IONIAN UNIVERSITY

Faculty of Information Science and Informatics

Department of Informatics

Ph.D. Thesis

Analysis of Probabilistic Information Dissemination in Wireless Networks Employing Algebraic Graph Theory Elements

George Koufoudakis

20 February 2019

Supervisor

Konstantinos Oikonomou, *Associate Professor*,
Ionian University

Advisory Committee Members

Konstantinos Oikonomou, *Associate Professor*,
Ionian University
Nektarios Kozyris, *Professor*,
National Technical University of Athens
Dimitrios Tsoumakos, *Associate Professor*,
Ionian University

Reviewing Committee Members

Konstantinos Oikonomou, *Associate Professor*,
Ionian University
Nektarios Kozyris, *Professor*,
National Technical University of Athens
Spyros Sioutas, *Professor*,
University of Patras
Vassilios Dimakopoulos, *Associate Professor*,
University of Ioannina
Georgios Kormentzas, *Associate Professor*,
University of Aegean
Dimitrios Tsoumakos, *Associate Professor*,
Ionian University
Theodoros Andronikos, *Associate Professor*,
Ionian University

To my daughter Vasiliki

&

to the memory of my sister Maria.

Abstract

Recent technological advances like the fifth generation mobile phone system (5G), cloud/fog computing are expected to support numerous applications and wide deployment of network devices (e.g., Internet of Things, Wireless Sensor Networks), that is forecasted to reach 25 billion devices by 2020. The expected increment in human-type and machine-type communications introduce a wide variety of communication characteristics with different requirements regarding data rate, latency, mobility and reliability. In such environments, several applications and services that are available in different network locations, regularly need to disseminate information for various purposes like data collection or to reveal useful features (e.g., sink node location, routing information, service and resource discovery, discovery of network functions, etc). Consequently, information dissemination mechanisms need to be adapted in order to meet certain requirements (e.g., guarantee low-latency, save valuable network resources, etc).

Blind flooding, one of the simplest broadcast mechanisms proposed in the literature, is suitable for the wired, small scale network paradigm of the early '80s and capable of reaching all network nodes in a deterministic manner. However, it requires a large amount of usually redundant messages in the network thus, consuming network resources (e.g., energy, bandwidth etc).

An increasing volume of research attention has been observed recently regarding *probabilistic flooding*. Probabilistic flooding can be seen as a suitable alternative for *pruning* transmissions by employing some fixed probability for message forwarding among neighbor nodes. The basic idea is to employ small values for the previously mentioned *forwarding probability* in order to reduce message transmissions that would not result in *coverage* increment (i.e., the percentage of nodes that have received the information message). Due to its probabilistic nature, probabilistic flooding cannot deterministically provide a global outreach, rather it guarantees it

with high probability.

This thesis revisits probabilistic flooding adopting elements from algebraic graph theory. More specifically, probabilistic flooding is modeled in order to obtain an analytical expression of the coverage in the form of a polynomial. The largest roots of this polynomial are shown to be a satisfactory approximation of the *threshold probability* (i.e., the minimum value of the forwarding probability that allows for global outreach with high probability). Moreover, the polynomial's roots are also used to confirm existing results in literature, and –based on certain observations– a novel algorithm is proposed for the estimation of the threshold probability.

Subsequently, the coverage polynomial is further analyzed and fundamental properties of probabilistic flooding –not covered in the literature– are revealed. In particular, the previously mentioned polynomial is further studied, and a new polynomial that depends on the *largest eigenvalue* and the *principal eigenvector* of the network's adjacency matrix is derived. Furthermore, analytical expressions regarding (i) coverage; (ii) a lower bound of the forwarding probability allowing for global network outreach; and (iii) a lower bound of the termination time are also derived. It is shown that threshold probability is inversely proportional to the largest eigenvalue of the network's adjacency matrix and the probability of a node to receive the information message is proportional to the eigenvector centrality of the initiator node and itself for large values of time t. Based on the analytical results, a new probabilistic flooding policy, i.e., *m-Probabilistic Flooding* is proposed and it is shown that the requirements for global outreach are independent of the underlying network's spectral properties.

The results of the analysis that took place in the context of this thesis can be applied to the newly emerging environment for information dissemination purposes. As shown by the simulation results, the analytical findings capture the behavior of probabilistic flooding and they have given a way to study such problems in modern wireless network environments, by employing elements of algebraic graph theory.

Περίληψη

Οι πρόσφατες τεχνολογικές εξελίξεις, όπως τα συστήματα ασύρματης επικοινωνίας πέμπτης γενιάς (5G), η τεχνολογία νέφους/ομίχλης, αναμένεται να υποστηρίξουν πολυάριθμες εφαρμογές και πληθώρα συσκευών δικτύου (π.χ. Διαδίκτυο των Πραγμάτων, Ασύρματα Δίκτυα Αισθητήρων) έως το 2020.

Η αναμενόμενη αύξηση των επικοινωνιών ανθρώπου και μηχανής εισάγει μια ευρεία ποικιλία χαρακτηριστικών επικοινωνίας με διαφορετικές απαιτήσεις σε σχέση τον ρυθμό μετάδοσης, την καθυστέρηση, την κινητικότητα και την αξιοπιστία. Σε τέτοια περιβάλλοντα, πολλές εφαρμογές και υπηρεσίες που είναι διαθέσιμες σε διαφορετικές τοποθεσίες δικτύου, διαχέουν τακτικά πληροφορίες για διάφορους σκοπούς, όπως η συλλογή δεδομένων ή για να ανακαλύψουν χρήσιμα χαρακτηριστικά του δικτύου (π.χ. τοποθεσία κόμβου-συλλέκτη, πληροφορίες δρομολόγησης, ανακάλυψη υπηρεσιών, πόρων, λειτουργιών κ.λπ.). Κατά συνέπεια, οι μηχανισμοί διάδοσης πληροφοριών πρέπει να προσαρμοστούν προκειμένου να ικανοποιηθούν συγκεκριμένες απαιτήσεις (π.χ. εγγύηση καθυστέρησης, εξοικονόμηση πολύτιμων πόρων δικτύου κ.λπ.).

Η πλημμυρίδα, ένας από τους απλούστερους μηχανισμούς διάδοσης πληροφοριών που προτάθηκαν στη βιβλιογραφία, είναι κατάλληλη για μοντέλα ενσύρματων δικτύων μικρής κλίμακας που υπήρχαν στις αρχές της δεκαετίας του ΄80 και ικανή να επισκεφτεί όλους τους κόμβους του δικτύου ντετερμινιστικά. Ωστόσο, απαιτεί ένα μεγάλο αριθμό συνήθως περιττών μηνυμάτων, καταναλώνοντας τους πόρους του δικτύου (π.χ. ενέργεια, εύρος ζώνης κ.λπ.).

Αυξανόμενη ερευνητική προσοχή έχει παρατηρηθεί πρόσφατα όσον αφορά την πιθανοτική πλημμυρίδα. Η πιθανοτική πλημμυρίδα μπορεί να θεωρηθεί ως μια εναλλακτική λύση για τον περιορισμό των περιττών μεταδόσεων θέτοντας μια σταθερή πιθανοτήτα στην προώθηση μηνυμάτων μεταξύ των γειτονικών κόμβων. Η βασική ιδέα είναι η χρήση μικρών τιμών για την προαναφερθείσα *πιθανότητα προώθησης* προκειμένου να μειωθούν οι μεταδόσεις μηνυμάτων που δεν συνεισφέρουν στην αύξηση της κάλυψης

(δηλ. Το ποσοστό των κόμβων που έχουν λάβει το μήνυμα). Λόγω του πιθανοτικού χαρακτήρα της, η πιθανοτική πλημμυρίδα δεν μπορεί να εγγυηθεί την κάληψη του δικτύου ντετερμινιστικά αλλά με μεγάλη πιθανότητα.

Στη συγκεκριμένη διατριβή, μελετάται η πιθανοτική πλημμυρίδα με στοιχεία αλγεβρικής θεωρίας γράφων. Πιο συγκεκριμένα, η πιθανοτική πλημμυρίδα μοντελοποιείται προκειμένου να εξαχθεί μια αναλυτική έκφραση της κάλυψης με τη μορφή πολυωνύμου. Οι μεγαλύτερες ρίζες αυτού του πολυώνυμου, δίχνεται ότι αποτελούν μια ικανοποιητική προσέγγιση της *ελάχιστης πιθανότητας προώθησης* (δηλ., η ελάχιστη τιμή της πιθανότητας προώθησης που επιτρέπει την πλήρη κάλυψη του δικτύου με μεγάλη πιθανότητα). Επιπλέον, οι ρίζες του πολυώνυμου χρησιμοποιούνται για την επιβεβαίωση των υπαρχόντων αποτελεσμάτων στη βιβλιογραφία και –με βάση συγκεκριμένες παρατηρήσεις– προτείνεται ένας νέος αλγόριθμος για την εκτίμηση της ελάχιστης πιθανότητας προώθησης.

Ακολούθως, το πολυώνυμο κάλυψης αναλύεται περαιτέρω και αποκαλύπτονται θεμελιώδεις ιδιότητες της πιθανοτικής πλημμυρίδας, οι οποίες δεν καλύπτονται στη βιβλιογραφία. Συγκεκριμένα, το προαναφερθέν πολυώνυμο μελετάται περαιτέρω και προκύπτει ένα νέο πολυώνυμο που εξαρτάται από την *μεγαλύτερη ιδιοτιμή* και το αντίστοιχο ιδιοδιάνυσμα του πίνακα γειτνίασης του δικτύου. Επιπλέον, εξάγονται αναλυτικές εκφράσεις σχετικά με (i) την κάλυψη, (ii) το χαμηλότερο όριο της πιθανότητας προώθησης που επιτρέπει την πλήρη κάλυψη του δικτύου, και (iii) ένα κάτω όριο του χρόνου τερματισμού. Δείχνεται ότι η ελάχιστη πιθανότητα προώθησης είναι αντιστρόφως ανάλογη της μεγαλύτερης ιδιοτιμή του πίνακα γειτνίασης του δικτύου και η πιθανότητα ενός κόμβου να λάβει το μήνυμα είναι ανάλογη της ιδιοδιανυσματικης κεντρικότητάς του αλλά και της ιδιοδιανυσματικής κεντρικότητας του κόμβου από τον οποίο ξεκίνησε η πιθανοτικη πλημμυρίδα, για μεγάλες τιμές του χρόνου t. Με βάση τα αναλυτικά αποτελέσματα, προτείνεται μια νέα πολιτική πιθανοτικής πλημμυρίδας, που ονομάζεται m-Probabilistic Flooding και αποδεικνύεται ότι οι απαιτήσεις για καθολική κάλυψη είναι ανεξάρτητες από τις φασματικές ιδιότητες του δικτύου.

Τα αποτελέσματα της ανάλυσης που πραγματοποιήθηκαν στο πλαίσιο αυτής της διατριβής μπορούν να εφαρμοστούν στα επερχόμενα περιβάλλοντα για τη διάδοση πληροφοριών. Όπως φαίνεται από τα αποτελέσματα της προσομοίωσης, τα αναλυτικά ευρήματα πλησιάζουν ικανοποιητικά τη συμπεριφορά της πιθανοτικής πλημμυρίδας επιτρέποντας τη μελέτη διάχυσης της πληροφορίας σε σύγχρονα περιβάλλοντα ασύρματων δικτύων, με χρήση στοιχείων από την αλγεβρική θεωρία γράφων.

Acknowledgments

Undertaking this PhD has been a truly life-changing experience for me that would not have been possible without the support and guidance that I received from many people.

I would like to express my special appreciation and thanks to my advisor Dr. Konstantinos Oikonomou, for being an always supportive mentor for me. Without his guidance and constant feedback this PhD would not have been implemented. Moreover, I would like to thank the Advisory Committee Members, Prof. Nektarios Koziris and Dr. Dimitrios Tsoumakos for their guidance and the Reviewing Committee members Prof. Spyros Sioutas, Dr. Vassilios Dimakopoulos, Dr. Georgios Kormentzas and Dr. Theodoros Andronikos. I would like also to thank Prof. Sonia Aïssa her the fruitfull collaboration all these years.

I thank my labmate at NMSLab, Dr. Georgios Tsoumanis for the time working together, and for the fun time we have had all these last years and my collegues at the NOC of the Ionian University.

I would like to thank my family: my parents Panagiotis and Vasiliki my sisters Maria and Natasa and my brother John for the encouragement and for supporting me throughout my life.

Last but not least, I would like to thank my daughter Vasiliki for her care and kindness and her patience during all these years.
Thank you, Vasiliki...

Parts of this research were supported by project “A Pilot Wireless Sensor Networks System for Synchronized Monitoring of Climate and Soil Parameters in Olive Groves,” (MIS 5007309) which is partially funded by European and National Greek Funds (ESPA) under the Regional Operational Programme “Ionian Islands 2014-2020”.

Contents

List of Figures

List of Figures

List of Tables

Chapter A

Introduction

INFORMATION dissemination is a crucial part of today's networks and it is expected to play an increasingly significant role in the forthcoming wireless networks of the numerous applications, the wide deployment and the large number of network devices. This introductory chapter aims to help the reader recall important milestones from the evolvement of today's networks, present some of the most important wireless network technologies, highlight the trends and the upcoming ones, before moving to literature review and the presentation of the motivation and contribution of this thesis.

A.1 Historical Milestones

The development of the first electronic computers, highlight the necessity of communication. The ARPANET project [1] funded by the *Advanced Research Projects Agency* (ARPA) of the United States Department of Defense in the early '60s was one of the first packet switching networks which made data transmission possible in 1965. The aim was to create a network that would offer communication between remote networks, even if intermediate systems were temporarily non-operational. The transmitted data are assembled into packets. Each packet encapsulates a part of the data along with the necessary information allowing to reach its destination, and upon successful reception, packets are reassembled into data. Until the early '70s, several packet switching networks have been developed such as the NPL network, Tymnet, Merit Network and CYCLADES, and the necessity of reconciling became

apparent. In 1973, the "Internetting research project" was launched to allow computers from different networks to communicate resulting in the development of the *Internet Protocol* (IP) [2].

In addition, the *Transmission Control Protocol* (TCP) [2] was also developed and complemented the Internet Protocol under the TCP/IP model stack. TCP defines how to establish and maintain a reliable, ordered, and error-free network connection between applications running on hosts over an IP network. IP packets, due to network congestion or traffic load balancing, may be lost or delivered out of order. TCP is able to detect such problems and rearranges the packets or requests for retransmissions.

In 1980, the *User Datagram Protocol* (UDP) was developed for transmitting messages (*datagrams*) between hosts over the IP network. UDP uses a connectionless mechanism for its operation and provides checksums for data integrity. However, there is no guarantee of correct delivery and message ordering.

Both IP and TCP form the basis of today's communications and, despite the existence of other protocols (e.g., UDP), the so-called TCP/IP protocol stack is dominant. The TCP/IP protocol stack partitions the communication system within a host into four layers [3]. The "Application" layer, handles the communication among the processes on the same or different hosts. It includes functions such as connection establishment and termination, character code translation, data encryption and compression. The "Transport" layer, includes functions for data segmentation, traffic control, acknowledgment, error detection, packet retransmission and packet ordering. The "Internet" layer, offers data exchange across network boundaries. It provides addressing and routing functions for delivering the data to the destination host [4]. The IP is the core protocol in the particular layer. The "Network Access" layer, is the lowest layer of the TCP/IP protocol stack. It encapsulates the packets of the Internet layer into *frames* and is responsible for transmitting frames between the hosts on the same network. The particular layer includes functions for line coding, bit and frame synchronization, error detection and flow control. Figure A.1 presents the layers of the TCP/IP protocol stack, which is the dominant networking model stack until now.

In 1983, the TCP/IP protocol stack, adopted as the standard protocol for ARPANET and in 1985, the United States *National Science Foundation* (NSF) funded the *National Science Foundation Network* (NSFNET) project, in order to

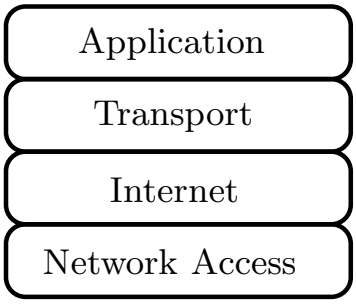


Figure A.1: *TCP/IP Layers.*

connect the supercomputing centers across the United States of America with the universities and the research centers. In the late '80s, several universities across the world, connect their networks to the NSFNET which became known as *INTERNET*. By the early '90s, the digital transmission became possible through the analog telephone network using a dial-up device called modem (modulator - demodulator), that covert the analog to digital signals and vice versa. The dial-up internet access offered connection speeds between 2400bps and 56Kbps.

A.2 Wireless Networks

More than 50 years ago, the United States Army employed radio signals for data transmission during World War II. *SIGSALY* [5], a heavily encrypted wireless data transmission technology was developed for securing the communications among the allies. In the late '60s, computers spread rapidly in academic institutions and in 1968, a group of researchers at the University of Hawaii, under the leadership of Norman Abramson started the development of the ALOHAnet [6], the first public wireless data packet network, that offered wireless connectivity to the campuses of the university. ALOHAnet employed *Ultra High Frequency* (UHF) radio signals for its operation and used a novel method for medium access control so-called *ALOHA random access* [7].

In the early '80s, the frequencies for mobile network systems became available, and in 1985 the *Federal Communications Commission* (FCC) permitted the use of several bands of wireless spectrum without a government licence, allowing the development of wireless local area networking such as WiFi. ALOHA channels

were used in the first generation mobile phone systems for signaling and control purposes [8]. In late '80s, the *European Telecommunications Standards Institute* (ETSI), was working on the *Global System for Mobile communications* (GSM) and expanded the use of ALOHA channels in mobile telephony. Additionally, the SMS message texting was implemented in the second generation mobile phone systems (2G) and in the early '00s ALOHA channels were added to the third generation mobile phone systems (3G) introducing the *General Packet Radio Service* (GPRS) that combines the slotted-ALOHA random access [9] and the Reservation ALOHA scheme [10].

A.2.1 Standard Technologies

Over the past two decades, several technologies regarding wireless communications have been developed, considering the different requirements of each environment. The widespread use of personal computers and mobile phones, the introduction of sensor networks and the need for short range device to device communications, lead to the development of different wireless technologies such as WiFi, WiMax, Bluetooth, Zigbee and Cellular communications, capable of exchanging data and meet the requirements of the particular environments. For example, cellular communications have been developed taking into account the need for wide coverage and the mobility of the users, while the focus of Bluetooth and Zigbee is set on the short range communication and the preservation of energy resources.

WiFi

Between 1980 and 1990, personal computers became more pervasive. The increasing demand for internet access, led to the need for wireless connectivity. Several vendors started to offer proprietary wireless solutions to their customers. In 1997, based upon Ethernet, the *Institute of Electrical and Electronics Engineers* (IEEE) introduced the 802.11 standard [11] in order to overcome the interoperability issues of the custom vendor solutions. The data rate of 802.11 standard was limited to 2 Mbps and several aspects such as encoding have not been addressed, leading to the development of incompatible equipment among different vendors [12].

The advantages in *Direct Sequence Spread Spectrum* (DSSS) technology [13], allowed for the development of the 802.11b standard, so-called Wireless Fidelity,

or WiFi operating at 2.4GHz with a maximum data rate of 11Mbps. However, the 802.11b employs the *Carrier Sense Multiple Access with Collision Avoidance* (CSMA/CA) [14] Media Access Control (MAC) Protocol, which limits the maximum throughput to a certain lower value [15].

At the same time, the 802.11a was also released. The particular standard operates at the 5GHz band and offers a data rate of 54Mbps using the *Orthogonal Frequency Division Multiplexing* (OFDM) [16] modulation. Given the operation of 802.11a at the 5GHz band, its effective range is slightly smaller than that of the 802.11b since the signal is absorbed more readily by solid objects. However, the adopted OFDM modulation is advantageous over DSSS, since the interference is significantly reduced in multipath environments [17] (e.g., indoor places).

In 2003, IEEE standardized the 802.11g that extended throughput to 54Mbps in the 2.4GHz band. Similarly to 802.11a standard, OFDM is the modulation scheme of 802.11g and due to the adoption of the CSMA/CA MAC protocol, its maximum theoretical throughput is limited to 31.4Mbps [18].

The need for higher data rates in communication systems, lead to the standardization of the 802.11n. The particular standard adopts the MIMO-OFDM modulation [19] and offers a theoretical data rate of 288.8 Mbps in the 2.4GHz band and 600Mbps in the 5GHz band, using multiple antennas to increase throughput. The most significant improvements are the increment of the channel bandwidth from 20MHz to 40MHz and the addition of the *Frame Aggregation* feature [20] to the MAC layer.

On the unlicensed 60GHz band, the *Wireless Gigabit Alliance* (WiGig) introduced the 802.11ad standard, that offers 6.7Gbps data rate with 2160MHz channel bandwidth. The support for beamforming [21], enables reliable communication in 10 meters range and it is backward compatible with the previous 802.11 standards.

In 2013, the 802.11ac had been introduced and similarly to 802.11n, uses multiple antennas and adopts the MIMO-OFDM modulation. It supports a maximum channel bandwidth of 160MHz and allows for even higher data rates of up to 3466.8Mbps in the 5GHz band. Table A.1 summarizes the presented 802.11 standards along with their properties.

Table A.1: *IEEE 802.11 standards.*

Standard	**802.11**	**802.11b**	**802.11a**	**802.11g**	**802.11n**	**802.11ad**	**802.11ac**
Year introduced	1997	1999	1999	2003	2009	2012	2013
Maximum data rate (Mbps)	2	11	54	54	600	6757	3466.8
Frequency band (GHz)	2.4	2.4	5	2.4	2.4 & 5	60	5
Channel bandwidth (MHz)	22	22	20	20	20, 40	2160	20, 40, 80, 160
Modulation	DSSS	DSSS	OFDM	OFDM	MIMO-OFDM	OFDM	MIMO-OFDM
Antenna configuration	Single	Single	Single	Single	Multiple	Single	Multiple

WiMAX

Worldwide Interoperability for Microwave Access (WiMAX), is an interoperability specification based on IEEE 802.16 and has been standardized in 2002. It has significantly more benefits than WiFi since it offers higher coverage (40km - 50km) and includes advanced mobility enhancements. It provides a wireless alternative to several broadband access technologies (e.g., xDSL, Cable) and can be used for last-mile broadband access. The first version of 802.16 standard, addresses the Line Of Sight (LOS) access using spectrum ranges between 10GHz and 66GHz. Several updates have been evolved to the particular standard such as 802.16a, 802.16c, 802.16d, 802.16e and 802.16m [22]. Its initial version offers a data rate of 40Mbps and this value was later increased up to 1Gbps and 100Mbps for fixed and mobile stations respectively [23].

