indicator used in technical analysis of stocks, known as MACD (i.e. Moving Average Convergence Divergence) is adapted to identify bursty topics determined by keywords in social media. The objective of the improved MACD indicator is to turn two trend-following indicators, a short and a longer period moving average of keywords, into a momentum oscillator. The trend momentum is calculated by subtracting the long from the short moving average. When the value of the trend momentum changes from negative to positive, the topic is becoming popular, whereas the level of popularity is decreasing when the value changes from positive to negative.

4.1.2 Inference and prediction of influence networks

The second category includes methods of finding, via prediction or inference, who influenced/infected whom, namely, the paths through which information propagates. Information diffusion models can be coarsely classified to explanatory and predictive [59].

The goal of **explanatory models** is to infer the underlying network over which information propagates, given the times users learn a piece of information. Correlations between the users' infection times are studied in [63] in order to infer the underlying network over which contagions spread. Assuming a static network topology and that activated nodes transmit the contagion to each of its neighbors independently with some probability, the propagation probability between two nodes decreases with the difference of their activation times. An iterative algorithm based on submodular function optimization is designed to find the spreading cascade that maximizes the likelihood of the observed data. An extension of this algorithm, which considers dynamic networks that change over time is introduced in [64], where the diffusion process is modelled as a spatially discrete network of continuous, conditionally independent temporal processes that occur at different rates. The likelihood of a node infecting another node at a given time is modeled via a probability density function depending on pairwise activation times and transmission rates. The proposed method infers the structure of the diffusion graph and the transmission rates between pairs of nodes by formulating a convex maximum likelihood problem, which is solved with stochastic gradient descent. In [65], the diffusion inference problem is treated as a problem of recommendations. Given the information cascade, which is modeled as a sequence of tuples (user id, timestamp of infection) as well as users' historical data, the behavioral features of the users are derived, that is, their relevance on different topics, as well as textual features of their exchanged messages. By considering the propagated information as an item for recommendation, the previously described features are exploited by a Recurrent Neural Network to solve the recommendation problem and infer the users' diffusion relationship.

Based on the fact that the dissemination data observed in an OSN is limited due to privacy restrictions, an explanatory model of information diffusion is introduced in section 4.2. The issue of partially observed information cascades is addressed by focusing on the structural properties of an OSN in order to determine a monitoring cover, that is, a minimum set of nodes that one has to monitor (e.g., users with public profiles, who also tend to be very popular in the network) in order to infer different content and diffusion dynamics that spread concurrently and independently across a network. Edge coloring and statistical learning are then applied to trace the corresponding diffusion paths.

In contrast to explanatory models, **predictive models** aim to predict the outcome of a specific diffusion process based on temporal or spatial network characteristics [59]. Widely used predictive *graph-based models* of information diffusion are the sender-oriented Independent Cascade (IC) probabilistic model and receiver-oriented Linear Threshold (LT) model. In the IC model, node u at time t becomes active (receives information from the diffusion process) and has one single chance of activating each inactive neighbor v at time $t+1$, with a probability p_{uv}. The process continues until no more activations can take place.

The LT model is the most popular of the threshold models used in studying diffusion in networks. In this model, a threshold value or a set of threshold values are used to determine ranges of values where the behavior predicted by the model changes significantly. Each edge (u, v) of the network is affiliated with a weight w_{uv}, and each node v has a threshold t_v. Node v is activated if the fraction of its active neighbors exceeds t_v [55].

The explanatory model employed in this dissertation, is the one of Random Walks. Given a graph and a starting node, one of its neighbors is selected at random to serve as a new starting point for a random neighbor selection. The sequence of points created by repeating the described random process is referred to as random walk on a graph [66]. Random walks are one of the most fundamental types of stochastic processes and are widely used to model information diffusion and interactions among entities in networks of various structures. They can be distinguished to three main types: discrete-time random walks, node-centric continuous-time random walks and edge-centric continuous-time random walks [67]. In section 4.2, several discrete-time random walks with non-adaptive walkers are employed to model the information sharing between the users of an OSN.

Non-graph based models of information propagation, as opposed to the graph based models, do not assume the existence of a specific graph structure and are mostly based on epidemic spread processes. In epidemics transmission, users are infected with a virus while others are susceptible to it. The virus can spread from infected to susceptible users, in a similar way that information diffuses from communicators to recipients. Therefore, users are classified into several classes and the variation in the amount of users of each class, due to state transitions, is studied. The epidemic models are expressed by differential equations and the most representative deterministic models are the SI (Susceptible-Infected), SIS (Susceptible-Infected-Susceptible) and SIR (Susceptible-Infected-Removed) [12, 13]. In the SI and SIS models [68], susceptible users switch to class I with a fixed probability. SIS additionally assumes that the users in I may switch to S with a fixed probability. In the case of SIR, the users in I permanently switch to the R class.

4.1.3 Identify influential users

Social influence is described in [69] as follows:

Definition 4. *Given two individuals u,v in a social network, u **influences** or exerts power over v, that is, u may change the opinion of v in a direct or indirect way.*

From the viewpoint of information diffusion, influence can be defined as the importance of each user in propagating information. Finding the most influential spreaders of information in an OSN is of great importance since it ensures efficiency in the process of information sharing, including recommendations. The most important features for measuring user influence are based on network structure, user interactions and user attributes. Centrality measures including degree, closeness, betweenness, eigenvector and Katz centrality [70] are conceived as evaluation metrics for characterizing the influence of network nodes with respect to the overall network structure. For instance, influential users of recommender systems operating in an OSN can be considered the ones who exhibit high similarity to a great portion of users, having at the same time high out-degree in the social network [71].

The problem of selecting the k most influential nodes of a network is known as Influence Maximization and was first presented in [72] as an algorithmic technique for viral marketing. It was then formulated in [73] as a discrete optimization problem that was proved to be NP-hard and by using an analysis framework based on submodular functions, a greedy strategy with approximation guarantees was presented. Apart from this greedy strategy, many heuristic and hybrid algorithms have been introduced to address the influence maximization problem in static and dynamic networks [74, 75]. In this dissertation, a problem closely related to influence maximization is addressed from the viewpoint of recommendations where the information sharing between the users of an OSN is taken into consideration. These are known as Social Recommender Systems and the problem addressed in section 4.4 is the one of item-user relevance maximization under user complex constraints. The proposed methodology is designed to select item-user tuples, that is, a set of items for assignment to a set of OSN users with susceptibility to information overload so that the network's total relevance to the former is maximized through inferred users interactions.

4.2 Sensing and monitoring of information diffusion in Online Social Networks

As discussed in section 4.1.2, inferring an influence network is a complex problem of prominent importance. Considering that sensing and monitoring information diffusion in OSNs typically requires significant sensing resources, an inference approach for an information diffusion process through efficient monitoring is proposed and analyzed in this thesis. Information is considered to belong to different classes and it is characterized by different spreading dynamics based on its popularity. The designed framework utilizes social network analysis metrics in order to reduce the sensing resources that would be required in an otherwise exhaustive monitoring approach, while employing statistical learning and probabilistic inference for maintaining the accuracy of information tracking, whenever needed. An edge coloring scheme is defined, based on which it is possible to keep track of information diffusion. The information is assumed to spread in the network according to various biased random walks that represent the dynamics of the considered classes of information. Statistical learning is employed for the inference of those cases where backtracking leads to multiple potential choices for information paths. The operation and efficacy of this approach is demonstrated in characteristic online social networks, such as distributed wireless (spatial) and scale-free (relational) topologies, and conclusions are drawn on the impact of topology on information spreading.

4.2.1 Models of the OSN and the diffusion process

In a platform of social networking, users form social bonds and interact by sharing information. An *information class* is defined as the cumulative information content that spreads according to the same mechanism over the platform. Examples include all information regarding a specific sport, news on politics, seasonal advertisements, etc., [76, 57]. Before the diffusion mechanism is explained in detail in subsection4.2.1.2, the considered communication network model is presented. The notation employed in this section is summarized in Table 4.1.

Table 4.1: Notation table for the problem of information diffusion monitoring.

Symbol	Interpretation
$G(V, E)$	Physical connectivity graph
v_i	Node $v_i \in G$
n	Number of nodes $n = \|V\|$ of the physical topology
$\mathbf{A}$	Adjacency matrix $\mathbf{A} = [\mathbf{a_{ij}}]$
$G'(V, E')$	Diffusion graph
T	Delaunay triangulation
I	Set of information classes $I = \{I_1, I_2, I_3, I_4\}$
I_i	Information class i
$d(v_i)$	Node degree of v_i
C	Set of colors $C = \{C_1, C_2, ..., C_k\}$
c_i	Color $i \in C$
S	Set of sensors
s_i	Sensor $s_i \in S$
L	Size of square deployment region for RGG
R	Transmission radius of RGG
d	Clustering parameter of SF network
q_j	Sequence of colored edges traversed j
G_{s_i}	1-ego network of node s_i

4.2.1.1 OSN model

Graph $G(V, E)$ is a representation model for the physical interconnection of an OSN, where V is the set of nodes, $|V| = n$, and E the set of links between pairs of nodes. Two representative physical interconnection topologies are considered, one of the Random Geometric Graph (RGG) type representing spatial networks, and one of scale-free (SF) type being representative of relational networks [23]. The RGG is a proper model for distributed mobile devices that form ad-hoc networks and necessarily have to rely on multihop routing of the information exchanged [23, 77], while SF is more representative of wired information network topologies forming in online social networks.

The $n \times n$ matrix $\mathbf{A} = [\mathbf{a_{ij}}]$, with $a_{i,j} \in \{0, 1\}$ for $i, j = 1, ..., n$ is the adjacency matrix of physical topology G, where $a_{ij} = 1$ whenever link (i, j) exists and zero otherwise. The physical network is considered undirected, and thus, $a_{ij} = a_{ji}, \forall i, j$. G represents the interconnection of physical devices that constitute the communication substrate for the exchange of information.

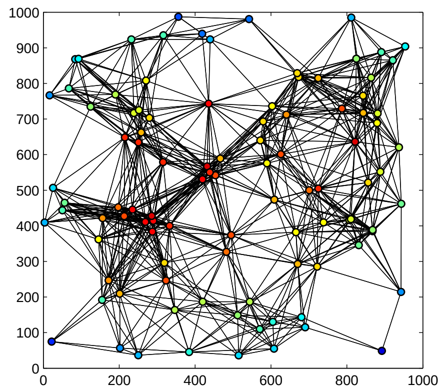
For the RGG, the network is assumed to be deployed over a square region of side L. Each node of the RGG is considered to have a transmission radius R, dictated by a specific maximum transmission power, selected uniformly for all nodes. Thus, a common R is employed by all user devices [77]. Transmission effects at the physical layer are not taken into account. They could be implicitly considered via an “effective radius” multiplicative factor, cumulatively modeling all such effects [78]. Also, the retransmission effects at the medium access layer that emerge during MAC frame exchange between neighboring nodes are not considered. Such retransmissions could have some effect on the temporal evolution of the information spreading process over the analyzed network. When retransmissions occur, the typically employed TCP protocol at the Network layer will ensure the correct exchange of information, with some additional delay. The eventual outcome reduces to a delayed information propagation effect, which does not fundamentally differ from the one studied, apart from a delay offset factor. Since this factor is small and almost uniform across all transmission links in the network, it can be disregarded without harming analysis.

The SF topology is characterized by parameter d, denoting the d connections each newcomer node creates with d pre-existing nodes of the initial network using the preferential attachment rule. Such process is employed for deriving the final SF topology [23]. The higher d is, the more clustered the SF network will be.

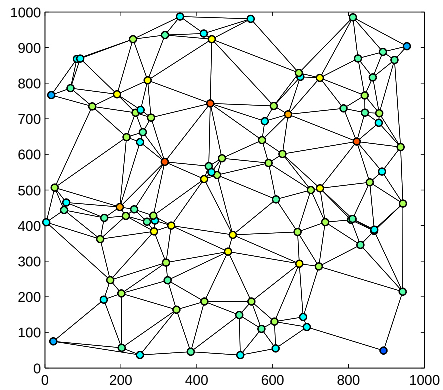
4.2.1.2 Information diffusion model

The considered information diffusion process assumes a set of I, $|I| = m$, classes of information spreading over the network, denoted as $I_1, I_2, ..., I_m$. For simplicity, $m = 4$ is considered, but extension to more classes of information is straightforward. The classes of information are assumed to be spreading in parallel over a spanning subgraph $G' = (V, E')$ of G, $|E'| < |E|$ and $E' \subset E$. This means that the flow of information does not necessarily follow the whole of the physical graph, but most frequently, it employs a subgraph of it. This is due to the fact that information diffusion is fundamentally a probabilistic process [57, 79]. This subgraph is referred to as diffusion graph.

In the case of the spatial network (RGG), the corresponding diffusion network is obtained through a “thinning” process that maintains the original properties of the communication network. This can be achieved by a Delaunay triangulation $T = (V, E_T)$ of the vertex set of G [35], where $E' = E_T \cap E$, an example of which is shown in Fig. 4.1. The diffusion subgraph obtained through this process will be a planar graph [35], which is a rather desired property for this case of RGG communication networks, since planarity ensures that users of small distance in the initial topology (distance between nodes represents the strength of social ties between users) are connected. In the case of the relational network (Barabasi-Albert model of SF networks [23]), the corresponding diffusion network is obtained by a conditioned *deterministic preferential pruning* of edges: In order to preserve the scale-free properties of the initial network in the diffusion network, where the former is parametrized by a number d, which is the number of each newcomer's connections with d pre-existing nodes [23], the network of parameter $\frac{d}{2}$ is used as a basis for filtering out edges that are incident to minimum degree vertices as follows: The nodes of both networks are sorted in order of increasing degree value and for each ordered pair the difference between their corresponding degree values is computed. For positive resulting values and in the case where the network remains connected, the edges incident to nodes of minimum degree are successively removed. This results in a network where the degree centrality of its nodes inherits both the degree distribution of the initial and the employed basis d, and follows a power-law distribution, as shown in Fig. 4.2.



(a)



(b)

Figure 4.1: Initial network and its corresponding diffusion graph. (a)Random Geometric physical graph: $|V| = 100$, $|E| = 845$, $R = 250$m. (b)Random Geometric diffusion graph: $|V| = 100$, $|E| = 271$.

Each information class is assumed to spread independently over the diffusion network according to a biased random walk [80], each with a different bias. Random walks have been extensively used as sufficient models for information diffusion [67, 80, 81]. Also, for each class a seed node $v \in V$ is chosen uniformly at random across the network, to serve the

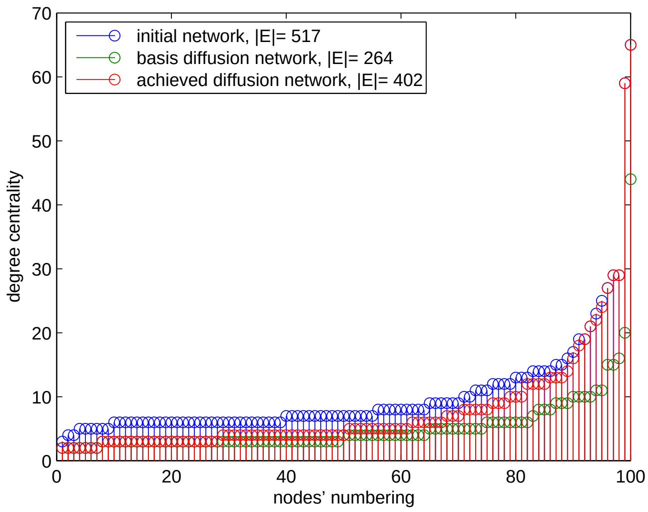


Figure 4.2: Degree centrality of the scale-free diffusion network resulting from preferential pruning of edges of the initial scale-free topology.

role of information seed in the diffusion process. Each biased random walk corresponds to a different information class as follows.

Information class I_1: The first information class is considered to be an unbiased random walk, which will also serve as a benchmark diffusion model. Thus, according to I_1, at each time a node informs one of its neighbors in the diffusion network with equal probability. The transition probability matrix of this information class is given by $T_1 = [p_{ij}^{(1)}]$, where

$$p_{ij}^{(1)} = \begin{cases} 1/d(v_i), & v_i v_j \in E' \\ 0, & \text{otherwise,} \end{cases} \tag{4.1}$$

and $d(v_i) = \sum_{i=1}^{n} a_{ij}$ is the degree centrality of node v_i of G'.

Information class I_2: The second class is considered as a shortest distance biased random walk with transition matrix $T_2 = [p_{ij}^{(2)}]$, where

$$p_{ij}^{(2)} = \begin{cases} \frac{d(v_i, v_j)^{-1}}{\sum_{v_j \in N'_{u_i}} d(v_i, v_j)^{-1}}, & v_i v_j \in E' \\ 0, & \text{otherwise,} \end{cases} \tag{4.2}$$

where N'_{v_i} is the neighborhood of v_i in G' and $d(v_i, v_j)$ is the Euclidean distance between v_i and v_j.

Information class I_3: The third class is considered as a closeness centrality-biased random walk with transition matrix $T_3 = [p_{ij}^{(3)}]$, where

$$p_{ij}^{(3)} = \begin{cases} \frac{C_{v_j}}{\sum_{v_j \in N'_{u_i}} C_{v_j}}, & v_i v_j \in E' \\ 0, & \text{otherwise,} \end{cases} \tag{4.3}$$

where C_{v_j} is the closeness centrality of v_j in G'.

Time is considered slotted, so that the information diffusion process proceeds in discrete time steps over G'.

4.2.2 Information diffusion sensing and inference

As discussed above, all information classes start spreading at the same time independently through an edge-colored network. Information spreading over the network is tracked through a subset of the initial network's nodes, which are considered to have sensing capabilities. Each time that an information class I_i hits a sensor node s_k, it leaves a trace in the form of a set of triples (I_i, t, q), where t is the time at which I_i hits s_k and q is the sequence of colors of the edges traversed between two successive hits to sensor nodes from class I_i. The diffusion inference is performed via backtracking, which operates as follows. For a trace in sensor node s_k, starting from the last element of the color sequence q, which corresponds to an edge incident to s_k and using 2 different inference methods discussed below, each color is mapped with respect to the physical network's structure to an edge of the diffusion network.

4.2.2.1 Sensor placement

The sensor placement scheme takes into account the degree centrality, i.e., node degree. The corresponding pseudocode is provided in Algorithm 1. The algorithm begins by sorting all nodes according to their degree centrality in a descending order. The highest ranked node is chosen as the first sensor-monitor node. Then each of the remaining ordered nodes is sequentially added to the set of monitors as long as its neighborhood does not include any of the already designated monitor nodes. This problem is equivalent to finding a minimum maximal independent set of the network [82].

Algorithm 1: Degree-based Monitor Placement

Input: Network Topology $G(V, E)$

Output: A subset of nodes $S \subset V$ as sensors

Sort all nodes $v_i \in V, i = 1, ..N$, where $N = |V|$, in order of decreasing degree centrality score. Obtain V as $V = \{v_1 \succeq v_2 \succeq ... \succeq v_N\}$, where " $\succeq$ " indicates the ordering of decreasing degree centrality score.

Initialize the set of sensors: $S = \{\emptyset\}$;

Insert the node with the highest degree to S: $S = \{v_1\}$;

for $i = 2 : N$ **do**

  if $N_{u_i} \cap S = \emptyset$, N_{u_i} *is the neighborhood of node* u_i, **then**

    $S \leftarrow S \cup \{u_i\}$;

  else

    return S;

4.2.2.2 Edge coloring

In order to perform efficient information diffusion inference via backtracking, a scheme that takes into account the structural properties of the physical graph G colors its edges. Since most of the times the existing OSNs are characterized by massive sizes, an efficient and usable coloring scheme that can be easily implemented is needed [83]. Algorithm 2 provides the corresponding pseudocode.

The presented approach capitalizes on the observation that a large-scale network can be partitioned into local subnetworks defined by the ego-networks of nodes [84], and thus, develop an ego-network-based edge-coloring scheme. Assuming there exist $|S|$ sensor nodes, $C = \{c_1, c_2, ..., c_{|S|}\}$ colors are considered. For each sensor s_i, $i = 1, 2, ...|S|$, its 1-ego network G_{s_i} is constructed and all edges incident to s_i are assigned the color c_i. The number of required colors is reduced by assigning the same color to edges that belong to disjoint ego-networks. For each edge that color reuse cannot be applied, a new color is assigned, which is also added to set C. The last two steps of color-reuse and new-color-assignment are repeated for the remaining edges until all edges of the network are colored.

Fig. 4.3 provides a quantitative idea of the average number of monitors and colors determined by the proposed scheme for information sensing and path backtracking. Furthermore, Fig. 4.4 provides tangible examples of the coloring and monitoring schemes for two specific

Algorithm 2: Ego-network-based Edge Coloring

Input: Network Topology $G(V, E)$ and sensors $S = \{s_1, ..., s_{|S|}\} \subset V$
Output: Edge-Colored Network $G(V, E, c)$, $c : E \to C$ is a mapping from edges to colors

Initialize the set of colors with $|S|$ colors: $C = \{c_1, ..., c_{|S|}\}$.
for $i = 1 : |S|$ **do**
    Find the 1-ego network of sensor $s_i : G_{s_i}(V_{s_i}, E_{s_i})$;
    for *all edges* $s_i u_j \in E_{s_i}$ **do**
        $c(s_i u_j) = c_i$;
        $C_{s_i} = \{c_i\}$; C_{s_i} is the set of the colors for the edges of s_i ego-network.
for *each edge* $uw \in E : c(uw) = 0$; **do**
    Compute the set G_{uw} of the ego-networks $G_{s_i}, i = 1, ..., |S|$ to which the vertices u and w belong.
    Compute its corresponding set of colors $C_{uw} \subset C$;
    Compute the set of allowed colors for the edge $uw : C'_{uw} = C \setminus C_{uw}$;
    if $C'_{uw} \neq \emptyset$ **then**
        Assign at random an element of C'_{uw} to uw: $c(uw) = c_{uw}$
        $C_{uw} \leftarrow C_{uw} \cup \{c_{uw}\}$;
        if $c_{uw} \cap \{c_1, ..., c_{|S|}\} = \emptyset$ **then**
            $C \leftarrow C \cup \{c_{uw}\}$;
        else
            return C;
    else
        Assign a new color to uw: $c(uw) = c_{uw}$;
        $C_{uw} \leftarrow C_{uw} \cup \{c_{uw}\}$;
        $C \leftarrow C \cup \{c_{uw}\}$;

instances of a random geometric and a scale-free graph, respectively.

By observation of the panels in Fig. 4.4, it can be deduced that the average path length between pairs of monitors is expected to be greater in RGG topologies and lower for SF, owe to the power-law degree distribution of the latter where monitor-hub nodes provide short distances to many other nodes. However, this does not mean that the average number of sensors required reduces as well. The fact that there are many nodes with very low degree, not properly covered by the initially chosen as monitors node-hubs means that additional sensors are required. This is verified by Fig. 4.3(a), where it is observed that the average number of sensors in RGG topologies (with constant radius R and in square region of side L) is independent of the increase in the number of nodes, whereas in SF topologies, the number of sensors increases linearly to the number of nodes (20% of the nodes).

Due to the power-law structure of the SF topology, noticeable diversity in the coloring of edges is observed between the RGG and SF topologies. Almost half of the edges of SF

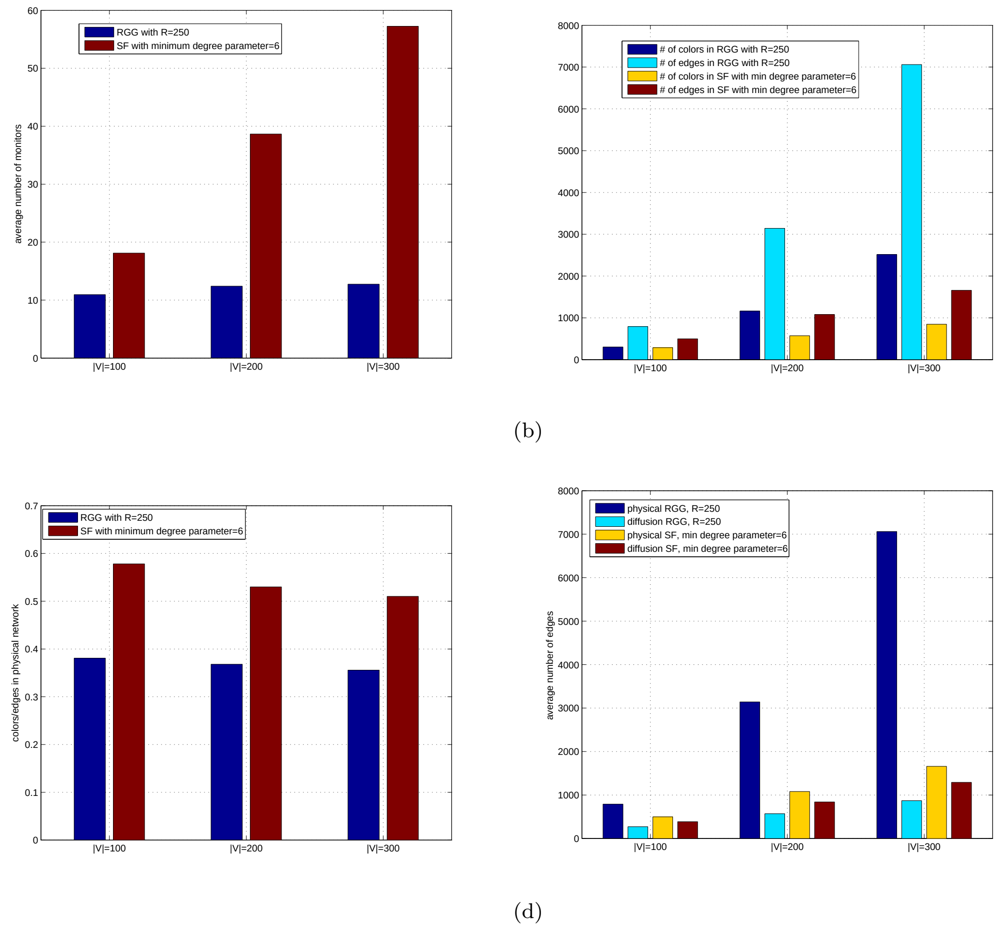
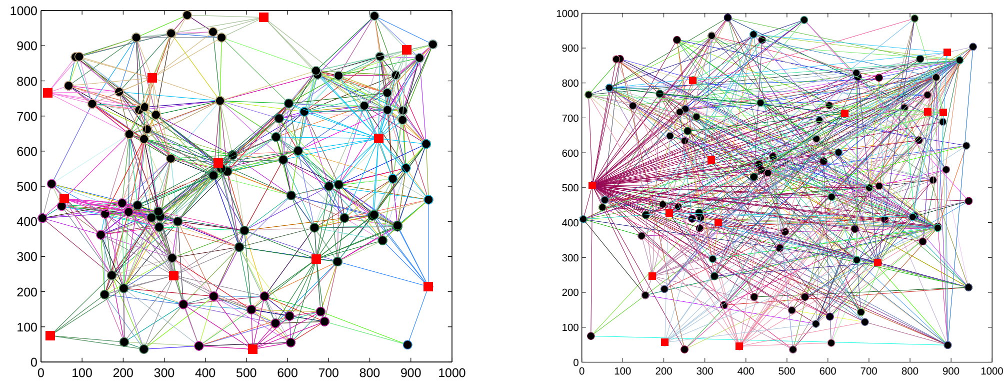


Figure 4.3: Average values of monitor nodes, colors and edges of physical and diffusion networks determined by the edge coloring approach. In panels (a) and (c) the results refer to physical topologies for both RGG and SF. Panel (b) depicts the numbers of colors and edges in the y-axis, as noted in the legend.

are incident to sensor nodes, thus inheriting the color defined by the sensor's ego-network they belong to. Also, color reuse cannot be applied in SF networks at the same extent as in RGG, since most sensors' ego-networks are not disjoint as in the case of RGG networks that are characterized by a distributed topology. The latter, combined with the aforementioned observations affect the inference processes in a way that will be discussed in the following.