The updated version of 802.16 released in 2004, added the 2GHz - 11 GHz range specifications for the physical layer operation and in 2005 the *Scalable Orthogonal Frequency Division Multiple Access* (SOFDMA) [24] was adopted by the 802.16e

standard as opposed to the OFDM in 802.16d. More recent versions of the particular standard include the MIMO technology, increasing the coverage and the bandwidth efficiency of WiMAX.

In order to support high peak data rates and offer significantly improved *Quality of Service* (QoS) comparing to previously MAC implementations, the MAC layer of the WiMAX is designed from the ground up. The *Subscriber Station* (SS) competes only once for gaining access to the network and after that, the station is allocated a time slot by the *Base Station* (BS). The time slot can be resized by the BS in order to meet the network's QoS requirements but remains assigned to the same SS. The MAC layer is responsible for organizing the upper layer's packets, so-called *MAC service data units* (MSDUs), into *MAC protocol data units* (MPDUs) and transmit them over the air. The frame aggregation mechanism allows for the aggregation of multiple MPDUs into a single burst and multiple MSDUs into a single MPDU to reduce overheads. Moreover, MSDUs may be fragmented into smaller ones and sent across multiple frames.

WiMAX includes encryption and authentication mechanisms that are handled by the MAC layer. Cryptographic schemes such as *Advanced Encryption Standard* (AES) and *Triple Data Encryption Standard* (3DES) are used to encrypt user data. Moreover, the device and user authentication framework are based on the *Internet Engineering Task Force* (IETF) Extensible Authentication Protocol (EAP) which supports a variety of authentication mechanisms.

The WiMAX architecture is depicted in Figure A.2 and includes three logical parts i.e., (i) the Subscriber Stations used by the end users, (ii) the *Access Service Network* (ASN) consisting of Base Stations and ASN gateways that form the radio access network at the edge and (iii) the Connectivity Service Network (CSN) that includes the Operational Support System (OSS) and provides IP connectivity, mobility and roaming services between ASNs through *Mobile IP Home Agent* (MIP-HA), the core IP network functions and connection with other networks such as PSTN.

Bluetooth

Bluetooth is a wireless communication protocol designed for short range communications among devices with low power consumption, based on low cost transceivers [25].

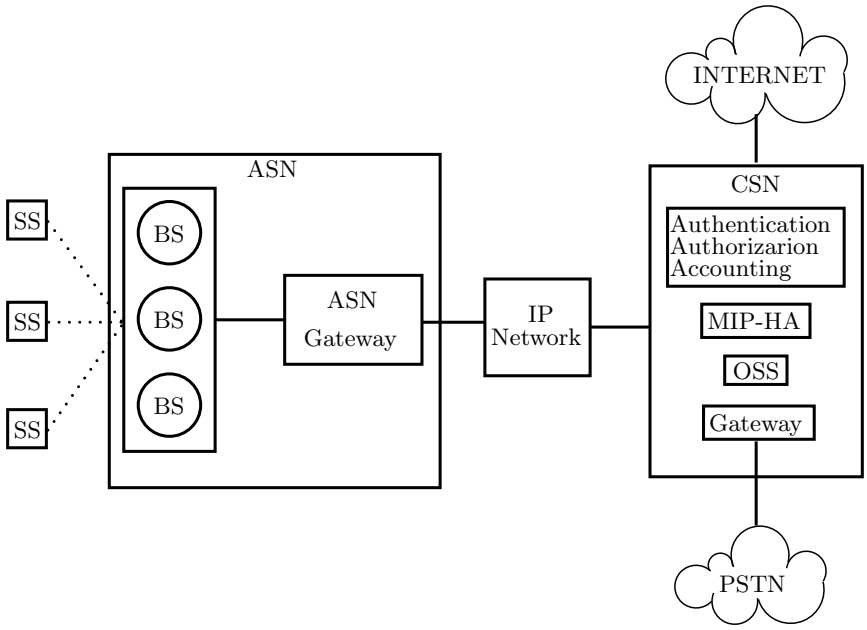


Figure A.2: *WiMAX architecture.*

Its physical and MAC layer has been standardized at 2002 by IEEE 802.15.1 and the range of included applications is known as *Wireless Personal Area Network* (WPAN). It operates at 2.4GHz band using the *Frequency Hopping Spread Spectrum* (FHSS) technology [26] where the radio signals are transmitted by switching the carrier among many frequency channels. The maximum data rate of the initial version of Bluetooth is limited to 1Mbps; however, this value was increased to 3Mbps in later versions.

Bluetooth follows a master/slave architecture and the data are transmitted into packets. Bluetooth defines two different topologies: the piconet [27] and scatternet [28]. A piconet is an ad-hoc network formed by a master and up to seven slave Bluetooth devices. Slave devices are synchronized by the master's clock and communicate in a point to point manner under the control of the master. However, a master's transmission may be either point to point or point to multipoint. A scatternet is defined as a collection of various piconets that overlap in space and time. A slave device may be a member of two or more piconets at the same time whereas, a master device in only one of them. Figure A.3 depicts a Bluetooth scatternet consisting of two piconets. Master devices are denoted by M1 and M2 and the slaves

with S1-S9. Slave devices S1, S2, S3, S4, S7 and S8, S9 belongs to piconet 1 and piconet 2 respectively whereas S5 and S6 are members of both piconets.

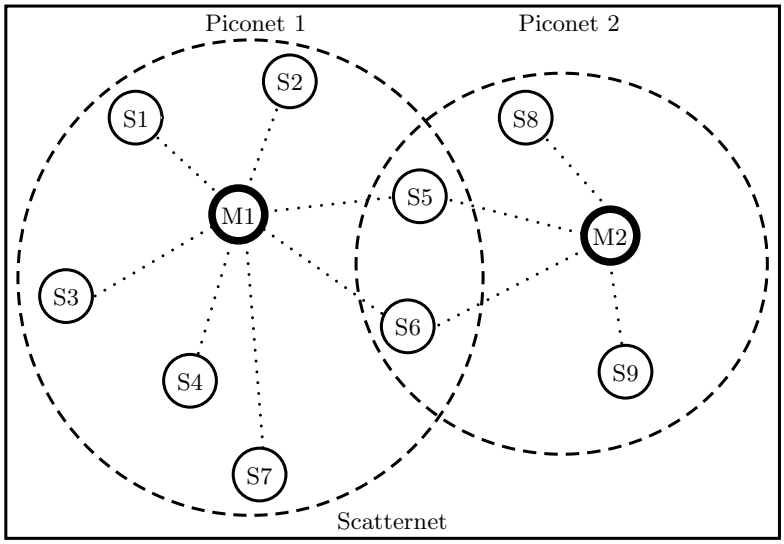


Figure A.3: *Example of a Bluetooth scatternet consisting of two piconets.*

In order to participate in a piconet, a device should interpret certain Bluetooth profiles which specify the general behavior of the communication [29]. Bluetooth profiles are used to control and configure the communication from the beginning and may also include information about dependencies on other profiles.

Packet transmission is based on the master's clock, having time intervals of $312.5\mu s$. Two consecutive time intervals form a time slot and two consecutive slots form a time slot pair. In order to avoid collisions, the Bluetooth protocol is based on the *Time Division Duplexing* (TDD) technology [30], where the master device starts the transmission at even time slots and the slave devices at the odd ones as depicted in Figure A.4. Moreover, for the throughput increment, packets are allowed to be three or five times slots long.

To preserve the energy resources, a slave device may be in standby mode where it is not participating in the piconet and listens every 1.28 seconds or 2.56 seconds for new messages or, in hold mode where only the internal timer is running for synchronization purposes [31].

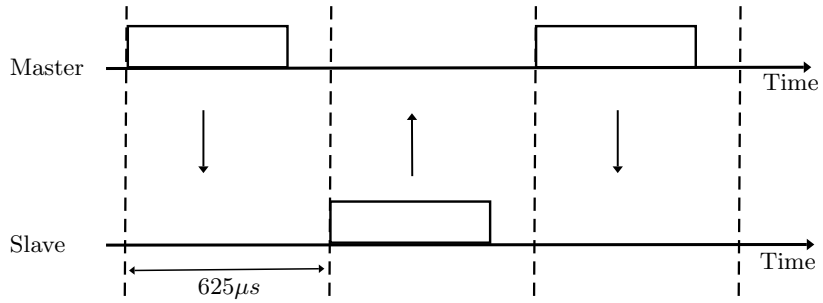


Figure A.4: *Example of a Bluetooth communication between a master and a slave device using single slot packets.*

Zigbee

Zigbee is a low cost and low power wireless communication protocol and its physical and MAC layers have been standardized by IEEE 802.15.4. It offers wireless connection to small scale projects in a simpler manner than other WPANs, such as Bluetooth. It is applicable for home and building automation systems [32], traffic management systems [33], and several consumer and industrial systems [34] that require low data rate and short range wireless data transfer [35]. Its effective coverage lays between 10 meters and 100 meters, depending on the transceiver and the environmental characteristics. However, data can be transmitted over longer distances in a multi hop manner through the network.

For the physical layer, the standard employs the CSMA/CA mechanism and specifies three wireless bands i.e., 868MHz, 915MHz and 2.4GHz that offer data rates of 20kbps, 40kbps and 250kbps respectively [36] whereas, the channel modulation is based on the DSSS scheme. In 2006, the revised version of 802.15.4 improves the maximum data rates to 250kbps for the 868MHz and 915MHz bands and announces the support for *Binary Phase Shift Keying* (BPSK) [37] and *Offset Quadrature Phase Shift Keying* (OQPSK) [38] modulation schemes.

The MAC layer of Zigbee allows for two different node types within the network i.e., *Full Function Device* (FFD) and *Reduced Function Device* (RFD). An FFD can serve as the coordinator of the network (i.e., selects the channel, the security policy and several network parameters) capable of communicating with any other device whereas, an RFD may only communicate with a single FFD and can never act as a

coordinator [35].

IEEE 802.15.4 standard, allows for either mesh or star topologies. However, any Zigbee network needs at least one FFD to act as the coordinator that sets the network's parameters. In star topologies, the devices communicate with the network coordinator whereas, in mesh topologies, an FDD is allowed to communicate with any other FFD within its range and an RFD is only allowed to communicate with a single FFD. Figure A.5 depicts a star and a mesh Zigbee topology consisting of a coordinator and several FFDs and RFDs.

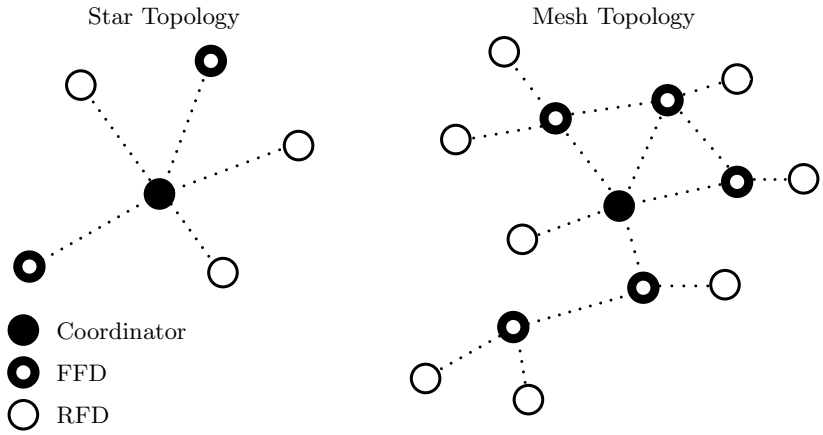


Figure A.5: *Example of a star and a mesh Zigbee topology.*

Cellular

Cellular networks have been introduced to support voice transmission using mobile phone devices. In 1947, the hexagonal cells were proposed by the engineers of Bell Labs [39] for mobile phones in vehicles. Each cell contains at least one transceiver that operates at different frequency band from its neighbor cells to avoid interference. Wide geographic areas can be covered by joined cells –instead of a single high power transceiver– enabling many portable devices to communicate with the base stations. Figure A.6 depicts a cellular network of nine joined cells that operate in three different frequency bands (i.e., F1, F2 and F3).

Several cellular networks have been deployed over the last years, allowing for connectivity with PSTN and internet. The *first generation mobile phone system*

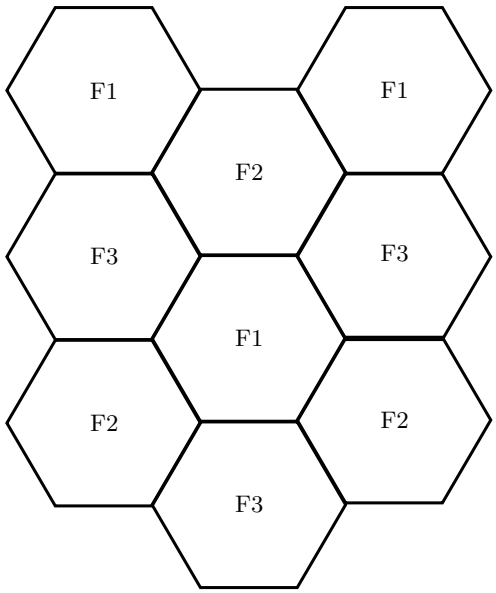


Figure A.6: *Example of a cellular network with nine cells and three different frequency bands.*

(1G) was deployed in 1979 for voice transmission over analog signals [40]. However, the connection to base station was initialized through digital signals in a different frequency band.

One of the most common first generation systems is the *Advanced Mobile Phone Service* (AMPS) that employs the *Frequency Division Multiple Access* (FDMA) technology [41], where two 25MHz frequency bands were allocated, one for the signal transmission and the other for the signal reception. Each band is split in two smaller bands of 12.5MHz allocated to different operators [42]. The channels were spaced 30KHz apart, allowing for 21 control and 395 voice channels.

In order to improve the system capacity, voice quality, and network coverage, the second generation mobile phone system (2G) based on the *Global System for Mobile communications* (GSM) standard, was introduced in 1991. The GSM, initially described a circuit switched network for voice transmission [43] however, later versions include data packet communications such as *General Packet Radio Services* (GPRS) [44] and *Enhanced Data rates for GSM Evolution* (EDGE). [45]. The 2G technology is mostly based on the *Time Division Multiple Access* (TDMA) technology [46], where a common frequency channel is divided into separate time slots

allocated to different users. In this way, the collisions are minimized since each user is allowed to transmit within a specific time slot.

The increasing demand for mobile broadband services and higher transmission rates lead to the development of the third generation mobile phone systems (3G) in 2001, offering *Mobile Broadband* (MBB) access. The 3G technology is based on the *Code Division Multiple Access* (CDMA) technology [47] where several users can transmit over a single frequency channel at the same time. Under CDMA, each user is assigned to a pseudo-random code that is used to modulate its own signal. At the receiver side, several signals arrive at the same time over the same frequency channel which are separated using the particular pseudo-random codes.

In 2009, the fourth generation mobile phone system (4G) was released resulting in a further improvement of the transmission rate. The 4G technology is employing all-IP network traffic, eliminating the circuit switching mechanism of the previous generations (1G, 2G and 3G) used for voice transmission. There are two different versions of the 4G technology, the LTE-FDD and the LTE-TDD. The LTE-FDD is based on the *Frequency Division Duplex* (FDD) technology [48], where two different frequency channels are used for uplink and downlink respectively and the LTE-TDD is based on the TDD technology, where a single frequency channel is divided into different time slots for uplink and downlink.

The proliferation of the mobile devices and the need for even higher data rate and reduced latency, lead to the proposal of the fifth generation mobile phone system (5G) which is expected to be available in 2020. Several frequency bands will be employed for the particular technology allowing for a data rate of up to 20Gbps. The *Non-Orthogonal Multiple Access* (NOMA) technology is proposed to be included in 5G where several users are allowed to transmit using the same resource such as time, frequency, or spreading code [49].

A.2.2 IoT, WSN and 5G

The growth of wireless broadband access and the technological advantages in embedded hardware design have enabled the extensive development of the *Internet of Things* (IoT) where devices such as home, industrial or health appliances, embedded with sensors and actuators can collect and exchange data [50]. The IoT, extends the internet connectivity beyond personal computers, mobile phones and tablets to

a wide range of everyday objects that can communicate and interact over the internet [51]. An IoT system consists of communication enabled smart devices with embedded processors, sensors and actuators that collect, transmit and act on data they acquire from the environment, without human intervention [52]. Each IoT device transmits its data directly to the internet or through a gateway as depicted in Figure A.7. The applications regarding IoT systems are applied in consumer, commercial, industrial and infrastructure domains. IoT devices are a crucial part of various applications such as, home/building automation, wearable technology, health care, transportation systems, environmental monitoring and agricultural systems. In the industrial domain, IoT integrates various manufacturing devices, enabling the rapid response to product demands and the optimization of the supply chain. It is expected that IoT installations will reach 25 billion devices by 2020 [53] resulting in a huge amount of data with different semantic, type, place and frequency characteristics. The centralized paradigm does not meet the storage and process requirements for such amount of data thus, a cloud approach may be employed [54]. Moreover, the requirements for low latency and high mobility that several IoT installations need, push the cloud functionality to the edge of IoT network, the so-called *Fog Computing* [55].

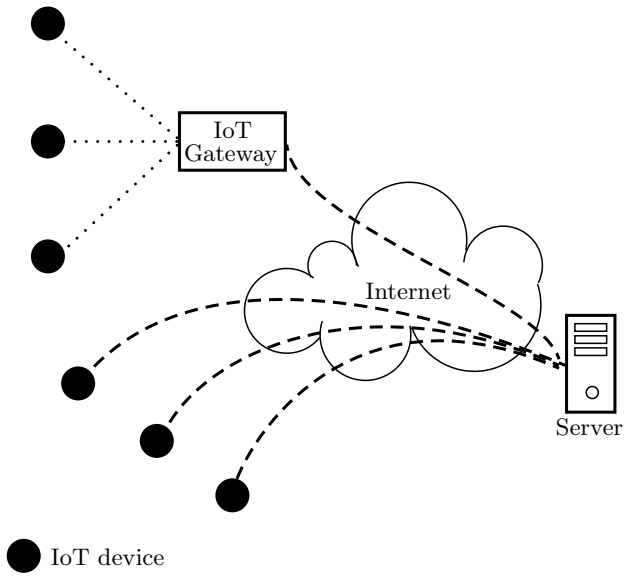


Figure A.7: *IoT network architecture.*

The data rate increment of the wireless transmission and the development of energy efficient processors, allowed for the introduction of *Wireless Sensor Networks* (WSNs), consisting of battery-powered sensor nodes that are spatially distributed in a wide area capable of sensing environmental conditions such as temperature, humidity, light and seismic waves [56]. The main difference with the IoT paradigm is that, the sensed data are often transmitted in a multi-hop manner towards a certain node (i.e., sink node) which can store them locally or transmit them to a central location [57] (e.g., a collection server) as depicted in Figure A.8. The WSN

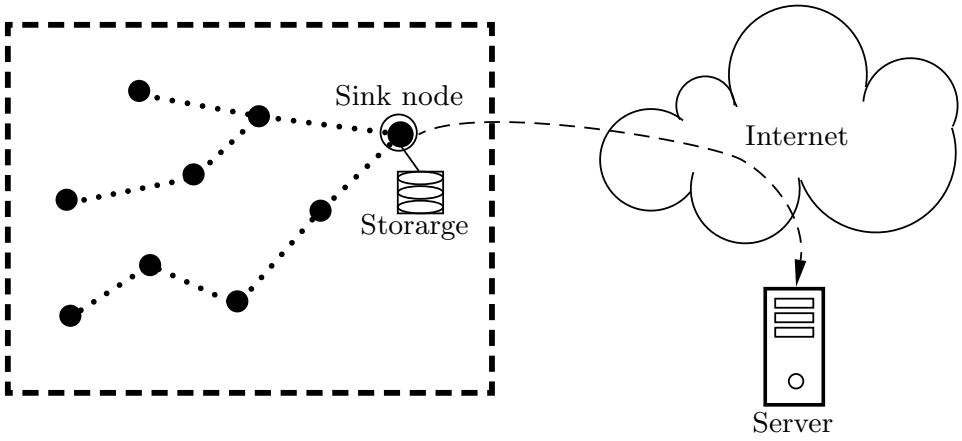


Figure A.8: *A WSN architecture.*

installations have increased rapidly due to the development of low cost sensors that offer scalability and adaptability [58]. In particular, WSNs may be extended by simply adding more sensor nodes that can quickly adapt to the network topology changes due to their inbuilt routing and relay capabilities.

WSNs are ofter deployed in areas where the battery replenishment may not be feasible. In order to prolong the sensor's lifetime, several energy harvesting methods may be employed such as, solar panels or wind turbines [59]. However, such methods do not guarantee energy replenishment deterministically [60] and several sensors may stop operating temporary, resulting in network topology changes. In order to cope with this problem, the network protocol stack should be power and routing aware. Figure A.9 depicts the sensor's protocol stack which, apart from the standard network layers (i.e., physical, data link, network, transport and application layer), includes additional cross-layer planes for coordinating the sensing tasks and reduce

the overall power consumption such as, power, mobility and task management [61].

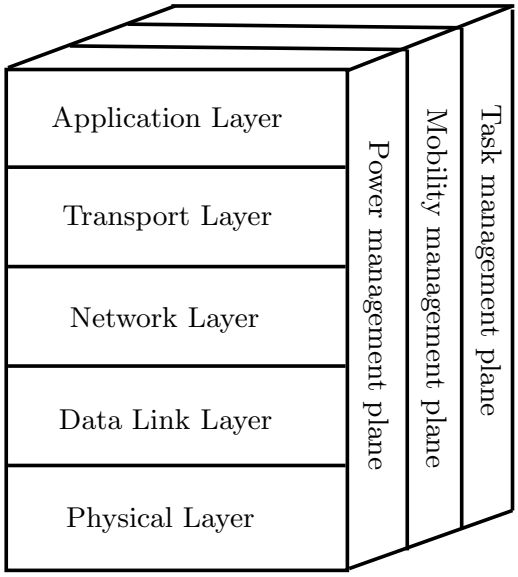


Figure A.9: *Sensor's protocol stack.*

The power management plane, attempts to reduce the consuming energy. For example it may turn off the transceiver periodically to avoid the reception of duplicated messages. The mobility management plane, undertakes the movement detection and updates the routing table accordingly to maintain the route to the sink node. The task management plane, balances the sensing tasks to the nodes such that only the necessary sensor nodes are assigned to certain sensing tasks in a particular area.

WSNs, can be classified in two categories, namely structured and unstructured [62]. Typically, structured WSNs consist of a small number of sensor nodes and are easy to manage while, in unstructured WSNs several sensors nodes are deployed in ad hoc manner and the management task is more complicated. Moreover, the control of a WSN may follow the centralized, decentralized or distributed model as depicted in Figure A.10. In centralized control model, a single node has global knowledge of the network and taking decisions regarding the node functionality (e.g., whether a node should be operating or not). In decentralized control model, the nodes are divided into clusters containing a cluster head node and the nodes functionality is determined by the cooperation of the head nodes. In distributed control model, there is

no central control and all nodes interact each other for network-wide decision of the node functionality.

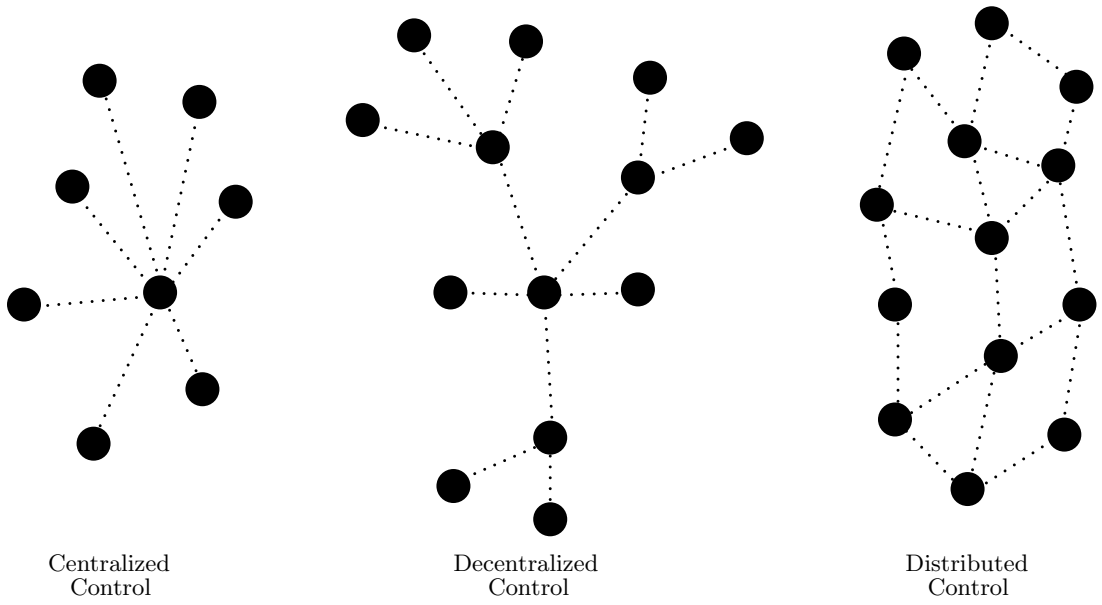


Figure A.10: *Control categories of WSNs.*

In order to simplify the control in large scale WSNs, the *Software Defined Networking* (SDN) [63] for WSNs environments has been proposed in the literature forming the *Software Defined Wireless Sensor Networks* (SD-WSNs) [64]. The main principles of SDN are the separation of the control and the data planes, the logical network centralization and the abstraction of the physical networks from the services [65]. The control layer interacts with the control applications by an *Application Programming Interface* (API) and with the network devices by the OpenFlow protocol [66] as depicted in Figure A.11.

In a SD-WSN, a particular node acts as the network controller and defines how the data are delivered among the nodes by defining data flows. Consequently, the nodes do not have to maintain routing tables since, the controller undertakes the routing decisions.

The increasing number of IoT and WSN installations and the proliferating demand for video and virtual reality traffic, require high data rates of many Gbps. The introduction of the fifth generation mobile phone systems (5G), will allow the

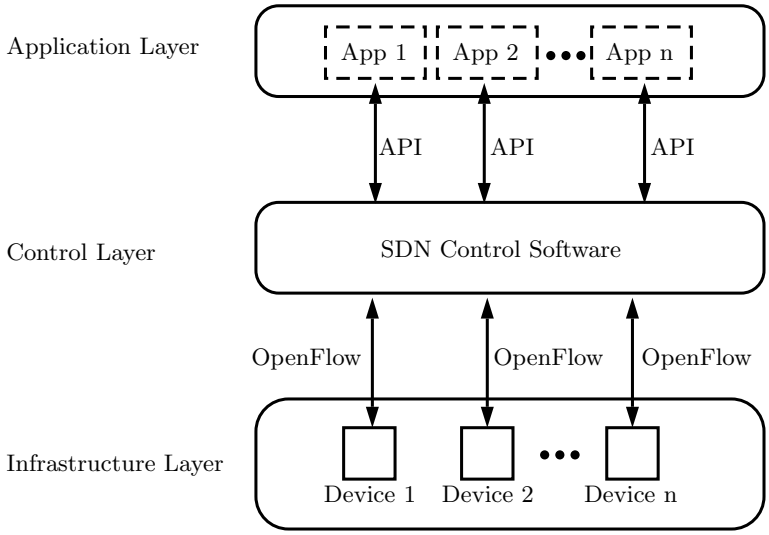


Figure A.11: *SDN architecture.*

wireless networks to meet the particular requirements [67]. 5G enables the interconnection of many WSN and IoT devices and allows the control and the digital observation to be moved from a local to a system-wide level. Moreover, access to information and data sharing are available anytime to anyone since the human-centric communications are expanded to include machine-centric communications also, giving more flexibility than the previous generations. The 5G system includes three generic communication services that is, *Extreme Mobile BroadBand* (xMBB) [68], *Massive Machine-Type Communication* (mMTC) [69] and *Ultra-reliable Machine-Type Communication* (uMTC) [70]. The xMBB combines low-latency and high data rate communication with very high coverage and smooth performance degradation as the number of mobile subscribers increases. The mMTC allows for the connectivity of billions of devices where the data rate is prioritized over wide area coverage and deep penetration. The uMTC provides low-latency and ultra-reliable communication for services that require high availability and reliability with low latency such as industrial manufacturing applications.

Today's networks include various costly hardware appliances that provide different services. The 5G system in order to yield high flexibility and to reduce the particular costs, involves SDN and *Network Function Virtualization* (NFV). A *VFN Infrastructure* (NFVI), provides multiple *Virtualized Network Functions* (VNF) that

may consist of various virtual machines running different processes in order to provide meaningful network services to the end users such as intrusion detection, intrusion prevention and routing, replacing the custom hardware appliances. The initialization, monitoring and proper operation of the VNFs are managed by an orchestration framework. The NFVI combines the software implementations of network functions, the hardware appliances and the virtualization management framework as it is depicted in Figure A.12. NFV and SDN are complementary although,

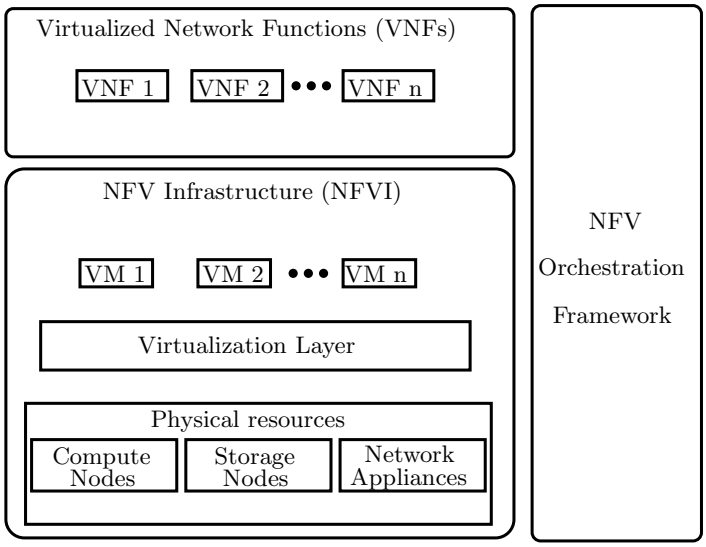


Figure A.12: *NFV framework.*

are not necessarily dependent on each other. NFV provides a flexible infrastructure where the SDN can run, allowing the flow based configuration of VNFs. NFV and SDN are novel architectures allowing for a new way to deploy a mobile network. Recent research projects on 5G have addressed the logical architecture design by defining VNFs and inter-VNF interfaces [71], [72]. This allows for a flexible placement of VNFs considering the opportunities and the limitations of the network and also only the necessary VNFs are applied thus, avoiding any overhead.

The 5G architecture is designed to meet the increasing demand in mobile and wireless traffic. In particular, mobile traffic is predicted to increase about 250 times over the decade 2010–2020 [73] and also, the number of communicating machines is forecasted to reach 25 billion devices by 2020 [53]. The expected increment in human-type and machine-type communications will lead to a wide variety of com-

munication characteristics with different requirements regarding cost, data rate, latency, mobility and reliability.

In such environments, several applications and services that are available in different network locations, regularly need to disseminate information for various purposes like data collection or to reveal useful features (e.g., sink node location, routing information etc). Consequently, any information dissemination method should meet the particular communication characteristics and use minimum network resources.

A.2.3 Information Dissemination

Information dissemination is a key enabler for the operation of several vital processes in wireless networks such as routing [74], service discovery [75], network management [76] etc, that need to spread data from a certain node (i.e., sink node) to all nodes within the network.

The requirements that should be met for efficient information dissemination are: reliability, efficiency and scalability [77]. In particular, a piece of information that needs to be disseminated, should be delivered to almost all of the nodes in the network (full coverage). However, it is not trivial to obtain full coverage, especially in environments where the topology may change arbitrarily (e.g., mobility environments). Moreover, spreading a message over the network, introduces a large amount of transmissions that occupy the wireless medium and consume energy resources. For example, the energy storage capacity in WSN environments is limited since sensor nodes rely on battery for their operation. Therefore, to prolong the network's lifetime, efficient dissemination methods may be employed to minimize the termination time and the redundant transmissions. In other environments such as *Industrial Internet of Things* (IIoT) and vehicular networks where reliability and minimum delay constraints need to be satisfied, dissemination methods should be able to spread information messages quickly to all network nodes. Furthermore, a satisfactory information dissemination method should adjust to any scale regarding the number and density of the nodes.

Different information dissemination techniques may be followed according to the category of the network topology. In particular, for the case of structured network topologies, information dissemination can take advantage of the global network knowledge while for the unstructured network topologies different approaches may

be followed.

In structured network topologies, dissemination methods are classified into plain-based and hierarchy-based. In plain-based methods, all network nodes are equivalent and the dissemination process takes advantage of the global network knowledge to spread the information message, minimizing the redundant transmissions and energy consumption [78], [79]. Such methods rely on the distance between the nodes or, their exact location using *Global Positioning System* (GPS). For example, the authors in [80] propose a dissemination method where any node that received the information message from a source node, rebroadcasts if its distance from the source node is larger than a certain threshold. Their aim is to increase the coverage area since, if the receiver is close to the sender, message rebroadcasting does not offer any significant coverage increment. Figure A.13 depicts two adjacent nodes (i.e., A and B) with transmission range r and distance between the nodes is denoted by d. The covered area by node A and node B is denoted by S_A and S_B respectively. The area that is covered by both nodes is denoted by $S_{A\cap B}$ and the area that is covered only by node B is denoted by $S_{B\setminus A}$. Initially, node A broadcasts the information message to the nodes within the area S_A, including node B. Subsequently, node B rebroadcasts the information message to nodes within area S_B. Nodes within area $S_{A\cap B}$ discard the message since it is already received from node A and only the nodes within area $S_{B\setminus A}$ receive the message for the first time. Clearly, the further nodes A and B are, the more nodes are covered by the rebroadcast of node B.

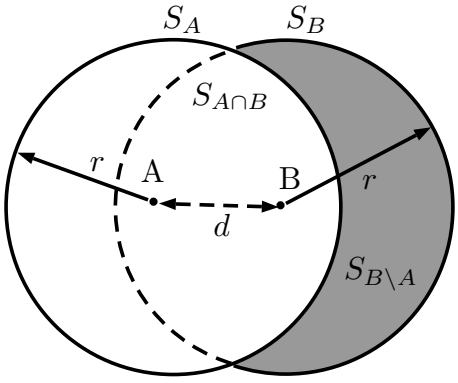


Figure A.13: *Coverage area of two adjacent nodes.*

In hierarchy-based methods, nodes are divided into clusters containing a cluster-

head node [81], [82]. Clusterheads form a backbone network and the dissemination method benefits from the knowledge of cluster topology of each clusterhead, to spread the information message within the network [83], [84]. Several algorithms have been proposed in the literature for cluster construction. For example, the authors in [85] propose a cluster construction algorithm where each node has a unique ID and nodes with the minimum ID are elected as clusterheads within the particular area. Member nodes in the cluster that can communicate with member nodes of a different cluster are called gateways. Figure A.14 depicts a network with three clusters, each one having a clusterhead and one or more gateways. Under hierarchy-based information dissemination methods, clusterheads broadcast the information message since they maintain connectivity with any cluster member. Apparently, cluster gateways may also rebroadcast the particular message to propagate it further to adjacent clusters.

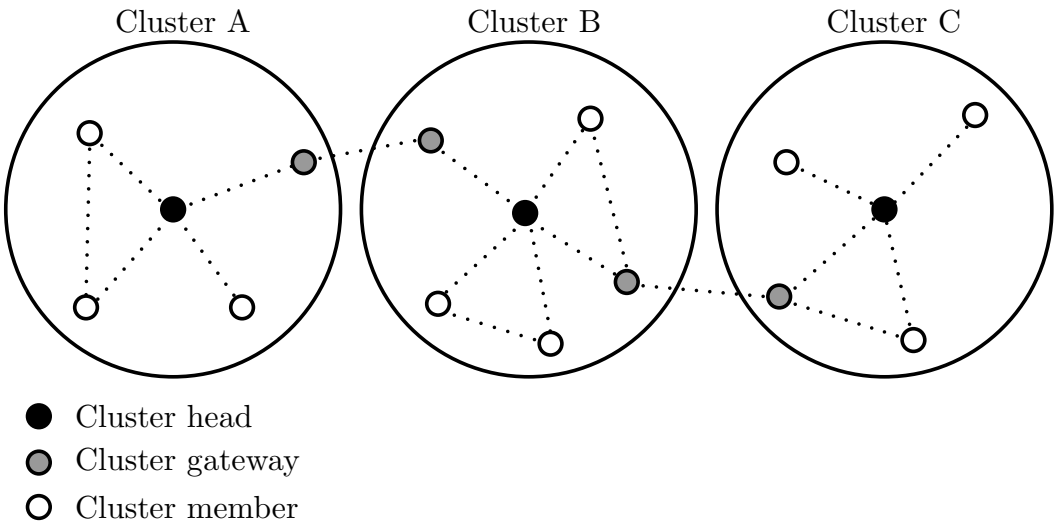


Figure A.14: *Example of a network with three clusters.*

In unstructured network topologies, dissemination methods rely on locally optimal strategies to spread the information message within the network since there is no knowledge regarding the underlying network topology. Consequently, such methods offer an approximation to the global optimal solution however, there is no additional overhead for determining the network topology as in the case of plain-based and hierarchy-based methods. The particular methods are classified into

negotiation-based and non-negotiation-based methods, based on whether a negotiation technique is employed [86].

Negotiation-based methods aim to mitigate the broadcast storm problem [80] while acquiring high reliability [87]. In such methods, before transmitting the information message, neighbor nodes exchange negotiation messages to elect a forwarder node. The forwarder node, is the only one that forwards information dissemination messages within the particular area.

The most common family of information dissemination methods that leverage a negotiation technique is the *Sensor Protocols for Information via Negotiation* (SPIN) [88]. The information message that needs to be spread over the network, is usually of a large size. Therefore, in SPIN-based methods, a source node that has the information message sends an advertisement packet (i.e., ADV) of a small size to its neighbors containing only message meta-data. All neighbor nodes that receive this meta-data for the first time, send back a message request packet (i.e., REQ) and, the source node transmits data packets (i.e., $DATA$) containing the information message to the neighbor nodes that the request received from. Figure A.15 depicts an example of the negotiation technique in SPIN-based methods. At the beginning only node 1 has the information message and sends ADV packets to its neighbors (i.e., node 2, node 3 and node 4) (Figure A.15-a). When nodes 2, 3 and 4, receive an ADV packet, they send back to node 1 a REQ packet (Figure A.15-b) and node 1 transmits data packets (DATA in Figure A.15-c). Subsequently, nodes 2, 3 and 4, transmit ADV packets to their neighbors (Figure A.15-d). Apparently, redundant ADV packets are also transmitted which are discarded (e.g., node 4 receives ADV packets from node 2 and node 3 and discards them). Nodes 5 and 6, send back REQ packets since, they had not receive the particular message in past (Figure A.15-e). Finally node 2 and node 3 send the information message to nodes 5 and 6, respectively (Figure A.15-f).

Non-negotiation-based methods aim to approximate the global optimal solution using only local knowledge of the network (i.e., the knowledge of nodes about their neighbors). Such methods are commonly used when the determination of the underline network topology is difficult (e.g., large scale environments) or insignificant (mobility environments). Blind flooding [89], is a commonly used method where each node that receives the information message sends it to all neighbors apart the one received from. Blind flooding maintains a simple control logic and its termination

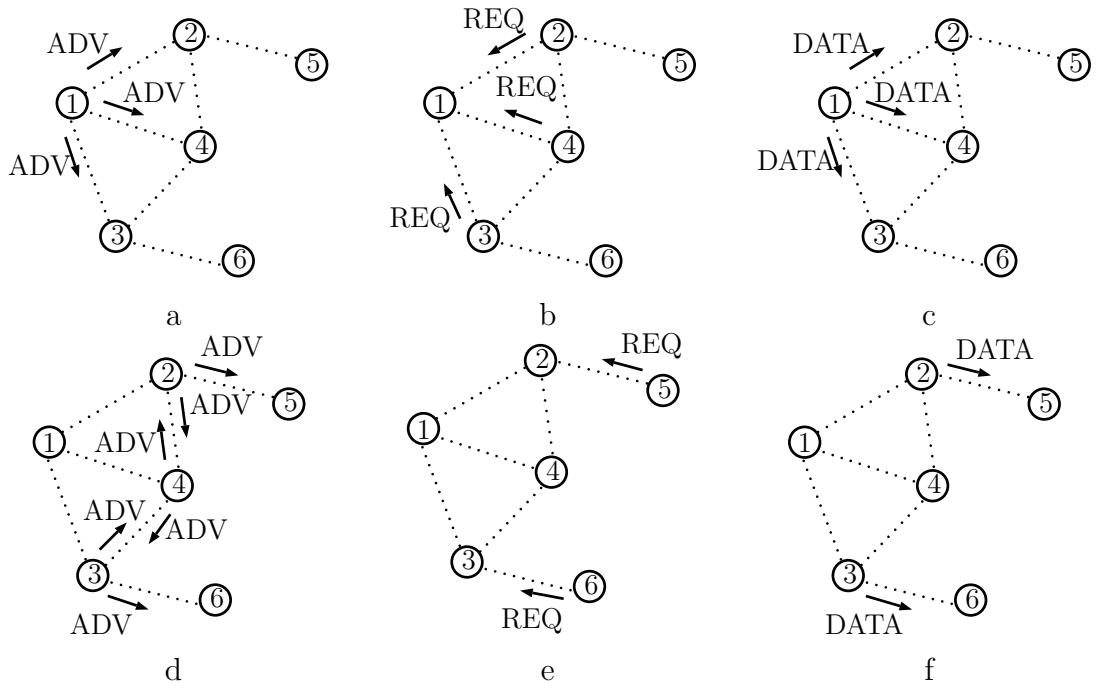


Figure A.15: *Example of the negotiation technique in SPIN-based methods.*

time is smaller than in negotiation-based methods where ADV and REQ packets has to be transmitted before the information message. However, blind flooding introduces a large amount of redundant messages resulting in the broadcast storm problem thus, undermining the energy resources. Several methods that aim to mitigate the broadcast storm problem such as probabilistic flooding [90], have been proposed in literature and are extensively discussed in Section A.3.

A.3 Past Related Work

This thesis focuses on probabilistic information dissemination in modern wireless networks that are inherently dynamic and large-scale and particularly, probabilistic flooding. In the following sections, past related work regarding probabilistic flooding in various network environments is extensively discussed. Moreover, works on relative scientific areas are also presented.