4.2.2.3 Proposed inference mechanism

In the event that information of class I_i begins from a non-sensor node and immediately (in one hop) hits a sensor (monitor), a case depicted in Fig. 4.5(i) with blue color, it is



(a) (b)

Figure 4.4: Examples of coloring and monitoring schemes. (a) An edge-colored Random Geometric network with sensors (red squares). $|V| = 100$, $|E| = 845$, $R = 250$m, $|S| = 12$, $|C| = 344$. (b) An edge-colored scale-free network with sensors (red squares). $|V| = 100$, $|E| = 496$, minimum-degree-parameter $d = 6$, $|S| = 13$, $|C| = 304$.

impossible to infer accurately the traversed edge, since all adjacent edges will have the same color. To address this scenario, statistical learning is employed, based on the frequencies of the edges' traversal by information class I_i. The relative frequencies are defined as follows:

$$f_{s_1 v_i} = \frac{\#\text{ of times } s_1 v_i \text{ was traversed}}{\sum_{v_j \in N_{s_1}} \#\text{ of times } s_1 v_j \text{ was traversed}}, \tag{4.4}$$

where $N_{s_1} = \{v_j : s_1 v_j \in E_{s_1}\}$ is the set of neighbors of sensor node s_1. In case where $f_{s_1 v_i} = f_{s_1 v_j}$, a frequency-oriented centrality metric is defined for the 1-hop neighbors of monitor s_1, as

$$C_{f_{v_i}} = \sum_{v_j \in N_{v_i}} f_{v_i v_j}, \tag{4.5}$$

with $N_{v_i} = \{v_j : v_i v_j \in E_{v_i}\}$ and E_{v_i} being the set of edges of node's v_i ego-network. According to the latter, the edge that is incident to s_1 and the node with the highest $C_{f_{v_i}}$ is selected.

In some cases mentioned above, the information obtained by a color sequence is insufficient for an injective mapping from colors to edges. For these, two methods are employed:

- **Deterministic**: Statistical learning of the edges' use frequencies.

  From the set of candidate edges, the algorithm chooses the one that is most frequently

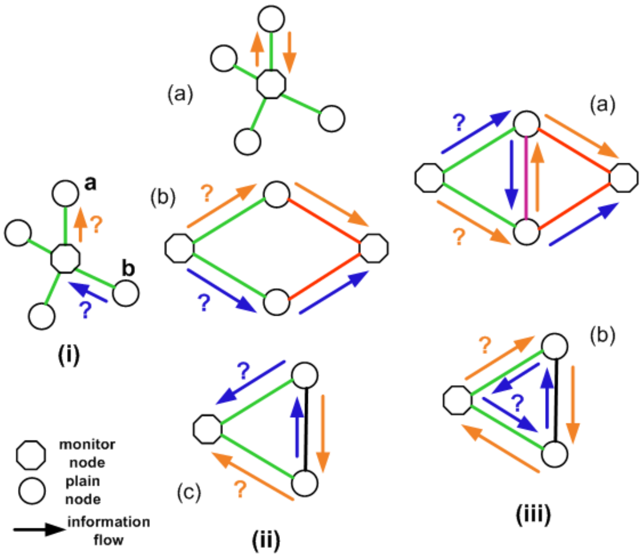


Figure 4.5: Cases of ambiguous color sequences of length 1, 2 and 3.

traversed by the corresponding information class, until the time the sequence is monitored.

- **Probabilistic**: An edge is selected at random from the candidate set with uniform distribution.

  From the set of candidate edges $E_{candidate} = \{s_i u_j\}$, where s_i is a monitoring node and $u_j \in V_{s_i}$, each candidate edge is selected with probability $\frac{1}{|V_{s_i}|}$.

Learning is needed in some cases where the coloring sequence is of length two. Such cases are shown in Fig. 4.5(ii), including one edge traversed twice, two edges between two monitors and two edges connecting a non-monitor node to a monitor one. The latter, denoted as the triangle case, may arise in sequences of length greater than two, as depicted in Fig. 4.5(iii) for the simple case of sequences of length three, which correspond to traversal of edges that form cycle paths in the physical topology. At this point, it is noted that the cases of length two are the toughest to infer, even with learning, as will be shown in section 4.2.3. As the number of such cases increases, the accuracy of the proposed approach will slightly decrease, rendering them a determining factor for its performance.

The aforementioned cases of length 3 can emerge as sub-patterns in sequences of greater lengths consisting of multiples of 3 colors including the depicted traversed triangle structures. An example of such a sequence regards the topology in Fig. 4.5.(iii)(b), where the edge between the two neighbors of monitor node can be potentially traversed repeatedly before eventually visiting the monitor again.

4.2.3 Evaluation results of information diffusion inference

The described information inference approach is evaluated over synthetic and real networks and the corresponding evaluation results are presented below. At first, definitions are provided regarding the evaluation measures employed and then the obtained results are analyzed.

4.2.3.1 Evaluation metrics

Two measures are used for the inference methods' performance evaluation, widely used in information retrieval: **precision** and **recall** [85]. These measures are employed to quantify the performance both at the level of network structure (physical topology) and at the flow level (diffusion network) of interactions among nodes.

A sequence of colors q_i corresponds to a set of edges E_{q_i} of the diffusion network (which is undirected). The corresponding inferred edges are denoted as $\hat{E}_{q_i}$. **Precision** is then defined as the fraction of inferred edges, which are present in the diffusion network and **recall** as the fraction of edges of the diffusion network, which are inferred.

At the level of network structure (physical topology), precision and recall are expressed as follows:

$$Precision_{structure} = \frac{|\hat{E}_{q_i} \cap E_{q_i}|}{|\hat{E}_{q_i}|}, \tag{4.6}$$

$$Recall_{structure} = \frac{|\hat{E}_{q_i} \cap E_{q_i}|}{|E_{q_i}|}. \tag{4.7}$$

In order to present the formulas of precision and recall at the interaction flow-level (diffusion topology), the interactions' flow is represented by a set of ordered directed edges $E^D_{q_i}$, where $|E^D_{q_i}| = k$, k is the length of q_i and $\hat{E}^D_{q_i}$ respectively. At interactions' flow level the two features examined are the edges' direction and the number of times an edge is traversed, as it is inferred by the color sequence q_i. Therefore, precision and recall at the interaction flow-level are expressed as follows:

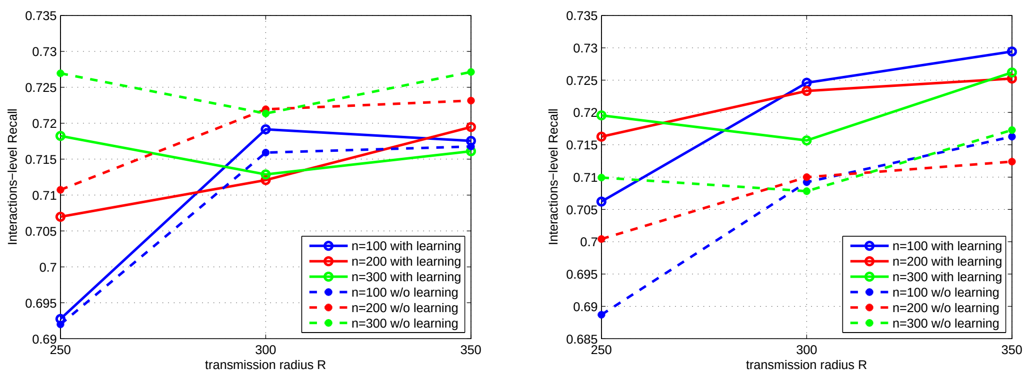
$$Precision_{interaction} = \frac{|\hat{E}^D_{q_i} \cap E^D_{q_i}|}{|\hat{E}^D_{q_i}|}, \tag{4.8}$$

$$Recall_{interaction} = \frac{|\hat{E}^D_{q_i} \cap E^D_{q_i}|}{|E^D_{q_i}|}. \tag{4.9}$$

4.2.3.2 Evaluation over synthetic networks

Evaluation results are presented for two cases. The first regards scenarios where monitoring stops when all the random walks reach all sensors (denoted as hitting time scenarios), while the second regards scenarios where monitoring stops at the cover time of all random walks, namely when all random walks cover the whole network (denoted as cover time scenarios).

The results of the two inference methods are applied in both spatial and relational networks. Fig. 4.6 presents the recall scores at the interaction level (information diffusion) for spatial (RGG) networks with respect to devices increasing their transmission radius, corresponding to networks of increasing densities.

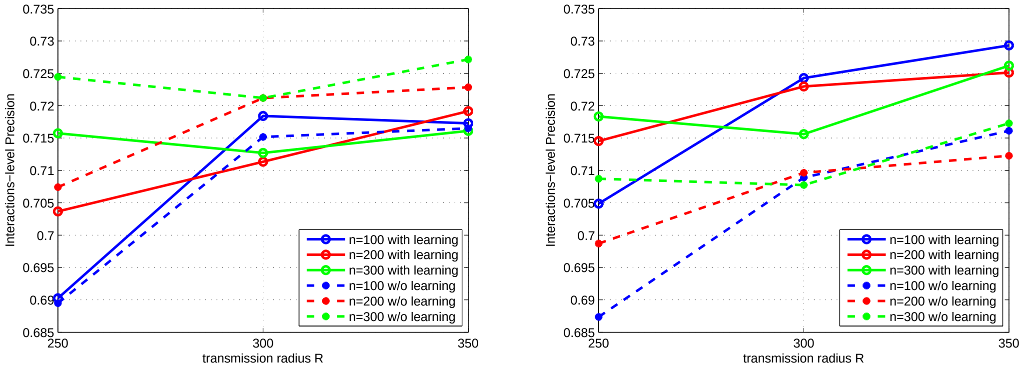


(a) (b)

Figure 4.6: Recall at interactions'-level in RGG: (a) Hitting time scenarios (b) Cover time scenarios.

The corresponding scores for the precision metric computed at the interaction level of the analyzed spatial (RGG) types of networks are shown in Fig. 4.7. By comparing Figs. 4.6 and 4.7, a general trend can be deduced, namely that the proposed learning increases the accuracy of the inference scheme for these types of networks.

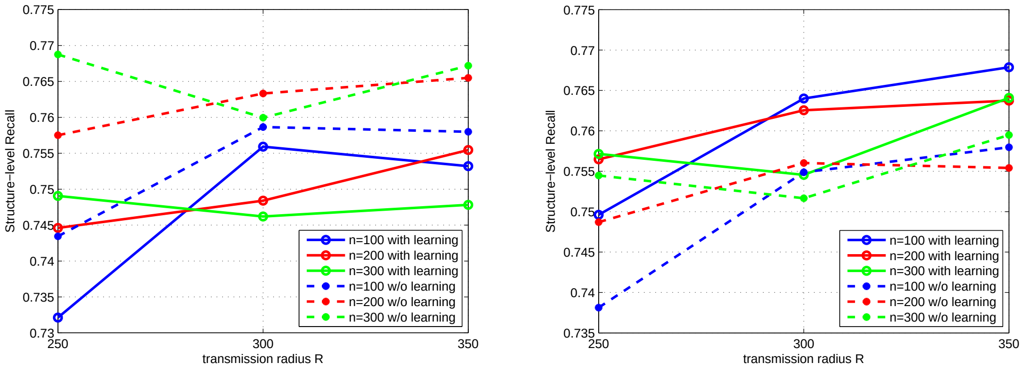
Similar results are presented in Figs. 4.8 and 4.9. Both precision and recall at the



(a)

(b)

Figure 4.7: Precision at interactions'-level in RGG: (a) Hitting time scenarios (b) Cover time scenarios.

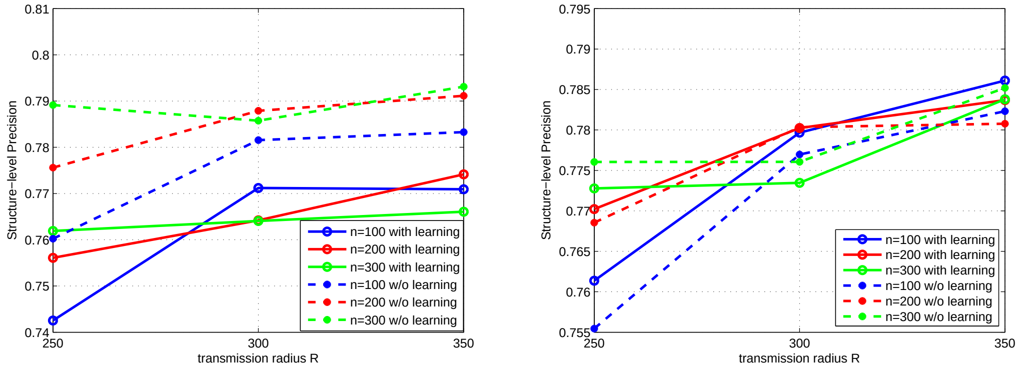


(a)

(b)

Figure 4.8: Recall at structure-level in RGG: (a) Hitting time scenarios (b) Cover time scenarios.

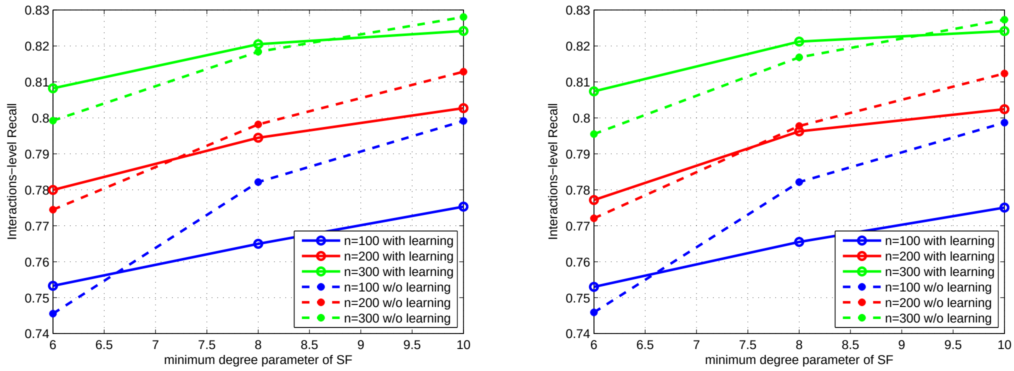
physical (network structure) level follow the same trends, namely both metrics increase in general as learning is applied, signifying better accuracy. However, it should be noted that in some cases the differences between the learning and probabilistic scheme are asymptotic, indicating that for some practical scenarios the simple probabilistic approach may be more suitable. Such cases are the ones where most of the employed sequences are of length 2 (or sequences including such patterns, as explained before), for which inference is tougher and less accurate. Another factor in favor of the probabilistic scheme is the cold-start problem experienced in the learning mechanism: When the first monitored sequences of colors cannot provide an injective mapping to edges (ambiguous cases of length 2 and 3, as



(a) (b)

Figure 4.9: Precision at structure-level in RGG: (a) Hitting time scenarios (b) Cover time scenarios.

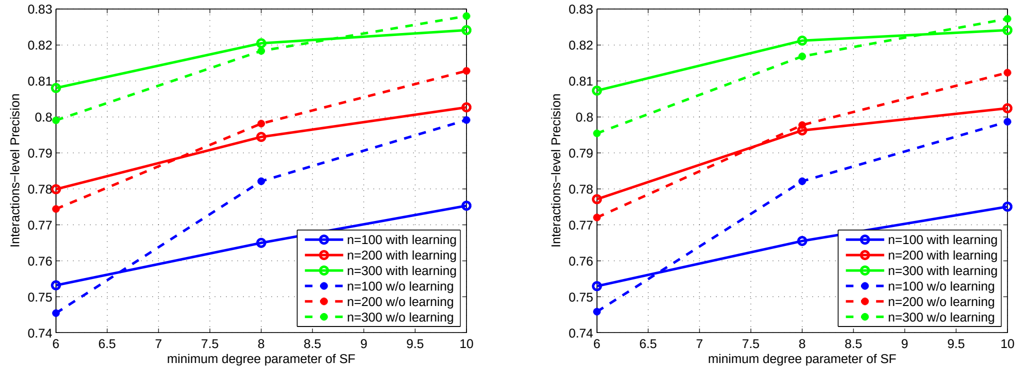
discussed above), there is no sufficient information for inference.



(a) (b)

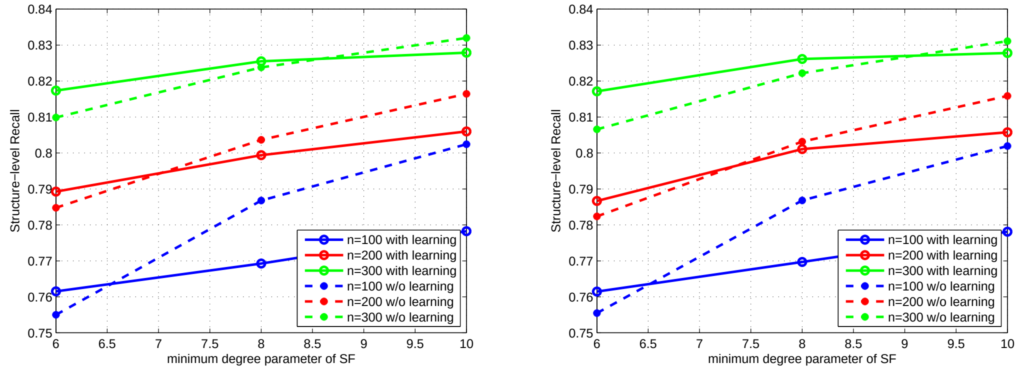
Figure 4.10: Recall at interactions'-level in SF: (a) Hitting time scenarios (b) Cover time scenarios.

Figs. 4.10 and 4.11 present the recall and precision scores at the interactions level (diffusion graph) respectively, for scale-free topologies and for both hitting and cover time scenarios. Two main observations can be made: Contrary to RGG, the performance of learning at hitting and cover time remains constant and probabilistic inference is slightly more accurate in the case of topologies with small number of nodes and high minimum degree parameter d. The first observation can be attributed to the emerging hub nodes of the network (most of which become sensors due to the sensor placement scheme applied):



(a) (b)

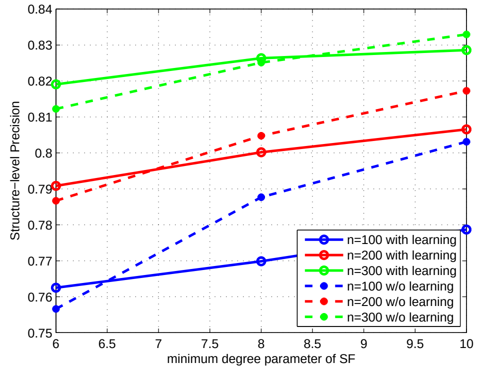
Figure 4.11: Precision at interactions'-level in SF: (a) Hitting time scenarios (b) Cover time scenarios.



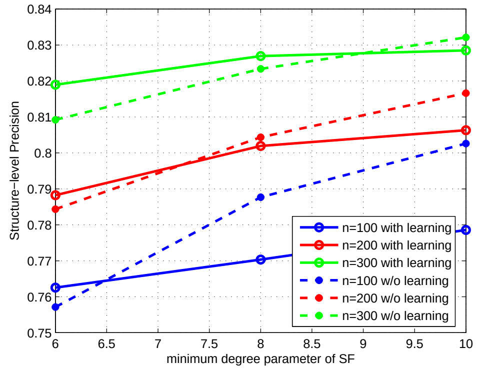
(a) (b)

Figure 4.12: Recall at structure-level in SF: (a) Hitting time scenarios (b) Cover time scenarios.

A sensor node may reach another sensor in 2 hops through a node belonging to a relatively large set of intermediate nodes (whose non-sensor neighbors have degree values close to the minimum degree parameter of the diffusion network), which belong to both sensors' neighborhoods. In addition to the former, due to the diffusion mechanisms applied, 52% of the monitored sequences in SF topologies are of length 2 or 3, thus inference with learning relies on 48% of the sequences (whereas in diffusion in RGG, 33% of the sequences are of length 2 or 3). This means that the learning approach, will fail in more cases, leading to a decrease of accuracy. Since the number of edges increases sublinearly to the nodes' increase, the performance of inference with learning improves at sparser topologies (of minimum



(a)



(b)

Figure 4.13: Precision at structure-level in SF: (a) Hitting time scenarios (b) Cover time scenarios.

degree parameter d=6 and $|V| = 300$), slightly exceeding the performance of the probabilistic scheme.

Figs. 4.12 and 4.13, provide the corresponding results on recall and precision regarding the physical network (structure-level) inference of scale-free topologies. Similar trends as those explained for Figs. 4.10 and 4.11 can be observed, namely that the performance of learning remains constant, and that probabilistic inference is slightly more accurate in the case of topologies with small number of nodes and high minimum degree parameter d.

4.2.3.3 Evaluation over real datasets

The performance of the proposed approach is investigated by tracking information diffusion in realistic conditions, using the popular *Higgs Twitter Dataset* [86], which consists of five directed networks (social, retweet, mention, reply, and activity network). The graph $G_p(V, E_p)$ represents the physical network, i.e., the network of social relationships among twitter users and it is a subnetwork of the *social network* of the dataset with $|V| = 6,943$, where V is the set of nodes with the greatest number of interactions in the *social network* and E_p is the set of directed edges between these nodes. From the *activity network*, only the *retweet* category of interactions between the nodes of V is selected to construct the diffusion network $G_d = (V, E_d)$. The fact that $E_p \supset E_d$ is not straightforward. A twitter user u who does not follow v, therefore (v, u) does not exist in the physical network, may retweet v through one of their common connections, Thus, the edge (v, u) may exist in the diffusion network. The latter should be added to the physical network so as to model the potential of information sharing between the two users. Also, the relationship $|E_p| > |E_d|$, is consistent with the assumption of the diffusion network being modeled as a spanning subgraph of the physical network. The direction of links in both networks is reversed in order to model properly the information flow.

Physical and diffusion network. Table 4.2 provides cumulatively the features of the physical and diffusion network obtained from the Higgs Twitter Dataset that is employed for evaluation.

Table 4.2: Elements of the physical and diffusion network obtained from the Higgs Twitter Dataset.

Physical network features		Diffusion network features	
Feature	Value	Feature	Value
Number of Nodes	6,943	Number of Nodes	6,943
Number of Edges	223,006	Number of Edges	28,769
Number of Monitors	2,429	Number of interactions	41,522
Number of Colors	55,669		

Monitor placement. Applying the methodology on a directed network entails a revision in the monitor placement methodology, which considers both the in-degree and out-degree of the nodes. It is desired that the nodes selected as monitors receive a lot of information, but at the same time they also forward significant amounts of information. Thus, nodes with high values of in- and out-degree become the best candidates for monitoring, followed by the ones with high in-degree and low out-degree.

Information classes. Even though the topic included in the dataset (information) is unique, different classes of the diffused information are assumed based on its dynamics. The dataset provides only the observed actions of users, e.g., for users u_1, u_2, u_3: $u_1 \rightarrow u_2$ and $u_2 \rightarrow u_3$ (user u_3 retweets user u_2 who retweeted u_1). Taking into consideration the timestamps of retweets and meeting the condition that both u_2, u_3 do not interact with other users in the time period defined by the timestamps, it is concluded that the information flow is $u_1 \rightarrow u_2 \rightarrow u_3$. These sequential interactions will define an information class. The above criteria lead to the inference of 37,977 information classes. The large number of

Table 4.3: Evaluation of inference over real datasets.

Color sequences' length, number, precision and recall			
Length of sequences	Number of sequences	Precision	Recall
1	21092	0.0653	0.0653
2	1091	0.9743	0.9743
3	130	1	1
4	14	1	1
5	4	1	1
Monitorable interactions: 22.331		Number of information classes: 37.977	

classes is justified by the nature of *retweeting* as explained above (retweets between non-followers/followers without knowledge of the intermediary), which results in the creation of information classes containing sequences of unary length. As shown in Table 4.3, this is the most frequent type of information class we derive from this dataset.

Evaluation of inference. Due to the large number of information classes, the small number of interactions that each monitor node records (on average 9 interactions/monitor) and the unary length of the majority of the color sequences (leading to the cold start problem), statistical learning on the edges' use frequencies cannot be applied. Thus, only the probabilistic inference approach on interactions' flow level is examined and evaluated, since the network is directed. The results are cumulatively presented in Table 4.3.

As presented in subsection4.2.2.3, for the color sequences of unary length, where information begins from a non-sensor node and immediately (in one hop) hits a sensor, it is impossible to infer accurately the traversed edge. This is due to the coloring scheme which dictates the same color to the sensor's adjacent edges. However, even in these cases, performance is slightly improved compared to the corresponding one employing a purely probabilistic inference approach. On the contrary, the accuracy score of our inference scheme is, as expected, very high in the remaining sequences of greater length, signifying its effectiveness in realistic scenarios as well.

Monitoring the interplay of a small amount of OSN users can therefore result in an accurate inference of the diffusion network. The knowledge of users' interactions in terms of the diffusion dynamics, that is, socio-awareness, can be integrated to a Recommender System operating over the OSN in order to increase users' engagement to the platform, since, as claimed in [87], recommendations from friends are preferred compared to recommendations generated by designated systems. In such systems, which will be described in detail in section 4.3, information diffusion can be considered an auxiliary mechanism for recommendations in order to reduce the promotional cost of redundant items by which the Recommender System is encumbered.

4.3 Socio-aware Recommender Systems

A cumulative definition of a recommendation problem is given in [88] and it is based on the notion of utility function denoting the usefulness of an item for a user.

Definition 5. *(Recommendation Problem) Given a user $u \in U$, where U is the set of all users, and an item $i \in I$, where I is the set of all recommendable items, let $f(u, i)$ be a utility function that measures the usefulness of item i to user u, $f : U \times I \to R$, where R is a totally ordered set of non-negative real numbers or integers. The goal of the recommendation problem is to find for each user $u \in U$ the item $i' \in I$ over all items in I that maximizes the expected utility for this specific user u, namely an item such that* $\forall u \in U : i'_u = \arg\max_{i \in \mathcal{I}} f(u, i)$.

The two fundamental entities of a RS, i.e., the users and the items, form two distinct spaces, the user space and the item space respectively. The features in the user and item space are typically represented by the components of associated feature vectors, namely the user feature vector and the item feature vector. The utility function measures the value that an item has for a user, in the form of a user rating. User ratings are defined on a subset of space $U \times I$, and thus, the objective of a recommendation problem is to predict the missing values of f in the corresponding user-item utility matrix [88]. In contrast to this rating-based approach, other approaches denoted as preference-based techniques formulate the recommendation problem with the objective to predict the users' relative preference ordering between items, as implied by past rating values.

Recommender systems are quite diverse in terms of their function, varying in terms of their application domain, the knowledge used and the computational model adopted [89]. For this reason, several classifications emerge, with the model of numerical user ratings for items to be the most widely accepted for computing recommendations. According to this model, given explicit or implicit user feedback, a learning algorithm is applied to filter the user's attributes and predict personalized recommendation options. Regarding this approach, recommendation techniques can be classified into two broad categories: *content-based* and *collaborative filtering.* A Content-based RS aims at recommending items similar to items that a user has already chosen, while collaborative-filtering approaches, locate first a user with features similar to the user for which a recommendation is searched, and select some items among the ones that the similar user selected, providing them as recommendations to the first user [71]. Both of these approaches have several limitations, a summary of which can be found in [90, 91]. In general, a content-based system is highly dependent on the availability of the descriptive data. In addition, it is susceptible to over-specialization and exhibits the users' cold-start problem [88], which also appears as a main drawback along with sparsity and scalability issues in collaborative systems. Furthermore, collaborative-filtering systems fail to categorize users who have common preferences with more than one group of similar users, which may result in inaccurate recommendations, a problem known as the gray sheep problem [92].

The Recommendation Systems [91, 93] that are based on the bilateral relation between RS and Social Networks (SN) are referred to as Social Recommender Systems (SRSs). By leveraging the properties and features of the entities that constitute the network, as quantified by appropriate SNA metrics, they produce more accurate and efficient results. Contrary to traditional recommendation systems, SRSs take into consideration different relations among users, e.g., in the form of *trust* [90, 91], which is defined as the subjective expectation of one's future behavior [94]), or *influence*, using either structural analysis (location-based SNA) or behavioral analysis (interaction-based SNA) [95]. The correlation between ratings and social structure of recommendations is obtained by quantifying the above features. This results in the definition of new, social-based similarity measures, which alleviate the problems of sparsity and cold-start [96]. Trust relationships increase RS coverage when the application of traditional similarity measures is impossible (e.g., Pearson correlation coefficient can only be applied to users with high overlap in items' ratings [97]), while users with no previous ratings may receive accurate recommendations by connecting to a "trusted" or "influential" user of the network [98].

An aspect of socio-awareness that can be exploited for improving the performance of RSs is information diffusion, or features of this process as it evolves over a social network of users. These systems are known as Information Diffusion Aware Recommendation Systems (IDARS). The integration of the sharing mechanism of information in recommender techniques may tackle key problems of previous recommendation systems, such as the long-tail effect and information overload. It can also increase system efficiency in big data environments by combining the similarity-based predictions of traditional recommendation methods with the social-based predictions developed for the information propagation mechanisms.

Approaches on information diffusion-aware RS can be segregated into two categories according to the incorporation of the diffusion mechanism in recommender systems. An issue which comes as a consequence of the power-law distribution of product sales in online commerce, such as Amazon and Netflix [99], is the *long-tail problem*: Although a small set of products are extremely popular (e.g., hits/blockbusters), the less common products (niche products) exceed in aggregate the market share of the former [99, 100]. Therefore, it is essential for a RS to provide suggestions of long-tail products. This serves not only retailers, who experience greater profit from sales of otherwise unpopular items, but also individuals, who have access to more diverse products, which in turn increases their engagement to the online platforms.

Until now, due to data sparsity, RS weakly enhance item diversity. From the perspective of information diffusion, estimating users' influence and inferring how influence propagates in the network, may address the long-tail problem in RS, as follows: Recommending long-tail items to the most influential users of the network leads to the wide adoption of the former through the diffusion mechanism. In [101], a collaborative innovation diffusion-aware recommendation mechanism is proposed to improve novel knowledge sharing (innovation diffusion) between the users of a corporate portal, where 80% of the total traffic comes from 2% of unique pages. Recommendation of long-tail information to a portal user u who visits a page p is obtained based on the browsing behavior, during a specific time period, of a small portion of users (reference users) who are the first to access page p before user u. The preferences of the reference users and user's u browsing history are conjoined and evaluated with the metrics of *precision* and *long-tail precision*, which is defined as a measure of the novelty of webpages, and then sorted in order to suggest to user u the highest ranked ones.

The second category includes the methodologies that tackle the problem of information overload [102]. The exponential growth of information in online systems accompanied by the limited processing ability of users, impacts decision-making and increases the need for mechanisms of relevant information retrieval and delivery. Traditional RSs handle this problem by filtering large volume of generated data according to users' similar preferences, disregarding users' social relations in a network. From this point of view, different models of information diffusion serve either as an alternative [103], or an extension [104, 105, 106] of the filtering mechanism in RS, adding social criteria (influence, trust, homophily) in recommendations.

In [104] the correlation between information diffusion modelling and collaborative filtering is highlighted with a joint ranking-oriented model of recommendations, where it is assumed that both mechanisms make predictions of users' preferences from a different perspective. Using a latent factor model and assuming that users are highly influenced by their friends, the latent factor of user u who adopts item i will be very close to the ones who have adopted item i in the past. The predicting score of ratings is computed by taking into consideration users' similarity and the linear combination of several temporal and structural features of information cascades. In [105] the IDARS DifRec uses a graph-based model of recommended items' expected diffusion in a network to predict conflicting suggestions between RS and social diffusion. Redundant suggestions, namely, information that users will eventually receive through the activity of their friends in the OSN (e.g., likes, shares, retweets) are withheld in order to optimize recommendations in terms of the overall items-users relevance score and avoid information overload. This optimization problem is mapped to the Maximum Weighted Independent Set problem, which is proved to be NP-hard and it is addressed by a heuristic greedy algorithm. Although DifRec marks a step towards addressing the problem of information overload by reducing significantly the amount of duplicate recommendations, there are some aspects of information diffusion not considered, which may lead to users refraining from any activity or even lose interest for the OSN platform. For instance, the recommendation of many different products may ruin a user's decision making. Also, being targeted by multiple recommendations of the same item can eventually lead to frustration [107]. The difference in tolerance displayed by each user to repetitive advertisements of the same item or towards the promotion of different items is taken into account when allocating content in the form of recommendations in [106]. Furthermore, methodologies that minimize redundant recommendations may consume a large amount of resources from the recommendation engine, which can be critical in platforms with millions or billions of users. This is a key issue investigated insection 4.4. Realizing the significance of monitoring and tracking of information propagation in efficient recommendations, the information diffusion inference scheme proposed insection 4.2 is exploited to design an IDARS. By integrating the predicted information sharing to recommendations, minimal amounts of resources are utilized to place content that best match user preferences and respect their tolerance to information.

4.4 Efficient socio-aware recommendations under complex user constraints

As discussed in section 4.3, both explanatory and predictive modeling approaches of information diffusion are employed in IDARS to enhance their performance. Inspired by this and from the viewpoint of coverage, the mechanism of direct content assignments of the RS is combined with the one of indirect recommendations, driven by the information flow in the OSN, in order to provide an ℓ-cover of assignments to the network. An ℓ-cover is defined as a set of assignments to users who maximize both the spread of recommendations and the total user-to-item relevance, so that each user in the network is covered (recommended either directly or indirectly) by at least ℓ items. The parameter ℓ is chosen to be a moderate number of items that respects the users' limited ability for information filtering and decision making [108]. Recommendations from the perspective of coverage are mostly item-oriented and typically, not adopting socio-awareness. In [109, 110], the problem of recommendations is related to the Maximum Coverage Problem. In [109], users are distinguished to different types according to their preferences to items, where preferences can vary over time. A multi-armed bandit method is used to compute the set of items to satisfy all different types of users, whereas in [110] a greedy algorithm is designed to obtain a list of K products that maximize the probability of purchase in e-commerce sites. Closer to the notion of coverage used in the approach proposed in this dissertation, but from an item-centric viewpoint, in [17], an item is considered to be d-covered if it is recommended to at least d users and the trade-off between item coverage and accuracy of recommendations is studied. Nevertheless, the aforementioned approach overlooks the social features of users and their susceptibility to information overload. This dissertation fills in this gap by designing an iterative greedy algorithm for socio-aware recommendations based on the covering ability of a user, which is conceived as an alternative concept of influence and it is determined by her structural (degree centrality) and behavioral features (similarity between users in terms of user-to-item relevance).

4.4.1 Model of the OSN and the influence network

4.4.1.1 Online Social Network

An OSN is modeled as a directed labeled graph $G(V, E, w_V)$, which will be referred to as system graph, with a set of users $V = \{u_1, ..., u_n\}$, and a set of edges E representing the relations between individuals. Edge $(u_1, u_2) \in E$ dictates that user u_2 follows user u_1, so that information is flowing from u_1 to u_2. Users in V will be recommended items from the set $I = \{i_1, ..., i_k\}$ through two different mechanisms, a direct one, where items are recommended to users by the IDARS, and an indirect, where users recommend items to one another through information sharing. The term assignment is only used for direct recommendations. The influence of a node $u \in V$ to its neighbors in G, is represented by w_u, so that $w_V = \{w_u\}_{u \in V}$. For each user $u \in V$ and each item $i \in I$ the relevance of item i to user u is denoted as $r_{u,i}$.

4.4.1.2 Influence network

User interactions in the OSN regarding who influences whom, are captured by item-specific influence graphs as a result of the predictive diffusion model proposed in [105]. For item i the probabilistic graph $G_P^{(i)} = (V, E, p^{(i)})$ is defined in order to compute the Reliable Sets of its nodes: Consider a direct recommendation of item i to user u. The Reliable Set of $u \in V$ associated with i, $R_u^{(i)}$, is defined as the set of user's u one-hop or multiple-hops neighbors, who are expected to be indirectly recommended of item i via u. The probabilistic graph's edge probabilities are based on the fact that the influence graph cannot be the same for all items: a high influence score in a pair of users in the system graph does not imply that users' preference is similar for all items, thus, the probability graph should be item-dependent. For this, the edge probabilities of $G_P^{(i)}$ are computed based on the influence between users and the item-to-user relevance $r_{u,i}$, which results in item-determined Reliable Sets. The edge probabilities are:

$$p^{(i)}(u,v) = r_{v,i} \cdot w_u, \quad \forall (u,v) \in E. \tag{4.10}$$

A Monte Carlo-based simulation method is employed to compute the Reliable Sets. It is applied separately, for each probabilistic graph, to each of its nodes. In particular, for item i, the input of the Monte Carlo algorithm is a node u and the probabilistic graph $G_P^{(i)} = (V, E, p^{(i)})$. Then, in each iteration t, a graph $G^{(t)} = (V, E^{(t)})$ is produced. After all the necessary (user-specified) iterations have been completed, all nodes that are reachable from node u in more than a (user-specific) fraction of iterations, are included in its Reliable Set $R_u^{(i)}$. The influence graph of item i, $G_n^{(i)} = (V, E^{(i)})$, is produced by joining the Reliable Sets of all the nodes of the network. An edge is added between each node u and each of the corresponding nodes that lie in $R_u^{(i)}$. The closed Reliable Set of u for item i, which is the Reliable Set of u for i that includes u, (for simplicity reasons, throughout the dissertation, the term Reliable Set refers to the closed Reliable Set) is therefore formally defined as:

$$R_u^{(i)} = \{u\} \cup \left\{ \bigcup_{(u,z) \in E^{(i)}} z \right\}. \tag{4.11}$$

Then, for item i, the relevance score $E_{v,i}$ of the set $R_u^{(i)}$ is defined as the sum of the relevance scores $r_{z,i}$, $\forall z \in R_u^{(i)}$:

$$E_{v,i} = \sum_{z \in R_u^{(i)}} r_{z,i}. \tag{4.12}$$

Table 4.4 summarizes the basic notation employed in the conidered system of OSN and the influence network model presented in section 4.4.1.

4.4.2 IDARS problem statement and analysis

In [105], the problem under examination is the one of allocating (i.e., recommending directly) at most ℓ items to each user so that the total relevance score of the network is maximized. This is realized by taking into consideration the underlying information diffusion process, meaning that an item cannot be directly recommended to neighboring nodes in the influence graph. In this dissertation, the items' allocation problem is investigated by relaxing this restriction, in order to achieve efficient (using less resources) diffusion of items in the

Table 4.4: Notation employed in the problem of information diffusion aware recommendations

Parameter	Interpretation
G	The graph representing the OSN (system graph)
$V = \{u_1, ..., u_n\}$	Nodes of graph G, denote users of the OSN
E	Edges of graph G, directed, denote follow relationships
w_u	Influence of user $u \in V$ to its neighbors
$I = \{i_1, ..., i_k\}$	Set of items available for recommendation in the OSN
$r_{u,i}$	Relevance score of item $i \in I$ to user $u \in V$
$G_P^{(i)}(V, E, p^{(i)})$	Probabilistic weighted graph
$p^{(i)}(u, v)$	Weight of the edge (u, v) in $G_P^{(i)}$
$G_n^{(i)}$	Influence graph of item i produced from G, has the same nodes as G
$E^{(i)}$	Edges of influence graph $G_n^{(i)}$
$R_u^{(i)}$	Reliable set of user $u \in V$ (inc. u) associated with i
$E_{u,i}$	Relevance score of $R_u^{(i)}$ when recommending item $i \in I$ to user $u \in V$
$c(R_u^{(i)})$	The cost of assigning item $i \in I$ to user $u \in V$
$N_{in}^{(i)}(u)$	The set of in-neighbors of user u
$\mathbf{U}$	Universe of all possible assignments u_i^j
$\mathbf{F}$	Family of all the reliable sets $R_u^{(i)}$
U_j	Class of all possible assignments to user u_j
$H_{i,j}$	Inter-user diversity for users u_i, u_j
N	Novelty
G_{rev}	Bipartite graph of users and books
G_u	Real social network
G_U	The largest component of G_u
$N_{ego}(u)$	Set of users reachable from u in 1 or 2 hops in G_U
$rt(u, b)$	Normalized rating of user u to book b
$d(u, v)$	Length of the shortest path between u and v in G_U

network under a specific coverage objective, where each user should be covered by at least ℓ out of k total items. This corresponds to a generalization of the Minimum Weighted Set Cover Problem, called the Minimum Weighted Partition Set Cover Problem [111]. In the unweighted version of this problem, the input is a set system (A, B), where $A = \{a_1, ..., a_n\}$ and $\mathbf{B} = \{B_1, B_2, ..., B_m\}$ is a collection of subsets of A. Also, r subsets $C_1, ..., C_r$ of A and r integers $d_1, ..., d_r$ are considered. The goal is to find a sub-collection $\mathbf{B'} \subseteq \mathbf{B}$ such that, for $i = 1, ..., r$, the number of elements from C_i covered by $\mathbf{B'}$ is at least d_i, that is

$$\left| \bigcup_{D \in \mathbf{B'}} D \cap C_i \right| \geq d_i, \quad \forall i = 1, ..., r. \tag{4.13}$$

A special case of the Minimum Weighted Partition Set Cover Problem, called Minimum Weighted ℓ-Cover, is formulated, where the subsets C_i, $i = 1, ..., r$, form a partition of A and $d_i = \ell$, $\forall i = 1, ..., r$. In particular, the universe of $n \times k$ elements:

$$\mathbf{S} = \{u_1^{(1)}, u_1^{(2)}, ..., u_1^{(k)}, ..., u_n^{(1)}, u_n^{(2)}, ..., u_n^{(k)}\}, \tag{4.14}$$

is considered to be the set of all possible assignments of items in I to users in V, with $u_j^{(i)}$ indicating the assignment of item i to user u_j. The classes

$$U_j = \bigcup_{m=1}^{k} u_j^{(m)}, \quad j = 1, ..., n, \tag{4.15}$$

form a partition of $\mathbf{S}$ and they represent all the candidate assignments of items to a specific user. Likewise, a family of $n \times k$ subsets of $\mathbf{S}$,

$$\mathbf{F} = \{R_{u_1}^{(1)}, R_{u_1}^{(2)}, ..., R_{u_1}^{(k)}, ..., R_{u_n}^{(1)}, ..., R_{u_n}^{(k)}\} \tag{4.16}$$

is considered. This models the eventual outcomes of the diffusion process. The set $R_{u_j}^{(i)}$ represents the recommendations obtained by assigning item i to user u_j. Item i will reach the neighbors of u_j in $G_n^{(i)}$ as determined by the corresponding Reliable Set.

The cost of recommendations is a function associated with $\mathbf{F}$ defined as $c : \mathbf{F} \to \mathbb{R}$ with

$$c(R_u^{(i)}) = \frac{1}{E_{u,i} + 1}, \forall i \in I, \ \forall u \in V. \tag{4.17}$$

At this point, it should be noted that the choice of the cost function depends on the objective(s) that the RS aims to meet. In the case examined, the problem of recommendations is treated as an assignment problem on influence graphs with coverage constraints. This leads to modeling the recommendation cost as the cost of assignments, that is, the cost of direct recommendations. It is assumed that indirect recommendations, namely, the ones realized by users' information sharing, have zero cost from the recommender system's perspective, since the diffusion of an item does not require the involvement of the recommender system (it happens due to the users' interplay and thus does not use any of the resources of the recommender system). The assignment cost of item i to user u, $c(R_{u_j}^{(i)})$, is defined as a function of the relevance score of the neighbors of u in the influence graph of i. Thus, the cost of assigning item i to user u is determined by both the influence of node u over her neighbors and their relevance to item i.

The goal is to find a minimum weighted ℓ-cover, that is, a collection of subsets $\mathbf{F}' \subset \mathbf{F}$ bearing the minimum total cost such that for $e \in \mathbf{F}'$

$$\left| \bigcup_{\mathbf{e} \in \mathbf{F}'} \mathbf{e} \cap U_j \right| \geq \ell, \quad \forall j = 1, ..., n. \tag{4.18}$$

By stating that the network is ℓ-covered by items means that each user received at least ℓ different items, equivalently, each user is ℓ-covered. Eventually, this means that removing an assignment from the network will result in some users receiving less than ℓ items.

The ℓ-Coverage Problem presented above can be formulated as a Nonlinear Integer Programming Problem due to nonlinear constraints. Let $\mathbf{X}$ be a matrix of zeros and ones with element x_{ij} representing the direct assignment of item i to user u_j. In particular:

$$x_{ij} = \begin{cases} 1, & \text{if } R_{u_j}^{(i)} \text{ is in the } \ell\text{-cover}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.19}$$

The problem of finding a collection of ℓ different items per user that minimizes the total cost of recommendations in the network, is formulated as follows:

P1:

$$\arg\min_{i,j} \sum_{i,j} w_{ij} \cdot x_{ij}, \tag{4.20}$$

subject to:

$$\sum_i f(x_{ij}) \geq \ell, \quad \forall j = 1, ..., n, \tag{4.21}$$

where $w_{ij} = c(R_{u_j}^{(i)})$ is the cost of assigning item i to u_j and

$$f(x_{ij}) = \begin{cases} 1, & \text{if } \sum_{k: u_j \in R_{u_k}^{(i)}} x_{ik} \geq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.22}$$

Constraint (4.21) ensures that each user will receive recommendations of at least ℓ different items. The mapping $f(x_{ij})$ captures the recommendation of item i to user u_j. Its value is set equal to 1 if the item was received either directly from the IDARS or indirectly from one or multiple sources, and zero otherwise.

We formulate an Integer Linear Program (ILP) equivalent to P1 in order to apply an LP-based Branch and Bound method for its solution. Thus, constraint (4.21) is linearized and two new constraints (4.28), (4.29) are added. These are necessary when the integer restrictions are relaxed. The non-linear mapping $f(x_{ij})$ is relaxed by introducing a new binary variable y_{ij}, which is defined as follows:

$$y_{ij} \geq \frac{\sum_{r=1}^{n} x_{ir}\lambda_{rj}^{(i)}}{\sum_{r=1}^{n} \lambda_{rj}^{(i)}}, \tag{4.23}$$

$$y_{ij} \leq \frac{\sum_{r=1}^{n} x_{ir}\lambda_{rj}^{(i)}}{\sum_{r=1}^{n} \lambda_{rj}^{(i)}} + 0.\overline{9}, \tag{4.24}$$

where

$$\lambda_{rj}^{(i)} = \begin{cases} 1, & \text{if } u_j \in R_{u_r}^{(i)}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.25}$$

Hence, the corresponding ILP is formulated as:

P2:

$$\arg\min_{i,j} \sum_{i,j} w_{ij} \cdot x_{ij}, \tag{4.26}$$

subject to:

$$\sum_{i} y_{ij} \geq \ell, \quad \forall j = 1, ..., n, \tag{4.27}$$

$$\sum_{i,j} x_{ij} \geq \ell, \tag{4.28}$$

$$\sum_{i} x_{ij} \leq \ell, \quad \forall j = 1, ..., n. \tag{4.29}$$

Constraint (4.28) ensures that at least ℓ direct assignments in total are made by the Recommender System, whereas constraint (4.29) sets an upper bound to the number of direct assignments per user.

4.4.3 Greedy Algorithm: CoveR

Due to the fact that it is computationally difficult to produce optimal results by solving the problem P1 described in section 4.4.2, the implementation of an approximation algorithm that produces results close to the optimal solution, in terms of the total defined cost of recommendations, is justified. A greedy algorithm is adopted for the Minimum Weighted ℓ-Cover Problem. Let X_t be the set of all possible assignments at the beginning of iteration t, with $X_1 = \mathbf{S}$, since no assignments of items to users have been made yet. In iteration t, the algorithm selects a user to assign an item, that is, the set from $\mathbf{F}$ that minimizes the ratio:

$$\frac{c(R_u^{(i)}) \cdot |R_u^{(i)}|}{|R_u^{(i)} \cap X_t|}, \quad \forall u \in V, i \in I. \tag{4.30}$$

An important distinction between the ℓ-Coverage and the Minimum Weighted Partition Set Cover problem should be underlined: In the ℓ-Coverage, the cost of a set $R_u^{(i)}$ is determined by the relevance score of its elements. Nevertheless, the relevance score of an item-user pair that appears in more than one of the selected sets should contribute to the total cost once, no matter how many times user u was exposed to i (the relevance score $r_{u,i}$ is included at most once in the total relevance score). Thus, the proposed method should penalize the assignments that cause duplicate recommendations to users. CoveR addresses this issue by assuming that every set $R_u^{(i)}$ has an initial covering ability $|R_u^{(i)}|$ of cost $c(R_u^{(i)})$ and the covering ability of the set at iteration t equals to $|R_u^{(i)} \cap X_t|$. By selecting the set with the lowest ratio in Eq. (4.30), the method shows preference to sets of low cost that are also capable of covering large parts of the network. Intuitively, CoveR endorses the assignment of items to users that have substantial influence to uncovered users, while minimizing the total cost of recommendations, so that every user acquires at least ℓ items. A pseudocode for the algorithm is provided in Fig. 4.14.

CoveR operates as follows: Having computed the cost of each possible assignment, the algorithm picks the set $R_u^{(i)}$ that minimizes the ratio in Eq. (4.30). Then, $R_u^{(i)}$ is included in $\mathbf{F}'$, which denotes the set of recommendations that are carried out by the RS. It should be noted that each class U_j has either one element in common with $R_u^{(i)}$ or none. All the classes of non-empty intersection with $R_u^{(i)}$ are checked to see whether they are ℓ-covered. Consider $v^{(i)} \in R_u^{(i)}$ and the class associated with v, $U_v = \cup_{i=1}^{k} v^{(i)}$. If U_v is not ℓ-covered (line 12 in pseudocode), only $v^{(i)}$ is removed from X_t, thus the candidate assignments for the next iteration are $X_{t+1} = X_t \setminus v^{(i)}$. If class U_v is ℓ-covered (line 10 in pseudocode), the algorithm removes from X_t all the elements of this class, so that the associated user will not be highly preferred for assignments in future iterations. In this case, $X_{t+1} = X_t \setminus \cup_{i=1}^{k} v^{(i)}$. The algorithm terminates when $X_t = \emptyset$.

Algorithm CoveR

Inputs: The family $\mathbf{F}$ of sets $\{R_u^{(i)}\}_{u \in V}^{i \in I}$, the relevance score $r_{u,i}$, $\forall u, i$, the universe of assignments, $\mathbf{S}$, the classes $U_j = \cup_{i=1}^{k} u_j^{(i)}$, $\forall j = 1, ..., n$ and ℓ, the minimum number of items to be recommended to a user.

Outputs: A binary matrix, $M = [m_{iu}]_{|I| \times |V|}$ of item-user recommendations, the set of assignments $\mathbf{F'}$, the total Relevance Score, $Score$.

1: Initialization: $\mathbf{X} \leftarrow S$ is the set of candidate assignments, $m_{iu} = 0$, $\forall i \in I$, $\forall u \in V$, $\mathbf{F'} = \emptyset$, $Score = 0$.
2: **for** each $u \in V$, $i \in I$ **do**
3: $\quad E_{u,i} = \sum\limits_{v \in R_u^{(i)}} r_{v,i}$
$\quad c(R_u^{(i)}) = \frac{1}{1+E_{u,i}}$
4: **end for**
5: **while** $\mathbf{X} \neq \emptyset$ **do**
6: $\quad$ Select the set $R_{u_j}^{(i)}$ that minimizes the ratio $\frac{c(R_u^{(i)}) \cdot |R_u^{(i)}|}{|R_u^{(i)} \cap X|}$ and has non-empty intersection with $\mathbf{X}$. If two or more sets have the same ratio, pick one uniformly at random.
7: $\quad \mathbf{F'} = \mathbf{F'} \cup R_{u_j}^{(i)}$
8: $\quad$ **for** each $v \in R_{u_j}^{(i)}$ **do**
9: $\quad\quad m_{iv} = 1$
10: $\quad\quad$ **if** $\sum_{i \in I} m_{iv} \geq \ell$ **then**
11: $\quad\quad\quad \mathbf{X} = \mathbf{X} \setminus \cup_{i=1}^{k} v^{(i)}$ % *remove all the elements of the class associated with user v*
12: $\quad\quad$ **else**
13: $\quad\quad\quad \mathbf{X} = \mathbf{X} \setminus v^{(i)}$ % *remove only $v^{(i)}$ among the elements in user's v class*
14: $\quad\quad$ **end if**
15: $\quad$ **end for**
16: **end while**
17: **for** each $m_{iu} \in M$ **do**
18: $\quad$ **if** $m_{iu} == 1$ **then**
19: $\quad\quad Score = Score + r_{u,i}$
20: $\quad$ **end if**
21: **end for**

Figure 4.14: Algorithm CoveR: Greedy algorithm for recommendations, ensuring that the network is ℓ-covered by recommendations.

4.4.4 Approximation ratio of CoveR

The performance of Cover is analyzed in terms of the cost of the produced assignments. In order to formally define how close the solution of CoveR is to the optimal solution, an approximation guarantee is provided. The approximation ratio sets an upper bound on the deviation of a solution generated by CoveR from the optimal solution.

Proposition 4.4.1

CoveR approximates the ℓ-Coverage Problem with performance ratio satisfying

$$\frac{c_{greedy}}{c_{opt}} \leq \frac{\Delta}{\delta} \cdot H(\Delta), \tag{4.31}$$

where $\delta = \min_{u \in G_n} deg^{(out)}(u) + 1$ and $\Delta = \max_{u \in G_n} deg^{(out)}(u) + 1$ denote the minimum and maximum out-degree of the influence graph G_n when self-edges are added to all of its nodes, thus the minimum and maximum cardinality of the sets in $\mathbf{F}$ correspondingly.

Proof. Let $\{(i_1, u_{j_1}), (i_2, u_{j_2}), ..., (i_s, u_{j_s})\}$ be the tuples of indices of the sets that the optimal solution uses to form an ℓ-cover of $\mathbf{S}$ with the minimum cost c_{opt}. To simplify the notation, the sets of this ℓ-cover are denoted by $\{O_1, ..., O_s\}$ and their corresponding costs by $\{o_1, ..., o_s\}$, where $c_{opt} = o_1 + ... + o_s$. Without loss of generality [112], the sets are assumed to be disjoint and that the following equality holds:

$$\sum_{x=1}^{s} |O_x| = \ell \cdot n, \tag{4.32}$$

which means that the optimal ℓ-cover manages to cover each class U_i, $i = 1, ..., n$ with ℓ items. Thus, exactly ℓ items are recommended to each user.

Let $\{(i_1, u_{j_1}), (i_2, u_{j_2}), ..., (i_m, u_{j_m})\}$ be the tuples of indices of the sets from $\mathbf{F}$ that the solution of CoveR uses to form an ℓ-cover of $\mathbf{S}$ with cost c_{greedy}. For simplicity purposes, the sets of this ℓ-cover are denoted by $\{R_1, ..., R_m\}$ and their corresponding costs by $\{r_1, ..., r_m\}$, where $c_{greedy} = r_1 + ... + r_m$. It is assumed that in iteration $i, i \in \{0, ..., m-1\}$, the algorithm chooses the set of index i.