Blind flooding [89], one of the first broadcast mechanisms proposed in the literature, permits each node to retransmit a received information message, to all of its neighbors (apart from the one it has been received from). Although, it is a simple

mechanism suitable for the wired, small scale network paradigm of the early '80s and capable of reaching all network nodes in a deterministic manner (i.e., *global outreach*), it requires a large amount of usually redundant messages in the network, also known as the broadcast storm problem [80], thus, consuming valuable network resources (e.g., energy, bandwidth etc). An increasing volume of research attention has been observed recently regarding *probabilistic flooding* (e.g., vehicular networks [91], energy balancing [92], underwater broadcasting [93], advanced probabilistic flooding [94], nanonetworks [95], [96]). Probabilistic flooding can be seen as a suitable alternative for *pruning* transmissions by employing some fixed probability for message forwarding among neighbor nodes [90]. The basic idea is to employ small values for the previously mentioned *forwarding* probability in order to exclude any redundant links that would not result in *coverage* (i.e., the percentage of nodes that have received the information message) increment. Due to its probabilistic nature, probabilistic flooding cannot deterministically provide a global outreach, rather it guarantees it *with high probability*.

A.3.1 Probabilistic Flooding

Crisóstomo et al. [97] consider various topologies of stochastic nature in order to derive suitable values for the threshold probability. They model probabilistic flooding as a process of creating a subgraph by choosing network nodes, randomly. It is shown that the probability of global outreach under probabilistic flooding is equal to the probability that the subgraph is a connected dominating set of the network.

Saeed et al. [95], propose probabilistic flooding for nanonetworks [96]. The simulations on two and three dimensional grid structures verify the existence of phase transition phenomena and it is shown that the minimum forwarding probability that allows for global outreach depends on the number of nodes within the network, the node density and the transmission model.

Probabilistic flooding is also studied for underwater environments by Koseoglu et al. [93]. In particular, they investigate the performance of probabilistic broadcast for *autonomous underwater vehicles* (AUV) networks where the AUVs are not aware of locations of their neighbors. It is shown that the number of transmissions is significantly reduced thus, shorter backoff intervals may be used which results in lower average and maximum delay. Furthermore, the reduction of the transmissions

results in lower energy consumption and mitigates the broadcast storm problem.

Betoule et al. [98], propose a distributed routing algorithm for topology discovery suitable for *Internet Service Provider* (ISP) transport networks, that is inspired by opportunistic algorithms used in ad hoc wireless networks. They also propose a simple control plane, that is able to discover multiple paths toward the same destination which is based on probabilistic flooding. It is shown that this technique limits the amount of control messages significantly. The simulation results performed on four realistic topologies, confirm the analytical findings and indicate that the particular approach may be of high practical interest.

Wireless Sensor Networks

Probabilistic flooding has been also proposed in the area of Wireless Sensor Networks. Oikonomou et al. [99] study probabilistic flooding in random graphs [100] and asymptotic expressions regarding the forwarding probability, the number of transmissions and the coverage are derived. In particular, they model probabilistic flooding as a process of creating a random overlay network and the minimum value of the forwarding probability that achieves global outreach is considered as the probability of the overlay network to be connected. A comparison regarding the number of messages under probabilistic flooding and blind flooding is carried out and it is shown that there is a significant message reduction under probabilistic flooding especially for large networks. Moreover, certain aspects of probabilistic flooding are also revealed by the simulations results.

Lichtblau and Dittrich [101], propose a probabilistic breadth-first search method to approximate the reachability of Network-Wide broadcast protocols, while Badreddine and Butucaru [102] employ probabilistic flooding in their broadcast protocol for wireless body area networks.

Koufoudakis et al. [60], propose a modified version of probabilistic flooding capable of dealing with the idiosyncrasies of the nodes in energy harvesting environments. In such environments, when the remaining energy of a node falls below a particular threshold, the node becomes non-operational and thus is not participating in the network. As soon as the energy collection process replenishes the node's battery, the node becomes operational again. Under the proposed probabilistic flooding mechanism, when a node receives an information message, selects an arbitrary subset of

its neighbors and waits until all the nodes in the selected subset become operational in order to transmit the message. Simulation results performed in geometric random graphs considering environmental conditions obtained from a realistic dataset, demonstrate the efficiency of the proposed mechanism and it is shown that global outreach can be achieved at the expense of increased termination time.

Tameemi et al. [103], propose a modified version of probabilistic flooding that uses lower message overhead without compromising network connectivity for interference limited cognitive radio networks. In the particular scheme, each node selects its own forwarding probability according to the number of neighbors. The proposed scheme is compared with blind flooding and probabilistic flooding and it is shown that it outperforms both in terms of the number of transmissions and network coverage.

Peer to Peer Networks

Probabilistic flooding has been extensively studied in the area of peer-to-peer networks. Gaeta and Sereno [104] propose a novel search strategy for unstructured and decentralized peer-to-peer networks based on probabilistic flooding. In the proposed algorithm, a peer forwards a search query to its neighbors, according to a forwarding probability that depends on the distance from the query originator and the number of neighbors of both the sender and the receiver. The simulations are based on a generalized random graph model that takes into consideration the variability of the established application level connections and the distribution of resources. Moreover, an analysis regarding the capability of peers to relay queries is also included. The simulations performed on the proposed model and real topologies obtained from the Gnutella application, support the claims of the analysis and it is shown that the number of messages is significantly reduced under the proposed strategy, compared to the case of blind flooding.

Banaei-Kashani and Shahabi [105], introduce a novel method to model, analyze and design unstructured peer-to-peer networks, based on the theory of criticality and complex systems. They provide extended examples of how this method can enhance the analytical understanding of the unstructured peer-to-peer networks and the ability to design efficient mechanisms. In this context, they propose a probabilistic flooding search mechanism that improves the efficiency of blind flooding.

Tsoumakos and Roussopoulos [106], propose an adaptive probabilistic flooding search algorithm that feedback new searches from the previous ones. In particular, the proposed algorithm utilizes probabilistically directed walkers for efficient searching in unstructured peer-to-peer networks. The simulation results on various topologies demonstrate the efficiency of the proposed adaptive probabilistic flooding search algorithm and it is shown that the number of messages is significantly reduced while achieving a high success rate.

Margariti and Dimakopoulos [94], propose a probabilistic flooding strategy for unstructured peer-to-peer networks that adapts the forwarding probability according to the estimation of the covered nodes at each timestep and the popularity of the requested resource. The simulations performed on random graphs topologies [100], power-law topologies and real network traces, demonstrate the efficiency of the proposed adaptive probabilistic flooding search strategy. It is shown that the particular strategy achieves high success rates while the number of redundant messages is kept low.

Han and Liu [107], propose a scalable file sharing protocol for mobile peer-to-peer networks which employs probabilistic flooding to spread shares. They evaluate the particular protocol by comprehensive trace-driven simulations and it is shown that it achieves a high degree of anonymity with lower cryptography processing overhead.

Banaei-Kashani et al. [108], propose a decentralized and interoperable discovery service architecture with semantic-level matching capability based on Gnutella peer-to-peer network. The proposed architecture employs probabilistic flooding to propagate the query messages. In particular, each peer that receives a query message decreases its TTL value by one and forwards to its neighbors with a fixed probability. Any duplicate message or any message with zero TTL is discarded.

Wireless Ad hoc Networks

In the area of wireless ad hoc networks, Li et al. [92], propose a correlation-based probabilistic flooding algorithm. In particular, the algorithm tracks the aggregate acknowledgments in order to decide whether to retransmit a packet based on the link correlation between the wireless nodes. The simulations were performed on geometric random graphs and it is shown that the energy consumption can be significantly reduced while mitigating the energy hole problem.

The link correlation technique is also considered by Jianping et al. [109]. They propose coverage-aware and energy-balanced probabilistic flooding algorithm which takes into account the link correlation between one-hop neighbors and reduces the number of acknowledgment packets in the network. It is shown that the network's lifetime is prolonged and the congestion is reduced.

Drabkin et al. [110], propose a novel protocol which combines the advantages of probabilistic flooding, counter-based broadcast, and lazy gossip. In particular, the proposed protocol includes a probabilistic flooding phase complemented by a counter based forwarding and a gossip based mechanism. Thus, the messages that were not transferred by the probabilistic flooding phase can be delivered by the counter based forwarding and the gossip mechanism while maintaining low communication overhead. The simulations performed on non-sparse network topologies, confirm the high delivery ratio and the improved latency of the algorithm.

Reina et al. [111], propose a hybrid flooding scheme for wireless ad hoc networks which adapts its probability to the conditions of the network. They combine different flooding methods in order to cope with the broadcast storm problem encountered by the blind flooding. They take into account the density of the nodes within the network. Moreover, a forwarding zone criterion capable of dealing with the node mobility is considered. They compare the proposed scheme with probabilistic flooding and blind flooding. The simulation results validate the efficiency of the proposed scheme and it is shown that the number of messages is reduced while ensuring high network coverage even on high mobility scenarios.

Palmieri [112], present an adaptive probabilistic flooding strategy taking into account the broadcast propagation in reactive network discovery. The broadcast propagation behavior is modeled as harmonic spherical waves from a single source and the determination of the forwarding probability is based on how the intensity of a wave attenuates with its distance from the source. The proposed strategy has been analyzed within the *Ad hoc On-demand Distance Vector* (AODV) routing environment, however, it is shown to be applicable in any reactive routing scheme that relies on broadcast diffusion for route discovery. Simulation results demonstrate the capability of the proposed strategy to reduce the routing overhead significantly and increase the packet delivery ratio regardless of the network density, the network load and the node mobility.

Agarwal et al. [113], propose an energy efficient route discovery method for *mobile*

ad hoc networks (MANET) based on the probabilistic flooding algorithm using a fixed forwarding probability. Analytical and algorithmic approaches for determining a sufficient value regarding forwarding probability are derived. Simulation results demonstrate that the proposed method outperforms the existing routing protocols based on the AODV protocol concerning average control packets received and energy consumption.

Vehicular Networks

Probabilistic flooding is also considered as a suitable alternative to blind flooding in the area of *Vehicular Ad-hoc Networks* (VANET). Xeros et al. [114], propose a scheme which is based on the application of epidemic routing within the hovering area and probabilistic flooding outside the hovering area. Vehicles outside the hovering area that receive the information message serve as bridges for the partitioned uninformed areas resulting in high network coverage. The forwarding probability of the proposed scheme outside the hovering area is adaptive and varies according to the number of vehicles within the hovering area. Simulation results confirm that the proposed scheme outperforms other candidate hovering protocols.

Mylonas et al. [91], propose a speed adaptive probabilistic flooding for vehicular ad hoc networks. The particular scheme is based on probabilistic flooding in order to mitigate the broadcast storm problem introduced by blind flooding. The probability that a node forwards an information message to its neighbors depends on the average vehicle speed within the area of interest. Simulations are based on realistic datasets and it is shown that it achieves high network coverage and low latency of message delivery with low overhead.

Saeed et al. [115], introduce simple models for designing effective probabilistic flooding schemes for vehicular ad-hoc networks. Their analytical framework offers a generic tool which can be used in different scenarios regarding the number of road lanes, the transmission range and the node density. They also derive analytical expressions regarding the probability of all vehicles receiving the information message as a function of the forwarding probability. The evaluation of these models is based on realistic datasets and it is demonstrated the significant performance improvements against blind flooding in terms of network coverage and end to end delay.

A.3.2 Gossip Approaches

Besides the traditional flooding schemes, there are many approaches based on gossip algorithms for information dissemination in ad hoc networks [116]. Gossip-based techniques are widely used in wireless sensor network environments (e.g., [117]) using a flooding-like concept. Chang et al. [118] propose a probabilistic and opportunistic flooding algorithm that controls rebroadcasts and retransmissions opportunistically. Each node rebroadcasts the received message to a fraction of its one-hop neighbors considering the link error rate in order to maximize the expected delivery probability. Simulation results confirm the efficiency of the proposed algorithm and it is shown that even for high link error rate, high reliability can be achieved.

Cartigny and Simplot [119], propose probabilistic broadcast schemes that aim to reduce the number of broadcasts in ad-hoc wireless networks and hence, the medium occupation. In order to decrease the number of messages and maintain a high network coverage, those nodes located at the radio border of the sender should rebroadcast with higher probability than the others. Simulation results demonstrate the efficiency of the proposed probabilistic broadcast scheme and it is shown that the number of rebroadcasts can be significantly low.

Qi and Dharma [120], propose a probabilistic broadcast scheme that dynamically adjusts the rebroadcasting probability according to the node distribution and node movement based on local information, without requiring the assistance of distance measurements or location determination devices. They compare the proposed scheme with the AODV protocol and a fixed probabilistic approach and it is shown that it the number of rebroadcasts is significantly reduced.

Lysiuk and Haas [121], considering Continuum Percolation Theory, propose a gossip model where the broadcast probability is fixed and inversely proportional to the average node degree. Moreover, a distributed gossiping scheme is also proposed, in which the nodes set the gossiping probability according to the number of own neighbors. It is shown that both schemes are scalable since, for fixed network area, the expected number of transmissions is independent of the number of the network nodes. Simulations confirm the analytical findings and demonstrate the efficiency of the proposed schemes.

Kyasanur et al. [122], propose a decentralized probabilistic gossip protocol which takes in to account the wireless losses and node failures. This protocol dynamically

adapts the rebroadcast probability according to the network topology. Simulation results demonstrate the benefits of the proposed protocol over existing protocols. It is also shown that it can successfully achieve application's reliability requirements in all topologies, and significantly reduce overheads when a lower degree of reliability is sufficient.

Shi and Shen [123], propose a gossip-based routing algorithm that adapts the rebroadcast probability according to the number of neighbors of each node. Moreover, a node that receives a duplicate message less than a fixed number of times, it rebroadcasts again the particular message instead of discarding it.

Khelil et al. [124], propose a novel adaptive gossip algorithm that takes into account both node density and node mobility within the network. The proposed algorithm implements a heuristic to distribute the messages across network partitions and uses an adaptive rebroadcast probability for the message distribution within network partitions. Simulation results demonstrate the effectiveness of the proposed algorithm in different topologies with various node density and mobility levels.

Ryu et al. [125], propose dynamic probabilistic broadcast schemes where the rebroadcast probability is determined dynamically according to the number of one-hop and two-hop neighbors of each node. Simulation results demonstrate the efficiency of the proposed schemes and it is shown that the number of redundant messages is kept low while maintaining high reachability.

Beraldi [126], propose a new class of gossip algorithms which direct the gossip process towards a specific node. The broadcast probability varies according to the position of the node that broadcasts the information message. The position of each node is determined by sending beacon frames periodically which are used for the calculation of the time elapsed since a node met the destination and the dwell time of a node with the destination. The analytical model is also presented along with extensive simulations that confirm the suitability of the proposed protocol.

Hu [127], propose a probabilistic broadcast algorithm that ensures higher reliability at lower message complexity than previous works in the area. The simulations performed in geometric random graphs and it is shown that the proposed algorithm achieves satisfactory latency performance for high reliability scenarios.

Sasson et al. [128], study probabilistic broadcast algorithm considering the phase transition phenomenon observed in percolation theory and random graphs. Simu-

lations are based on both ideal and realistic scenarios and it is demonstrated that probabilistic broadcast enhances the successful packet delivery in dense networks.

A.3.3 Epidemic Approaches

An important aspect revealed in this thesis is the close relationship between probabilistic flooding and epidemic-based approaches focusing on the epidemic threshold. Wand et al. [129] model the virus propagation in real and synthesized network topologies. Under reasonable approximations, they prove that the epidemic threshold for a network is closely related to the largest eigenvalue of its adjacency matrix. It is shown that their epidemic threshold model, comprehends existing thresholds in the literature for the cases of Erdős-Rényi, Barabási-Albert and regular graphs.

Kephart and White [130], study the virus spreading in computer networks. They employ a standard epidemiological model on directed graphs and combine analytical and simulation results to study its behavior. It is shown that the epidemic threshold is closely related to the inverse of the average node degree of the network.

Pastor-Satorras and Vespignani [131], study epidemic thresholds in scale-free networks. They analyze computer virus infections obtained from real datasets in order to find the average lifetime and the persistence of viral strains on the Internet. They also propose a model for infection spreading on scale-free networks.

Van Mieghem et al. [132], analyze the influence of the network characteristics on virus spreading. They introduce an intertwined Markov chain-model and it is shown that the sharp epidemic threshold is equal to the inverse of the largest eigenvalue of the corresponding adjacency matrix.

Socievole et al. [133], study the relationship between the viral conductance [134] and the largest eigenvalue of the adjacency matrix. They described the influence of the particular metrics on network robustness with respect to the spread of SIS epidemics. Simulations performed on real Internet backbones and on Erdős-Rényi, Barabási-Albert, Watts–Strogatz and complete bipartite graphs. It is shown that the viral conductance and the largest eigenvalue of the adjacency matrix are highly correlated and the latter is sufficient for comparing network robustness.

A.4 Contribution

The main contribution of this thesis is the efficient information dissemination in wireless networks using probabilistic flooding. In particular, probabilistic flooding is modeled and analyzed in order to derive analytical expressions regarding the minimum value of the forwarding probability that allows for global outreach of the network (i.e., the threshold probability).

A.4.1 Problem Definition and Motivation

As already mentioned, the number of communicating devices is expected to increase rapidly, reaching about 25 billion devices by 2020 [53]. The 5G architecture is capable of enabling large scale WSN and IoT installations and offers flexibility regarding resource allocation. Any application or service is accessible by anyone at any time with minimum delay overhead. In order to meet the 5G requirements regarding low-latency, mobility and reliability, any application, service or WSN and IoT installation should be able to operate efficiently and use reasonable network resources.

Information dissemination plays a crucial role in modern network environments being an integral part of various vital processes. Flooding-like mechanisms are widely used by various processes (e.g., service discovery, data collection, routing) in order to disseminate information on the network. Blind flooding [89], an example depicted in Figure A.16, is a common mechanism for reliable information dissemination capable of reaching all network nodes in a deterministic manner (i.e., global outreach). However, it adopts a significant amount of redundant transmissions, leading to the broadcast storm problem [80], thus, unnecessarily using resources.

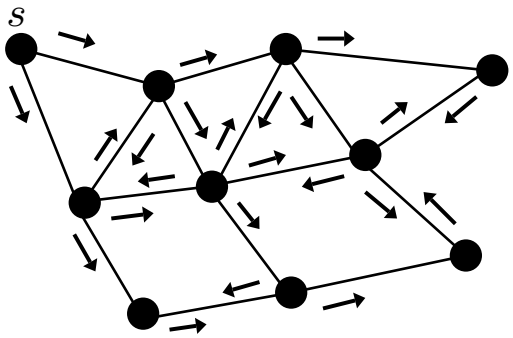


Figure A.16: *An example of the blind flooding mechanism.*

A.4.2 Proposed Approaches and Solutions

In this thesis, probabilistic flooding is proposed as a suitable alternative to blind flooding in order to prune redundant transmissions that do not result in network coverage and waste valuable network resources. In particular, probabilistic flooding is analytically modeled using elements from the area of algebraic graph theory. The binomial approximation is employed in order to proceed with the analysis and an analytical expression regarding coverage is derived in the form of a polynomial. The higher roots of the particular polynomial are shown to be a satisfactory approximation to the forwarding probability. Moreover, the analysis regarding the polynomial's roots confirms existing results in the literature and based on certain observations, a novel algorithm is proposed for the estimation of the threshold probability.

The derived coverage polynomial is further analyzed and fundamental properties of probabilistic flooding not covered in any previous work are revealed. In particular, a new polynomial that depends on the largest eigenvalue and the principal eigenvector of the network's adjacency matrix is introduced. In addition, (i) analytical expressions regarding coverage, (ii) a lower bound of the forwarding probability allowing for global network outreach, and (iii) a lower bound of the termination time are derived. It is shown that the threshold probability is inversely proportional to the largest eigenvalue of the network's adjacency matrix and the probability of a node to get the information message is proportional to the eigenvector centrality of the initiator node and itself for large values of time t. Analytical findings reveal that,

- initiating probabilistic flooding from a node with higher eigenvector centrality,

results in higher network coverage;

- nodes with higher eigenvector centrality are more probable to get covered and
- termination time lower bound decreases as the initiator's node eigenvector centrality increases.

Moreover, a new probabilistic flooding policy called m-Probabilistic Flooding is proposed and it is shown that the requirements for global outreach are independent of the underlying network's spectral properties. In particular, under m-Probabilistic Flooding, each node selects a fixed number of its neighbors arbitrarily, to transmit the information message.

Simulations are performed on *Erdős-Rényi* [100] (ER) and *Geometric Random Graph* [135] (GRG) topologies. For the simulation purposes, the simplified model of circular radio communication is considered as it is a common practice in the literature (e.g., [91, 92, 93, 118, 128]). The particular model does not take into account certain aspects like radio irregularities, e.g., [136], etc. However, the work presented here is based on the adjacency matrix and therefore, it is not affected by the considered model.

For the evaluation of the analytical expressions, a C++ program is developed in order to generate ER and GRG topologies. In GRG topologies, nodes are uniformly and independently distributed in the $[0, 1] \times [0, 1]$ plane and a link between any pair of nodes exists if and only if the Euclidean distance between the said pair is smaller than the connectivity radius r_c. In ER topologies the existence of a link between two nodes depends on the *connection probability* p. GNU Octave [137] with Octave networks toolbox [138] are also used to evaluate the analytical expressions. Unless otherwise stated, the presented results are average values over 30 independent runs.

A.4.3 Contribution Summary

Probabilistic flooding is studied here in order to reveal its fundamental properties that enable the efficient information dissemination in modern wireless networks that are dynamic and typically large-scale. Using algebraic graph theory elements, coverage is modeled in the form of a polynomial and it is shown to be closely related to the corresponding graph's spectrum. A new algorithm for the estimation of the threshold probability is proposed. Moreover, analytical expressions regarding coverage,

a lower bound of the forwarding probability allowing for global network outreach, and a lower bound of the termination time are derived. It is shown that probabilistic flooding efficiency is closely related to the spectral properties of the network's graph and a new probabilistic flooding policy called m-Probabilistic Flooding is also proposed.

A.5 Thesis Structure

The current chapter of this thesis presents a short history of the internet and its dominant protocols that enable reliable and fast packet transmission. Moreover, the most important wireless technologies (i.e., WiFi, WiMAX, Bluetooth and Zigbee) and the four generations of mobile systems (i.e., 1G, 2G, 3G and 4G) are also presented. Furthermore, WSN, IoT and the next generation mobile system (i.e., 5G) architectures are discussed and the focus is set on the requirements for wide coverage, high data rate, low latency, mobility and reliability.