At the end of iteration t of the greedy algorithm, let Z_t denote the set of the remaining elements in uncovered classes and z_t to be the number of elements that remain to be selected by CoveR for all the classes to be ℓ-covered. It should be highlighted that $|Z_t| \geq z_t$ at every iteration. In every iteration $(t+1)$, CoveR selects the tuple of indices (i, u_j) that minimizes the ratio

$$\frac{r_{t+1} \cdot |R_{t+1}|}{|R_{t+1} \cap Z_t|}. \tag{4.33}$$

Hence, the minimum ratio set that is greedily picked at iteration $(t+1)$ must have a ratio that is at most the minimum among the sets of the optimal solution after iteration t. Formally,

$$\frac{r_{t+1} \cdot |R_{t+1}|}{|R_{t+1} \cap Z_t|} \leq \frac{o_x \cdot |O_x|}{|O_x \cap Z_t|}, \forall x \in \{1, ..., s\} : O_x \cap Z_t \neq \emptyset. \tag{4.34}$$

All fractions of the form

$$\frac{o_x \cdot |O_x|}{k_x}, \ \ \forall x \in \{1, ..., s\}, \ \ \forall k_x \in \{1, ..., |O_x|\} \tag{4.35}$$

are considered. These fractions model, for each set O_x, $x = 1, ..., s$, all its possible contributions in coverage (a set may cover from 1 to $|O_x|$ elements at iteration t). Due to Eq. (4.32), there are $z_0 = \ell \cdot n$ fractions in total, since there are $|O_x|$ fractions for each $x \in \{1, ..., s\}$. The fractions are sorted in non-increasing order $f_1 \geq f_2 \geq ... \geq f_{z_0}$. Then, from Lemma 1 in [112], the following inequality holds:

$$\frac{r_{t+1} \cdot |R_{t+1}|}{|R_{t+1} \cap Z_t|} \leq f_{z_t}, \forall t \in \{0, ..., m-1\}. \tag{4.36}$$

Consequently, from Lemma 2 in [112], it holds that:

$$r_1 \cdot |R_1| + ... + r_m \cdot |R_m| \leq o_1 \cdot |O_1| \cdot H(|O_1|) + ... + o_s \cdot |O_s| \cdot H(|O_s|), \tag{4.37}$$

where $H(i) = 1 + ... + \frac{1}{i}$, is the i^{th} harmonic number. The term $|O_i| \cdot H(|O_i|)$ is the sum of all possible contributions in coverage, from 1 to $|O_i|$ elements, of the set O_i. The minimum and maximum out-degree of the influence graph G_n are denoted by $\delta = \min_{u \in G_n} deg^{(out)}(u) + 1$ and $\Delta = \max_{u \in G_n} deg^{(out)}(u) + 1$, correspondingly. Likewise, in the case of item-specific influence graphs, the minimum and maximum out-degree of nodes among all the influence graphs are selected. In δ and Δ the degree values are increased by 1, thus, a self-edge is added in every vertex of the influence graph in order to model the minimum and maximum cardinality of the sets in $\mathbf{F} = \{R_1, ..., R_m\}$. Combining the above with inequality (4.37) and the facts that

$$|R_i| \geq \delta, \ \ \forall i \in \{1, ..., m\}, \tag{4.38}$$

$$|O_i| \cdot H(|O_i|) \leq \Delta \cdot H(\Delta), \ \ \forall i \in \{1, ..., s\}, \tag{4.39}$$

the following holds:

$$\begin{aligned} r_1 \cdot \delta + ... + r_m \cdot \delta \leq r_1 \cdot |R_1| + ... + r_m \cdot |R_m| \leq \\ o_1 \cdot |O_1| \cdot H(|O_1|) + ... + o_s \cdot |O_s| \cdot H(|O_s|) \leq \\ o_1 \cdot \Delta \cdot H(\Delta) + ... + o_s \cdot \Delta \cdot H(\Delta). \end{aligned} \tag{4.40}$$

By the transitivity property, from inequality (4.40), the following is obtained:

$$(r_1 + ... + r_m) \cdot \delta \leq (o_1 + ... + o_s) \cdot \Delta \cdot H(\Delta). \tag{4.41}$$

Therefore, it holds that:

$$c_{greedy} \leq \frac{\Delta}{\delta} \cdot H(\Delta) \cdot c_{opt}. \tag{4.42}$$

□

4.4.5 Evaluation of CoveR

The evaluation of CoveR's performance is realized via modeling and simulation, in terms of various metrics. The obtained results are compared with the ones produced by DifRec and a baseline Recommender System, which provides to every user the ℓ items of her highest relevance without taking into consideration the users' interactions (it is therefore diffusion unaware). The results of CoveR are also compared with the ones of an LP-based Branch and Bound (BnB) method [113] applied to the Integer Linear Problem (Problem P2) described in section 4.4.2. The networks used for the experimental evaluation are both synthetic and real. Synthetic networks are modeled by scale-free (SF) and small-world (SW) graphs, which are the most appropriate for representing the structure and evolution of the typically observed social networks [23, 114]. The results are averaged over 25 distinct topologies for each configuration presented in Table 4.5. The user-item relevance $r_{u,i}$ follows a uniform distribution between 0 and 1.

The performance of CoveR is evaluated by measuring the total relevance score, which, combined with the average number of direct and indirect recommendations, reflects the quality of users' experience and the system's efficiency to recommendations. The number of direct and indirect recommendations models the resources consumed by the recommender system and the extent of the exploitation of the diffusion process by the RS, respectively. Another important metric employed is coverage, which is defined in terms of the average number of items recommended to each user. Coverage quantifies the impact of diffusion awareness in recommendations by measuring the informational burden on users. Small values of coverage (but above the coverage goal) imply the absence of information overload to the network, therefore, greater user satisfaction and engagement to the platform. Additional metrics for assessing CoveR's performance are novelty and inter-user diversity [115], as well as the global diversity of recommendations, which is defined as the average number of recommended items in the network.

The following figures display the results that correspond to $\ell = 5$, as the most indicative ones. It should also be noted that in all the stacked bar plots, both primary bars and sub-bars start from $y = 0$, in the vertical axis y.

By the results presented in the following subsections, it is evident that the BnB method yields results very close to the ones produced by CoveR, validating its theoretically-obtained performance results. It should also be mentioned that the search for the optimal solution via BnB, oftentimes and especially when the search space increases (e.g., in networks with 300 nodes and beyond), is time consuming, so the results presented are the feasible solutions achieved by BnB.

Table 4.5: Summary of Experimental Options and Choices for CoveR

Parameters	**Available Options**			
# Users	100	200	300	
# Items	20	40	100	
ℓ	1	2	3	5
Network type	SF	SW		

4.4.5.1 Results on synthetic networks

Efficiency of recommendations. The notion of efficiency in RSs is introduced with respect to the achieved relevance score combined with the number of direct and indirect recommendations. An efficient approach will manage to make few assignments, thus, few direct recommendations, however of great impact in terms of relevance score, by highly exploiting the users' interplay that is translated to many indirect recommendations. CoveR can be more efficient than DifRec in that sense, as shown by the values of the average relevance score per assignment, which is presented in Table 4.6. This is due to a major difference in the mechanism of assignments: As opposed to CoveR, in DifRec, in order to avoid redundant recommendations, it is forbidden for any two neighboring nodes to be directly recommended of the same item, which may result in coverage failures, as shown in Fig. 4.15.

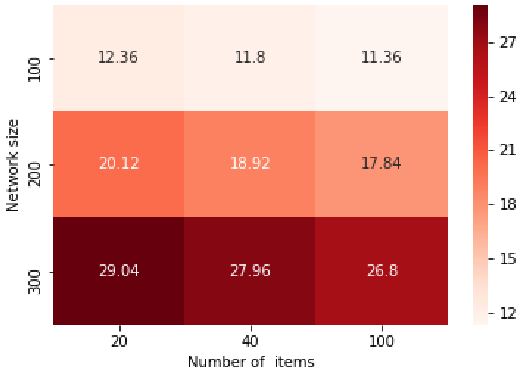


Figure 4.15: Average number of users who did not receive recommendations in SF networks of varying size for ℓ=5 with DifRec. Similar results are observed in SW networks.

Relaxing this constraint in CoveR, allows balancing the exploitation of the behavioral features (information diffusion) with the structural characteristics (users' social ties) of the network. Especially in networks of highly skewed degree distribution (SF topologies), which models the extreme heterogeneity of users in terms of influence (users may be classified to highly influential or insignificantly influential), CoveR's assignments that are equal to 18% of those made by DifRec (Fig. 4.16) lead to a major diffusion effect (Fig. 4.17). In turn, this results in a high total relevance score depicted in Fig. 4.18, outperforming DifRec. Even when heterogeneity smoothens, as represented by the degree distribution of SW networks, meaning that influencers of multiple scales are present in the OSN, CoveR achieves diffusion results comparable to DifRec (Fig. 4.19) with approximately 25% of DifRec's direct recommendations (Fig. 4.20). These are also reflected in the relevance score, as well as the relevance score per assignment, in Fig. 4.21 and Table 4.6 respectively. Regarding the baseline model, the results in Table 4.7 show that it achieves the highest relevance scores in both scale-free and small-world topologies. This is anticipated, as in the baseline approach, each node is simply recommended of its top ℓ preferred items. This demands $|V| \cdot \ell$ direct assignments from the information-unaware recommender system, contrary to CoveR, which makes less than $|V|$ assignments in all the examined cases, outperforming the former in terms of the average relevance score per assignment (Table 4.6). This further results in each user, u, to receive a great number of indirect recommendations (Table 4.7), at least $\ell \cdot |\bigcap_{i\in I} N_{in}^{(i)}(u)|$, where $N_{in}^{(i)}(u)$ is the set of in-neighbors of u for item i. This amount of recommendation probes is also reflected in the number of items per user that takes values in the interval $[3.6 \cdot \ell, 13.1 \cdot \ell]$, which are significantly larger than CoveR or DifRec, and it is likely to ruin the decision making process of a user [102] and therefore, her engagement to the platform. Furthermore, it is observed that the number of indirect recommendations in CoveR is independent of the available number of items, implying a more steady and predictable performance in different settings, as the diffusion process is not disrupted by either a possible abundance or lack of items. All of the above suggest that CoveR is able to produce efficient recommendations, meaning that a few targeted assignments of items spread throughout the network and all users end up with recommendations of high overall relevance.

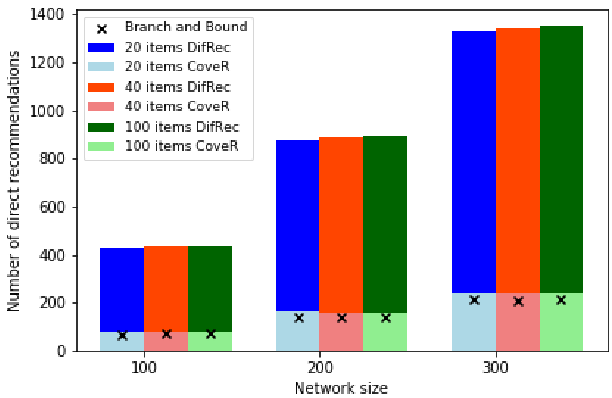


Figure 4.16: Average number of direct recommendations in SF networks of varying size for ℓ=5 with CoveR, DifRec and Branch and Bound.

Coverage of network and recommendation diversity. CoveR sets the lower bound in

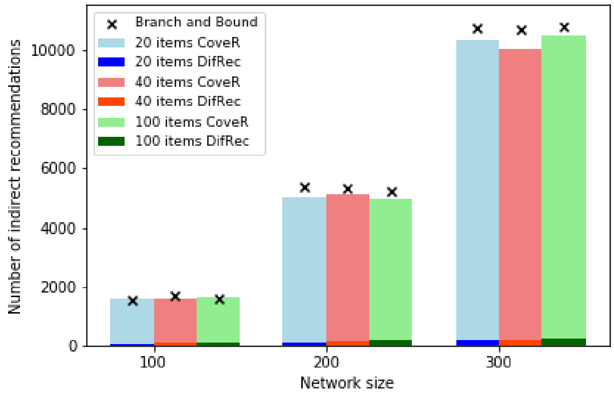


Figure 4.17: Average number of indirect recommendations in SF networks of varying size for ℓ=5 with CoveR, DifRec and Branch and Bound.

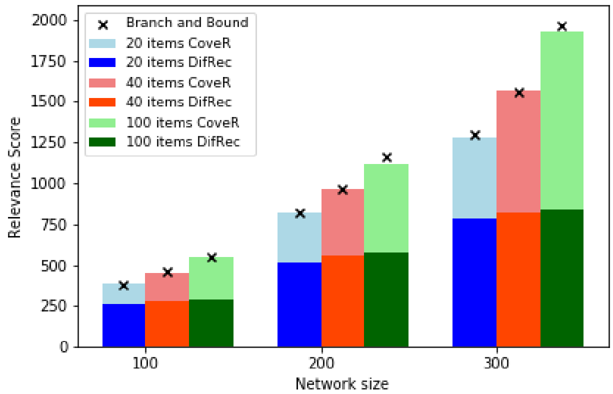


Figure 4.18: Relevance Score of recommendations in SF networks of varying size for ℓ=5 with CoveR, DifRec and Branch and Bound.

recommendations equal to ℓ, namely, it ensures that at least ℓ items will be recommended to each user of the network, either directly or indirectly. On the contrary, DifRec fails to make recommendations to approximately 10% of the users in both SF and SW networks (as shown in Fig. 4.15). In the examined topologies, CoveR also bounds from above the number of recommendations per user to $2.6 \cdot \ell$, as validated by Figs. 4.22 and 4.23. For moderate values of ℓ, this is a manageable amount of information for a user to process [116]. In DifRec the upper bound seems to be topology-dependent reaching $3.66 \cdot \ell$ in SW networks (Fig. 4.23). Moreover, the BnB method also provides a solution that manages to assign at least ℓ items per user, which is anticipated due to constraint (4.27) of the ILP problem P2 in section 4.4.2.

Table 4.6: Average relevance score per assignment with CoveR, DifRec and Baseline Recommender System

	CoveR		**DifRec**		**Baseline**	
(users,items)	SF	SW	SF	SW	SF	SW
(100,20)	4.78	7.65	0.61	1.63	1.66	1.79
(100,40)	5.67	9.39	0.66	1.92	2.78	3.45
(100,100)	6.92	11.23	0.66	2.32	4.17	6.67
(200,20)	4.98	4.98	0.59	1.59	1.72	1.79
(200,40)	6.05	5.98	0.63	1.87	3.12	3.33
(200,100)	6.98	6.99	0.64	2.22	5.26	6.67
(300,20)	5.34	5.06	0.59	1.56	1.75	1.79
(300,40)	6.44	6.07	0.61	1.84	3.22	3.23
(300,100)	8.02	6.93	0.62	2.17	5.88	6.67

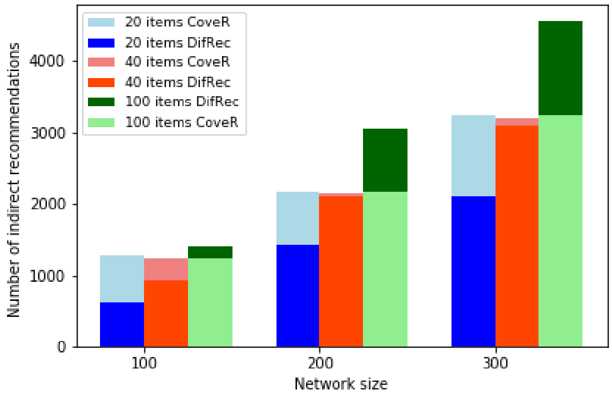


Figure 4.19: Average number of indirect recommendations in SW networks of varying size for ℓ=5 with CoveR and DifRec.

In addition to achieving the ℓ-coverage of the network, i.e., leaving no user with less than ℓ recommendations, CoveR also manages to recommend various different items across the OSN. This is an important aspect of any recommender system, as it affects both the users who wish to be recommended of a variety of items of their preference, and the retailers who wish to see their products advertised and gaining revenue. Figs. 4.24 and 4.25 present the number of different items that were recommended at least once in the network for SF and SW topologies, respectively. From these plots, it is observed that CoveR manages to assign numerous items throughout the network, oftentimes the majority of the ones available, ensuring diversity in recommendations. This capability of CoveR is further stressed by the fact that it is achieved by making significantly less assignments to users than DifRec, as it can be verified by Figs. 4.16 and 4.20 that display the direct recommendations performed in each experimental scenario. In this way, CoveR constitutes a less intrusive solution for

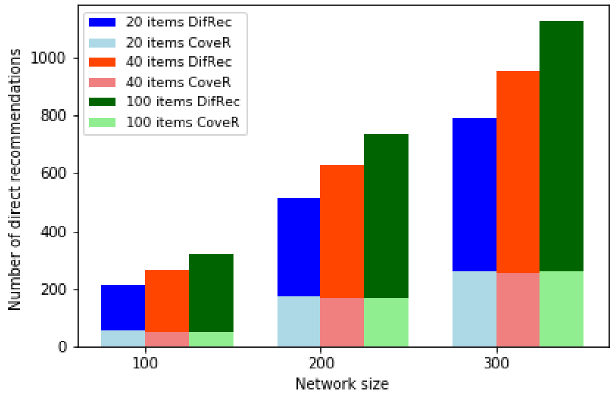


Figure 4.20: Average number of direct recommendations in SW networks of varying size for ℓ=5 with CoveR and DifRec.

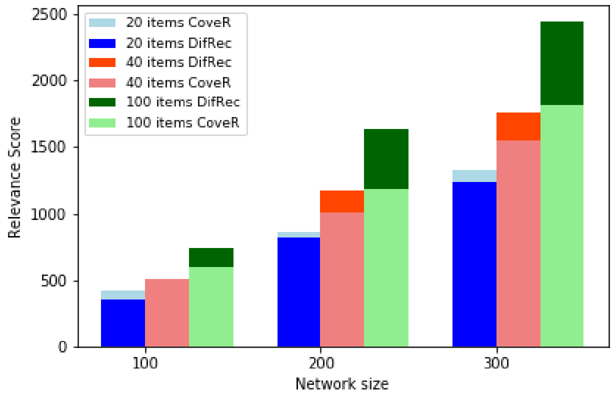


Figure 4.21: Relevance Score of recommendations in SW networks of varying size for ℓ=5 with CoveR and DifRec.

IDARS but competitively effective from the viewpoint of diverse recommendations. This is also confirmed by the values of inter-user diversity and novelty [115], presented in Fig. 4.26. Given users u_i and u_j, the corresponding inter-user diversity is given by the difference between the top-l places of their recommendation lists, as measured by the Hamming distance H_{ij}:

$$H_{ij}(l) = 1 - \frac{Q_{ij}}{l}, \tag{4.43}$$

where $Q_{ij}(l)$ is the number of common items in the top-l places of the recommendation lists of u_i and u_j. Given the top-l items' recommendation lists, Q_u, $\forall u \in V$, novelty, N, is

Table 4.7: Summary of Results in a baseline Recommender System

	items/user		**relevance score**		**indirect recs**	
(users,items)	SF	SW	SF	SW	SF	SW
(100,20)	18.22	18.98	837.68	900.35	4660.8	15179
(100,40)	29.91	35.55	1389.91	1699.09	4660.8	15179
(100,100)	43.36	68.95	2050.68	3327.79	4660.8	15179
(200,20)	18.73	18.79	1730.22	1785.87	15032.6	41931.8
(200,40)	34.08	34.64	3120.90	3294.45	15032.6	41931.8
(200,100)	57.41	70.18	5252.47	6636.98	15032.6	41931.8
(300,20)	18.67	18.52	2612.71	2677.78	29311.8	72059.8
(300,40)	35.04	33.28	4810.30	4795.98	29311.8	72059.8
(300,100)	65.72	68.09	8893.19	9689.11	29311.8	72059.8

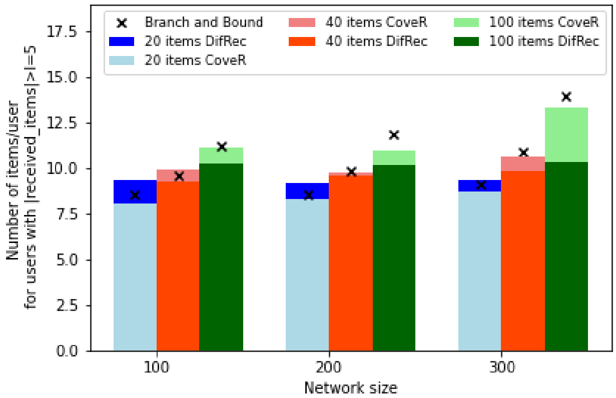


Figure 4.22: Average number of items-per-user for users who received more than ℓ=5 recommendations in SF networks of varying size with CoveR, DifRec and Branch and Bound.

defined as follows:

$$N = 1 - \frac{1}{|V| \cdot l \cdot |I|} \sum_{u \in V} \sum_{i \in Q_u} K_i, \tag{4.44}$$

where K_i is the ranking of item i when items are arranged in increasing order of relevance to the corresponding user.

4.4.5.2 Results on a real network

In order to demonstrate the applicability of CoveR in real world scenarios, the LibraryThings dataset [117, 118], is employed. Users of the platform can form friendships and review book titles, also awarding them star ratings, ranging from 0 to 5 stars. This dataset can be employed in the evaluation of an IDARS system. In order to obtain the influence graph that

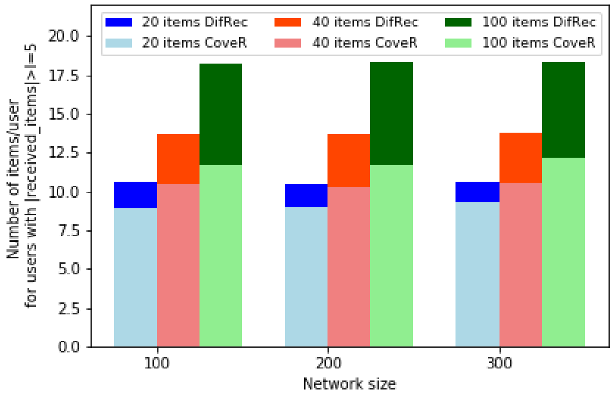


Figure 4.23: Average number of items-per-user for users who received more than ℓ=5 recommendations in SW networks of varying size with CoveR, DifRec and Branch and Bound.

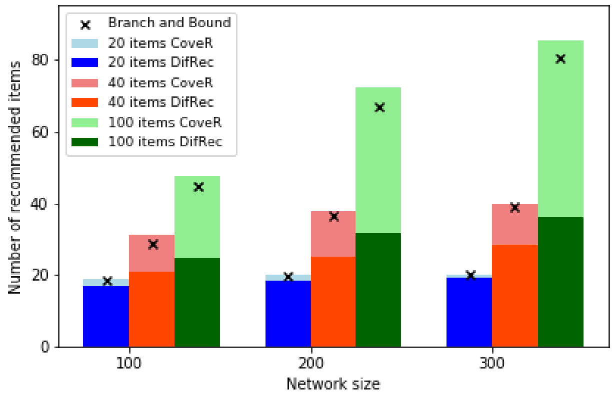


Figure 4.24: Average number of recommended items in SF networks of varying size for ℓ=5 with Cover and DifRec.

is necessary for the operation of both CoveR and Difrec, the following procedure is executed: First, the dataset is cleaned of any review entries that did not contain star ratings. Then, a bipartite graph is formed. The graph contains books and users and its edges are formed by connecting each user with her reviewed books. The largest connected component, graph G_{rev}, is identified and the top k books (i.e., the k books having the most reviews) are found. Moreover, the social network, G_u, denoting the friendships among users, is formed. In the next step, an undirected graph is created by joining those users that have reviewed the same book and are also neighbors in the social graph. Once again, the largest connected component, G_U, is identified. Regarding the relevance scores for each user-book pair, the star ratings are used. If a user has explicitly rated a book, its relevance score is equal to

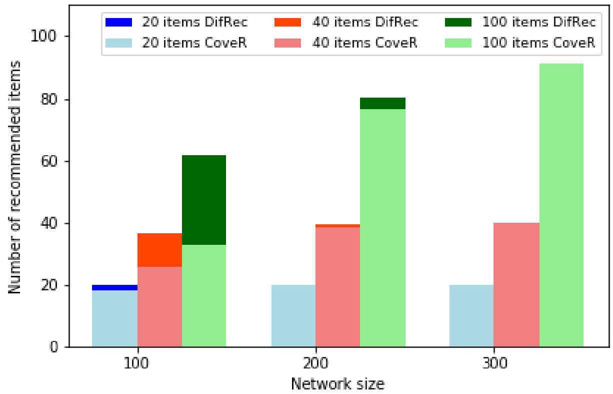


Figure 4.25: Average number of recommended items in SW networks of varying size for ℓ=5 with Cover and DifRec.

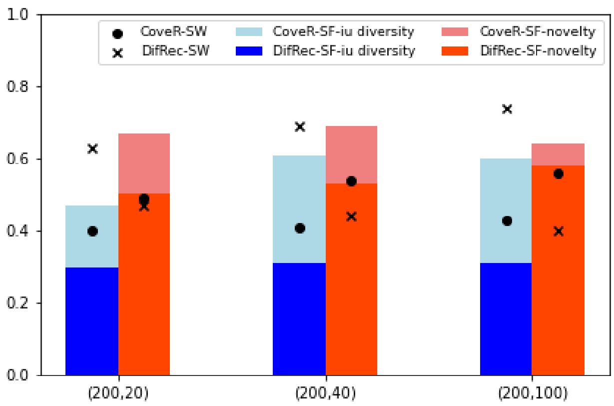


Figure 4.26: Average novelty and inter-user (iu) diversity in SF and SW Networks with CoveR and DifRec.

its normalized rating in $[0, 1]$. If user u has not reviewed book b, the set of nodes who have rated b and are reachable from u in either one or two hops in G_U, denoted as $N_{ego}(u)$, are detected and the relevance score of user u to book b, is given by the formula:

$$r_{u,b} = \frac{1}{|N_{ego}(u)|} \cdot \sum_{v \in N_{ego}(u)} \frac{rt(v,b)}{d(u,v)}, \tag{4.45}$$

where $rt(v, b)$ is the normalized rating of user v for book b and $d(u, v)$ is the shortest path length between nodes u, v. Also, for every node v that lies in $N_{ego}(u)$ and assists in the calculation of the relevance score, a directed edge is assumed to join them starting from the more popular (i.e., the one with the highest degree in G_U) and ending to the less popular one. In this way, the directed influence graph is computed.

In this experiment, an influence graph containing 1,322 nodes and 27,601 edges is used as input to CoveR and Difrec. The value K of DifRec and the target value ℓ of coverage in CoveR, are set equal to 5. Among the books in the platform, the top 50 reviewed books are selected. The relevance scores are obtained by the procedure described earlier and are considered to reflect the preference of the users to these books. The results are summarized in Table 4.8. It can be seen from this table, that although the relevance score achieved by DifRec is higher than the one achieved by CoveR, DifRec's direct recommendations surpass more than 400% those of CoveR, while achieving only about 39% higher total relevance score. The ratio of relevance score per direct recommendation is significantly higher for CoveR than DifRec. This highlights the suitability of CoveR for covering completely the real network with at least ℓ recommendations per user, while achieving significantly large relevance score by relying heavily on the information diffusion happening in the social network, indicated by the number of indirect recommendations that are higher in the case of CoveR. Finally, from the items per user ratio it is derived that CoveR manages to not only achieve a better ratio of relevance score per direct recommendation (assignment), but also keep the number of distinct items per user close to the ℓ parameter. This indicates the accurate identification of the subset of users for direct recommendations that result in the coverage of the network with at least ℓ items per user, without overloading the users.

Table 4.8: Average relevance score per assignment with CoveR and DifRec in the Real Network.

	Direct Recs.	**Indirect Recs.**	**relevance score**	**rel. score/ dir.recs.**	**items/ user**
CoveR	787	27003	4224.05	5.34	8.16
DifRec	3613	13259	5884.63	1.63	13.81

The fact that online social networks and platforms of streaming services are mostly accessed by users via their mobile devices, a recommender system such as CoveR may exploit, besides the knowledge on users' interactions and preferences, users' mobility pattern in order to derive local communities and encourage collaboration in content sharing at the physical world. According to the users' physical and social ties and by acknowledging the impact of recommendations in users' requests for content, the problem of content placement at edge caching networks as well as content sharing via Device-to-Device (D2D) communication is investigated under various objectives in chapter 5.

Chapter 5

Socio-aware content allocation in physical and cyber-physical networks

This chapter focuses on methodologies for content placement in physical networks, such as mobile edge caching networks and cyber-physical networks, which are formed by mobile social networks or platforms of streaming services and mobile edge caching networks.

5.1 Mobile Edge Caching

Various bandwidth-craving applications, mostly live and on-demand video streaming, have lately become widespread. Most of the Internet's traffic is streaming-related [119], and the corresponding content is accessed from wireless/mobile devices [4]. Mobile Edge Caching (MEC) is a technology that utilizes edge servers as cache nodes to store popular content closer to the user. This has emerged as a promising solution for solving several of the associated challenges, e.g., achieving small end-to-end delay [120], avoiding core network congestion [121], improving user Quality of Experience (QoE) [122], etc., which can further aid in decongesting backhaul links of the communication network [6]. In MEC, content requests are issued by User Equipments (UEs) and the requested content is delivered by one of the cache-enabled edge nodes. A caching scheme can be proactive or reactive depending on whether the caching decision is realized before or after a request for content is made. Proactive caching of popular content is suggested in [123] to alleviate backhaul congestion by avoiding duplicate data transfers and in [124] to improve user QoE. Also, proactive caching can leverage network information, such as users' content preferences, social features and mobility patterns in order to improve caching efficiency. In general, the caching problem can be divided into the selection of the caching locations (i.e., where to cache) and the placement of content to caches (i.e., what to cache).

Caching Location. Content can be cached at Macro Base Stations (BSs), Small Base Stations (SBSs), Femto Base Stations (FBSs), Pico Base Stations (PBSs) and UEs. In the case of UE caching, also known as Device-to-Device (D2D) caching, UEs provide cumulatively a large low-cost cache space [125]. The D2D communication paradigm [126] promotes offloading content requests from the main network, where a device can store content and share it with its neighboring devices, if requested. The BS usually monitors the caching status of each UE and directs requests so that they can be satisfied by cache-enabled devices. In [127], mobile caching via D2D connectivity is compared to local caching at the radio access network edge and the results show that in dense networks, D2D caching may serve more user requests through cache-assisted D2D communication, whereas edge caching issues higher cache hits because of the great capacity of storage units in Small Base Stations (SBSs). Regarding larger topologies, the peer-to-peer caching policy in [128] ensures the delivery of content throughout a country even when the ability to connect to the Internet is rather limited. The knowledge on users' interactions and social features can also be exploited in order to serve requests locally and disseminate content via D2D communication [129].

Content placement. The design of a content placement strategy requires awareness of content popularity and size, the locations of existing replicas in the caching topology, user preferences and mobility patterns. It can be formulated under various objectives including cache hit ratio maximization [130], which is defined as the ratio of the number of the requested cached files over the total number of cached files, traffic offloading maximization [131], enhancement of user QoE [132], content delivery delay minimization [133], energy efficiency [134]. The effect of user mobility on content placement is investigated in [132, 135]. In [135], the problem of encoded content allocation is studied in a femto caching network with mobile users and a distributed approximation algorithm minimizes the probability of using the main base station for content delivery. An extension of this is proposed in [132] with a mobility-aware content allocation strategy in D2D caching networks, where helper caches are not fixed as in [135]. Content placement is treated as a problem of monotone submodular maximization over a matroid constraint and a greedy approximation algorithm is designed for its solution. In section 5.5, the users' mobility pattern is leveraged to derive the topology of a heterogeneous caching network (HetNet) [136] consisting of a BS, SBSs and UEs, based on which a fraction of the UEs is selected to assist in caching. The content stored in the cache-enabled devices should be delivered to the users within a given deadline, otherwise, it is retrieved by the core network via the BS.

5.2 Mobile Social Networks

Mobile Social Networks (MSNs) are considered the intersection of mobile communication networks with online social networks and belong to the category of cyber-physical networks. MSNs are defined as heterogeneous networks formed by the interactions of individuals with similar interests or objectives through their mobile devices (smartphones, tablets) within virtual communities [137]. Mobile applications leverage users' features, behaviors and ties acquired by online social networks to create native communities and encourage users' collaboration. On the other hand, online social networks exploit mobile features and accessibility to enrich their knowledge on users' social relationships and to empower the concept of real-time web [138], respectively.

Due to the evolution of mobile devices, which are currently equipped with sensing modules (cameras, accelerometers, etc.), global position system receiver (GPS) and multiple wireless interfaces (4G and 5G cellular, WiFi, WiFi Direct, Bluetooth), traditional social networks are extended with location awareness and automatic processing of sensed data [139]. Moreover, multiple radios enable the formation of opportunistic networks, where users exchange information in an ad-hoc manner (i.e., wireless mobile ad-hoc networks). Based on the exploited network infrastructures, MSNs may be classified to web-based, decentralized and hybrid, which is a combination of the former. Web-based MSNs are mostly dependent on centralized communication structures (WiFi and cellular) with the most prominent paradigm to be the online social networking platforms accessed through mobile browsers and smartphone applications (e.g., Facebook, Twitter, Foursquare, etc.). On the contrary, Decentralized MSNs are primarily based on opportunistic networks formed by users who share information using wireless technologies such as WiFi-Direct and Bluetooth. Even though opportunistic networks are characterized by unstable topologies with sparse connections or even disconnected components, their main advantage is that they can be used for data transmission when this cannot be realized via the centralized structure (e.g., in the subway). Users of a hybrid MSN, which is the most recent trend, can access information from the content provider via a centralized server and share data by forming opportunistic networks where they communicate directly with each other without connecting to the network infrastructure. Data dissemination is determined by several factors including users' mobility and social features (popularity, similarity, trust, willingness to share content, etc.). Therefore, identifying influential users to assign content to, along with grouping users of similar interests via community detection are deemed crucial in order to disseminate content efficiently in the MSN [140].

The inherent proximity-driven sharing ability of mobile devices and the delay-tolerant nature of many cellular contents make offloading cellular traffic produced in MSNs a promising way to alleviate the traffic load at the backhaul [141] without upgrading the cellular network, which is considered an expensive solution of low financial return [142]. Opportunistic traffic offloading leverages the opportunistic network formed by mobile devices in order to offload traffic data. In opportunistic offloading some mobile users cache cellular contents (e.g., videos, movies) that are expected to be requested by other users from the core network and share these contents with encountering mobile users via opportunistic device-to-device communication. Since there is no stable path among mobile users, content delivery heavily depends on the mobility of users. Proper incentives can make users willing to experience a tolerable delay in order to consume cellular contents, and to assist in their storing and forwarding [143, 144]. Current research studies focus on storing popular content in geographical floating circles [145] or fixed user equipment [5, 146], so that future queries can be served without requiring communication via cellular links. In section 5.5, this problem, combined with mobility-aware content caching in heterogeneous cache networks is investigated from the perspective of recommendations and content dissemination via D2D communication between the users of a MSN.

5.3 Mobile Edge Caching and recommendations in Mobile Social Networks

Recently, the interplay between mobile edge caching and recommendations is of great research interest. On the one hand, caching at the edge may improve the streaming experience of the user and release important network resources for the operator. On the other hand, user demands are significantly affected by the Recommender Systems operating on MSNs and platforms of streaming services [10], which in turn determine the caching policies. Both the content access cost, such as the delay experienced for the content delivery, as well as the quality of recommendations, are important factors for the users' engagement to the platform. The recent tendency of Content Providers partnering with Internet Service Providers in order to form their own Content Delivery Networks, such as Netflix Open Connect [147], allows to jointly handle content caching and recommendations towards optimizing user QoE. The impact of recommendations on user content requests is investigated in [7], where YouTube's caching efficiency is increased by shaping the video demands of its users through recommendations with a reordering of viewers' related lists so that cached content is presented above non-cached content. Following up this line of research, in [8], the impact of the recommendation position on the performance of cache-aware recommendations is examined in the case of sequential content requests, where consecutive requests are not independent and are modeled by a Markovian traversal model of the content catalogue [148], which resembles PageRank manipulation [149]. The idea of optimizing caching policies by taking into account recommendations is also investigated under a generic network setting in [150] where network and content provider collaboration is not required. The CABaRet algorithm proposed in [150] leverages available information provided by a RS, and returns cache-aware recommendations. Measurements over the YouTube service show that CABaRet increases the cache hit ratio significantly. The trade-off between caching efficiency and quality of recommendations is studied in [9], [151] and [3]. The objective in [9] is to maximize the cache hit rate by forming user demands towards cached content via recommendations. This is achieved by a heuristic algorithm which first places content in caches driven by user preferences and then makes recommendations that promote the cached content. This is further examined by taking into account the associations of mobile users to a caching network of

SBSs with limited service capacity in [151], where a methodology is developed to maximize the cache hit rate while guaranteeing a minimum quality of service and quality of recommendations for the users of the Content Provider (CP) platform. In [3], the joint problem of content allocation in caches and recommendations is studied as a problem of QoE maximization and a polynomial-time algorithm with approximation guarantees is proposed for its solution. This problem, under the same objective is further investigated in section 5.5 from the viewpoint of user mobility, delivery delay tolerance and D2D caching, whereas the methodology developed in section 5.4 leverages unilaterally the information of users' social bonds and content preferences as predicted by an RS, in order to satisfy their actual content requests in a heterogeneous caching network, rather than shaping user demands in order to be served by limited resources.

In the following section, proactive caching is studied as a socio-aware content allocation problem in mobile networks to which users are connected in order to consume content generated in Mobile Social Networks and platforms of streaming services. The BS caching approach assisted by UEs is related to the approaches in [129, 152, 153]. The problem under examination is the one of cooperative, socio-aware caching at individual devices that serve as cache nodes from which other nodes acquire data. In [129], a proactive caching mechanism is proposed that capitalizes on the spatial and social features of network users. Using metrics derived from Social Network Analysis, a set of influential users is selected to participate in content caching, in order for this to be diffused to their social connections via D2D communications. In [152], the problem of fair cooperative caching is studied in Mobile Social Networks. Cache nodes are selected based on an overlapping neighborhood centrality measure and a heuristic algorithm is designed to derive the caching scheme. Also, in [153], historical data of downloaded content is used to predict users' future content requests based on which the former are organized into classes of similar users. The caching approach presented in section 5.4 combines the above to develop a cache node selection scheme. It relies on the predictive model of a Recommendation Engine to derive similarities between users. The similarity among users determines a partition of the network into homogeneous in taste and geographically close users and the problem of cache node selection is addressed independently in each resulting group based on user social features.

5.4 CAUSE: A base station caching scheme aided by user equipment

The problem of limited utilization of User Equipment (UE) for Base Station (BS) assisted caching and content sharing through $D2D$ communication is addressed by splitting it into two subproblems. The first determines the number of assisting UE caches according to user similarity and proximity, while the second allocates items for caching in BS and UE so that the cache hit ratio is maximized. Finding the number of assisting UE caches is mapped to a community detection problem, while item placement is formulated as a cache hit ratio maximization problem, which corresponds to the NP-hardMultiple Knapsack Problem. A two-step algorithm, referred to as CAUSE, solves the standard Knapsack Problem successively for every available cache, using a Dynamic Programming based approach. The performance of CAUSE is evaluated through simulation and analysis over synthetic networks. The obtained results indicate that BS caching assisted by UE can be beneficial in terms of the cache hit ratio, for both users and wireless network operators, outperforming a vanilla scheme in which the only cache between a user and the core network is the Base Station's cache.

By exploiting features from Complex Network Analysis [23], users are grouped in communities via clustering, based on their common preferences and location, so that popular content for the users in each community is stored in selected devices and it is then distributed via D2D communication to the rest of the users in this community. Extending content caching at users' devices with D2D communication can result in a higher cache hit ratio. With the D2D communication cache sharing and by using communities of users with similar interests, additional content requests can be satisfied locally for the less popular content in general, which however can be quite desired within a specific community. This way the BS caching can still serve the requests for the content of broader interest.

5.4.1 Model of content catalogue and user network

5.4.1.1 Similarity graph

A similarity graph $H(V, E_H)$ is considered with $V = \{u_1, ..., u_a\}$ being the users of an online platform (e.g., an Mobile Social Network, a streaming platform, etc.). A Recommender System that operates on this platform, aims to recommend to users $u \in V$ items from the set $I = \{i_1, .., i_b\}$. Items can be any type of multimedia content or information data. The preference of user u for item i is predicted by the RS and expressed with a request probability $p(i, u)$, while a similarity measure (e.g., the Jaccard or Cosine similarity, etc.) [2] is used to derive users of similar preferences.

The edges in E_H connect the users that have a similarity score above a certain threshold. This threshold can be specified according to the specific application and controls the density of the similarity graph with low values resulting in dense graphs, whereas high thresholds result in sparser graphs, in which two users are unlikely to be considered similar if they do not have identical taste in items.

5.4.1.2 System cache memories

Each item $i \in I$ has a finite size $M_i \in [0, 1]$. This is a normalized measure of the item's actual size, usually measured in Megabytes. Each cell has a BS with a cache of capacity $c_{BS} \in \mathbb{R}_{>0}$. Furthermore, a user may employ, if selected and is willing to, a smaller cache of limited capacity c_{UE} in her User Equipment (UE, e.g., a smart device) for which it holds that $c_{UE} < c_{BS}$. In the following, one cell with a single BS is considered.

5.4.1.3 Location-based graph

Every user $u \in V$ has an effective radius R_D that denotes the maximum distance at which she can transmit or receive data (e.g., messages, videos, etc.). At any given time, each user belongs to a single cell, which means that she is associated with a BS, communicating with it directly in order to access items that are stored in its cache. Also, it is assumed that nodes lying in each other's radius are able to exchange content directly through D2D communication. At any given time, these nodes and their radii form a graph $L(V, E_L, R_D)$. An edge $(u, v) \in E_L$ exists if $d(u, v) \leq R_D$, where $d(u, v)$ is the Euclidean distance between nodes u and v. In order to model the resulting network in the following simulations, since every node has the same effective radius, a Random Geometric Graph (RGG) is employed to model users' devices and their connections. A similar approach is adopted in [154].

5.4.1.4 System graph

The similarity graph H, is obtained by evaluating metrics on the underlying online platform. The location-based graph L, is a snapshot of the D2D network formed in the cell at a given time. Their intersection results in the System Graph $G(V, E)$ with $E = E_H \cap E_L$. The edges of this graph denote the connections of the users that are closely related in terms of item preferences and whose physical proximity allows D2D communication. An example of the employed system model is displayed in Fig. 5.1.

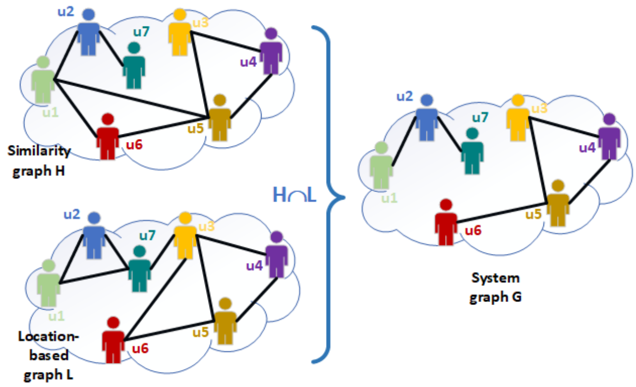


Figure 5.1: The system graph G obtained as $G = H \cap L$ from the similarity graph H and the location-based graph L.

5.4.1.5 User requests model

In the above model, the users request items available through the online platform. Based on the assumed prediction model of user demands made by the RS, each item-user pair is assigned a request probability $p(i, u)$. Two models of recommendations-driven user requests are examined, a deterministic and a probabilistic one. In the deterministic model, the sequence of requests made by user u follows the decreasing order of $p(i, u)$ for $i \in I$. In the probabilistic method, user u chooses at random the items that will be requested with a probability proportionate to $p(i, u)$ for every item i. This is achieved by applying fitness proportionate selection with the fitness of each item i being the probability $p(i, u)$. Also, equivalently to [9], the global utility of item i is expressed as:

$$U_g(i) = \sum_{u \in V} p(i, u). \tag{5.1}$$

5.4.2 The problem of BS caching assisted by UE

By exploiting users' information provided by the RS, one may predict their future behavior in terms of requested content and thus optimize caching efficiency, as quantified by cache hit ratio. Caching efficiency may be further increased by D2D caching. The problem of BS caching assisted by UE can be divided into two subproblems. The first is to determine the number of assisting caches (UE caches) required, based on users' similarity and proximity. The second problem is to allocate the items in both BS and UE caches so that the cache hit ratio is maximized.

5.4.2.1 Selection of assisting UE caches

The problem of finding the number of UE caches is mapped to a problem of graph partitioning and coverage. Detecting communities in the system graph is equivalent to identifying groups of users who are similar and geographically close. It should be noted that the terms community and cluster are used interchangeably throughout this section. A partition of the network is obtained by detecting communities with a modularity maximization method [155]. A high value of modularity indicates a good community structure, thus, the partition corresponding to the maximum value of modularity on a graph is expected to be of good quality.

Degree-based UE selection with m CHs per cluster. Assuming that the outcome of modularity maximization is a set of k clusters $\mathbf{A} = \{A_1, ..., A_k\}$, the number of users who will be delegated content for caching within each cluster, referred to as ClusterHeads (CHs), is set to $m \in \mathbb{N}_{>0}$ with $m < min_i(|A_i|)$, where $|A_i|$ is the number of users in cluster A_i. Thus, the number of UE caches will be $m \cdot k$. The UE cache selection scheme takes into account the degree centrality of the nodes in the subgraphs defined by the clusters. At first, all users of cluster $A \in \mathbf{A}$ are sorted according to their degree centrality in descending order. The node of highest degree is chosen as the first UE cache. Then, until m UE caches are selected in the cluster, the user with the highest covering ability in the community, that is, the user with the highest number of non-common neighbors to the already selected CHs, is chosen as the next UE cache in A, expressed as:

$$\underset{u_i \in A \setminus A^{CH}}{\arg\max}\ N[u_i|A] \setminus \Big(\bigcup_{u_j \in A^{CH}} N[u_j|A] \Big), \tag{5.2}$$

where A^{CH} is the set of UE caches in community A, $A^{CH} \subset A$, and $N[u_i|A]$ is the closed neighborhood of u_i in cluster A, that is, the set of one-hop neighbors of u_i in A, including u_i. An illustrative example is shown in Fig. 5.2 for the case of $m = 1$.

Degree and capacity-based UE selection with varying number of CHs per cluster. Due to the resolution limit of modularity maximization, communities may differ significantly in size. Depending on the size of each community, a different number of UE should be selected for content caching so that the majority of network users are covered with cached content, that is, they can have access to content cached in UE. As in the previous UE selection method, assuming that the outcome of modularity maximization is a set of k clusters $\mathbf{A} = \{A_1, ..., A_k\}$, $\lceil p \cdot |A_i| \rceil$ is specified to be the number of CHs in cluster A_i, where $|A_i|$ is the number of users in community A_i. Parameter $p \in [0, 1]$ is determined through simulations to ensure that CHs cover most of the users in their community, i.e., they are connected, in aggregate, to most of the users in their community. This UE cache selection method takes into account both the nodes' degree centrality in the respective communities' subgraphs, and their storage capacity c_j, $j = 1, ...|V|$. A centrality measure is defined to give prominence to CHs with a good ratio of capacity to degree, while achieving low overlap between their neighborhoods in the corresponding community. Let A_i be the community under examination and C_{A_i} to be the set of its CHs. Community A_i forms the network $G_{A_i}(V_{A_i}, E_{A_i})$, with E_{A_i} to be the set of edges between the users in V_{A_i}. Given the average weight $\overline{M}$ of items in the content catalogue, the average number of items that can be assigned for caching to u_j is equal to

$$\tilde{c_j} = \frac{c_j}{\overline{M}}. \tag{5.3}$$

The closed free neighborhood of u_j in community A_i, is the set of the neighbors of u_j in community A_i, who are not connected to any other CHs of A_i, denoted by

$$N_F[u_j|A_i] = \{u_k : (u_k, u_j) \in E_{A_i},\ (u_k, a) \notin E_{A_i},\ \forall a \in C_{A_i}\}.$$

The average number of items per user in $N_F[u_j|A_i]$ is

$$\frac{\tilde{c_j}}{|N_F[u_j|A_i]|}. \tag{5.4}$$

The centrality of node u_j is defined as:

$$f(u_j) = \begin{cases} c_j - \frac{c_j}{|N_F[u_j|A_i]|}, & \text{if } lb \leq \frac{\tilde{c_j}}{|N_F[u_j|A_i]|} \leq ub, \\ 0, & \text{otherwise.} \end{cases} \tag{5.5}$$

The lower and upper bounds lb, ub are determined by the UE capacities and the density of the networks under examination. The centrality metric f rewards users of high degree, when the average number of items per user in the closed free neighborhood of u_j is at least lb and less than or equal to ub. Also, it penalizes small degree and large capacity, i.e., users who have few or no neighbors in their community. Then, the iterative scheme for CH selection in a cluster is formulated as follows: For community A_i, while the number of CHs does not exceed the predetermined threshold $\lceil p \cdot |A_i| \rceil$, the scheme selects among the users in A_i', which is the subset of users in A_i with non-zero centrality values, the user who maximizes f, i.e.,

$$\arg\max_{u_j \in A_i'} f(u_j). \tag{5.6}$$

Each time a CH is selected, the centrality scores for the rest of the candidate users in the corresponding community are recalculated.

5.4.2.2 Item allocation in BS and UE caches

Based on the preferences of users to items, the items should be allocated in both BS and UE caches to maximize the cache hit ratio. Let $\mathbf{X}$ be a $(m \cdot k + 1) \times |I|$ matrix of zeros and ones to represent the items stored in caches, with the elements in the first row of $\mathbf{X}$, $(x_{1i})_i$

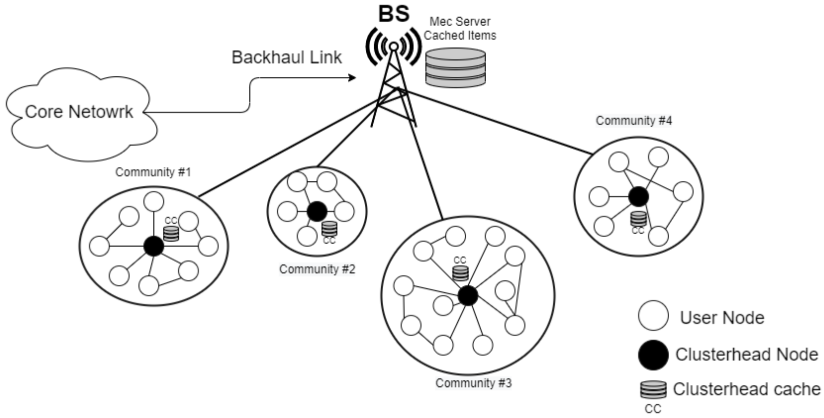


Figure 5.2: Partition of a network into clusters with one CH per cluster. For simplicity reasons, the edges between communities are ommited.

to represent the items cached in the BS. In particular:

$$x_{ci} = \begin{cases} 1, & \text{if item } i \text{ is cached in } C_c, \\ 0, & \text{otherwise,} \end{cases} \tag{5.7}$$

where $\mathbf{C} = \{C_1, ..., C_{m \cdot k+1}\}$ is the set of caches with $C_1 = C_{BS}$, $C_2 = C_{UE}^1$,..., $C_{m \cdot k+1} = C_{UE}^{m \cdot k}$.

An element of the $|I| \times |V|$ matrix of user requests $\mathbf{R}$, as predicted by the RS is expressed as

$$r_{iu} = \begin{cases} p(i,u), & \text{if item } i \text{ is requested by user } u, \\ 0, & \text{otherwise,} \end{cases} \tag{5.8}$$

and the $|V| \times |\mathbf{C}|$ matrix $\mathbf{\Lambda}$ of adjacencies between users and caches, given by the solution of the CH selection problem, with

$$\lambda_{uc} = \begin{cases} 1, & \text{if user } u \text{ is connected to cache } C_c, \\ 0, & \text{otherwise.} \end{cases} \tag{5.9}$$

The problem of finding a collection of different items to store in caches that maximizes the cache hit ratio is formulated as follows:

$$\max_{\mathbf{X}} \sum_{c=1}^{m\cdot k+1} \sum_{u=1}^{|V|} \sum_{i=1}^{|I|} x_{ci} \cdot r_{iu} \cdot \lambda_{uc}, \tag{5.10}$$

subject to:

$$\sum_{c=1}^{m\cdot k+1} x_{ci} \leq 1, \quad \forall i = 1, ..., |I|, \tag{5.11}$$

$$\sum_{i=1}^{|I|} x_{1i} \cdot M_i \leq c_{BS}, \tag{5.12}$$

$$\sum_{i=1}^{|I|} x_{ji} \cdot M_i \leq c_{UE} \quad \forall j = 2, ..., m \cdot k + 1. \tag{5.13}$$

Constraint (5.11) ensures that there is no overlap between the items stored in different caches, namely, an item may belong to at most one cache. Constraints (5.12), (5.13) capture the cache storage capacity limit of BS and UE respectively.

This problem corresponds to the Multiple Knapsack Problem [156], which is a generalization of the standard Knapsack Problem (KP) from a single knapsack to $m \cdot k + 1$ knapsacks with different capacities (BS cache, UE caches). The objective is to assign each item to at most one of the knapsacks such that none of the capacity constraints are violated and the total profit of the items stored into knapsacks is maximized. The Multiple Knapsack Problem is known to be NP-hard[156], thus a heuristic is employed for the item allocation problem in BS and UE caches. The overall operation of this approach is presented in Fig. 5.3.

5.4.2.3 Greedy algorithm CAUSE

A two-step algorithm is proposed in order to allocate items to caches with respect to user preferences. The algorithm is referred to as CAUSE and its pseudocode is presented in Fig. 5.4. CAUSE solves the standard Knapsack Problem successively for every available cache as follows: At first, given the items' global utility, U_g, content is stored in the BS cache by applying the Dynamic Programming algorithm presented in [157] (§8.2) to acquire an $1 - \epsilon$, $\epsilon > 0$ approximation of the optimal solution of the KP. The second step of the algorithm is the placement of items in UEs. For this, a local utility, U_l, of the uncached items is computed. This utility function models the importance/popularity of items in the community-defined neighborhood of CHs. Therefore, the local utility of item i, determined by CH u that belongs to community A, is given by the expression:

$$U_l(i,u,A) = \sum_{v \in N[u|A]} p(i,v), \tag{5.14}$$

where $N[u|A]$ is the closed neighborhood of user u in community A. For each UE cache, based on the corresponding local utilities of items, we use the Dynamic Programming algorithm to obtain the corresponding item placement.

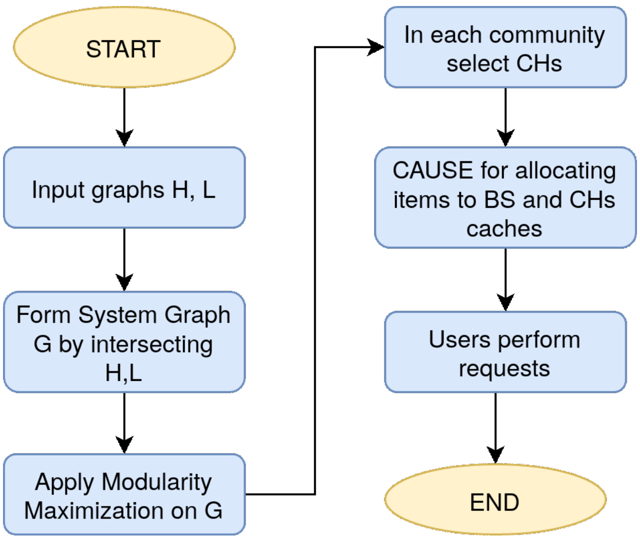


Figure 5.3: A flowchart of the overall framework.

5.4.3 Evaluation results for CAUSE

5.4.3.1 Simulation methodology

The performance of the proposed algorithm CAUSE is evaluated based on the metric of cache hit ratio achieved when the users of the system request items. The behavior of CAUSE is examined for both models of deterministic and probabilistic user requests, with respect to various parameters, namely the network size, the content catalogue size, the cache size of the BS and the CHs, the number of requests per user and the number of CHs determined by the degree-based approach. CAUSE is compared with a vanilla scheme in which the only cache between a user and the core network is the BS's cache.

Algorithm CAUSE
Input: Matrix $\mathbf{R}$ of request probabilities, set of caches $\mathbf{C}$, set of clusters $\mathbf{A}$, matrix $\mathbf{\Lambda}$ of adjacencies between users and caches.
Output: Binary matrix $\mathbf{X}$ of content placement $x_{c,i}$, where c is the index of a cache and $i \in I$.

for $i \in I$ **do**
  Compute global utility $U_g(i)$ from (5.1).
end for
Use Dynamic Programming Approximation Algorithm (DPAA) to compute the set of items I_{BS} to store in C_{BS}.
for $i \in I_{BS}$ **do**
  $x_{BS,i} = 1$
end for
$I = I \setminus I_{BS}$
$\mathbf{C} = \mathbf{C} \setminus C_{BS}$
for $A \in \mathbf{A}$ **do**
  for $u \in A$ that corresponds to cache $C_u \in \mathbf{C}$ **do**
    for $i \in I$ **do**
      Compute local utility $U_l(i, u, A)$ from (5.14).
    end for
    Use DPAA to compute the set of items I_u to store in C_u.
    for $i \in I_u$ **do**
      $x_{u,i} = 1$
    end for
    $I = I \setminus I_u$
    $\mathbf{C} = \mathbf{C} \setminus C_u$
  end for
  $\mathbf{A} = \mathbf{A} \setminus A$
end for

Figure 5.4: Algorithm CAUSE.