In order to meet the particular requirements, probabilistic flooding is proposed as a suitable alternative to the traditional blind flooding mechanism, to prune redundant transmissions that do not contribute to network coverage. A survey of the related works in epidemic spreading and gossip based approaches and also the information dissemination in peer to peer, wireless ad-hoc and vehicular networks is also given here along with the motivation and a brief description of the contribution of this thesis.

In Chapter B, the basic algebraic graph theory elements are presented and in Chapter C, probabilistic flooding is studied and coverage is analytically modeled in the form of a polynomial. The roots of the particular polynomial are discussed in Chapter D and an algorithm for the estimation of the threshold probability is proposed.

Chapter E reveals the fundamental properties of probabilistic flooding and it is shown to be closely related to the spectrum of the corresponding network's graph. In particular, a new polynomial is derived regarding coverage and the analysis focuses on certain properties of the adjacency matrix. These properties are exploited to derive analytical expressions regarding the behavior of probabilistic flooding. Moreover, for certain graph topologies, a lower bound regarding forwarding probability is derived which is independent of their spectral properties. Also, a new probabilis-

tic flooding policy called m-Probabilistic Flooding is proposed and it is shown that the requirements for global outreach are independent of the underlying network's spectral properties.

The contribution summary and the conclusions are drawn in Chapter F. The various theorems, lemmas and corollaries are included in the appendices apart from those that their proof was important to be included within the main part of this thesis. Additional information about the simulator, abbreviations, list of publications etc, is included in the rest of the appendices.

Chapter B

Algebraic Graph Theory Elements

ANY complex network can be represented by a *graph*. Graphs can be *undirected* or *directed* depending on the type of communication between the adjacent nodes. An undirected graph $G(V, E)$ is a finite set of *vertices* (i.e., V) and a pair of elements of V called *edges* (i.e., E). Similarly, a directed graph or *digraph* is a finite set of vertices and a pair of ordered elements of V (i.e., *directed edges*). An example of an undirected and a directed graph is presented in Figure B.1 and Figure B.2 respectively. The dots represent the vertices and the lines connecting them represent the edges of the graph.

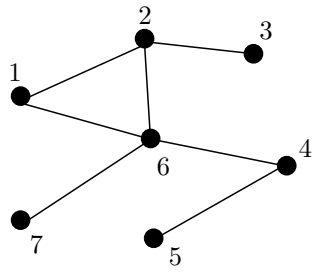


Figure B.1: *Undirected graph example.*

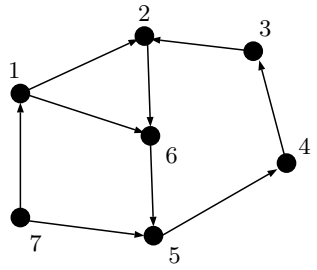


Figure B.2: *Directed graph example.*

A path in a graph is a sequence of edges connecting a sequence of distinct vertices. For example in Figure B.1 a possible path that connects vertices $1, 2, 6, 4$ and 5 is $(1, 2)(2, 6)(6, 4)(4, 5)$. A graph is said to *connected* if there exists a path among any pair of vertices, otherwise the graph is disconnected. For example the graph presented in Figure B.3 is disconnected since there is no path that connects vertices 1 and 4.

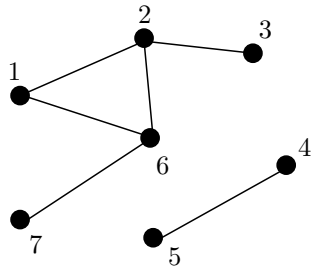


Figure B.3: *Disconnected graph example.*

Graphs can be further categorized based on the connections between vertices. The most common graph categories are: Complete, Regular, Bipartite, Complete Bipartite and Random graphs [139].

A complete graph is a graph in which any pair of vertices is connected by an edge. Clearly, a complete graph of n vertices denoted by K_n, has $\frac{n(n-1)}{2}$ edges. Figure B.4 depicts examples of complete graphs of 5,6 and 7 vertices (i.e., K_5, K_6 and K_7 respectively).

In regular graphs, all vertices have the same number of connections. The number of connections of a vertex characterizes further a regular graph. Thus, a reg-

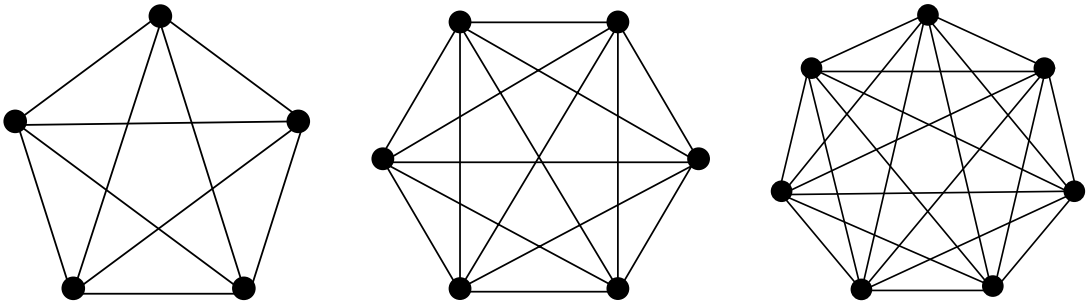


Figure B.4: *Complete graph examples.*

ular graph where each vertex has k connections is said to be a $k-$regular graph. A $k-$regular graph of n vertices has $\frac{kn}{2}$ edges. Figure B.5 depicts examples of $3-$regular, $4-$regular and $5-$regular graphs.

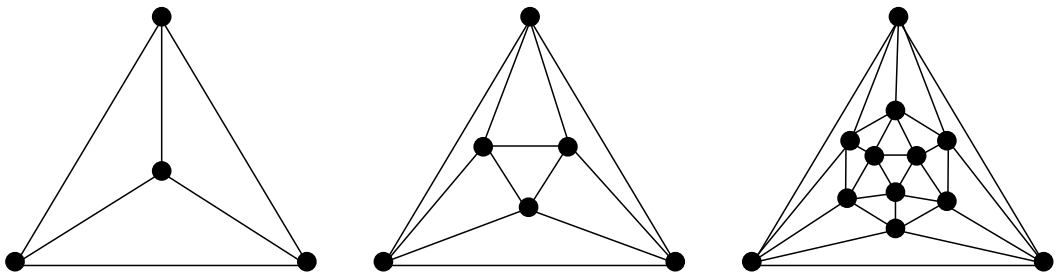


Figure B.5: *Regular graph examples*

In bipartite graphs, the vertices are divided into two sets U and V such that every edge connects a vertex that belongs to U with a vertex that belongs to V. Consequently, the neighbors of any vertex in set U belong to set V. Similarly, in complete bipartite graphs, all vertices of set U are connected to all vertices of set V. Figure B.6 depicts an example of a bipartite graph and a complete bipartite graph.

In random graphs, the edges that connect the corresponding vertices are generated randomly. There are various models for the generation of random graphs. For example, the edges in *Erdős-Rényi* [100] (ER) graphs are generated according to fixed probability p (i.e., *connection probability*). Clearly, the number of edges of the graph is a random variable with expected value given by, $\binom{n}{2}p$, where n is the

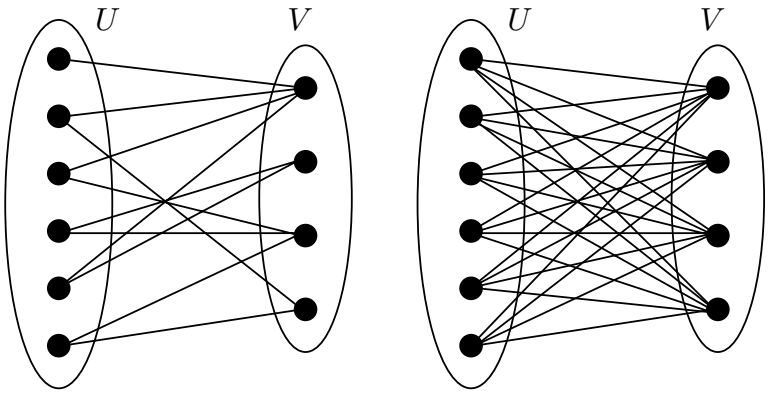


Figure B.6: *Example of a bipartite and a complete bipartite graph*

number of graph vertices (i.e., $n = |V|$). The number of the neighbors of each vertex is also a random variable and its expected value is given by, np.

In *Barabási-Albert graph* model, the probability of vertex j to connect with vertex i is given by, $\frac{d_i}{\sum_{x=1}^{x=n} d_x}$, where d_i is the number of neighbors of vertex i. The particular graphs are commonly used in the literature (e.g., [140], [141], [142], [143], [144]) to model several natural and human-made systems such as social networks, citation networks, peer-to-peer networks etc.

Geometric Random Graph [135] (GRG) is another model for the generation of random graphs. In GRG the vertices are independently and identically distributed in a plane. An edge between a pair of vertices exists if their Euclidean distance is less or equal than a fixed value r_c called *connectivity radius*. It is shown in [135] that a GRG is disconnected for any value of the connectivity radius less or equal than $(1-\epsilon)\sqrt{\frac{\log n}{\pi n}}$, where $\epsilon \to 0^+$. The particular graph model is widely used in the literature to model ad-hoc and wireless sensor networks (e.g., [91, 92, 93, 118, 128]).

B.1 The Adjacency Matrix

Any graph can be represented by a matrix such as the *Adjacency* ($\mathcal{A}$), the *Laplacian*, or the *Incidence* matrix [145]. This thesis focuses on the adjacency matrix of undirected graphs that is an $n \times n$ square matrix where n is the number of vertices

of the graph (i.e., $n = |V|$). The element at row i and column j can be either one or zero depending on whether the vertices i and j are connected or not i.e.,

$$(\mathcal{A})_{i,j} = \begin{cases} 1, & (i,j) \in E \\ 0, & (i,j) \notin E \end{cases}$$

For example, the corresponding adjacency matrix of the graph presented in Figure B.1 is given by,

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Since self-links are not considered in the following analysis, the diagonal elements of the adjacency matrix are always zero i.e., $(\mathcal{A})_{i,i} = 0, i \in [1,n]$. Consequently, the *trace* (i.e., the sum of the diagonal elements) of $\mathcal{A}$ equals zero, $tr(\mathcal{A}) = 0$. It is observed that the adjacency matrix is symmetric i.e., $\mathcal{A} = \mathcal{A}^T$, where $\mathcal{A}^T$ is the transpose of $\mathcal{A}$. Also the sum of row's (or column's) i elements equals the number of one-hop neighbors of vertex k (i.e., vertex k degree d_k) or,

$$\sum_{i=1}^{i=n} (\mathcal{A})_{i,k} = \sum_{i=1}^{i=n} (\mathcal{A})_{k,i} = d_k.$$

Consequently, the average degree $\bar{d}$ of the graph is given by,

$$\bar{d} = \frac{\sum\limits_{i=1}^{i=n} d_i}{n} = \frac{\sum\limits_{i=1}^{i=n}\sum\limits_{j=1}^{j=n} (\mathcal{A})_{i,j}}{n} = \frac{2L}{n},$$

where L is the number of links (i.e., $L = |E|$). Let Y denote the all ones $1 \times n$ vector such that $Y = [1, 1, 1...., 1]$. Then, the element at position i of the $1 \times n$ vector $\mathcal{D}$ given by,

$$\mathcal{D} = \mathcal{A}Y^T, \tag{B.1}$$

corresponds to the degree of vertex i i.e., $(\mathcal{D})_i = d_i$. Consequently, the number of links of the graph is given by,

$$L = |E| = \frac{1}{2}\sum_{i=1}^{i=n}(d_i) = \frac{1}{2}\sum_{i=1}^{i=n}(\mathcal{D})_i = \frac{1}{2}Y\mathcal{A}Y^T = \frac{1}{2}\sum_{i=1}^{i=n}\left(\sum_{j=1}^{j=n}(\mathcal{A})_{i,j}\right).$$

B.2 Graph Walks

A walk of length k in a graph from vertex v_0 to vertex v_k is a succession of k arcs such as $(v_0 \to v_1)(v_1 \to v_2)\dots(v_{k-1} \to v_k)$. A closed walk of length k is a walk that starts and returns to the same vertex after k hops i.e., $(v_0 \to v_1)(v_1 \to v_2)\dots(v_{k-1} \to)v_0$. A path is a walk in which all vertices are different [146].

A useful property of the adjacency matrix is that the number of walks of length k from vertex i to vertex j is given by, $(\mathcal{A})^k_{i,j}$. This can be proved by induction as follows. For $k = 1$ the number of walks from vertex i to vertex j is the number of links between i and j. Consequently, if i and j are adjacent, the number of walks equals one, otherwise equals zero i.e., $(\mathcal{A})_{i,j}$. A k-length walk from vertex i to vertex j consists of a $(k-1)$-length walk from vertex i to a vertex r which is adjacent to vertex j Let the number of walks of length $k-1$ from vertex i to vertex r be given by, $(\mathcal{A})^{k-1}_{i,r}$. The number of walks of length one from vertex r to vertex j equals one. It follows that the total number of walks of length k from vertex i to vertex j is given by,

$$\sum_{r=1}^{r=n}\left(\mathcal{A}^{k-1}\right)_{i,r}(\mathcal{A})_{r,j} = \left(\mathcal{A}^k\right)_{i,j}$$

Employing the rules of matrix multiplication the number of walks of length k from vertex i to vertex j is given by,

$$\left(\mathcal{A}^k\right)_{i,j} = \sum_{r_1=1}^{r_1=N}\sum_{r_2=1}^{r_2=N}\cdots\sum_{r_{k-1}=1}^{r_{k-1}=N}(\mathcal{A})_{i,r_1}(\mathcal{A})_{r_1,r_2}\dots(\mathcal{A})_{r_{k-1},j}$$

For example, considering the graph presented in Figure B.1, the number of walks of length two between vertex 1 and vertex 6 equals one (i.e., $(1 \to 2)(2 \to 6)$). Moreover, the number of walks of length three between the same vertices is five as seen next,

1. $(1 \to 2)(2 \to 1)(1 \to 6)$

2. $(1 \to 6)(6 \to 1)(1 \to 6)$
3. $(1 \to 6)(6 \to 4)(4 \to 6)$
4. $(1 \to 6)(6 \to 7)(7 \to 6)$
5. $(1 \to 6)(6 \to 2)(2 \to 6)$

and the corresponding matrices $\mathcal{A}^2$ and $\mathcal{A}^3$ are given by,

$$\mathcal{A}^2 = \begin{bmatrix} 2 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 3 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 4 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathcal{A}^3 = \begin{bmatrix} 2 & 4 & 1 & 1 & 1 & 5 & 1 \\ 4 & 2 & 3 & 1 & 1 & 6 & 1 \\ 1 & 3 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 2 & 5 & 0 \\ 1 & 1 & 0 & 2 & 0 & 0 & 1 \\ 5 & 6 & 1 & 5 & 0 & 2 & 4 \\ 1 & 1 & 1 & 0 & 1 & 4 & 0 \end{bmatrix} \text{ respectively}$$

and it is observed that $(\mathcal{A}^2)_{1,6} = 1$, $(\mathcal{A}^3)_{1,6} = 5$.

A graph is said to *connected* if there exists a walk among any pair of vertices explicitly,

$$\exists k \in \mathbb{N} : (\mathcal{A}^k)_{i,j} \neq 0 \text{ for } i, j \in [1, n]$$

The smallest integer k such that $(\mathcal{A}^k)_{i,j} \neq 0$ for each pair (i, j) is called *diameter* of G. Thus, the diameter is the length of the longest shortest path in G.

A useful matrix property for detecting graph connectivity is the matrix reducibility. A matrix $\mathcal{S}$ is reducible if and only if it can be placed into an upper-triangular form i.e.,

$$\mathcal{S} = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

otherwise, the matrix is irreducible. Since the adjacency matrix is symmetric, a reducible adjacency matrix is of the form,

$$\mathcal{A} = \begin{bmatrix} B & 0 \\ 0 & D \end{bmatrix}.$$

It is clear that the vertices are divided into two sets B and D. The vertices belonging to set B are not connected with any vertex of set D thus, the graph is disconnected. Consequently, the adjacency matrix corresponding to a connected graph is irreducible [145].

B.3 Graph spectrum

The spectrum of graphs is commonly used in graph theory to classify the structure and the properties of graphs. In order to comprehend graph's spectrum, the following definitions are presented. The characteristic polynomial of the adjacency matrix is by definition

$$\Pi_{\mathcal{A}}(\lambda) = \det(\mathcal{A} - \lambda I),$$

where det() denotes the determinant and I the $n \times n$ identity matrix. The characteristic polynomial can be written in the form of

$$\Pi_{\mathcal{A}}(\lambda) = \sum_{i=0}^{i=n} c_k \lambda^k = c_n \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0, \tag{B.2}$$

where c_i are the coefficients of the polynomial. Since the characteristic polynomial is of degree n it follows that,

$$\Pi_{\mathcal{A}}(\lambda) = \prod_{i=0}^{i=n} (\lambda_k - \lambda). \tag{B.3}$$

The roots of the particular polynomial (i.e., $\lambda_1, \lambda_2, \ldots \lambda_n$) are the eigenvalues of the adjacency matrix $\mathcal{A}$. From Equation (B.2) and Equation (B.3) for $\lambda = 0$ it follows that,

$$\Pi_{\mathcal{A}}(0) = \det(\mathcal{A}) = c_0 = \prod_{i=0}^{i=n} \lambda_k.$$

Hence, if $\det(\mathcal{A}) = 0$ there is at least one zero eigenvalue. It is shown [145] that, $c_n = (-1)^n$ and $c_{n-1} = (-1)^{n-1} tr(\mathcal{A}) = 0$. For example the characteristic polynomial of the adjacency matrix corresponding to the graph of Figure B.1 is given by,

$$\Pi_{\mathcal{A}}(\lambda) = -\lambda^7 + 7\lambda^5 + 2\lambda^4 - 10\lambda^3 - 2\lambda^2 + 3\lambda$$

and it is clearly observed that $c_7 = (-1)^7 = -1$, $c_6 = (-1)^6 tr(\mathcal{A}) = 0$ and $c_0 = \det(\mathcal{A}) = 0$.

The eigenvectors of an $n \times n$ square matrix $\mathcal{S}$ are a special set of non-zero vectors that satisfy,

$$\mathcal{S}\vec{x}_i = \lambda_i \vec{x}_i.$$

Similarly, the left eigenvectors satisfy,

$$\vec{y}_i \mathcal{S} = \lambda_i \vec{y}_i.$$

Each eigenvector $\vec{x}_i$, corresponds to an eigenvalue λ_i and it is observed that the product of a square matrix with its eigenvector results in a scalar multiple of the particular eigenvector. The normalized eigenvector U_i of $\vec{x}_i$ is given by,

$$U_i = \frac{1}{|\vec{x}|}\vec{x}.$$

Obviously,

$$U_i^T U_i = \frac{1}{|\vec{x}|}\vec{x}^T \frac{1}{|\vec{x}|}\vec{x} = \frac{1}{|\vec{x}|^2}\vec{x}^T\vec{x} = 1.$$

Let $\lambda_1 \geq \lambda_2 \geq \lambda_3, \ldots, \lambda_n$ denote the eigenvalues of the adjacency matrix $\mathcal{A}$ and $U_1, U_2, U_3, \ldots, U_n$ the corresponding normalized eigenvectors. Since graph G is connected, then matrix $\mathcal{A}$ is irreducible. According to the Perron-Frobenius theorem [145], an irreducible non-negative square matrix has a simple real positive largest eigenvalue λ_1 and the magnitude of any other eigenvalue does not exceed λ_1. Also, all the eigenvectors apart from the *principal* one (i.e., the one corresponding to λ_1), have at least one negative element. Similarly, all the normalized eigenvectors apart from the normalized principal eigenvector, have at least one negative element. Explicitly,

- $\lambda_1 > \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_n$;
- $\lambda_1 \geq |\lambda_2| \geq |\lambda_3| \geq ... \geq |\lambda_n|$;
- $\min((U_k)_i) < 0\,, k > 1\,, 1 \leq i \leq n$;
- $\min((U_1)_i) \geq 0\,, 1 \leq i \leq n$.

Only for the special case of the bipartite graph the magnitude of the minimum eigenvalue equals the largest one, i.e., $\lambda_1 = |\lambda_n|$. Thus, for non-bipartite graphs it follows that, $\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq ... \geq |\lambda_n|$.

For example, the eigenvalues of the adjacency matrix that corresponds to the graph of Figure B.1 are,
$(\lambda_1 =)2.4745, (\lambda_2 =)1.1143, (\lambda_3 =)0.5241, (\lambda_4 =)0, (\lambda_5 =) - 0.7615, (\lambda_6 =) - 1.3891$ and $(\lambda_7) = -1.9624$. Clearly,
$\lambda_1 > |\lambda_2| > |\lambda_3| > |\lambda_4| > |\lambda_5| > |\lambda_6| > 0$. The corresponding eigenvectors are,
$U_1^T = [0.4438\ 0.5022\ 0.2030\ 0.2879\ 0.1163\ 0.5960\ 0.2409]$,
$U_2^T = [-0.2130\ -0.3700\ -0.3321\ 0.6118\ 0.5490\ 0.1327\ 0.1191]$,

$U_3^T = [-0.0679\ 0.2696\ 0.5144\ 0.2206\ 0.4208\ -0.3052\ -0.5823]$,
$U_4^T = [-0.5774\ 0.0000\ 0.5774\ 0.0000\ -0.0000\ -0.0000\ 0.5774]$,
$U_5^T = [0.3037\ 0.0467\ -0.0613\ -0.5037\ 0.6615\ -0.2779\ 0.3650]$,
$U_6^T = [0.5486\ -0.6451\ 0.4644\ 0.1748\ -0.1258\ -0.1169\ 0.0842]$,
$U_7^T = [-0.1628\ -0.3461\ 0.1764\ -0.4582\ 0.2335\ 0.6657\ -0.3392]$.
It is observed that the elements of U_1^T are positive while, there is at least one negative element in any other eigenvector. For the case of the bipartite graph presented in Figure B.6 the eigenvalues of the corresponding adjacency matrix are, $2.5243, 2.0000, 1.0000, 0.7923, 0.0000, 0.0000, -0.7923, -1.0000, -2.0000$ and -2.5243. Obviously, the magnitude of the smallest eigenvalue equals the largest one, i.e., $2.5243 = |-2.5243|$.