Synthetic datasets are used to model item and user similarity. The content catalogue's size is assigned values from the set $\{500, 750, 1000, 1250, 1500\}$. Each item i of the catalogue has finite size, M_i, which follows a uniform distribution in $[0, 1]$. Both the size of the BS cache, c_{BS} and the size of UE caches, c_{UE}, are proportional to the average size of items, $\overline{M}$, with $c_{BS} = \overline{M} \cdot y$, $y \in \{27.5, 35, 42.5, 50, 57.5, 65, 72.5\}$ and $c_{UE} = \overline{M} \cdot z$, $z \in \{6, 9, 12, 15, 18, 21\}$, where y and z are indicative values for the number of items stored in BS and UE caches respectively. The networks used for the experimental setup are modeled by weighted Random Geometric Graphs (RGG) of size $|V| = \{100, 150, 200, 250, 300, 350, 400\}$, to model the users' spatial features. The weights are selected uniformly at random in $[0, 1]$ and model the similarity of users. Edge pruning is applied to RGG to create the system graph based on a similarity threshold. A moderate value is chosen for this threshold, in order to preserve 50% of the number of edges of the initial graph to the system graph. The cases of placing one and two CHs to each community of the system graph are examined. The results are averaged over 25 topologies for each configuration. Unless otherwise stated, the default values of the employed parameters in the simulations are: $|V| = 100$, $|I| = 500$, $c_{BS} = 40 \cdot \overline{M}$, $c_{UE} = 6 \cdot \overline{M}$.

In section 5.4.1.5 a deterministic and a probabilistic model of user requests are introduced. Based on these models, each user will make R demands for content. Consider a user u who belongs in community A, requesting item i. There are three possible ways for u to acquire i: (a) The item is stored in BS cache, therefore, a cache hit is recorded for BS cache, (b) the item is stored in the cache of user's u neighboring CH, thus, a cache hit is recorded for the corresponding CH cache and (c) the requested item is not cached and it will be retrieved by the core network. This is recorded as a cache miss. The simulation terminates when all users complete their requests. In the simulations, the number of requests per user, R, takes values from the set $\{2, 5, 8, 11, 14, 17, 20, 23, 26, 29\}$, whereas the default value is set to 5.

In the following section the impact of various network features to the achieved cache hit ratio is investigated. Each feature is examined separately under the existence of one and two CHs, while keeping the rest of the parameters at their default values.

5.4.3.2 Simulation results for CAUSE

Impact of network size, $|V|$ and catalogue size, $|I|$. Figs. 5.5, 5.6, show that a great percentage of users' content demands are satisfied locally either by BS or UE caches, in both the deterministic and probabilistic schemes of requests. As expected, the deterministic scheme yields a higher cache hit ratio (Figs. 5.5, 5.6, 5.7, 5.8), since the design of content placement is based on the most preferred items for every user, as predicted by the RS. It is also noticed that the overall cache hit ratio increases in small sized networks, where caches satisfy less requests. Equivalently, there is a small decrease in the overall cache hit ratio as the content catalogue size increases, namely, more items are available for the users, as depicted in Figs. 5.7 and 5.8. This is because user requests become highly diverse for the caches to handle. The importance of UE assisting caching is evident when selecting an additional CH in each community, which results in a $4 - 6\%$ increase in the total cache hit

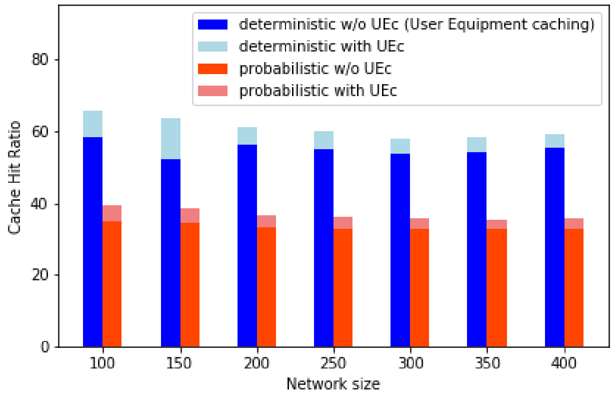


Figure 5.5: Cache hit ratio of systems in networks of varying size, with one CH per community.

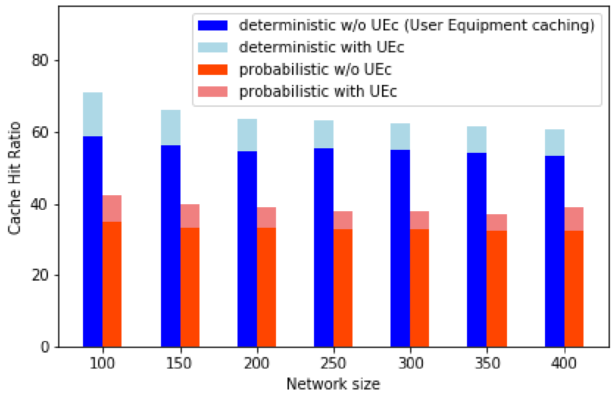


Figure 5.6: Cache hit ratio of systems in networks of varying size, with two CHs per community.

ratio. This is due to the fact that the second CH serves the users who were neither connected to the first CH, nor they could find the content of their preference cached elsewhere.

Impact of the number of requests per user, R. As depicted in Figs. 5.9 and 5.10, when the number of requests per user increases, there is an expected decrease in the achieved cache hit ratio. Despite this fact, CAUSE manages to improve this ratio both in the case of deterministic and probabilistic requests.

This improvement is even more noticeable in the presence of two clusterheads that succeed at providing more user requests with the appropriate content. The results presented may serve as a good indicator of pinpointing the best time to reorganize the caches in the system in order to achieve high cache hit ratios. In greater detail, it is observed that when

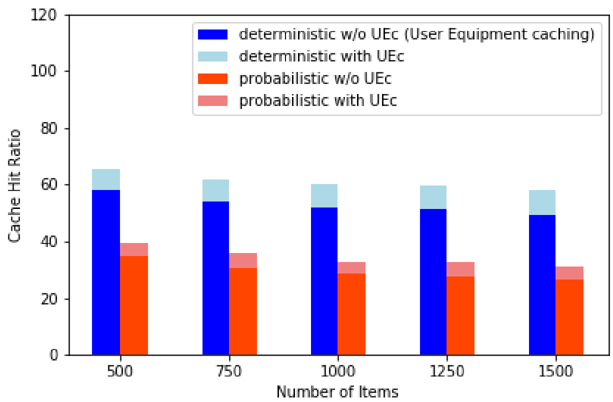


Figure 5.7: Cache hit ratio of systems with content catalogues of varying size, with one CH per community in the network.

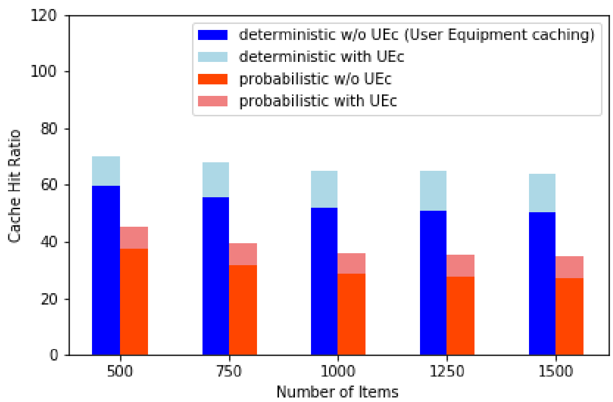


Figure 5.8: Cache hit ratio of systems with content catalogues of varying size, with two CHs per community in the network.

the number of requests per user is low (i.e., less than 10) the deterministic scheme is able to satisfy more than 50% of the total requests. If the cache is reorganized (i.e., rerunning CAUSE) every 8 requests made per user, then high cache hit ratios will be noted. The increase though would be at the expense of frequent communication and data exchange with the core network, which is a time-consuming solution. On the other hand, a solution could be the increase of the BS cache storage capacity, thus, an investment in hardware by the owner of the network. Therefore, the need for balance between cache sizes and frequency of reorganization of the cache memories of UEs and the BS is deemed crucial for the overall satisfaction of the users.

Impact of Base Station cache size, c_{BS}. The BS cache size is defined in terms of

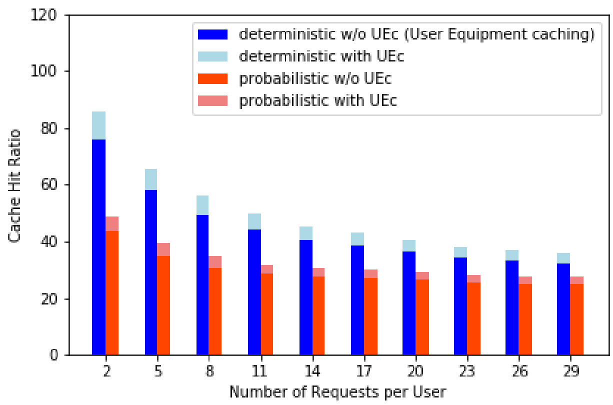


Figure 5.9: Cache hit ratio of systems with varying number of requests per user, with one CH per community in the network.

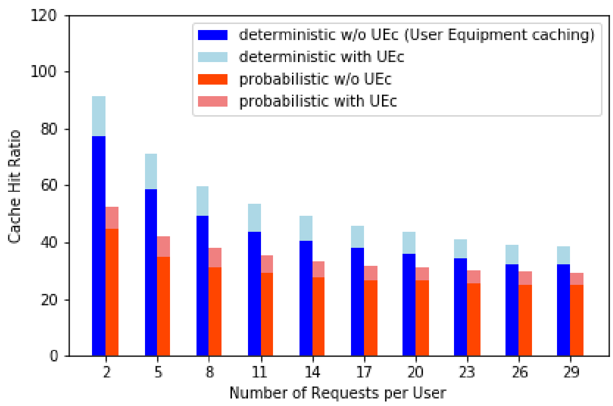


Figure 5.10: Cache hit ratio of systems with varying number of requests per user, with one CH per community in the network.

the expected number of items contained. Keeping all the other parameters at their default values and by changing the size of the BS's cache, it is noticed in Figures 5.11 and 5.12 that by increasing the cache memory of the BS there is an increase in the overall cache hit ratio.

This is expected, as larger caches allow for more items to be stored into them and give the chance for more cache hits at user requests, both on the deterministic and probabilistic request models. Concerning CAUSE's performance, it outperforms the simple vanilla scheme with no UE caching, albeit this improvement decreases as the BS's cache size increases. This is expected because the users will first request the item from the BS and if the BS fails to retrieve it, they will search for it in the UE cache. When the size of BS cache is large, the

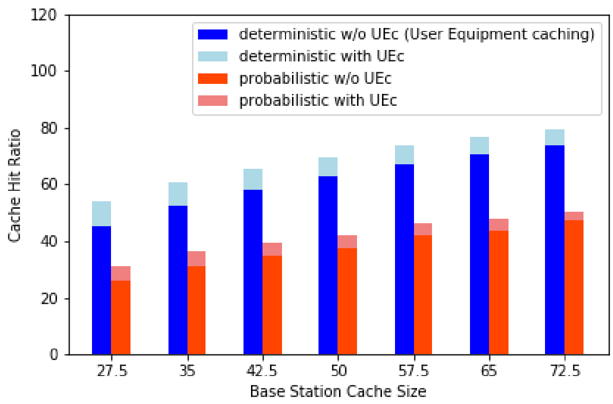


Figure 5.11: Cache hit ratio of systems with BS caches of varying size, with one CH per community in the network.

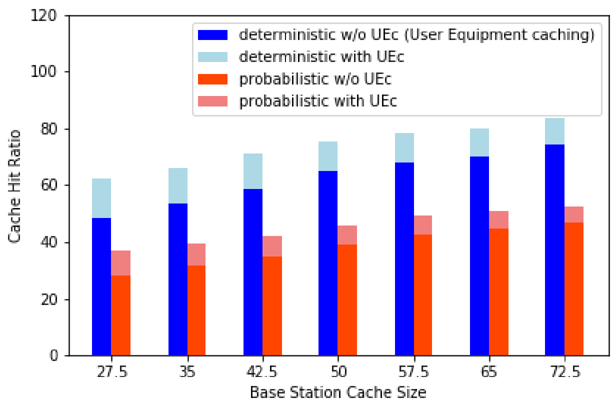


Figure 5.12: Cache hit ratio of systems with BS caches of varying size, with two CHs per community in the network.

UEs end up serving less requests, yet, large cache memories are not a viable solution as they require significant financial investment from the network owner. Thus, there is always an incentive for supplementary caching in order to keep the data exchange focused on the edge rather than the core network.

Impact of Clusterhead cache size, c_{UE}. Similarly to the BS cache size, the CH cache size is determined by the expected number of stored items. The averaged results of the performed simulations are presented in Figures 5.13 and 5.14.

The increase in the CH's cache leads to more cache hits, due to the CH's enhanced capability of serving user requests that the BS is incapable of satisfying. As the CH's cache size increases, the benefit of applying UE caching versus a simpler approach is more evident

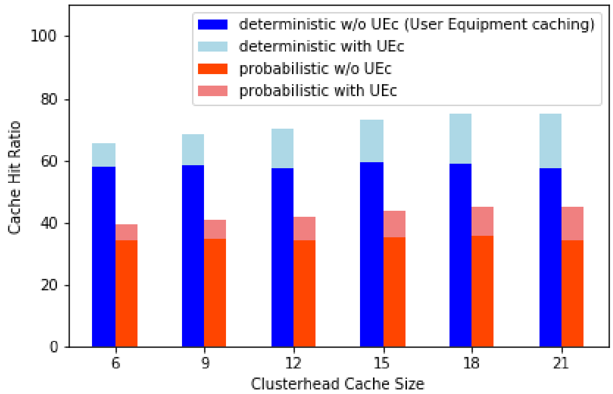


Figure 5.13: Cache hit ratio of systems with CH caches of varying size, with one CH per community in the network.

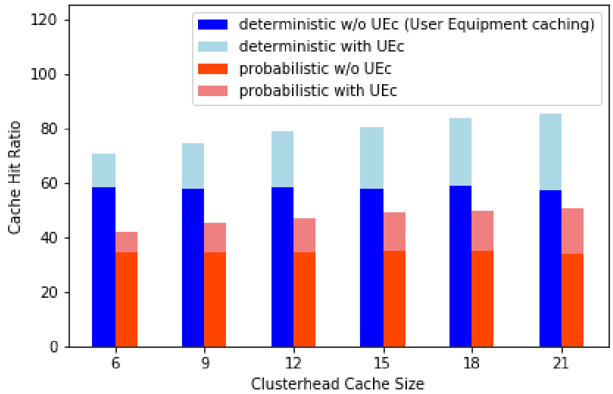


Figure 5.14: Cache hit ratio of systems with CH caches of varying size, with two CHs per community in the network.

regardless of the request model (deterministic or probabilistic). This difference is even more evident when there exist two CHs in every detected community. This result highlights the benefits of opting to cache more content in the network edge through the involvement of the users' smart devices. Of course, the allocated memory should not interfere or affect significantly other processes running on the device. Moreover, there must exist proper incentives for the user to willingly participate in this scheme (e.g., reduced cost for the periods when she assumes the role of CH, etc).

The contribution of the assisting UE caches to the efficiency of a caching strategy is highlighted through simulation results in the cases where two CHs exist in every community. In the default setup, the users having access to the first UE cache of their community, are approximately 59.4% of the total users, while those having access to the second UE cache are 46.9%, leading to around 90% of total users with access to at least one UE cache, since some of them are connected to both UE caches. Thus, by only setting two UE caches per community users are adequately connected for niche content sharing.

Taking into account the impact of recommendations in users' content demands, in the following section, the UE contribution in efficient MEC and content dissemination is further investigated from the perspective of user mobility and delay tolerance in content delivery.

5.5 Opportunistic offloading at the network edge and recommendations in MSNs

In a HetNet with small cells and UEs which offload traffic from the core network by storing and delivering data via D2D communication [158], users' mobility pattern can be exploited: (a) to derive for each user, the expected waiting time to encounter cache-enabled devices and (b) to determine a subset of the UEs that will cache content and participate in the offloading. Expressing the user QoE as a function of user-content relevance and its expected delivery delay, the joint problem of content placement in caching networks and recommendations in MSNs is formulated as a user QoE maximization problem, which is known to be NP-hard[159]. In order to address it, a heuristic algorithm that focuses on the content delivery delay is proposed and evaluated through simulation over synthetic datasets. The obtained results are presented in section 5.5.8 and are compared with a state-of-the-art polynomial-time approximation algorithm and show that the proposed algorithm balances efficiently the trade-off between the quality of the solution and the execution time.

5.5.1 Model of caches and content

5.5.1.1 Heterogeneous caching network

The HetNet under examination consists of a BS, k SBSs $S = \{s_1, ..s_k\}$ and n mobile users with smart devices (UEs), $U = \{u_1, ..., u_n\}$ with transmission ranges $r_t(BS) > r_t(s) > r_t(u)$ with $r_t(s_i) = r_t(s)$, $\forall s_i \in S$ and $r_t(u_j) = r_t(u)$, $\forall u_j \in U$. All UEs can communicate with the BS, while a UE can communicate with an SBS, if in range. Also, a UE may communicate with another UE in a D2D fashion in order to share cached content.

5.5.1.2 Content Catalogue

The content catalogue is modeled by the set $I = \{i_1, ..., i_m\}$ of m items with size z_i. The items' popularity is assumed to be known. It may be estimated by the Recommender System operating on the platform. The relevance of user u to content i is given by $r : \{U \times I\} \to [0, 1]$.

5.5.1.3 Cache memories

Each SBS is equipped with some storage capacity where content is cached in order to avoid backhaul congestion. UE have also some storage capacity $cap : \{S, U\} \to \mathbb{R}$. The capacity of each user $u \in U$ is less than that of a SBS, $cap(s) >> cap(u)$, $\forall s \in S$, $\forall u \in U$. Given the average size of an item $\bar{z}$ and the cache capacity of every user u, $cap(u)$, the average number of items that can be stored in every user's cache can be computed as $k(u) = \lfloor \frac{cap(u)}{\bar{z}} \rfloor$. Content is stored at a predefined number of UEs. This number is specified by the network operator, who gives a reward to the UEs participating in content offloading via D2D communication, which will be referred to as Cluster Heads (CHs). The set of CHs is denoted by $C = \{c_1, ..., c_g\}$.

5.5.1.4 Recommendation lists

Platforms of streaming services, such as Spotify and Youtube, employ Recommender Systems (RSs) to help its users decide on the consumption of the available content. Each user $u \in U$, who enters the platform, views a list of recommended content $\Gamma(u)$, where $\Gamma : U \to \mathbf{I}_\Gamma$ is a set-valued function and $\mathbf{I}_\Gamma$ is the family of the subsets of I with cardinality l. The recommendation list that results in the maximum aggregated relevance score for every user is denoted as $\Gamma_b(u)$.

5.5.2 Model of users mobility

5.5.2.1 Contact rate between users

The meeting events between (v, u), $\forall v, u \in U$ are given by a Poisson process with rate $\lambda_s(v, u)$. The meeting rates are drawn from an arbitrary probability distribution. It is assumed, for the sake of fairness in content sharing, that in a meeting event between u and v,

u will receive only one content from v, even if the meeting duration is long. Also, the expected inter-contact time coincides with the expected delay experienced by u in order to get content from v. The inter-contact distribution is exponential with rate $\lambda_s(v,u)$. Therefore, the expected inter-contact time for (v,u) will be $\frac{1}{\lambda_s(v,u)}$ when $\lambda_s(v,u) \neq 0$. Assuming that not all contacts between (v,u) lead to successful content exchange (e.g., battery depletion of a device), the expected delay will be greater than the expected inter-contact time. Given that the probability of successful content exchange is p_s, the number of successful meetings follows a Poisson distribution with rate $\lambda_{vu} = \lambda_s(v,u) \cdot p_s$. Therefore, the expected waiting time for successful content exchange between (v,u), when $\lambda_{vu} = \lambda_s(v,u) \cdot p_s \neq 0$ will be $\frac{1}{\lambda_{vu}} = \frac{1}{\lambda_s(v,u) \cdot p_s}$.

5.5.2.2 Expected delivery delay

The function of the expected delivery delay is defined as $f : \{U \times BS, U \times S, U \times U\} \rightarrow \mathbb{R}$. For every pair of users (v,u) $\forall v,u \in U$, in the case of $\lambda_{vu} = 0$, the expected delay experienced for content exchange is set to be $f(v,u) = \infty$. If $\lambda_{vu} > 0$, the expected delay is defined as follows.

$$f(v,u) = \begin{cases} \frac{1}{\lambda_{vu}}, & \text{if } v \neq u, \\ 0, & \text{if } v = u. \end{cases} \tag{5.15}$$

5.5.2.3 Expected number of efficient meetings between users

Considering the expected time between successful content exchange for each pair of users as the expected delay between users v and u, one can estimate the number of times two users will meet during the examined period of duration T. A non-symmetrical matrix $\mathbf{M} = (m_{vu}) \in \mathbb{N}^{|U| \times |U|}$ can be computed as follows:

$$m_{vu} = \begin{cases} \min\{k(v), \lfloor \lambda_{vu} T \rfloor, l\}, & \text{if } f(v,u) < T_s, \\ \min\{k(v), l\}, & \text{if } v = u, \\ 0, & \text{otherwise.} \end{cases} \tag{5.16}$$

Matrix $\mathbf{M}$ contains the expected number of efficient content exchanges between two nodes (since one content is delivered during a meeting) that will result in successful delivery of a content with tolerable delay, namely, for expected delay that is less than T_s. Therefore, the element m_{vu} gives the number of times that user v can deliver non replicated content to user u, which depends on the storage capacity of v, the number of meetings between (v, u) and the size of u's recommendation list. For example, if u is recommended of $l = 6$ items and user v can store on average $k(v) = 3$ items and meets 8 times user u during the time period under examination, the efficient meetings of v with u are $m_{v,u} = \min\{3, 8, 6\} = 3$, since user u can be delivered of all the items stored in v within the time duration of 3 meetings (one item per meeting).

5.5.2.4 Contact rates between users and SBSs

The meeting events for the pair (s, u), $\forall u \in U$, $\forall s \in S$ are also given by a Poisson process with rate $\lambda_{su} < \lambda_{vu}$. It is assumed that during a meeting event between s and u, u may receive more than one item from s, since user u will spend more time on average within an SBSs' transmission range, due to the fact that $r_t(s) > r_t(v)$. The expected number of items delivered to u during her meeting with s is equal to $n_d = \lfloor \frac{r_t(s)}{r_t(u)} \rfloor$. Therefore, the delay experienced for the delivery of n_d items is

$$f(s, u) = \begin{cases} \frac{1}{\lambda_{su}}, & \text{if } \lambda_{su} = 0, \\ \infty, & \text{otherwise.} \end{cases} \tag{5.17}$$

5.5.2.5 Expected number of efficient meetings between users and SBSs

Equivalently to m_{vu}, m_{su} represents the number of efficient meetings of (s, u) for content exchange and it is defined as follows.

$$m_{su} = \begin{cases} \min\{\lceil \frac{k(s)}{n_d} \rceil, \lfloor \lambda_{su} T \rfloor, \lceil \frac{l}{n_d} \rceil\}, & \text{if } f(s, u) < T_s, \\ 0, & \text{otherwise.} \end{cases} \tag{5.18}$$

Therefore, $|S|$ rows are added to matrix $\mathbf{M}$ to include the expected number of meetings between users and SBSs.

5.5.3 Opportunistic offloading of user requests

Content is delivered to the users in U from the CHs via D2D communication, the SBSs and the core network through the BS. An opportunistic offloading model [158] is assumed, where $u \in U$ can wait for an amount of time, let T_s, until she moves within range of a CH or a SBS in order to retrieve content $i \in \Gamma(u)$ from the corresponding caches. If i is stored in more than one caches, user u will retrieve it from the one of lowest delivery delay. If time T_s is reached, the operator delivers the content directly through the BS. For this, the delay experienced for the delivery of i to u by the BS is set to T_s.

5.5.4 User Quality of Experience

5.5.4.1 Quality of Recommendations (QoR)

As in [3], the QoR of user $u \in U$ is considered to be a function of her relevance to the items of the recommendation list $\Gamma(u)$:

$$\phi(u, \Gamma(u), r) = \sum_{i \in \Gamma(u)} r(u, i). \tag{5.19}$$

5.5.4.2 Quality of Service (QoS)

The QoS of user u is considered a function of the expected tolerable delay experienced by her in order to access the content of her recommendation list. It is the deviation of the total expected waiting time of u from the total maximum tolerable waiting time $|\Gamma(u)|T_s$. A higher deviation indicates a better QoS. For user u, $H_u = \{h \in \{C \cup S\} : m_{hu} > 0\}$ is the set of caches to which u is connected for efficient content exchanges. In the case where content i is not cached anywhere in the set H_u, thus it will be delivered to u by the BS with delay T_s, the extended set $H_u^\dagger = \{H_u \cup BS\}$ is defined. Its elements are sorted in increasing order of tolerable delay. Therefore, the last element in $H_u^\dagger$ will be the BS. The QoS of user u is defined as

$$\psi(u, f, \Gamma(u), H_u, \Omega_u) = \sum_{j=1}^{|H_u^\dagger|} (T_s - f((j)_u, u) \sum_{i \in \Gamma(u)} \left[\prod_{\nu=1}^{j-1} (1 - \omega_{(\nu)_u i}) \right] \cdot \omega_{(j)_u i}, \tag{5.20}$$

where $f((j)_u, u)$ is the tolerable delay experienced by u for content delivered by the j-th element of $H_u^{\dagger}$. The binary variable $\omega_{(j)_u i}$ indicates whether item i can be delivered via the element in the j-th position of $H_u^{\dagger}$. The product $\left[\prod_{\nu=1}^{j-1}(1-\omega_{(\nu)_u i})\right] \cdot \omega_{(j)_u i}$ ensures that the cache of lowest delay containing item i in $H_u^{\dagger}$ will be selected for the delivery of item i to u. The QoS of a user who receives the content of her recommendation list from the BS will be equal to 0.

5.5.4.3 Quality of Experience (QoE)

The QoE of user u is defined as a convex combination of the QoR and the QoS.

$$Q(u, \Gamma(u), r, f, H_u, \Omega_u) = a\psi(u, f, \Gamma(u), H_u, \Omega_u) + (1-a)\phi(u, \Gamma(u), r), \tag{5.21}$$

where $a \in [0, 1]$ is a parameter indicating the trade-off of the QoR and the QoS. In case of $a = 0$, the users are willing to wait for at most T_s in order to consume their favorite content. In case of $a = 1$, the users are willing to consume less relevant content in order to experience minimum delivery delay.

5.5.5 Formulation of the joint caching and recommendations problem

The problem of content caching and recommendations in heterogeneous networks is divided into two subproblems. The first problem concerns the selection of UEs that will assist in the opportunistic offloading, and the second deals jointly with the allocation of content to caches and the recommendations to users in order to maximize their QoE.

5.5.5.1 Mobility-aware clusterhead selection

The mobility-aware CH selection method, which will be referred to as CH-cover, selects UEs who act as local relays on behalf of the network operator. These devices should provide in aggregate cached content to most of the network users, while balancing the number of recommended content requests that each one will serve. The operator is considered to provide appropriate incentives to the selected UEs since offloading raises high battery consumption as well as storage and security issues.

For the selection of CHs, the binary matrix of efficient meetings is expressed as $\mathbf{B} = (b_{uv}) \in [0,1]^{n \times n}$, with $u, v \in U$ with elements

$$b_{uv} = \begin{cases} 1, & \text{if } m_{uv} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5.22}$$

This means that $b_{uv} = 1$ if u meets v with expected delay less than T_s. Given the matrices $\mathbf{B}$ and $\mathbf{M}$, the goal is to find the vector $\mathbf{x}$ of size n with its elements x_u, $u = 1, ..., n$ to be defined as follows.

$$x_u = \begin{cases} 1, & \text{if } u \text{ is a CH,} \\ 0, & \text{otherwise.} \end{cases} \tag{5.23}$$

$$\arg\max_{\mathbf{x}} \sum_{u \in U} \sum_{v \in U} x_u \cdot m_{uv}, \tag{5.24}$$

subject to:

$$\sum_{u \in U} x_u = g, \tag{5.25}$$

$$\sum_{u \in U} x_u \cdot b_{uv} \geq 1, \quad \forall v \in U, \tag{5.26}$$

$$\sum_{v \in U} x_u \cdot m_{uv} \leq c \quad \forall u \in U, \tag{5.27}$$

where $c \in (min_u \sum_{v \in U} x_u \cdot m_{uv}, max_u \sum_{v \in U} x_u \cdot m_{uv})$.
The objective is to find the set C of g users (constraint (5.25)) who will maximize the number of efficient content exchanges so that all the network users can access cached content from at least one CH (constraint (5.26)) and the load of efficient content exchanges is balanced between CHs. Balancing is determined by the parameter c (constraint (5.27)). For small contact rates or small values of g, the problem may be infeasible. In that case, constraint (5.26) is dropped and a set of g CHs is computed to cover maximally the set of the users.

We create a directed weighted graph $G(U, E, \mathbf{w})$ with $(u, v) \in E$, if $b_{uv} = 1$. The corresponding weight is $w(u, v) = m_{uv}$. Then, finding the set of g CHs can be mapped to the problem of finding a minimum weighted dominating set (or partial dominating set in the case of infeasibility) of cardinality g for graph G, which is reduced to the minimum weighted set cover (or weighted partial set cover respectively) problem as follows:

Let $\mathbf{A} = \{N_u\}_{u \in U}$ be a family of sets with N_u to be the set associated with u that consists of its one-hop neighbors. Each set has a value that is equal to

$$W(N_u) = \sum_{v \in N_u} w(u, v). \tag{5.28}$$

At first, the family of candidate sets for the set cover is defined as $\mathbf{F} := \mathbf{A}$. The sets $N_u \in \mathbf{A}$ for which it holds that $\sum_{u \in U} W(N_u) > c$, $c = P_g$, where P_g is the g-th percentile of W, are deleted from $\mathbf{F}$. For the sets in the family $\mathbf{F}$, an iterative method is applied, which greedily picks at iteration t the set N_u that minimizes the ratio

$$\frac{1}{|U_t \cap N_u| \cdot W(N_u)}, \tag{5.29}$$

where U_t is the set of uncovered elements of U at time t. For $t = 0$ it holds that $U_0 := U$. The node u is added to the set of CHs C. The procedure terminates at time t' at which $U_{t'} = \emptyset$ or $t' = g$.

- If $U_{t'} = \emptyset$ and $t' \leq g$, the elements that belong to C are removed from $\mathbf{F}$ and the iterative method is repeated.
- If $t' = g$ and $U_{t'} \neq \emptyset$, C is a minimum weighted partial set cover of U.

5.5.6 The QoE problem formulation

Given the set of caches (i.e., CHs and SBSs) to which user is connected for efficient delivery of recommended content, the objective is to maximize the total QoE of the network. For this, it is required to determine the users' recommendation lists $\Gamma(u)$ $u \in U$, namely, the user-item $|U| \times |I|$ binary matrix $\mathbf{\Gamma} = (\gamma_{ui})$, where $\gamma_{ui} = 1$ if $i \in \Gamma(u)$ and the $(|C| + |S| + 1) \times |I|$ binary matrix $\mathbf{\Omega} = (\omega_{ji})$ of the items' placement. The last row of $\mathbf{\Omega}$ concerns the items that can be delivered from the core network via the BS, thus, its entries are all equal to 1. The problem is formulated as follows:

$$\arg\max_{\mathbf{\Gamma},\mathbf{\Omega}} \sum_{u \in U} Q(u, \Gamma(u), r, f, H_u, \Omega_u), \tag{5.30}$$

subject to:

$$\sum_{i=1}^{|I|} \omega_{ji} \leq k(j) \quad \forall j = 1, ..., |C| + |S|, \tag{5.31}$$

$$\sum_{i=1}^{|I|} \omega_{(|C|+|S|+1)i} = |I|, \tag{5.32}$$

$$\sum_{i=1}^{|I|} \gamma_{ui} = l, \quad \forall u \in U. \tag{5.33}$$

Constraint (5.31) depicts the capacity limits of the network caches, while constraint (5.32) ensures that all the items can be delivered from the core network via the BS. Constraint (5.33) sets the size of the users' recommendation lists.

This problem is equivalent to the QoE problem with equal-sized content constraints presented in [3], which is proved to be NP-hard[133] and for which a greedy algorithm with approximation guarantees is designed. Even though the proposed approach provides a $\frac{1}{2}$-approximation for the QoE problem, it is based on an exhaustive evaluation of all the possible solutions. In order to avoid this, the problem is decomposed into a content placement problem in a network of heterogeneous caches and a network-friendly content allocation problem in the users of a MSN. For this, a heuristic approach is designed to balance the quality of the provided solution and the execution time.

5.5.7 Maximizing QoR in a minimum delay topology

At first, the QoE problem is solved for $a = 1$. In this case, the problem becomes a maximization problem of users' QoS and it is solved by computing a minimum delay network topology that captures cache-user connectivity. Based on this, items are placed to caches in order to maximize the users' QoR via cache-aware recommendations. Given the matrix $\mathbf{M}$ of efficient meetings and the corresponding sets $H_u^{\dagger}$, $\forall u \in U$, as well as the $(|C| + |S| + 1) \times |U|$

matrix $\mathbf{A} = (a_{iu})$ with the elements

$$a_{iu} = \begin{cases} 1, & \text{if } i \in \{c \in C : m_{cu} > 0\} \cup \{BS\}, \\ n_d, & \text{if } i \in \{s \in S : m_{su} > 0\}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.34}$$

The objective is to find the $(|C| + |S| + 1) \times |U|$ matrix $\mathbf{E} = (e_{iu})$ of minimum delay for the delivery of at least l contents per user. The problem is formulated as an Integer Linear Programming Problem as follows.

$$\arg\min_{\mathbf{E}} \sum_{i \in H_u^{\dagger}} f(i, u) \cdot e_{iu}, \tag{5.35}$$

subject to:

$$\sum_{i \in H_u^{\dagger}} a_{iu} \cdot e_{iu} \geq l, \quad \forall u \in U, \tag{5.36}$$

$$x_{iu} \in \{0, 1, .., m_{iu}\}, \quad \forall i \in H_u^{\dagger}, \tag{5.37}$$

where $m_{iu} = l$ and $f(i, u) = T_s$, if i refers to the BS.

By solving the above optimization problem, each user will have access to at least l items with minimum expected delivery delay. Given matrix $\mathbf{E}$, one may proceed sequentially to the placement of content to the network caches and the computation of the users' recommendation lists. The proposed framework is referred to as Minimum Delay (MD) framework and consists of the MDCP algorithm for content placement and the MDR algorithm for recommendations.

Content Placement. Each CH and SBS h will cache different items based on the expected minimum-delay efficient content exchanges with the users $u \in N_h = \{v \in U : a_{hv}e_{hv} > 0\}$ and their relevance score on items, r. The utility of each item for a cache device $h \in \{C \cup S\}$, $\Theta(h, i)$, is computed as a weighted average of the relevance scores of $u \in N_h$, promoting the more frequent contacts.

$$\Theta(h, i) = \frac{\sum_{u \in N_h} a_{hu} e_{hu} r(u, i)}{\sum_{u \in N_h} a_{hu} e_{hu}}. \tag{5.38}$$

Namely, the more items a user accesses from a cache memory, the higher impact her preferences will have on the content placement in this cache. The procedure is described in Algorithm MDCP (Algorithm 3).

Algorithm 3: Algorithm MDCP (Minimum Delay Content Placement)

Input: Matrices $\mathbf{A}$ and $\mathbf{E}$, relevance scores $r(u, i)$, $i \in I$, $u \in U$.
Output: Matrix of content placement $\mathbf{\Omega}$.

$I' \leftarrow I$
while $\exists h \in \{C \cup S\}$ *not full* **do**
    for $h \in \{C \cup S\}$ *not full* **do**
        for $i \in I'$ *with* $\omega_{hi} \neq 1$ **do**
            Compute $\Theta(h, i)$
    $\arg\max_{(h', i')} \Theta(h, i)$
    $\omega_{h'i'} = 1$
    $I' \leftarrow I' \setminus \{i'\}$

Recommendations. The list $\Gamma(u)$, $\forall u \in U$ is created by assigning to u at most $a_{hu}e_{hu}$ stored in $h \in N_u = \{h \in H_u^{\dagger} : a_{hu}e_{hu} > 0\}$, which are of highest relevance to her. The cache memories in N_u have been previously sorted in ascending order of content delivery delay. The procedure is described in Algorithm MDR (Algorithm 4).

Algorithm 4: Algorithm MDR (Minimum Delay Recommendations)

Input: Minimum delay network $G_m(V_d, E_m, w_m)$, matrix of content placement $\mathbf{\Omega}$, relevance scores $r(u, i)$, $i \in I$, $u \in U$.
Output: Matrix of recommendations $\mathbf{\Gamma}$.

for $u \in U$ **do**
    Create the items' list $I_f(u) = \{i \in I : i \text{ is cached in } h \in N_u\}$
    Sort $h \in N_u$ in ascending order of content delivery delay.
    $iterator = 0$
    for $h \in N_u$ **do**
        while $iterator < a_{hu}e_{hu}$ **do**
            $\arg\max_{i' \in I_f(u) : \omega_{hi'} = 1} r(u, i)$
            $\gamma_{ui'} = 1$
            $I_f(u) \leftarrow I_f(u) \setminus \{i'\}$
            $iterator = iterator + 1$

5.5.8 Evaluation of the MD framework

The MD framework is evaluated in terms of the achieved QoR, QoS and QoE scores. The obtained solutions are compared with the ones produced by the method proposed in [3], which will be referred to as JCR, for variable number of users $|V| = \{100, 200, 300, 400, 500\}$

and content catalogues of items $|I| = \{300, 1000\}$ with size $z_i = 1$, $\forall i \in I$. The relevance of items to users is drawn uniformly at random from the interval $[0, 1]$. The UEs have storage capacity $k(u) \in \{2, 3, 4\}$, $u \in U$. The number of SBSs is set to $|S| = 5$. All SBSs have memories of capacity $k(s) \in [5, 10] \subset \mathbb{N}$, $\forall s \in S$. The number of items that a user can get from an SBS during a single meeting is set equal to $n_d = 3$. Every user-user pair meets with rate chosen uniformly at random in the interval $[0.01, 0.4]$, while every user-SBS pair meets with rate chosen uniformly at random in the interval $[0.001, 0.15]$. The number of CHs g is equal to 5% of the users' number. The size of the user recommendation list is set to $l = 5$. The maximum tolerable delay is $T_s = 10$. The time of observation begins at $t = 0$ and the examined time period is $\Delta t = l \cdot Ts + 1 = 51$, indicating just over the maximum amount of time for a user to retrieve all of her recommendations from the BS.

The metric *n-QoR* is used as a measure of the deviation of the obtained QoR from the optimal QoR achieved by the recommendation lists of the whole network and it is defined as follows:

$$n\text{-}QoR = \frac{1}{|U|} \sum_{u \in U} \frac{QoR(u)}{QoR_{max(u)}}. \tag{5.39}$$

For user u, $QoR(u)$ is calculated based on Eq. (5.19) and divided by $QoR_{max}(u)$, which is the QoR score of u achieved when $\Gamma(u) = \Gamma_b(u)$. These scores are averaged across the whole network of users.

The metric *n-QoS* is given by the expression

$$n\text{-}QoS = \frac{1}{|U|} \sum_{u \in U} \frac{QoS(u)}{l \cdot T_s}, \tag{5.40}$$

where for every user $u \in U$ the achieved QoS score, computed by Eq. (5.20), is divided by the maximum total delay that would be experienced if every content recommended to u was retrieved from the BS. The *n-QoS* is the result of the averaging process across all the users.

Finally, for evaluating the QoE, the metric *n-QoE* is employed, which is given by the expression:

$$n\text{-}QoE = n\text{-}QoR + n\text{-}QoS. \tag{5.41}$$

5.5.8.1 Caching location

The fraction of recommended content retrieved by different locations is depicted in Fig. 5.15 for the MD approach and in Fig. 5.16 for the JCR approach. The MD approach utilizes more the SBSs by recommending to users more frequently content stored in their caches rather than content cached in CHs as depicted in Fig. 5.15, where at most 46% of the recommended content will be delivered via the CHs. This is explained by the fact that following the MD approach, the devices that determine the users' recommendation lists are the ones which deliver content at the minimum delay. Also, in the case of SBSs, it holds that during a single meeting of user u with SBS s, $\Gamma(u)$ can be filled with up to $n_d = 3$ items stored in s in contrast to the case of CHs, where only one content can be obtained during a meeting. Even though both methods increase the use of the CHs for the delivery of recommended content as the network size increases, in the JCR method, the CHs are utilized in greater extent delivering 57.8% of the recommended content to the network of 500 users, which may degrade the caching performance of the network, since, frequent $D2D$ transmissions result in higher energy consumption and reduce significantly the CHs limited lifetime.

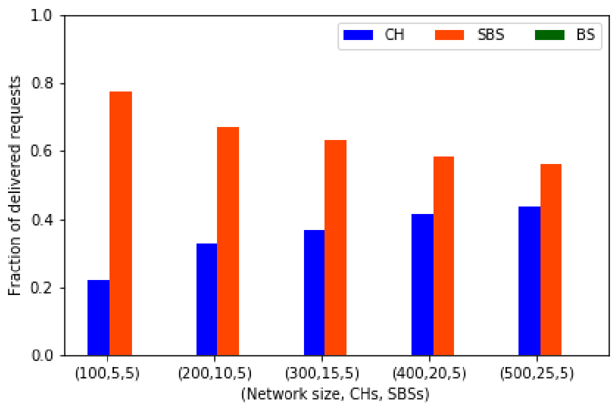


Figure 5.15: Location for the items recommended by the MD approach.

5.5.8.2 n-QoR, n-QoS and n-QoE scores

In Fig. 5.17 the MD and JCR approaches are compared in terms of their achieved *n-QoR*, *n-QoS* and *n-QoE* scores for a content catalogue of 300 items. It is observed that the

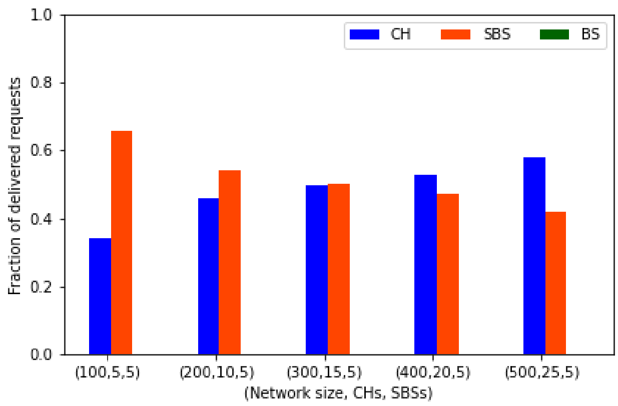


Figure 5.16: Location for the items recommended by the JCR approach.

achieved *n-QoR* for the MD approach is lower than that achieved by JCR in all the examined cases. This is expected since MD's objective is to compute recommendation lists of minimum delay. The results on the *n-QoS* for these methods are of similar quality. Regarding the JCR method, the execution time is quite high, as depicted in Table 5.1, contrary to MD, which is just a few seconds. Thus, the small loss in the *n-QoE* of MD is justifiable.

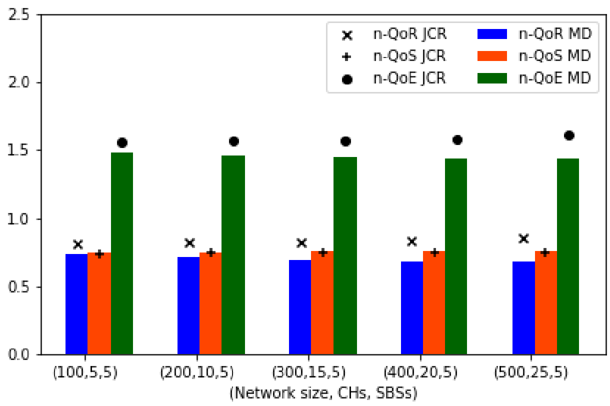


Figure 5.17: *n-QoR*, *n-QoS* and *n-QoE* scores for the MD and JCR methods.

Table 5.1: Execution time (sec) of MD and JCR approaches

Network Setup	**MD**	**JCR**
(100,5,5)	1.02	8278.30
(200,10,5)	2.59	39684.54
(300,15,5)	4.59	162617.48
(400,20,5)	7.17	346948.55
(500,25,5)	13.34	757482.49

Chapter 6

Conclusion

6.1 Summary of results

Chapters 3, 4 and 5 presented research progress on several coverage problems in complex networks, which more specifically focus on network monitoring, tracking and inference of information dissemination and socio-aware content allocation, under various constraints. The application domains included physical (Wireless Sensor Networks), cyber (Online Social Networks) and cyber-physical networks (Mobile Social Networks and Mobile Edge Caching Networks).

In physical networks, such as WSNs, coverage is considered an important performance metric, which reflects how well the network monitors a field of interest. Coverage can be measured in different ways depending on the application. In the case considered in section3, sensors monitor an entire area. Assuming the sensors have the ability to modify their sensing ranges dynamically, the objective was to maximize the area covered by randomly dispersed sensors, while reducing their sensing range as much as possible, resulting in low energy consumption, despite the presence of convex opaque obstacles. This problem was addressed with a framework that capitalized on the notion of the visibility polygon. Two algorithms were designed, a centralized and a distributed one in order to maximize the ratio of covered area to consumed energy, while ensuring a minimum coverage percentage. Simulation results showed that these schemes maintain high coverage percentages of the feasible coverable region, that is, the area covered by the deployed sensors when operating at maximum sensing range, while significantly reducing the associated sensing energy.

Monitoring in Online Social Networks is the process of tracking users' interplay as captured by information sharing. Understanding the flow of information in an OSN plays an important role in social network applications, including recommendations. In networks of billions of users, a great deal of resources is required to infer its diffusion dynamics. Inspired by the notion of coverage in WSNs, in section4.2 a monitoring cover of an OSN is determined, that is, a minimum set of nodes in a social graph, whose activity has to be monitored in order to recover the information propagation graph, in terms of who influences whom in the OSN. Finding a monitoring cover was treated as a variation of the Minimum Vertex Cover problem and a greedy methodology was introduced for its solution, followed by a graph coloring and two backtracking schemes, a deterministic and a probabilistic one, for the inference of the information diffusion graph. The operation and efficacy of the proposed framework was demonstrated in real and synthetic online social networks, such as distributed wireless (spatial) and scale-free (relational) topologies, and conclusions were drawn about the impact of topology on the information spreading inference, with probabilistic inference being more accurate in the case of sparser topologies with small number of nodes and high minimum degree.

Knowledge on the spread dynamics of information was integrated to recommendations in OSNs to increase the diversity of the recommended content and avoid redundancy in suggestions of items that users may attain through the activity of their connections in the OSN. Motivated by the design of IDARS for accurate and efficient recommendations with respect to the limited cognitive capacity of its users, in section4.4, the problem of content allocation to users was studied from a user coverage perspective. An IDARS was designed to utilize minimal amounts of resources for recommendations that best match user preferences and respect their capacity to information. This recommendations problem was formulated as one of computing a Minimum Weighted ℓ-Cover, which is a generalization of the well-studied Minimum Weighted Set Cover Problem. An ℓ-cover was defined as a set of assignments to users who maximize both the spread of recommendations and the total user-to-item relevance, so that each user in the network is covered by at least ℓ items. In order to solve the ℓ-Coverage problem (find an ℓ-cover for the OSN), greedy algorithm CoveR was proposed and proved to be an $O(\frac{\Delta}{\delta}H(\Delta))$-approximation for the ℓ-Coverage problem, where

Δ, δ are the maximum and minimum degree of the network respectively and $H(\Delta)$ is the Δ^{th} harmonic number. CoveR's performance was evaluated through extensive simulations on both synthetic and real networks. The obtained results indicated that the quality of its solution is comparable to the one obtained by the Branch and Bound method, while at the same time outperformed other state-of-the-art information diffusion-aware recommendation heuristics.

Finally, socio-aware content allocation was studied in cyber and cyber-physical networks in Chapter 5. The information on users' features, as acquired by RSs operating in Mobile Social Networks (MSNs) and platforms of streaming services, along with the users' mobility patterns were leveraged to derive local communities of users and encourage their collaboration in local content sharing. According to their physical and social ties and by acknowledging the impact of recommendations in content requests, the problem of content placement at heterogeneous caching networks and content sharing via Device-to-Device (D2D) communication was investigated under various objectives.

In section5.4, allocating items to a physical network, i.e., a caching network of BS and UEs, was formulated as a cache hit ratio maximization problem and the solution was approximated by a dynamic programming based approach. The obtained results highlighted the contribution of D2D caching in the network's caching efficiency. This was further investigated in section5.5 from the perspective of user mobility and delay tolerance in content delivery via a D2D-based opportunistic offloading scheme, which was studied jointly with cache-aware recommendations. Content placement was aimed in two dimensions, the physical dimension, determined by the caching network and the cyber dimension determined by the users of the platform where the RS operates. Multiple criteria based on users' mobility patterns were proposed to determine the user equipment participating in the offloading.

Expressing the user QoE as a function of user-content relevance and its expected delivery delay, the joint problem of caching and recommendations was treated as a user QoE maximization problem and it was addressed by the MD framework that solved sequentially the problems of (a) finding the minimum delay content delivery network, (b) content placement in cache-enabled devices and (c) cache-aware recommendations, ensuring that each user will be recommended of highly preferred content with minimum delivery delay. The results of the MD framework were compared with a state-of-the-art polynomial-time approximation algorithm and showed that MD balances efficiently the trade-off between the quality of the solution and the execution time.

6.2 Insights for future research

As discussed in Chapter 5, content services, the majority of which employ RSs to allocate content to its users, are responsible for most of the mobile traffic. This is why network-friendly recommendations, such as the cache-aware recommendations analyzed in Sections 5.4, 5.5 are considered a promising solution for improving the caching efficiency as well as the quality/cost of content delivery [160].

Following up current research works claiming that the recommendation position of content affects user demands [7, 8], methodologies of network-friendly recommendations in MSNs and platforms of streaming services (Youtube, Netflix) reorder, for each user, her personalized list of recommended content so that cached content at the network edge is presented above the corresponding non-cached content [9]. This may result in users viewing items in different order or even viewing different items (e.g., the posts' order of their newsfeed in facebook/twitter, their list of video recommendations in Youtube) when they access the platform through a fixed connection, compared to the items viewed via a mobile connection, driven by the availability of the content in the corresponding network caches and the associated access cost, which seems to be greater in mobile networks. For example, the backhaul connection for Small Base Stations is often wireless, which makes the delivery of non-cached content costly, since backhaul transmission to a SBS is required.

In order to reduce backhaul accesses, popular content is cached at the mobile edge servers and, due to cache-awareness, it is strongly promoted in recommendations. On the other hand, niche content, which refers to the information which is not widely popular [161], due to the limited storage resources of the cache-enabled servers, is highly unlikely to be cached at the mobile edge servers within a timeliness-preserving time frame. Awareness of the switching pattern of users connection from a fixed to a mobile network and vice-versa, may be leveraged to facilitate the diffusion of niche content in dynamic networks formed by users' network switching. Network switching-aware recommendations can be formulated as an optimal control problem of content allocation over a given time horizon such as the one in [11], where the information diffusion process is modeled as a Susceptible-Infected epidemic and the recruitment of susceptible nodes to the infected class is used to speed up the dissemination process.

Another important issue in the joint caching-recommendations paradigm is the one of fairness, which can be defined in several ways depending on the system and its involved entities.

From the perspective of **D2D cooperative caching**, in [14] the fairness of individuals is captured as a measure of users' equal opportunity to access MSN data. In [15], fairness is investigated in terms of load balancing in a caching network of heterogeneous peer edge devices. Caching is considered fair if less content is stored in cache-enabled nodes with fewer resources.

From the viewpoint of **recommendations**, fairness can be defined with respect to user preferences, which in the literature is referred to as c-fairness (consumer-fairness) [16] and the provider, which is known as p-fairness [17]. In cp-fairness aware RSs, fairness is provided to both consumers and providers creating recommendations lists that satisfy criteria for all the network entities [18, 19, 20, 21]. In [160] different measures of fairness are defined in order to study the extent of unfairness in network-friendly recommendations with respect to content producers.

Combining the different notions of fairness, the joint D2D caching-recommendations problem can be formulated as a content allocation problem in a cyber-physical network of caches and users with multiple complex constraints concerning: (a) load balancing for the selection of the UEs for content caching based on the devices' capabilities (e.g., limited storage capacity, battery life) and users' social/behavioral features, such as influence, trust and willingness to share content, for example, users with large cache memories who are not willing to share content should not make good candidates for content caching and forwarding (b) user coverage in content placement, that is, ensuring that every user has access to at least a specific number of items of high relevance and good streaming quality, (c) item coverage in content placement and recommendations, i.e., every content should be recommended to at least a miminum cardinality set of users with small access/delivery cost (smaller cost when cached to an edge server close to the user).

These fairness-driven criteria can become much more complex to satisfy in an information-diffusion aware setup, where the direct assignment of a content to a specific user can make it available for consumption to other users with whom he/she is opportunistically or socially (via the MSN) connected and affect their decision making or even shape their future content demands as in [148].

Bibliography

[1] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[2] Lillian Lee. Measures of distributional similarity. *arXiv preprint cs/0001012*, 2000.

[3] Dimitra Tsigkari and Thrasyvoulos Spyropoulos. User-centric optimization of caching and recommendations in edge cache networks. In *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pages 244–253. IEEE, 2020.

[4] Mauro Conti, Qian Qian Li, Alberto Maragno, and Riccardo Spolaor. The dark side (-channel) of mobile devices: A survey on network traffic analysis. *IEEE communications surveys & tutorials*, 20(4):2658–2713, 2018.

[5] Yong Li, Mengjiong Qian, Depeng Jin, Pan Hui, Zhaocheng Wang, and Sheng Chen. Multiple mobile data offloading through disruption tolerant networks. *IEEE Transactions on Mobile Computing*, 13(7):1579–1596, 2013.

[6] Jeffrey G Andrews. Seven ways that hetnets are a cellular paradigm shift. *IEEE communications magazine*, 51(3):136–144, 2013.

[7] Dilip Kumar Krishnappa, Michael Zink, Carsten Griwodz, and Pål Halvorsen. Cache-centric video recommendation: an approach to improve the efficiency of youtube caches. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11(4):1–20, 2015.

[8] Theodoros Giannakas, Thrasyvoulos Spyropoulos, and Pavlos Sermpezis. The order of things: Position-aware network-friendly recommendations in long viewing sessions.

In *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pages 1–8. IEEE, 2019.

[9] Livia Elena Chatzieleftheriou, Merkouris Karaliopoulos, and Iordanis Koutsopoulos. Caching-aware recommendations: Nudging user preferences towards better caching performance. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2017.

[10] Carlos A Gomez-Uribe and Neil Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19, 2015.

[11] Kundan Kandhway and Joy Kuri. Optimal resource allocation over time and degree classes for maximizing information dissemination in social networks. *IEEE/ACM Transactions on Networking*, 24(5):3204–3217, 2016.

[12] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[13] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[14] Dongsheng Wei, Konglin Zhu, and Xin Wang. Fairness-aware cooperative caching scheme for mobile social networks. In *2014 IEEE International Conference on Communications (ICC)*, pages 2484–2489, 2014.

[15] Yaodong Huang, Xintong Song, Fan Ye, Yuanyuan Yang, and Xiaoming Li. Fair and efficient caching algorithms and strategies for peer data sharing in pervasive edge computing environments. *IEEE Transactions on Mobile Computing*, 19(4):852–864, 2020.

[16] Bora Edizel, Francesco Bonchi, Sara Hajian, André Panisson, and Tamir Tassa. Fairecsys: mitigating algorithmic bias in recommender systems. *International Journal of Data Science and Analytics*, 9(2):197–213, 2020.