Since the adjacency matrix is a real symmetric matrix, according to the spectral theorem [147], it follows that its eigenvectors are orthogonal. For example, by multiplying eigenvectors U_1 and U_2 of the previous example, it follows that

$$U_1^T U_2 = 0.4438 \cdot (-0.2130) + 0.5022 \cdot (-0.3700) + 0.2030 \cdot (-0.3321) + 0.2879 \cdot 0.6118 \\ +0.1163 \cdot 0.5490 + 0.5960 \cdot 0.1327 + 0.2409 \cdot 0.1191 = 0$$

The largest eigenvalue of the adjacency matrix plays a crucial role in many applications such as virus spreading (e.g., [132], [130], [129], [133]), etc. Thus, sharp bounds and exact or approximate expressions regarding λ_1 are desirable. A classical lower bound of λ_1 is obtained from Rayleigh's inequalities i.e.,

$$\lambda_1 = \sup_{\vec{x} \neq \vec{0}} \frac{\vec{x}^T \mathcal{A} \vec{x}}{\vec{x}^T \vec{x}},$$

where $\vec{x}$ is any vector and the maximum is attained *iff* $\vec{x}$ is the principal eigenvector of $\mathcal{A}$. Let Y be the all ones vector (i.e., $Y = [11 \ldots 1]$) then for $\vec{x} = Y$ it follows that, $\lambda_1 \geq \dfrac{Y^T \mathcal{A} Y}{Y^T Y}$ or,

$$\lambda_1 \geq \frac{2L}{n} = \bar{d}, \tag{B.4}$$

and the maximum is attained for the case of regular graphs since Y is an eigenvector of any regular graph corresponding to the eigenvalue $\lambda_1 = d_R$ [148] where, d_R is the regular graph's degree. For example, the adjacency's matrix largest eigenvalue of the graph presented in Fib. B.1 is $\lambda_1 = 2.4745$. The particular graph has seven vertices and 7 edges thus, $\bar{d} = \dfrac{7}{7} = 1$. Obviously, $\lambda_1 = 2.4745 > \bar{d} = 1$. For the case

of the 4−regular graph presented in Figure B.5 it is, $\lambda_1 = \bar{d} = 4$. An upper bound regarding λ_1 is given by,

$$\lambda_1 \leq \sqrt{\frac{2L(n-1)}{n}} = \sqrt{\bar{d}(n-1)}, \tag{B.5}$$

where the equality is attained for the case of the complete graph K_n. Any vertex of a complete graph of n vertices has exactly $n-1$ neighbors. Thus, from Equation (B.5) it follows that $\lambda_1 = n-1$. Clearly, a complete graph K_n is also a regular graph R_{n-1} of n vertices. Thus, from Equation (B.4) it follows that $\lambda_1 = n-1$.

A measure of significant importance in graph theory is the influence of each vertex within the graph [149], [150], [151] [152]. In particular, a vertex that is placed at a central point of a graph and has a large number of neighbors may be more important than a vertex that is placed at a less central point of the graph and is connected to a single neighbor. In order to cope with vertex influence, several centrality measures have been introduced in the literature. The vertex centrality plays a crucial role in a wide variety of scientific areas such as transportation networks (e.g., [153], [154]), communication networks (e.g., [155], [156], [157]), social networks (e.g., [158], [159], [160]), game theory (e.g., [161]) and biochemical networks (e.g., [162], [163]). Different centrality measures may be considered based on the needs of each application and on the underlying graph's category (directed or undirected). The simplest centrality measure is the *degree centrality* which assigns an importance score to each vertex based on the number of neighbors. For the case of directed graphs, degree centrality is distinguished in *in-degree* and *out-degree* centrality according to the number of incoming and outgoing links of the particular vertex respectively. The degree centrality can be obtained by the $1 \times n$ vector $\mathcal{D}$ given by Equation (B.1). Often instead of the degree centrality, the *normalized degree centrality* is used, which is given by the division of the degree centrality by the maximum possible number of neighbors (i.e., $n-1$). For example, the degree centrality and the normalized degree centrality of the vertices within the graph presented in Figure B.1 are depicted in Table B.1.

Table B.1: *The degree centrality and the normalized degree centrality of Figure B.1.*

Vertex	1	2	3	4	5	6	7
Degree centrality	2	3	1	2	1	4	1
Normalized degree centrality	0.3333	0.5	0.1667	0.3333	0.1667	0.6667	0.1667

For the case of the directed graph presented in Figure B.2, the in-degree and out-degree centralities along with the normalized ones are depicted in Table B.2.

Table B.2: *The in-degree and the out-degree centralities along with the normalized ones of Figure B.2.*

Vertex	1	2	3	4	5	6	7
In-degree centrality	1	2	1	1	2	2	0
Out-degree centrality	2	1	1	1	1	1	1
Normalized in-degree centrality	0.1667	0.3333	0.1667	0.1667	0.3333	0.3333	0
Normalized out-degree centrality	0.3333	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667

The degree centrality is the simplest measure of vertex connectivity and is useful for applications that depend on identifying vertices based on their degree. However, it may be inefficient for applications that are concerned with optimal facility placement into a network in order to minimize the communication or distance costs (e.g., [164], [165], [166]). For the particular applications, the *closeness centrality* measure may be considered as more suitable. Closeness centrality of a vertex within a connected graph is defined as the inverse of the sum of the shortest path distances from the particular vertex to all other vertices of the graph. Explicitly, closeness centrality of vertex u (i.e., $c_c(u)$) is given by,

$$c_c(u) = \frac{1}{\sum_{v \neq u} x(u,v)},$$

where $x(u,v)$ is the shortest path distance between vertices u and v. Thus, the higher the closeness centrality of a vertex, the shorter its distance from the other vertices within the graph. Often instead of the closeness centrality, the *normalized closeness*

centrality is used, which is given by the product of the closeness centrality with the maximum possible shortest distance within the graph (i.e., $n-1$). Explicitly,

$$\hat{c}_c(u) = \frac{n-1}{\sum_{v \neq u} x(u,v)}.$$

For example, the closeness centrality and the normalized closeness centrality of the vertices within the graph presented in Figure B.1 are depicted in Table B.3.

Table B.3: *The closeness centrality and the normalized closeness centrality of Figure B.1.*

Vertex	1	2	3	4	5	6	7
Closeness centrality	0.0909	0.1	0.0667	0.0909	0.0625	0.125	0.0769
Normalized closeness centrality	0.5455	0.6	0.4	0.5455	0.375	0.75	0.4615

Various applications focus on the number of shortest paths passing through certain vertices. For example, WSN recharging policies (e.g., [165], [167], [168]) focus on determining the network nodes that suffer from the energy hole problem and replenish their batteries. In particular, the nodes that relay a large number of messages in the network are expected to drain their batteries faster than the nodes relaying fewer messages. Such applications concern about the *betweenness centrality* of the vertices. The betweenness centrality of vertex u (i.e., $c_b(u)$) is defined as the fraction of total shortest paths passing through vertex u over the total shortest paths within the graph. Explicitly,

$$c_b(u) = \sum_{v \neq u \neq h} \frac{g_{vh}(u)}{g_{vh}},$$

where g_{vh} denotes the shortest path between vertices v and h and $g_{vh}(u)$ denotes the number of those paths that pass through vertex u. Clearly, betweenness centrality scales with the number of pairs of vertices. Therefore, instead of the betweenness centrality, the *normalized betweenness centrality* may be used, which is given by the division of the betweenness centrality by the number of pairs of vertices not including u. For the case of undirected graphs the number of pairs of vertices not including u equals $\frac{(n-1)(n-2)}{2}$ thus, the normalized betweenness centrality is given by,

$$\hat{c}_b(u) = \frac{2c_b(u)}{(n-1)(n-2)}.$$

For example, the betweenness centrality and the normalized betweenness centrality of the vertices within the graph presented in Figure B.1 are depicted in Table B.4.

Table B.4: *The betweenness centrality and the normalized betweenness centrality of Figure B.1.*

Vertex	1	2	3	4	5	6	7
Betweenness centrality	0	5	0	5	0	11	0
Normalized betweenness centrality	0	0.2381	0	0.2381	0	0.5238	0

The natural extension of the degree centrality is the *eigenvector centrality*. Although, vertices of high degree do not necessarily have a high eigenvector centrality. Moreover, a vertex with high eigenvector centrality does not necessarily have a large number of neighbors. In this centrality measure, a vertex poses high eigenvector centrality if it is connected to other vertices of high eigenvector centrality too. The particular centrality is obtained by the principal eigenvector of the adjacency matrix. Thus, the element i of the principal eigenvector corresponds to the eigenvector centrality of vertex i.

It is shown that the degree centrality is obtained by the product of the adjacency matrix with the all ones vector (i.e., Equation (B.1)). For the case of the (non-bipartite) undirected graph presented in Figure B.1, vector $\mathcal{D}$ equals

$$\mathcal{D} = [2\ 3\ 1\ 2\ 1\ 4\ 1].$$

The element i of the particular vector corresponds to the number of one-hop neighbors of vertex i. The product of the adjacency matrix by the vector $\mathcal{D}$ results in a new $1 \times n$ vector $\mathcal{D}_2$ given by,

$$\mathcal{D}_2 = \mathcal{A}\mathcal{D} = [7\ 7\ 3\ 5\ 2\ 8\ 4].$$

The element i of vector $\mathcal{D}_2$ corresponds to the number of the two-hop neighbors of vertex i. Similarly, the element i of vector $\mathcal{D}_3 = \mathcal{A}\mathcal{D}_2 = [15\ 18\ 7\ 10\ 5\ 23\ 8]$, corresponds to the number of the three-hop neighbors of vertex i. Consequently, vector $\mathcal{D}_i$ can be seen as "spread out" of the degree centrality. Moreover, for $i \geq 2$ vector $\mathcal{D}_i$ is given by the recurrence relation,

$$\mathcal{D}_i = \mathcal{A}\mathcal{D}_{i-1}$$

or,

$$\mathcal{D}_i = \mathcal{A}^{i-1}\mathcal{D}. \tag{B.6}$$

Since the eigenvectors of the adjacency matrix are orthogonal, vector $\mathcal{D}$ can be written as a linear combination of $U_1, U_2, \ldots, U_n$. Consequently,

$$\mathcal{D} = c_1U_1 + c_2U_2 + \cdots + c_nUn \tag{B.7}$$

Substituting Equation (B.7) to Equation (B.6) it follows that,

$$\mathcal{D}_i = \mathcal{A}^{i-1}(c_1U_1 + c_2U_2 + \cdots + c_nUn)$$

or,

$$\mathcal{D}_i = c_1\mathcal{A}^{i-1}U_1 + c_2\mathcal{A}^{i-1}U_2 + \cdots + c_n\mathcal{A}^{i-1}Un$$

or,

$$\mathcal{D}_i = c_1\lambda_1^{i-1}U_1 + c_2\lambda_2^{i-1}U_2 + \cdots + c_n\lambda_n^{i-1}U_n$$

or,

$$\mathcal{D}_i = \lambda_1^{i-1}\left[c_1U_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^{i-1}U_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^{i-1}U_n\right]$$

Consequently,

$$\lim_{i\to\infty}\mathcal{D}_i = \lambda_1^{i-1}\left[c_1U_1 + c_2\lim_{i\to\infty}\left(\frac{\lambda_2}{\lambda_1}\right)^{i-1}U_2 + \cdots + c_n\lim_{i\to\infty}\left(\frac{\lambda_n}{\lambda_1}\right)^{i-1}U_n\right].$$

According to the Perron-Frobenius theorem [145] $\lambda_1 \geq |\lambda_k|$, for $k \geq 2$ or, $\frac{|\lambda_k|}{\lambda_1} < 1$. Thus, $\lim_{i\to\infty}\left(\frac{\lambda_k}{\lambda_1}\right)^{i-1} \to 0$. Consequently, $\lim_{i\to\infty}\mathcal{D}_i = c_1\lambda_1^{i-1}U_1$. It is observed that for large values of i, vector D_i is a scalar multiple of the principal eigenvector. Thus, according to eigenvector centrality, the importance of a vertex depends on the number of ∞-hop neighbors. For example the eigenvector centrality of the vertices within the graph presented in Figure B.1 are depicted in Table B.5.

Table B.5: *The eigenvector centrality of Figure B.1*

Vertex	1	2	3	4	5	6	7
Eigenvector centrality	0.4438	0.5022	0.203	0.2879	0.1163	0.596	0.2409

Chapter C

Probabilistic Flooding Analytical Model

PROBABILISTIC flooding is analyzed in this chapter employing algebraic graph theory elements. An $n \times 1$ vector is introduced such that its i-th element corresponds to the probability that node i has received the information message. Subsequently, considering the binomial approximation, this particular vector is further studied and an analytical expression is derived regarding coverage in the form of a polynomial.

C.1 Network and System Definitions

Let the network topology $G(V, E)$ be represented by the set of nodes V and the set of links E among them. Let n be the number of nodes in the network, or $n = |V|$. For simplicity and without loss of generality let nodes be assigned unique integer numbers i, where $i = 1 \dots n$.

If a link (i, j) exists in the network among nodes i and j, then link (j, i) also exists (i.e., symmetric links). Let d_i denote the degree of node $i \in V$, or equivalently, the number of neighbor nodes of node i. The average degree of the network (i.e., $\bar{d}$) corresponds to the average of d_i's, thus, $\bar{d} = \sum_{\forall i \in V} d_i/n = \sum_{i=1}^{n} d_i/n$, or $\bar{d} = 2|E|/n$.

C.2 Elements of Probabilistic Flooding

Let s denote the *initiator* node of a probabilistic flooding activity in the network planning to disseminate a particular information message initially available at the initiator node. Let P_t be an $n \times 1$ vector at timestep $t \geq 0$, its i-th element be denoted as $(P_t)_i$. If node i has received a message under probabilistic flooding, i.e., node i is *covered* by probabilistic flooding, at any time step τ, $\tau \leq t$, then $(P_t)_i = 1$, otherwise $(P_t)_i = 0$. Obviously, the summation of all elements of P_t, i.e., $\sum_i^n (P_t)_i$, corresponds to the number of covered nodes at any time step t. Given that at time $t = 0$ only the initiator node s keeps the under dissemination information message, then $(P_0)_s = 1$ and $(P_0)_{\substack{i \\ i \neq s}} = 0$.

Probabilistic flooding is a simple information dissemination policy that can be easily implemented. All nodes forward the information message they receive to their neighbor nodes (excluding the node that the message was received from), according to a *forwarding probability* q that: (i) is fixed; (ii) can be set by the initiator node s; and (iii) can be part of the information message initially sent by node s and thus forwarded in the network. In case the particular message is received a second time by any node, it is ignored (not forwarded). It is assumed that message transmission, reception, and processing takes place within a single timestep. At timestep 0, only node s has the particular information message. Let network *coverage* at any time $t \geq 0$, denoted by C_t, be the fraction of covered network nodes at time t. Obviously, $C_t = 1/n \sum_{i=1}^{n} (P_t)_i$. An example of probabilistic flooding is presented in Figure C.1.

C.2.1 Termination Time

Reaching all network nodes or equivalently, $C_t = 1$, for some time t, can be trivial by setting $q = 1$; then probabilistic flooding reduces to blind flooding. However, the number of messages sent in the network can be high since it is upper bounded by $\sum_{i=1}^{n}(d_i - 1) + 1 = 2|E| - n + 1$ (i.e., twice the number of network links minus one for each node since it is not forwarded backward except for the initiator node). Let $\mathcal{T}(q, s)$ denote *termination time* for the particular initiator node s and forwarding probability q. Termination time $\mathcal{T}(q, s)$ is defined as the particular time instance that there was no probabilistic flooding activity for the first time in the network, or $P_{\mathcal{T}(q,s)} = P_{\mathcal{T}(q,s)-1}$ (all elements equal) and $P_t \neq P_{t-1}$, for any $0 < t < \mathcal{T}(q, s)$.

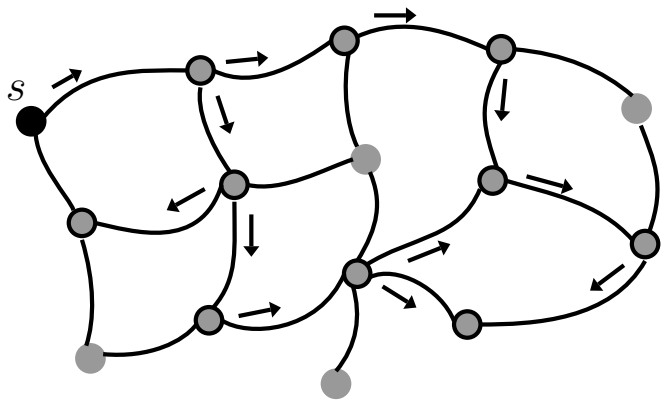


Figure C.1: *A probabilistic flooding example. The initiator node s (solid black) sends the first information message to a fraction of its neighbor nodes according to the forwarding probability as shown by the corresponding arrows. When these nodes are covered (black perimeter), they forward the information message, accordingly. Some nodes may eventually not be covered (no perimeter line).*

Clearly, the process is terminated when either all network nodes are covered or when the forwarded messages fail to advance further due to the probabilistic nature of this process. An easily derived upper bound of termination time, when $q = 1$, corresponds to the network's *diameter* D, thus $\mathcal{T}(q, s) \leq D$, always assuming a connected network.

It is interesting to investigate termination time when $q < 1$ always assuming that $C_{\mathcal{T}(q,s)} = 1$, even though due to the probabilistic nature of the process, $C_{\mathcal{T}(q,s)}$ may be less than but close to unity for large values of the forwarding probability q (i.e., $C_{\mathcal{T}(q,s)} \to 1$). Let $\check{q}(s)$ denote a *threshold* regarding the forwarding probability q for the initiator node s, such that $C_{\mathcal{T}(q,s)} \to 1$ for any $q \geq \check{q}(s)$, and $C_{\mathcal{T}(q,s)} < 1$ for any $q < \check{q}(s)$ ($\check{q}(s)$ will be referred to hereafter as the threshold probability). Therefore, for any $\check{q}(s) \leq q < 1$, $C_{\mathcal{T}(q,s)} \to 1$ is assured and termination time is expected to increase since the information message will not be forwarded (with probability $(1 - q)$) through certain links that may belong to shortest paths towards the rest of the nodes.

C.2.2 Number of messages

As the information message is forwarded, the number of transmissions increases with a significant effect on the network (e.g., energy consumption for the case of wireless networks). Let $\mathcal{Z}_t$ denote the number of messages sent until time t under probabilistic flooding. For $q = 1$, it is evident that $|E| \leq \mathcal{Z}_{\mathcal{T}(1,s)} \leq 2|E|$. For $\check{q}(s) \leq q < 1$, the total number of messages (i.e., the number of messages at termination time) is expected to decrease, thus, $\mathcal{Z}_{\mathcal{T}(q,s)} \leq \mathcal{Z}_{\mathcal{T}(1,s)} \leq 2|E|$. Eventually, the main goal of the analysis is to analytically investigate this threshold probability, considering the effects on network coverage, termination time and the number of sent messages.

C.3 Problem Formulation and Analysis

Let $\mathcal{A}$ denote the $n \times n$ adjacency matrix of $G(V, E)$. Assuming timestep $t \geq 2$ and a node j that has received the information message for the first time at the previous time step $t-1$, then $(P_{t-1})_j = 1$ and $(P_{t-2})_j = 0$ (node j was covered at time $t-1$ and not before). Consequently, for this particular node j, $(P_{t-1} - P_{t-2})_j = 1$. Eventually, $P_{t-1} - P_{t-2}$ is also an $n \times 1$ vector whose elements are either 1 or 0 indicating those nodes that have received the information message for the first time at time $t-1$.

According to probabilistic flooding, all nodes that receive the information message for the first time forward it to their neighbor nodes with probability q at the following timestep. Assume that node i has not received the information message yet. Its neighbor nodes correspond to those elements of the i-th row of $\mathcal{A}$ (or column, since $\mathcal{A}$ is symmetric) that are equal to 1. Given that $\mathcal{A}\,(P_{t-1} - P_{t-2})$ is an $n \times 1$ vector, its i-th element corresponds to the number of neighbor nodes of node i that have received the information message at timestep $t-1$ for the first time. These nodes are expected to forward the information message to node i, each with probability q. Consequently, *none* of them will forward it, with probability $(1-q)^{(\mathcal{A}(P_{t-1}-P_{t-2}))_i}$, thus node i will receive it with probability $1-(1-q)^{(\mathcal{A}(P_{t-1}-P_{t-2}))_i}$, at timestep t.

In order to derive a more tractable form, the binomial approximation is employed, $(1-q)^{(\mathcal{A}(P_{t-1}-P_{t-2}))_i} \approx 1 - q\,(\mathcal{A}\,(P_{t-1} - P_{t-2}))_i$, assuming small values of the forwarding probability q and of $|q\,(\mathcal{A}\,(P_{t-1} - P_{t-2}))_i\,|$. Eventually, let $\check{P}_t$ be an $n \times 1$

vector such that

$$\check{P}_t = q\mathcal{A}\left(P_{t-1} - P_{t-2}\right). \tag{C.1}$$

There is a close connection between $\check{P}_t$ and P_t, despite the fact that $(\check{P}_t)_i$ corresponds to the probability that node i receives the information message at timestep t, while $(P_t)_i$ corresponds to the fact that node i is covered (i.e, $(P_t)_i = 1$), or not (i.e., $(P_t)_i = 0$) at timestep t. Further investigation of $\check{P}_t$ may not be possible since it relies on $(P_{t-1} - P_{t-2})$, which is generally not available for each timestep. Therefore, in order to proceed with the analysis and in analogy to Equation (C.1), let $\mathcal{P}_t$ be an $n \times 1$ vector such that,

$$\mathcal{P}_t = q\mathcal{A}\left(\mathcal{P}_{t-1} - \mathcal{P}_{t-2}\right), \tag{C.2}$$

for any $t \geq 1$, where $\mathcal{P}_0 = P_0$ and $\mathcal{P}_{-1}$ is an $n \times 1$ zero vector. In the sequel, $(\mathcal{P}_t)_i$ is seen as the probability that node i has received the information message until time t. Therefore, instead of analyzing $\check{P}_t$, $\mathcal{P}_t$ is studied and the obtained results are later confirmed by the simulations.