[17] Iordanis Koutsopoulos and Maria Halkidi. Efficient and fair item coverage in recommender systems. In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure

Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), pages 912–918. IEEE, 2018.

[18] Robin Burke. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093*, 2017.

[19] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th acm international conference on information and knowledge management*, pages 2243–2251, 2018.

[20] Dimitris Sacharidis, Kyriakos Mouratidis, and Dimitrios Kleftogiannis. A common approach for consumer and provider fairness in recommendations. In *RecSys (Late-Breaking Results)*, pages 1–5, 2019.

[21] Gourab K Patro, Abhijnan Chakraborty, Niloy Ganguly, and Krishna Gummadi. Incremental fairness in two-sided market platforms: On smoothly updating recommendations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 181–188, 2020.

[22] Michele Ianni, Elio Masciari, and Giancarlo Sperlí. A survey of big data dimensions vs social networks analysis. *Journal of Intelligent Information Systems*, pages 1–28, 2020.

[23] Vasileios Karyotis, Eleni Stai, and Symeon Papavassiliou. *Evolutionary Dynamics of Complex Communications Networks*. CRC Press Taylor and Francis Group, 1 edition, 2013.

[24] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[25] Atif Sharif, Vidyasagar Potdar, and Elizabeth Chang. Wireless multimedia sensor network technology: A survey. In *2009 7th IEEE International Conference on Industrial Informatics*, pages 606–613. IEEE, 2009.

[26] Rahul Priyadarshi, Bharat Gupta, and Amulya Anurag. Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. *The Journal of Supercomputing*, pages 1–41, 2020.

[27] Abhishek Tripathi, Hari Prabhat Gupta, Tanima Dutta, Rahul Mishra, KK Shukla, and Satyabrat Jit. Coverage and connectivity in wsns: A survey, research issues and challenges. *IEEE Access*, 6:26971–26992, 2018.

[28] Ibrahiem MM El Emary and S Ramakrishnan. *Wireless sensor networks: from theory to applications*. CRC press, 2019.

[29] Azzedine Boukerche and Peng Sun. Connectivity and coverage based protocols for wireless sensor networks. *Ad Hoc Networks*, 80:54–69, 2018.

[30] Mathew D Penrose. The longest edge of the random minimal spanning tree. *The annals of applied probability*, pages 340–361, 1997.

[31] Xiaorui Wang, Guoliang Xing, Yuanfang Zhang, Chenyang Lu, Robert Pless, and Christopher Gill. Integrated coverage and connectivity configuration in wireless sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 28–39, 2003.

[32] Andrew Howard, Maja J Matarić, and Gaurav S Sukhatme. Mobile sensor network deployment using potential fields: A distributed, scalable solution to the area coverage problem. In *Distributed Autonomous Robotic Systems 5*, pages 299–308. Springer, 2002.

[33] Khalil Mougou, Saoucene Mahfoudh, Pascale Minet, and Anis Laouiti. Redeployment of randomly deployed wireless mobile sensor nodes. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5. IEEE, 2012.

[34] Guang Tan, Stephen A Jarvis, and Anne-Marie Kermarrec. Connectivity-guaranteed and obstacle-adaptive deployment schemes for mobile sensor networks. *IEEE Transactions on Mobile Computing*, 8(6):836–848, 2009.

[35] Mark De Berg, Otfried Cheong, Marc Van Kreveld, and Mark Overmars. Computational geometry: introduction. *Computational geometry: algorithms and applications*, pages 1–17, 2008.

[36] Mohammed Farsi, Mostafa A Elhosseini, Mahmoud Badawy, Hesham Arafat Ali, and Hanaa Zain Eldin. Deployment techniques in wireless sensor networks, coverage and connectivity: A survey. *IEEE Access*, 7:28940–28954, 2019.

[37] Fadi M Al-Turjman, Hossam S Hassanein, and Mohamad Ibnkahla. Quantifying connectivity in wireless sensor networks with grid-based deployments. *Journal of Network and Computer Applications*, 36(1):368–377, 2013.

[38] Mirjana Maksimović and Vladimir Milošević. Evaluating the optimal sensor placement for smoke detection. *Yugoslav journal of operations research*, 26(1):33–50, 2016.

[39] Yanheng Liu, Longxiang Suo, Dayang Sun, and Aimin Wang. A virtual square grid-based coverage algorithm of redundant node for wireless sensor network. *Journal of network and computer applications*, 36(2):811–817, 2013.

[40] Vasileios Karyotis and Symeon Papavassiliou. 8 topology control in cooperative wireless ad hoc networks. *Cooperative Wireless Communications*, page 167, 2009.

[41] Zongheng Zhou, Samir R Das, and Himanshu Gupta. Variable radii connected sensor cover in sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 5(1):1–36, 2009.

[42] Jiong Wang and Sirisha Medidi. Energy efficient coverage with variable sensing radii in wireless sensor networks. In *Third IEEE international conference on wireless and mobile computing, networking and communications (Wimob 2007)*, pages 61–61. IEEE, 2007.

[43] Jie Jia, Jian Chen, Guiran Chang, Yingyou Wen, and Jingping Song. Multi-objective optimization for coverage control in wireless sensor network with adjustable sensing radius. *Computers & Mathematics with Applications*, 57(11-12):1767–1775, 2009.

[44] Mohammadreza Razzazi and Azade Fotouhi. Coverage of wireless sensor networks in the presence of transparent obstacles. In *SoftCOM 2010, 18th International Conference on Software, Telecommunications and Computer Networks*, pages 209–213. IEEE, 2010.

[45] Haisheng Tan, Yuexuan Wang, Xiaohong Hao, Qiang-Sheng Hua, and Francis CM Lau. Arbitrary obstacles constrained full coverage in wireless sensor networks. In *International Conference on Wireless Algorithms, Systems, and Applications*, pages 1–10. Springer, 2010.

[46] Mrutyunjay Rout and Rajarshi Roy. Dynamic deployment of randomly deployed mobile sensor nodes in the presence of obstacles. *Ad hoc networks*, 46:12–22, 2016.

[47] Lu Dai and Bang Wang. Sensor placement based on delaunay triangulation for complete confident information coverage in an area with obstacles. In *2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC)*, pages 1–8. IEEE, 2015.

[48] GaoJun Fan and ShiYao Jin. Coverage problem in wireless sensor network: A survey. *Journal of networks*, 5(9):1033, 2010.

[49] Ji Li, Lachlan LH Andrew, Chuan Heng Foh, Moshe Zukerman, and Hsiao-Hwa Chen. Connectivity, coverage and placement in wireless sensor networks. *Sensors*, 9(10):7664–7693, 2009.

[50] Honghai Zhang, Jennifer C Hou, et al. Maintaining sensing coverage and connectivity in large sensor networks. *Ad Hoc Sens. Wirel. Networks*, 1(1-2):89–124, 2005.

[51] Mingming Lu, Jie Wu, Mihaela Cardei, and Minglu Li. Energy-efficient connected coverage of discrete targets in wireless sensor networks. In *International Conference on Networking and Mobile Computing*, pages 43–52. Springer, 2005.

[52] Hossam El Gindy and David Avis. A linear algorithm for computing the visibility polygon from a point. *Journal of Algorithms*, 2(2):186–197, 1981.

[53] Halsey Lawrence Royden and Patrick Fitzpatrick. *Real analysis*, volume 32. Macmillan New York, 1988.

[54] Chi-Fu Huang and Yu-Chee Tseng. The coverage problem in a wireless sensor network. *Mobile networks and Applications*, 10(4):519–528, 2005.

[55] Huiyuan Zhang, Subhankar Mishra, My T Thai, J Wu, and Y Wang. Recent advances in information diffusion and influence maximization in complex social networks. *Opportunistic Mobile Social Networks*, 37(1.1):37, 2014.

[56] Shin-Ming Cheng, Vasileios Karyotis, Pin-Yu Chen, Kwang-Cheng Chen, and Symeon Papavassiliou. Diffusion models for information dissemination dynamics in wireless complex communication networks. *Journal of Complex Systems*, 2013, 2013.

[57] Eleni Stai, Vasileios Karyotis, Antonia-Chrysanthi Bitsaki, and Symeon Papavassiliou. Strategy evolution of information diffusion under time-varying user behavior in generalized networks. *Computer Communications*, 100:91–103, 2017.

[58] Suman Banerjee, Mamata Jenamani, and Dilip Kumar Pratihar. A survey on influence maximization in a social network. *Knowledge and Information Systems*, 62(9):3417–3455, 2020.

[59] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. Information diffusion in online social networks: A survey. *ACM Sigmod Record*, 42(2):17–28, 2013.

[60] Jon Kleinberg. Bursty and hierarchical structure in streams. *Data mining and knowledge discovery*, 7(4):373–397, 2003.

[61] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 497–506, 2009.

[62] Rong Lu and Qing Yang. Trend analysis of news topics on twitter. *International Journal of Machine Learning and Computing*, 2(3):327, 2012.

[63] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(4):1–37, 2012.

[64] Manuel Gomez Rodriguez, Jure Leskovec, David Balduzzi, and Bernhard Schölkopf. Uncovering the structure and temporal dynamics of information propagation. *Network Science*, 2(1):26–65, 2014.

[65] Xiao Yang, Mianxiong Dong, Xiuzhen Chen, and Kaoru Ota. Recommender system-based diffusion inferring for open social networks. *IEEE Transactions on Computational Social Systems*, 7(1):24–34, 2019.

[66] L. Lovász. Random walks on graphs: A survey. In D. Miklós, V. T. Sós, and T. Szőnyi, editors, *Combinatorics, Paul Erdős is Eighty*, volume 2, pages 353–398. János Bolyai Mathematical Society, 1996.

[67] Naoki Masuda, Mason A Porter, and Renaud Lambiotte. Random walks and diffusion on networks. *Physics reports*, 716:1–58, 2017.

[68] Vasileios Karyotis and MHR Khouzani. *Malware diffusion models for modern complex networks: theory and applications.* Morgan Kaufmann, 2016.

[69] Dumitru-Clementin Cercel and Stefan Trausan-Matu. Opinion propagation in online social networks: A survey. In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)*, pages 1–10, 2014.

[70] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[71] Vasileios Karyotis, Margarita Vitoropoulou, Nikos Kalatzis, Ioanna Roussaki, Symeon Papavassiliou, O Khalid, S Khan, and A Zomaya. Efficient and socio-aware recommendation approaches for big data networked systems. *Big Data Recommender Systems*, 1:58–87, 2019.

[72] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66, 2001.

[73] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[74] Sancheng Peng, Yongmei Zhou, Lihong Cao, Shui Yu, Jianwei Niu, and Weijia Jia. Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, 106:17–32, 2018.

[75] Honglei Zhuang, Yihan Sun, Jie Tang, Jialin Zhang, and Xiaoming Sun. Influence maximization in dynamic social networks. In *2013 IEEE 13th International Conference on Data Mining*, pages 1313–1318. IEEE, 2013.

[76] Eleni Stai, Vasileios Karyotis, and Symeon Papavassiliou. User interest dictated information diffusion over generalized networks. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 1569–1574. IEEE, 2015.

[77] Paolo Santi. *Mobility models for next generation wireless networks: ad hoc, vehicular and mesh networks.* John Wiley & Sons, 2012.

[78] Christian Bettstetter and Christian Hartmann. Connectivity of wireless multihop networks in a shadow fading environment. *Wireless Networks*, 11(5):571–579, 2005.

[79] Zongqing Lu, Yonggang Wen, Weizhan Zhang, Qinghua Zheng, and Guohong Cao. Towards information diffusion in mobile social networks. *IEEE Transactions on Mobile Computing*, 15(5):1292–1304, 2015.

[80] Roberto Beraldi. Biased random walks in uniform wireless networks. *IEEE Transactions on Mobile Computing*, 8(4):500–513, 2008.

[81] Oliver C Ibe. *Elements of random walk and diffusion processes.* John Wiley & Sons, 2013.

[82] EL Lawler, JK Lenstra, and AHG Rinnooy Kan. Generating ell maximal independent sets. *NP-hardness, polynomial time algorithms, SIC 9, 3, 558-565 NPC MIS*, 1980.

[83] Ryan A Rossi and Nesreen K Ahmed. Coloring large complex networks. *Social Network Analysis and Mining*, 4(1):228, 2014.

[84] Linton C Freeman. Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291–304, 1982.

[85] Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan. *Introduction to information retrieval*, volume 39. Cambridge University Press Cambridge, 2008.

[86] Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1–20, 2016.

[87] Sogol Naseri, Arash Bahrehmand, and Chen Ding. An improved collaborative recommendation system by integration of social tagging data. In *Recommendation and Search in Social Networks*, pages 119–138. Springer, 2015.

[88] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.

[89] Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: introduction and challenges. In *Recommender systems handbook*, pages 1–34. Springer, 2015.

[90] Folasade Olubusola Isinkaye, YO Folajimi, and Bolande Adefowoke Ojokoh. Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3):261–273, 2015.

[91] Kanna Al Falahi, Nikolaos Mavridis, and Yacine Atif. Social networks and recommender systems: a world of current and future synergies. In *Computational Social Networks*, pages 445–465. Springer, 2012.

[92] Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[93] Johann Stan, Fabrice Muhlenbach, and Christine Largeron. Recommender systems using social network analysis: Challenges and future trends, 2014.

[94] Wanita Sherchan, Surya Nepal, and Cecile Paris. A survey of trust in social networks. *ACM Computing Surveys (CSUR)*, 45(4):1–33, 2013.

[95] Behnam Hajian and Tony White. Modelling influence in a social network: Metrics and evaluation. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 497–500. IEEE, 2011.

[96] Paolo Massa and Bobby Bhattacharjee. Using trust in recommender systems: an experimental analysis. In *International conference on trust management*, pages 221–235. Springer, 2004.

[97] Paolo Massa and Paolo Avesani. Trust-aware collaborative filtering for recommender systems. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 492–508. Springer, 2004.

[98] Farzad Eskandanian, Nasim Sonboli, and Bamshad Mobasher. Power of the few: Analyzing the impact of influential users in collaborative recommender systems. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, pages 225–233, 2019.

[99] Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, and Chen Chen. Challenging the long tail recommendation. *arXiv preprint arXiv:1205.6700*, 2012.

[100] Chris Anderson. *The long tail: Why the future of business is selling less of more.* Hachette Books, 2006.

[101] Masayuki Ishikawa, Peter Geczy, Noriaki Izumi, and Takahira Yamaguchi. Long tail recommender utilizing information diffusion theory. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 785–788. IEEE, 2008.

[102] Manuel Gomez Rodriguez, Krishna Gummadi, and Bernhard Schoelkopf. Quantifying information overload in social media and its impact on social contagions. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[103] Farman Ullah and Sungchang Lee. Social content recommendation based on spatial-temporal aware diffusion modeling in social networks. *Symmetry*, 8(9):89, 2016.

[104] Ye Pan, Feng Cong, Kailong Chen, and Yong Yu. Diffusion-aware personalized social update recommendation. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 69–76, 2013.

[105] Hossein Vahabi, Iordanis Koutsopoulos, Francesco Gullo, and Maria Halkidi. Difrec: A social-diffusion-aware recommender system. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1481–1490, 2015.

[106] Konstantinos Tsitseklis, Margarita Vitoropoulou, Vasileios Karyotis, and Symeon Papavassiliou. Socio-aware recommendations under complex user constraints. *IEEE Transactions on Computational Social Systems*, 8(2):377–387, 2021.

[107] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[108] Dirk Bollen, Bart P Knijnenburg, Martijn C Willemsen, and Mark Graus. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 63–70, 2010.

[109] Mahmuda Rahman and Jae C Oh. Graph bandit for diverse user coverage in online recommendation. *Applied Intelligence*, 48(8):1979–1995, 2018.

[110] Mikael Hammar, Robin Karlsson, and Bengt J Nilsson. Using maximum coverage to optimize recommendation systems in e-commerce. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 265–272, 2013.

[111] Tanmay Inamdar and Kasturi Varadarajan. On the partition set cover problem. *arXiv preprint arXiv:1809.06506*, 2018.

[112] Petr Slavik. Improved performance of the greedy algorithm for partial cover. *Information Processing Letters*, 64(5):251–254, 1997.

[113] Stephen P Bradley, Arnoldo C Hax, and Thomas L Magnanti. *Applied mathematical programming*. Addison-Wesley, 1977.

[114] Albert-László Barabási and Eric Bonabeau. Scale-free networks. *Scientific american*, 288(5):60–69, 2003.

[115] Linyuan Lü, Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. Recommender systems. *Physics reports*, 519(1):1–49, 2012.

[116] John Buchanan and Ned Kock. Information overload: A decision making perspective. In *Multiple criteria decision making in the new millennium*, pages 49–58. Springer, 2001.

[117] Chenwei Cai, Ruining He, and Julian McAuley. Spmc: socially-aware personalized markov chains for sparse sequential recommendation. *arXiv preprint arXiv:1708.04497*, 2017.

[118] Tong Zhao, Julian McAuley, and Irwin King. Improving latent factor models via personalized feature projection for one class recommendation. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 821–830, 2015.

[119] Cisco Public. Cisco global cloud index: Forecast and methodology 2015-2020. *white paper, Networking, Cisco Visual*, 2016.

[120] Xiaoge Huang, Ke Xu, Qianbin Chen, and Jie Zhang. Delay-aware caching in internet of vehicles networks. *IEEE Internet of Things Journal*, 2021.

[121] Francesco Pantisano, Mehdi Bennis, Walid Saad, and Mérouane Debbah. Match to cache: Joint user association and backhaul allocation in cache-aware small cell networks. In *2015 IEEE International Conference on Communications (ICC)*, pages 3082–3087. IEEE, 2015.

[122] Fahao Chen, Peng Li, Deze Zeng, and Song Guo. Edge-assisted short video sharing with guaranteed quality-of-experience. *IEEE Transactions on Cloud Computing*, 2021.

[123] Jingjing Yao, Tao Han, and Nirwan Ansari. On mobile edge caching. *IEEE Communications Surveys & Tutorials*, 21(3):2525–2553, 2019.

[124] Konstantinos Poularakis, George Iosifidis, Antonios Argyriou, and Leandros Tassiulas. Video delivery over heterogeneous cellular networks: Optimizing cost and performance. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 1078–1086. IEEE, 2014.

[125] Arash Asadi, Qing Wang, and Vincenzo Mancuso. A survey on device-to-device communication in cellular networks. *IEEE Communications Surveys & Tutorials*, 16(4):1801–1819, 2014.

[126] Muhammad Waqas, Yong Niu, Yong Li, Manzoor Ahmed, Depeng Jin, Sheng Chen, and Zhu Han. A comprehensive survey on mobility-aware d2d communications: Princi-

ples, practice and challenges. *IEEE Communications Surveys & Tutorials*, 22(3):1863–1886, 2019.

[127] Zheng Chen and Marios Kountouris. D2d caching vs. small cell caching: Where to cache content in a wireless network? In *2016 IEEE 17th international workshop on signal processing advances in wireless communications (SPAWC)*, pages 1–6. IEEE, 2016.

[128] Sean Oesch and Max Schuchard. Nation scale mobile ad hoc network for normally isolated topologies. In *2019 international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData)*, pages 156–163. IEEE, 2019.

[129] Ejder Bastug, Mehdi Bennis, and Mérouane Debbah. Living on the edge: The role of proactive caching in 5g wireless networks. *IEEE Communications Magazine*, 52(8):82–89, 2014.

[130] Francesco Pantisano, Mehdi Bennis, Walid Saad, and Mérouane Debbah. In-network caching and content placement in cooperative small cell networks. In *1st International Conference on 5G for Ubiquitous Connectivity*, pages 128–133. IEEE, 2014.

[131] Xiuhua Li, Xiaofei Wang, and Victor CM Leung. Weighted network traffic offloading in cache-enabled heterogeneous networks. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2016.

[132] Rui Wang, Jun Zhang, SH Song, and Khaled B Letaief. Mobility-aware caching in d2d networks. *IEEE Transactions on Wireless Communications*, 16(8):5001–5015, 2017.

[133] Karthikeyan Shanmugam, Negin Golrezaei, Alexandros G Dimakis, Andreas F Molisch, and Giuseppe Caire. Femtocaching: Wireless content delivery through distributed caching helpers. *IEEE Transactions on Information Theory*, 59(12):8402–8413, 2013.

[134] Frédéric Gabry, Valerio Bioglio, and Ingmar Land. On energy-efficient edge caching in heterogeneous networks. *IEEE Journal on Selected Areas in Communications*, 34(12):3288–3298, 2016.

[135] Konstantinos Poularakis and Leandros Tassiulas. Exploiting user mobility for wireless content delivery. In *2013 IEEE International Symposium on Information Theory*, pages 1017–1021. IEEE, 2013.

[136] Jeffrey G Andrews. Seven ways that hetnets are a cellular paradigm shift. *IEEE communications magazine*, 51(3):136–144, 2013.

[137] Xiping Hu, Terry HS Chu, Victor CM Leung, Edith C-H Ngai, Philippe Kruchten, and Henry CB Chan. A survey on mobile social networks: Applications, platforms, system architectures, and future research directions. *IEEE Communications Surveys & Tutorials*, 17(3):1557–1581, 2014.

[138] Nipendra Kayastha, Dusit Niyato, Ping Wang, and Ekram Hossain. Applications, architectures, and protocol design issues for mobile social networks: A survey. *Proceedings of the IEEE*, 99(12):2130–2158, 2011.

[139] Kiran K Rachuri, Christos Efstratiou, Ilias Leontiadis, Cecilia Mascolo, and Peter J Rentfrow. Metis: Exploring mobile phone sensing offloading for efficiently supporting social sensing applications. In *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 85–93. IEEE, 2013.

[140] Aizimaiti Xiaokaiti, Yurong Qian, and Jia Wu. Efficient data transmission for community detection algorithm based on node similarity in opportunistic social networks. *Complexity*, 2021, 2021.

[141] Mehul Motani, Vikram Srinivasan, and Pavan S Nuggehalli. Peoplenet: engineering a wireless virtual social network. In *Proceedings of the 11th annual international conference on Mobile computing and networking*, pages 243–257, 2005.

[142] Dianlei Xu, Yong Li, Xinlei Chen, Jianbo Li, Pan Hui, Sheng Chen, and Jon Crowcroft. A survey of opportunistic offloading. *IEEE Communications Surveys & Tutorials*, 20(3):2198–2236, 2018.

[143] Xuejun Zhuo, Wei Gao, Guohong Cao, and Sha Hua. An incentive framework for cellular traffic offloading. *IEEE transactions on mobile computing*, 13(3):541–555, 2013.

[144] Hamed Shah-Mansouri, Vincent WS Wong, and Jianwei Huang. An incentive framework for mobile data offloading market under price competition. *IEEE Transactions on Mobile Computing*, 16(11):2983–2999, 2017.

[145] Sheng Zhang, Jie Wu, Zhuzhong Qian, and Sanglu Lu. Mobicache: Cellular traffic offloading leveraging cooperative caching in mobile social networks. *Computer Networks*, 83:184–198, 2015.

[146] Marco Valerio Barbera, Julinda Stefa, Aline Carneiro Viana, Marcelo Dias de Amorim, and Mathias Boc. Vip delegation: Enabling vips to offload data in wireless social mobile networks. In *2011 international conference on distributed computing in sensor systems and workshops (DCOSS)*, pages 1–8. IEEE, 2011.

[147] Timm Böttger, Felix Cuadrado, Gareth Tyson, Ignacio Castro, and Steve Uhlig. Open connect everywhere: A glimpse at the internet ecosystem through the lens of the netflix cdn. *ACM SIGCOMM Computer Communication Review*, 48(1):28–34, 2018.

[148] Theodoros Giannakas, Pavlos Sermpezis, and Thrasyvoulos Spyropoulos. Show me the cache: Optimizing cache-friendly recommendations for sequential content access. In *2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pages 14–22. IEEE, 2018.

[149] Stefano Ermon, Carla Gomes, Ashish Sabharwal, and Bart Selman. Designing fast absorbing markov chains. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[150] Savvas Kastanakis, Pavlos Sermpezis, Vasileios Kotronis, and Xenofontas Dimitropoulos. Cabaret: Leveraging recommendation systems for mobile edge caching. In *Proceedings of the 2018 Workshop on Mobile Edge Communications*, pages 19–24, 2018.

[151] Livia Elena Chatzieleftheriou, George Darzanos, Merkourios Karaliopoulos, and Iordanis Koutsopoulos. Joint user association, content caching and recommendations

in wireless edge networks. *ACM SIGMETRICS Performance Evaluation Review*, 46(3):12–17, 2019.

[152] Dongsheng Wei, Konglin Zhu, and Xin Wang. Fairness-aware cooperative caching scheme for mobile social networks. In *2014 IEEE international conference on communications (ICC)*, pages 2484–2489. IEEE, 2014.

[153] Yanfeng Wang, Mingyang Ding, Zhiyong Chen, and Ling Luo. Caching placement with recommendation systems for cache-enabled mobile social networks. *IEEE Communications Letters*, 21(10):2266–2269, 2017.

[154] Maziar Nekovee. Worm epidemics in wireless ad hoc networks. *New Journal of Physics*, 9(6):189, 2007.

[155] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

[156] Hans Kellerer, Ulrich Pferschy, and David Pisinger. Multiple knapsack problems. In *Knapsack Problems*, pages 285–316. Springer, 2004.

[157] Vijay V Vazirani. *Approximation algorithms.* Springer Science & Business Media, 2013.

[158] Pavlos Sermpezis and Thrasyvoulos Spyropoulos. Offloading on the edge: Performance and cost analysis of local data storage and offloading in hetnets. In *2017 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pages 49–56. IEEE, 2017.

[159] Negin Golrezaei, Andreas F Molisch, Alexandros G Dimakis, and Giuseppe Caire. Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution. *IEEE Communications Magazine*, 51(4):142–149, 2013.

[160] Theodoros Giannakas, Pavlos Sermpezis, Anastasios Giovanidis, Thrasyvoulos Spyropoulos, and George Arvanitakis. Fairness in network-friendly recommendations. *arXiv preprint arXiv:2104.00959*, 2021.

[161] Sharad Goel, Andrei Broder, Evgeniy Gabrilovich, and Bo Pang. Anatomy of the long tail: ordinary people with extraordinary tastes. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 201–210, 2010.

Publications

Journal publications published

- M. Vitoropoulou, K. Tsitseklis, V. Karyotis and S. Papavassiliou, “CoveR: An Information Diffusion Aware Approach for Efficient Recommendations Under User Coverage Constraints,” *in IEEE Transactions on Computational Social Systems*, Vol. 8, No. 4, pp. 894-905, August 2021.

- K. Tsitseklis, M. Vitoropoulou, V. Karyotis & S. Papavassiliou, “Socio-aware Recommendations under Complex User Constraints”, *in IEEE Transactions on Computational Social Systems*, Vol. 8, No. 2, pp. 377-387, April 2021.

- M. Vitoropoulou, V. Karyotis & S. Papavassiliou, “Sensing and monitoring of information diffusion in complex online social networks”, *Peer-to-Peer Networking and Applications*, Vol. 12, pp. 604–619, May 2019.

Conference contributions

- M. Vitoropoulou, K. Tsitseklis, A. Karakoulias, V. Karyotis & S. Papavassiliou, “CAUSE: Caching Aided by USer Equipment”, *IEEE International Conference on Cyber, Physical and Social Computing (CPSCom)*, 2020.

- C. Tsanikidis, M. Vitoropoulou, V. Karyotis and S. Papavassiliou, “On the Energy-Efficient Coverage of Network Regions with Convex Opaque Obstacles,” *IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 135-141, Bologna, 2018.

- M. Vitoropoulou, K. Tsitseklis, V. Karyotis & S. Papavassiliou, “Caching, Recommendations and Opportunistic Offloading at the Network Edge”, *accepted by the 17th*

International Conference on Mobility, Sensing and Networking (MSN 2021) in September 2021.

Book Chapters

- V. Karyotis, M. Vitoropoulou, N. Kalatzis, I. Roussaki and S. Papavassiliou, "Efficient and Socio-aware Recommendation Approaches for Big Data Networked Systems", *Big Data Recommender Systems*, Vol. 1, Chap. 4, pp. 58-87, Jul. 2019.