The following theorem, reveals the relation among the vector $\mathcal{P}_t$, the initial conditions (i.e., P_0), the network topology (i.e., $\mathcal{A}$) and the forwarding probability (i.e., q).

Theorem 1 *Vector $\mathcal{P}_t$ can be written as,*

$$\mathcal{P}_t = \mathcal{B}_t P_0, \tag{C.3}$$

where $\mathcal{B}_t$ is a symmetric $n \times n$ matrix, given by

$$\mathcal{B}_t = \sum_{k=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{k+1} \binom{t-k+1}{t-2k+2} q^{t-k+1} \mathcal{A}^{t-k+1}. \tag{C.4}$$

Proof 1 *It is known that $\mathcal{P}_0 = P_0$. According to Equation* (C.2)*, $\mathcal{P}_1 = q\mathcal{A}\left(\mathcal{P}_0 - \mathcal{P}_{-1}\right)$. Given that*

$$\mathcal{P}_0 - \mathcal{P}_{-1} = \mathcal{P}_0 = P_0,$$

it follows that

$$\mathcal{P}_1 = q\mathcal{A}P_0.$$

For $t = 2$,

$$\mathcal{P}_1 - \mathcal{P}_0 = q\mathcal{A}P_0 - P_0.$$

Based on Equation (C.2),

$$\mathcal{P}_2 = q\mathcal{A}\left(q\mathcal{A}P_0 - P_0\right) = q^2\mathcal{A}^2P_0 - q\mathcal{A}P_0.$$

Following the same steps,

$$\mathcal{P}_3 = q^3\mathcal{A}^3P_0 - 2q^2\mathcal{A}^2P_0,$$

$$\mathcal{P}_4 = q^4\mathcal{A}^4P_0 - 3q^3\mathcal{A}^3P_0 + q^2\mathcal{A}^2P_0,$$

$$\mathcal{P}_5 = q^5\mathcal{A}^5P_0 - 4q^4\mathcal{A}^4P_0 + 3q^3\mathcal{A}^3P_0,$$

$$\mathcal{P}_6 = q^6\mathcal{A}^6P_0 - 5q^5\mathcal{A}^5P_0 + 6q^4\mathcal{A}^4P_0 - q^3\mathcal{A}^3P_0,$$

$$\mathcal{P}_7 = q^7\mathcal{A}^7P_0 - 6q^6\mathcal{A}^6P_0 + 10q^5\mathcal{A}^5P_0 - 4q^4\mathcal{A}^4P_0,$$

etc...

As it evolves, it becomes obvious that $\mathcal{P}_t$ *is actually a polynomial of matrix* $q\mathcal{A}$ *of order* t*. A closer look reveals that each such polynomial consists of* $\left\lfloor \frac{t}{2} \right\rfloor + 1$ *monomials of order* $t, t-1, ..., t-\left\lfloor \frac{t}{2} \right\rfloor - 1$*. The sign of each monomial's coefficient changes regularly for '+' to '−' and vice versa, starting from '+'. The absolute value of each coefficient is derived by studying matrix* $\mathcal{C}$ *(zeros are omitted for simplicity). Each element* $(\mathcal{C})_{r,k}$ *corresponds to the absolute value of the coefficient of the* k*-th monomial at time step* $t = r - 1$*.*

$$\mathcal{C} = \begin{pmatrix} 1 & & & & \\ 1 & & & & \\ 1 & 1 & & & \\ 1 & 2 & & & \\ 1 & 3 & 1 & & \\ 1 & 4 & 3 & & \\ 1 & 5 & 6 & 1 & \\ 1 & 6 & 10 & 4 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \; \mathcal{S} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 6 & 10 & 15 \\ 1 & 4 & 10 & 20 & 35 \\ 1 & 5 & 15 & 35 & 70 \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

After a closer look, $\mathcal{C}$ resembles the Pascal triangle and particularly the Pascal matrix $\mathcal{S}$, where each element $(\mathcal{S})_{r',k'} = \binom{r'+k'-2}{r'-1}$. It is easy to observe that each value $(\mathcal{C})_{r,k}$ of matrix $\mathcal{C}$ appears at the same column in matrix $\mathcal{S}$ ($k = k'$). The only difference is with respect to the indexing of lines (i.e., $r \neq r'$ apart from $r = 1$). However, a shift of two lines can be observed with respect to the column index when matrices $\mathcal{C}$ and $\mathcal{S}$ are compared. Therefore, the value of each element $(\mathcal{C})_{r,k} = \binom{r-k}{r-2k+1}$. Given that the r-th row of matrix $\mathcal{C}$ corresponds to time $t-1$, then for time step t, the k-th monomial of polynomial $\mathcal{P}_t$ is of order $t-k+1$, provided that $k \leq \lfloor \frac{t}{2} \rfloor + 1$. Similarly, the absolute value of the corresponding coefficient is given by $\binom{t-k+1}{t-2k+2}$ and the sign is given by $(-1)^{k+1}$. The summation of all these elements amounts to $\mathcal{B}_t$, which is trivial to show that is a symmetric $n \times n$ matrix and this completes the proof.

Corollary 1 *Matrix $\mathcal{B}_t$ can be written as $\mathcal{F}_t q^{\lfloor \frac{t+1}{2} \rfloor} \mathcal{A}^{\lfloor \frac{t+1}{2} \rfloor}$, where $\mathcal{F}_t$ is an $n \times n$ matrix.*

Proof 2 *According to Equation* (C.4)*, the monomial of $\mathcal{B}_t$ of the minimum order corresponds to $k = \lfloor \frac{t}{2} \rfloor + 1$ and it is given by $t - \lfloor \frac{t}{2} \rfloor$. If t is an even number, then $t - \lfloor \frac{t}{2} \rfloor = \frac{t}{2} = \lfloor \frac{t+1}{2} \rfloor$, else, $t - \lfloor \frac{t}{2} \rfloor = \frac{t+1}{2} = \lfloor \frac{t+1}{2} \rfloor$. Since $\mathcal{B}_t$ and $\mathcal{A}$ are $n \times n$ matrices, then $\mathcal{F}_t$ is also an $n \times n$ matrix.*

Given that P_0 is an $n \times 1$ vector such that $(P_0)_s = 1$ and $(P_0)_{\substack{i \\ i \neq s}} = 0$, then $\mathcal{B}_t P_0$ is an $n \times 1$ vector corresponding to the s-th column of $\mathcal{B}_t$ (or row since $\mathcal{B}_t$ is symmetric). Based on Equation (C.3), $\mathcal{P}_t$ is equal to the s-th column (or row) of $\mathcal{B}_t$.

Let $w_{\ell,i}$ denote the summation of the number of walks of size ℓ starting from node i.

Lemma 1 *The number of walks of size ℓ starting from node i is given by,*

$$w_{\ell,i} = \sum_{r_1=1}^{n} \sum_{r_2=1}^{n} \cdots \sum_{r_{\ell-1}=1}^{n} (\mathcal{A})_{i,r_1} (\mathcal{A})_{i,r_2} \ldots (\mathcal{A})_{r_{\ell-2},r_{\ell-1}} d_{r_{\ell-1}}. \tag{C.5}$$

Proof 3 *The number of walks [145] of size ℓ (hops) between node i and node j and is given by*

$$(\mathcal{A}^{\ell})_{i,j} = \sum_{r_1=1}^{n} \cdots \sum_{r_{\ell-1}=1}^{n} (\mathcal{A})_{i,r_1} \ldots (\mathcal{A})_{r_{\ell-2},r_{\ell-1}} (\mathcal{A})_{r_{\ell-1},j}.$$

Obviously,

$$w_{\ell,i} = \sum_{j=1}^{n} (\mathcal{A}^{\ell})_{i,j}$$

or,

$$w_{\ell,i} = \sum_{j=1}^{n} \sum_{r_1=1}^{n} \cdots \sum_{r_{\ell-1}=1}^{n} (\mathcal{A})_{i,r_1} \ldots (\mathcal{A})_{r_{\ell-2},r_{\ell-1}} (\mathcal{A})_{r_{\ell-1},j}$$

or equivalently,

$$\sum_{r_1=1}^{n} \cdots \sum_{r_{\ell-1}=1}^{n} (\mathcal{A})_{i,r_1} \ldots (\mathcal{A})_{r_{\ell-2},r_{\ell-1}} \sum_{j=1}^{n} (\mathcal{A})_{r_{\ell-1},j}.$$

Note that the number of neighbor nodes d_i corresponds to the summation of all elements of the i-th column (or row) of the adjacency matrix $\mathcal{A}$. Therefore,

$$\sum_{j=1}^{n} (\mathcal{A})_{r_{\ell-1},j} = d_{r_{\ell-1}},$$

and the lemma is proved.

Given that $(\mathcal{P}_t)_i$ corresponds to the probability that node i has received the information message until time t, then $\sum_{i=1}^{n} (\mathcal{P}_t)_i$ corresponds to the average number of nodes that have received the information message. Let $\mathtt{P}_t$ be the average summation of all elements of $\mathcal{P}_t$, or $\mathtt{P}_t = 1/n \sum_{i=1}^{n} (\mathcal{P}_t)_i$. As shown next, $\mathtt{P}_t$ is a polynomial of q and will be denoted as $\mathtt{P}_t(q)$.

Lemma 2 *$\mathtt{P}_t(q)$ is a polynomial of q and degree t,*

$$\mathtt{P}_t(q) = \frac{1}{n} \sum_{i=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{i+1} \binom{t-i+1}{t-2i+2} w_{t-i+1,s} q^{t-i+1}. \tag{C.6}$$

$\mathtt{P}_t(q)$ can also be written as a product of polynomials, one of which is $q^{\lfloor \frac{t+1}{2} \rfloor}$.

Proof 4 *Based on Equation* (C.3) *and Equation* (C.4)*, it is trivial to write $\mathtt{P}_t(q)$ as in Equation* (C.6)*. Based on Corollary 1, it is obvious that there exists a common factor $q^{\lfloor \frac{t+1}{2} \rfloor}$. The maximum degree of q is assumed when i takes the minimum value or $i = 1$. Since the power of each monomial is given by $t - i + 1$, then t corresponds to the maximum degree.*

Polynomial $\mathtt{P}_t(q)$ is actually an analytical expression that aims to capture the behavior of coverage $C(t)$ under probabilistic flooding for the particular value of the forwarding probability q. The aim in the sequel is to analyze further the analytical model presented in this section and particularly polynomial $\mathtt{P}_t(q)$ in order to derive an analytical expression regarding threshold probability $\breve{q}(s)$.

C.4 Simulation Results and Evaluation

This section presents the simulation results for the evaluation of the analytical model and the effectiveness of Theorem 1. The binomial approximation limitations are also highlighted and it is shown that the proposed analytical model captures the probabilistic flooding behavior satisfactorily for small values of the forwarding probability and of $|q\left(\mathcal{A}\left(P_{t-1}-P_{t-2}\right)\right)_i|$.

C.4.1 Model Evaluation

It is interesting to see how the number of sent information messages is reduced under probabilistic flooding when compared to the case that blind flooding is employed. Given that the number of sent information messages at termination time is denoted as $\tilde{\tilde{\mathcal{Z}}}_{\tilde{\mathcal{T}}(\tilde{q}(s),s)}$, the depicted in Figure C.2 message *ratio* is given by fraction $\frac{\tilde{\tilde{\mathcal{Z}}}_{\tilde{\mathcal{T}}(\tilde{q}(s),s)}}{\tilde{\tilde{\mathcal{Z}}}_{\tilde{\mathcal{T}}(1,s)}}$ ($\tilde{\tilde{\mathcal{Z}}}_{\tilde{\mathcal{T}}(1,s)}$ corresponds to the number of messages under blind flooding). As it can be observed from Figure C.2, as the connectivity radius increases (more links in the networks), the corresponding threshold probability $\tilde{q}(s)$ decreases and at the same time, the message ratio decreases and starting from around 0.6 it becomes close to zero for values of $r_c \geq 0.15$. Termination time ratio (i.e. $\frac{\tilde{\mathcal{T}}(\tilde{q}(s),s)}{\tilde{\mathcal{T}}(1,s)}$), on the other hand, increases as expected, since the information message is now forwarded through longer walks. Clearly, the experienced termination time ratio increment can be as large as 2.6 for dense topologies (e.g., $r_c \geq 0.5$) but this is compensated since the number of sent information messages is drastically reduced.

Figures C.3-{a,b,c} depict simulation and analytical results (i.e., Equation (C.4)) regarding coverage as a function of time for GRG topologies with a connectivity radius of $r_c = 0.3$, $r_c = 0.4$ and $r_c = 0.5$ respectively. It is observed that the analytical model is close to the simulation results for the initial and the final phase (i.e. coverage close to 1) of probabilistic flooding, whereas for the transition phase

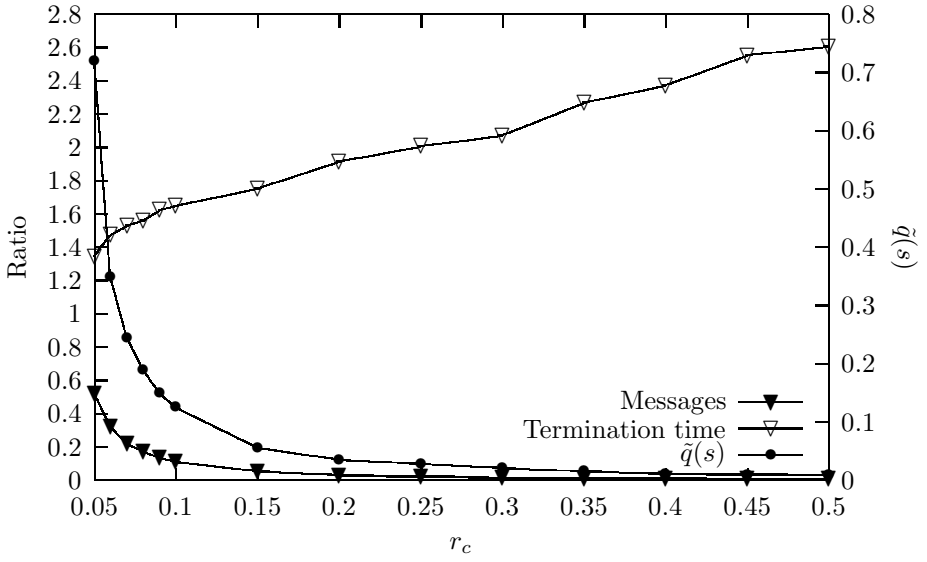


Figure C.2: *Ratio of messages and termination time of probabilistic flooding over blind flooding. The threshold probability $\tilde{q}(s)$ values are also depicted.*

(i.e. significant coverage increment) the difference may be significant. This is a consequence of the binomial approximation since, for the rapid increment of coverage it is satisfied that $|q\left(\mathcal{A}\left(\mathcal{P}_{t-1}-\mathcal{P}_{t-2}\right)\right)_i| \ll 1$.

For the case of GRG with connectivity radius $r_c = 0.3$ (i.e., Figure C.3-a) the most rapid increment of coverage is observed at timestep $t = 3$, i.e., $\mathcal{C}_t$ increases faster than C_t. This behavior is also observed in Figure C.3-b and Figure C.3-c at $t = 2$. The same behavior is observed for the case of ER topologies (i.e., Figures C.4-{a,b,c}) at $t = 3$.

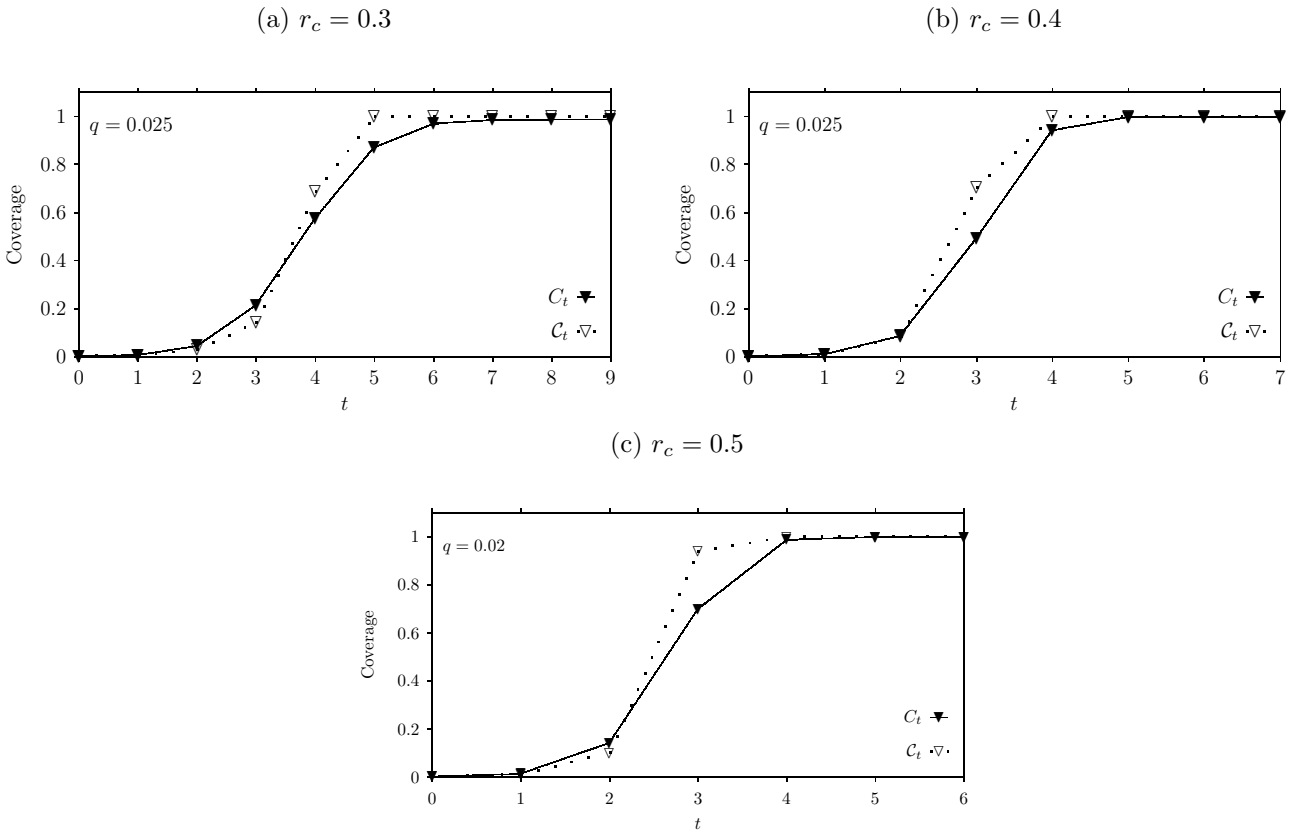


Figure C.3: *Simulation (C_t) and analytical ($\mathcal{C}_t$) results regarding coverage as a function of time for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5.

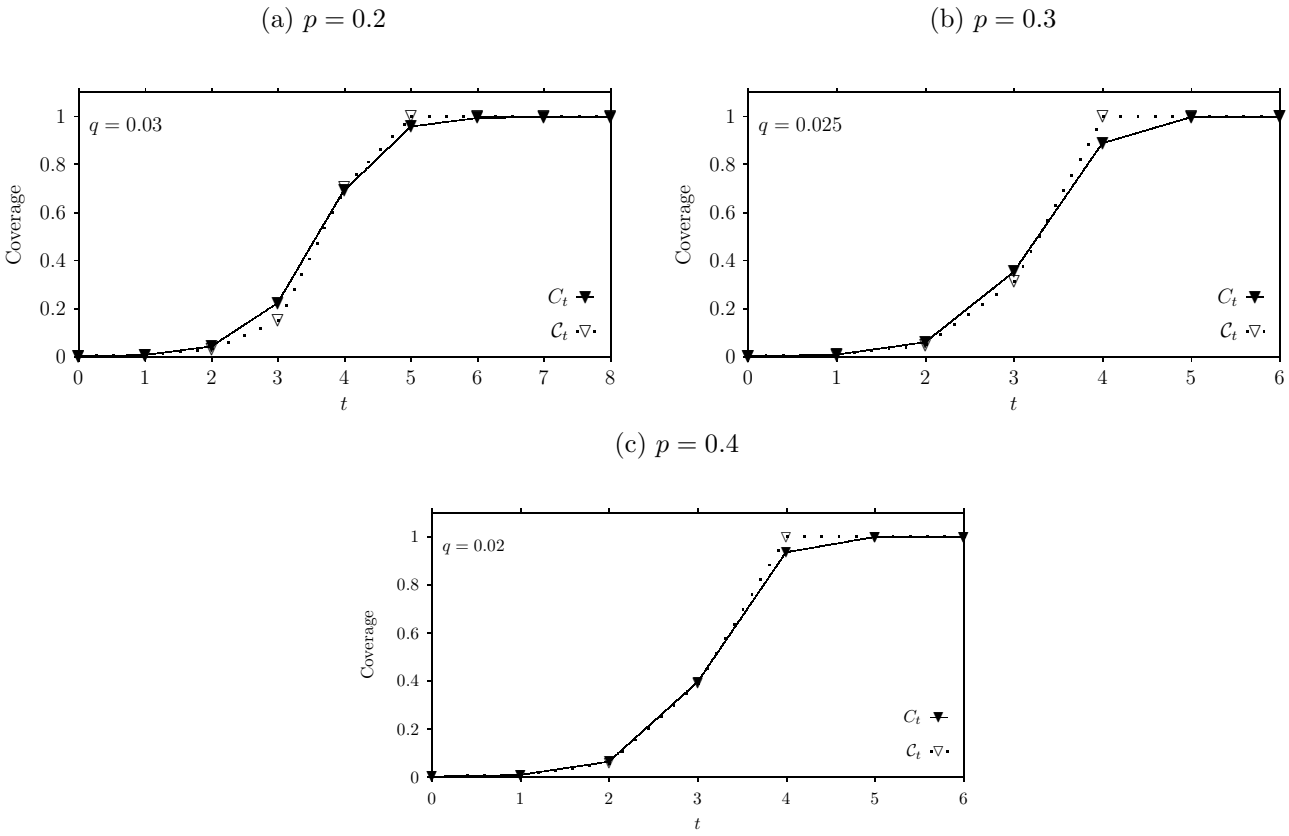


Figure C.4: *Simulation (C_t) and analytical ($\mathcal{C}_t$) results regarding coverage as a function of time for three ER topologies with connection probability* $p = 0.2, 0.3$ *and* 0.4.

C.5 Discussion

Probabilistic flooding is shown to be a suitable alternative to blind flooding, capable of pruning redundant transmissions by employing some fixed probability for message forwarding among neighbor nodes. The particular mechanism is analyzed using algebraic graph theory elements. An $n \times 1$ vector is introduced such that the i-th element corresponds to the probability that node i has received the information message. Subsequently, considering the binomial approximation, the particular vector is further analyzed and an analytical expression regarding coverage is derived in the form of a polynomial.

Termination time of probabilistic flooding (i.e., the particular time instance that there is no probabilistic flooding activity for the first time in the network) is also studied. It is shown that for any value of the forwarding probability larger than the threshold probability and less than one (i.e., $\check{q}(s) \leq q < 1$), termination time is expected to increase since the information message will not be forwarded (with probability $(1-q)$) through certain links that may belong to shortest paths towards the rest of the nodes.

Simulation results presented in Section C.4 confirm the analytical findings and it is shown that the analytical expression regarding coverage, captures the probabilistic flooding behavior for small values of the forwarding probability. However, for large values of the forwarding probability, the analytical results may deviate from and the simulation ones due to the use of the binomial approximation. It is also demonstrated that the number of messages under probabilistic flooding may be significantly smaller than under blind flooding, on the expense of increased termination time.

Chapter D

Threshold Probability Estimation

IN this chapter, the previously presented probabilistic flooding analytical model is further analyzed in order to obtain an approximation of the threshold probability. It is shown that the threshold probability can be satisfactorily approximated by the higher roots of Equation (C.6). The particular roots are also used to confirm the existing results in the open literature. Subsequently, based on certain observations a novel algorithm is introduced for the estimation of the threshold probability. The simulation results confirm the analytical findings and in addition, they demonstrate that the estimated value derived here regarding threshold probability, is suitable for achieving high coverage and reducing the number of transmitted information messages.

D.1 Threshold Probability Analysis

A first bound regarding the threshold probability $q(s)$ can be derived straightforwardly from Corollary 1. Based on Equation (C.4), $\mathcal{P}_t$ can be written as $\mathcal{P}_t = \mathcal{F}_t q^{\lfloor \frac{t+1}{2} \rfloor} \mathcal{A}^{\lfloor \frac{t+1}{2} \rfloor} P_0$. Analyzing this case following the steps presented in [129], it is concluded that for $q < \frac{1}{\lambda_1}$, probabilistic flooding fails to cover the entire network (λ_1 is the maximum eigenvalue of the adjacency matrix $\mathcal{A}$). Obviously, $q(s) > \frac{1}{\lambda_1}$, is a necessary (although not sufficient) condition for global outreach. This result is in accordance with the epidemic threshold given in [129].

In particular, for $t = 1$, polynomial $\mathtt{P}_t(q)$, as given by Equation (C.6), $\mathtt{P}_1(q) = \frac{1}{n} w_{1,s} q$. Asking for more than one nodes to be covered, or $\mathtt{P}_1(q) > \frac{1}{n}$, then $q > \frac{1}{w_{1,s}}$.

According to Equation (C.5), $w_{1,s} = d_s$. Therefore, $q > \frac{1}{d_s}$. Assuming a regular graph (the same degree $\bar{d}$ for all nodes), then $q > \frac{1}{d}$. The latter result corresponds to the epidemic threshold as derived in [130]. Clearly, $q(s) > \frac{1}{d}$.

Both bounds regarding threshold probability are already presented in the literature (i.e., [129] and [130]). However, the fact they are also derived here indicates that the analytical model introduced in the previous section is able to capture the behavior of coverage under probabilistic flooding. This motivates the following analysis of the roots of polynomial $\mathtt{P}_t(q)$ in order to shed more light on the boundaries of threshold probability. More specifically, for any time instance t, the focus is the particular root that is the largest and smaller than one denoted by $\rho_{s,t}$ for initiator node s and time $t > 2$ will be referred to hereafter as the *forwarding root*. Examining the roots of $\mathtt{P}_t(q)$ allows the derivation of the range of values of $q(s)$ for which $\mathtt{P}_t(q) > 0$ (i.e., a positive number of covered nodes at time t which is expected to be satisfied).

The following theorem shows that root $\rho_{s,t}$ does exist at time $t > 1$ and $\mathtt{P}_t(q) > 0$, for any $\rho_{s,t} < q \leq 1$.

Theorem 2 *There exists a range of values* $(\rho_{s,t}, 1]$ *such that if* $q \in (\rho_{s,t}, 1]$, *then* $\mathtt{P}_t(q) > 0$, *where* $\rho_{s,t}$ *is the maximum root of* $\mathtt{P}_t(q)$ *that is smaller than* 1 *for* $t > 1$.

Proof 5 *If* $q = 1$, *then at any time* t, $\mathtt{P}_t(1) > 0$. *Given that* $\mathtt{P}_t(q)$ *is a polynomial, it is a continuous function of* q. *Therefore, for* $q = 1 - \epsilon$, *where* $\epsilon \to 0^+$, $\mathtt{P}(q_s - \epsilon)$ *remains greater than* 0. *As* ϵ *increases, there will be a certain value of* ϵ *such that* $\mathtt{P}_t(1 - \epsilon) = 0$. *Apparently, this takes place at* $q = \rho_{s,t}$, *where* $\rho_{s,t}$ *is the largest root of* $\mathtt{P}_t(q)$ *provided that it is smaller than* 1.

The following corollary establishes the fact that the threshold probability cannot be smaller or equal to $\rho_{s,t}$.

Corollary 2 *If* $\rho_{s,t}$ *is a root of* $\mathtt{P}_t(q)$ *of odd multiplicity, then* $q_s > \rho_{s,t}$.

Proof 6 *If* $\rho_{s,t}$ *is a root of odd multiplicity, then* $\mathtt{P}_t(q_s) < 0$ *for* $q_s = \rho_{s,t} - \epsilon$, *where* $\epsilon \to 0^+$. *Consequently,* $q(0)$ *cannot be smaller than* $\rho_{s,t}$. *At the same time,* $q(0)$ *cannot be equal to* $\rho_{s,t}$ *since* $\mathtt{P}_t(q) > 0$ *should be satisfied for any* $q \geq q(s)$.

Theorem 2 and Corollary 2 draw the way for deriving further boundaries for $q(s)$. Since it is not possible for the general case to get an analytical expression

with respect to $\rho_{s,t}$, it is shown later using simulations that $\rho_{s,t}$ is a simple root (odd multiplicity) of polynomial $\mathtt{P}_t(q)$. Efficient numerical methods can be employed for these roots, as it is the case in the simulations section. On the other hand, analytical expressions regarding $\rho_{s,t}$ for $t = 2 \ldots 5$ can be derived. The analytical expressions are shown in Table D.1.

Table D.1: *Roots of* $\mathtt{P}_t(q_s)$ *for* $t = 2, 3, 4, 5$.

Root	Analytical Expression	Root	Analytical Expression
$\rho_{s,2}$	$\frac{w_{1,s}}{w_{2,s}}$	$\rho_{s,4}$	$\frac{3w_{3,s}+\sqrt{9w_{3,s}^2-4w_{4,s}w_{2,s}}}{2w_{4,s}}$
$\rho_{s,3}$	$\frac{2w_{2,s}}{w_{3,s}}$	$\rho_{s,5}$	$\frac{4w_{4,s}+\sqrt{16w_{4,s}^2-12w_{5,s}w_{3,s}}}{2w_{5,s}}$

It is interesting to study root $\rho_{s,2}$ and particularly, the fraction $\frac{w_{1,s}}{w_{2,s}}$, for the case of regular topologies as before. $\frac{w_{1,s}}{w_{2,s}}$ corresponds to the ratio of the number nodes one hop away from node s over the number of nodes two hops away from the initiator node s. Given the homogeneous scenario, then $\rho_{s,2} = \frac{\bar{d}}{\bar{d}^2} = \frac{1}{\bar{d}}$. Given that $q > \frac{1}{\bar{d}}$ should be satisfied, this is actually the same result as before for $t = 1$ and confirms similar results in the literature, e.g., [130], as already mentioned. Note that $\rho_{s,2}$ provides for $q > \frac{2}{\bar{d}}$. Obviously, $\bar{d} > 2$ is required in order for $q < 1$, which is an interesting result as follows. In particular, if $\bar{d} = 2$, this regular topology corresponds to the special case of a ring topology (blind flooding can be employed since there would be no information message savings under probabilistic flooding).

As t increases, it is obvious that root $\rho_{s,t}$ decreases. Let $\Delta(s,t) = \rho_{s,t} - \rho_{s,t-1}$, for $t > 2$, correspond to the *root difference* at time t and the particular initiator node. It is interesting to observe the behavior of $\Delta(s,t)$ as time increases. The analytical expressions shown in Table D.1 are used considering a regular topology in the following lemma to show a root difference decrement. The particular roots for $\rho_{s,6}$ and $\rho_{s,7}$ for the regular topologies are also derived as can be seen from the

proof of Lemma 3.

Lemma 3 *Assuming a regular topology such that* $\frac{w_{t-1,s}}{w_{t,s}} = \frac{1}{\bar{d}}$*, the corresponding root difference* $\Delta(s,t)$ *decreases for* $t = 3,4,5,6,7$*, for initiator node* s*.*

Proof 7 *The first step is to calculate* $\rho_{s,2}$*,* $\rho_{s,3}$*,* $\rho_{s,4}$ *and* $\rho_{s,5}$*. Given that* $\frac{w_{t-1,s}}{w_{t,s}} = \frac{1}{\bar{d}}$*, and* $w_{1,s} = \bar{d}$*, it is easy to derive that* $w_{2,s} = \bar{d}w_{1,s} = \bar{d}^2$*,* $w_{3,s} = \bar{d}^3$ *and eventually,* $w_{t,s} = \bar{d} = \bar{d}^3$*. Substituting these value to the analytical expressions of Table D.1, it is derived that* $\rho_{s,2} = \frac{1}{\bar{d}}$*,* $\rho_{s,3} = \frac{2}{\bar{d}}$*,* $\rho_{s,4} = \frac{3+\sqrt{5}}{2\bar{d}}$ *and* $\rho_{s,5} = \frac{3}{\bar{d}}$*.*

The next step is to calculate roots for $\rho_{s,6}$ *and* $\rho_{s,7}$*. For* $t = 6$ *and* 7*, and given Equation* (C.6)*, polynomial*

$$\mathtt{P}_6(q) = \frac{1}{n}\left(w_{6,s}q^6 - 5w_{5,s}q^5 + 6w_{4,s}q^4 - w_{3,s}q^3\right)$$

and

$$\mathtt{P}_7(q) = \frac{1}{n}\left(w_{7,s}q^7 - 6w_{6,s}q^6 + 10w_{5,s}q^5 - 4w_{4,s}q^4\right).$$

Assuming the homogeneous environment, equations $\mathtt{P}_6(q) = 0$ *and* $\mathtt{P}_7(q) = 0$ *are equivalent to*

$$x^3 - 5x + 6x - 1 = 0$$

and

$$x^3 - 6x^2 + 10x_4 = 0,$$

respectively, where $x = q\bar{d}$*. These particular values correspond to* $\rho_{s,6} \approx \frac{3.25}{\bar{d}}$ *and* $\rho_{s,7} \approx \frac{3.4}{\bar{d}}$

Consequently,

$$\Delta(s,3) = \rho_{s,3} - \rho_{s,2} = \frac{2}{\bar{d}} - \frac{1}{\bar{d}} = \frac{1}{\bar{d}},$$

$$\Delta(s,4) = \rho_{s,4} - \rho_{s,3} = \frac{3+\sqrt{5}}{2\bar{d}} - \frac{2}{\bar{d}} = \frac{\sqrt{5}-1}{2\bar{d}} \approx \frac{0.618}{\bar{d}},$$

$$\Delta(s,5) = \rho_{s,5} - \rho_{s,4} = \frac{3}{\bar{d}} - \frac{3+\sqrt{5}}{2\bar{d}} = \frac{3-\sqrt{5}}{2\bar{d}} \approx \frac{0.382}{\bar{d}},$$

$$\Delta(s,6) = \rho_{s,6} - \rho_{s,5} \approx \frac{3.25}{\bar{d}} - \frac{3}{\bar{d}} = \frac{0.25}{\bar{d}}$$

and

$$\Delta(s,7) = \rho_{s,7} - \rho_{s,6} \approx -\frac{3.4}{\bar{d}} - \frac{3.25}{\bar{d}} = \frac{0.15}{\bar{d}}.$$

Obviously,

$$\Delta(s,3) > \Delta(s,4) > \Delta(s,5) > \Delta(s,6) > \Delta(s,7),$$

since

$$\frac{1}{\bar{d}} > \frac{0.618}{\bar{d}} > \frac{0.382}{\bar{d}} > \frac{0.25}{\bar{d}} > \frac{0.15}{\bar{d}}.$$

The fact that root difference converges is observed not only for regular topologies but in the general case, as it will be shown later using simulation results. This motivates the introduction of the following algorithm that determines a *forwarding probability estimation* $\ddot{q}(s)$ for which a *root difference threshold* δ is satisfied.

Algorithm 1: Estimation of $\ddot{q}(s)$.

`Input` : Initiator node s, Polynomial $\mathtt{P}_t(q)$, Root difference threshold δ
`Output` : Value $\ddot{q}(s)$
`Initialize`: $\ddot{q}(s) \leftarrow \rho_{s,3}; t = 3; \Delta \leftarrow \ddot{q}(s) - \rho_{s,2};$
`while` $\Delta > \delta$ `do`
  $t \leftarrow t + 1;$
  $\ddot{q}(s) \leftarrow \rho_{s,t}$;
  $\Delta \leftarrow \ddot{q}(s) - \rho_{s,t-1}$;
`end`
`Return` : $\ddot{q}(s)$;

Algorithm 1 will be used next to derive the forwarding probability estimation and compare it against the forwarding probability threshold as given by the simulation results.

D.2 Simulation Results and Evaluation

This section presents the simulation results for the evaluation of the analytical expressions regarding threshold probability and for Algorithm 1. It is shown that the higher the value of the forwarding probability the closer the analytical expressions to the simulation results. This is a direct consequence of the employment of the binomial approximation as already mentioned in Section C.4.

D.2.1 Model Evaluation

Figures D.1-{a,b,c} depict simulation results regarding coverage as a function of time for three GRG topologies with connectivity radius $r_c = 0.3$, $r_c = 0.4$ and $r_c = 0.5$ respectively. Five different values regarding the forwarding probability q have been considered for all four depicted cases. First, a relatively small value such that the network is not fully covered. Second, a value that corresponds to the threshold probability $\tilde{q}(s)$ as derived by the simulation. Third, the estimated forwarding probability $\ddot{q}(s)$ as derived by Algorithm 1. Note that $\tilde{q}(s)$ and $\ddot{q}(s)$ may be significantly close. Finally, two other values, larger than the previous ones arbitrarily selected for comparison purposes. It is observed that, as time passes the higher the forwarding probability, the faster coverage increases until the point of convergence. For $q < \tilde{q}(s)$, not all network nodes are covered while for $q \geq \ddot{q}(s)$ all nodes (i.e, more than 95%) are covered. Note that in all cases $\ddot{q}(s)$ is close to $\tilde{q}(s)$, that is an indication of the effectiveness of the proposed analytical model.

A similar simulation setup is used for the model evaluation in ER topologies. Figures D.2-{a,b,c} depict simulation results regarding coverage as a function of time for three ER topologies with connection probability $p = 0.2$, $p = 0.3$ and $p = 0.4$ respectively. Five different values regarding the forwarding probability q have been considered for all four depicted cases. First, a relatively small value such that the network is not fully covered. Second, a value that corresponds to the threshold probability $\tilde{q}(s)$ as derived by the simulation. Third, the estimated forwarding probability $\ddot{q}(s)$ as derived by Algorithm 1 and finally two other values, larger than the previous ones arbitrarily selected for comparison purposes. It is also observed that in all cases $\ddot{q}(s)$ is close to $\tilde{q}(s)$.

Figure D.3 depicts $\tilde{q}(s)$ and $\ddot{q}(s)$ for various GRG topologies as functions of the connectivity radius (r_c). Clearly, for small values of the connectivity radius (r_c) (and correspondingly large values of $\tilde{q}(s)$ and $\ddot{q}(s)$), the estimation of $\ddot{q}(s)$ is not close to the obtained value $\tilde{q}(s)$. However, as the value of the connectivity radius (r_c) increases and for $r_c \geq 0.07$, $\ddot{q}(s)$ is almost identical to $\tilde{q}(s)$. The reason for this difference is based on the employment of the binomial approximation in the analytical model which requires small values of the forwarding probability q. As the network topology becomes less dense (fewer links, i.e., r_c decreases) evidently larger values of the forwarding probability should be employed to achieve full coverage. On

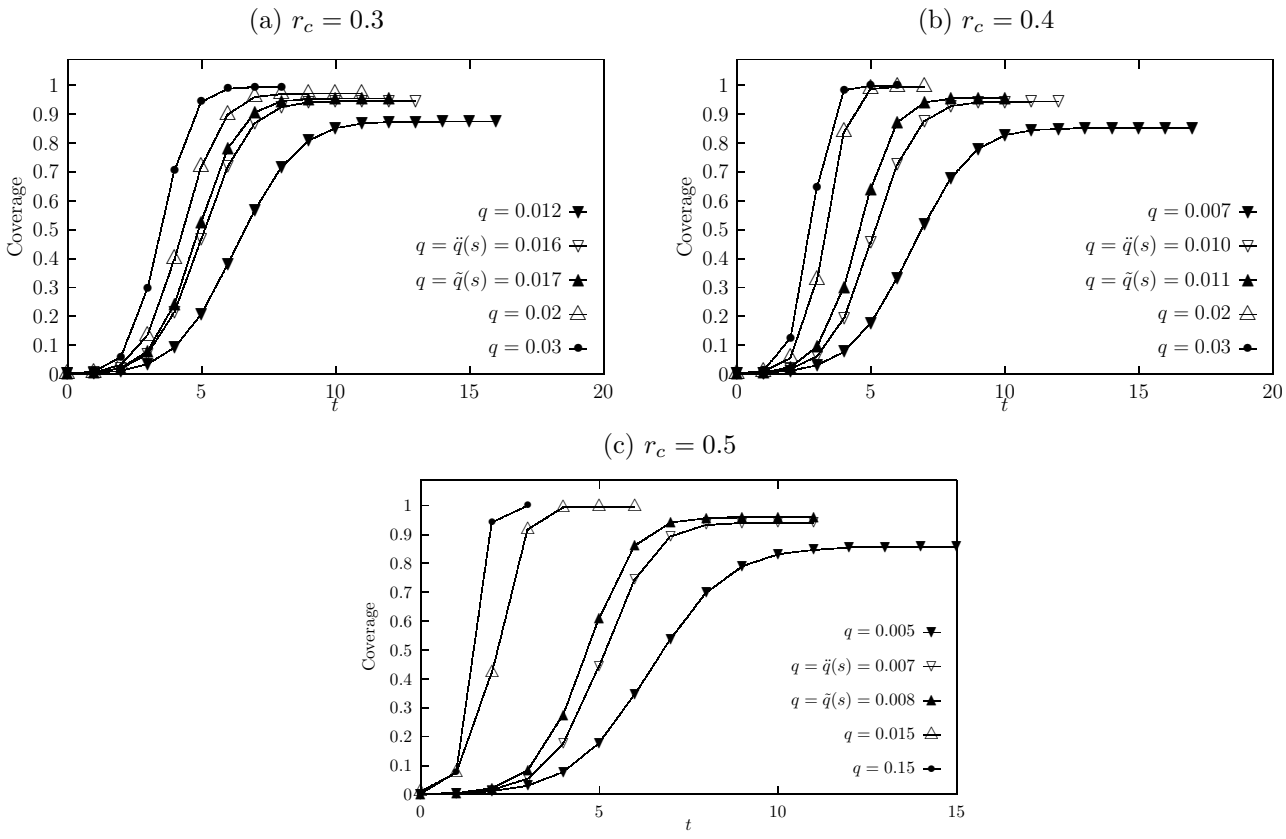


Figure D.1: *Coverage as a function of time for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5 *and five different values of the forwarding probability.*

the other hand, as r_c increases and the values of the threshold probability decrease, the analysis successfully captures the overall behavior and a $\ddot{q}(s)$ value close to $\tilde{q}(s)$ is obtained.

Figure D.4 depicts $\tilde{q}(s)$ and $\ddot{q}(s)$ for various ER topologies as functions of the connection probability p. It is also observed that for small values of the connection probability (p) (and correspondingly large values of $\tilde{q}(s)$ and $\ddot{q}(s)$), the estimation of $\ddot{q}(s)$ is not close to the obtained value $\tilde{q}(s)$. However, as the value of the connection probability (p) increases and for $p \geq 0.07$, $\ddot{q}(s)$ is almost identical to $\tilde{q}(s)$. As it is mentioned before, the reason for this difference s based on the employment of the binomial approximation in the analytical model.

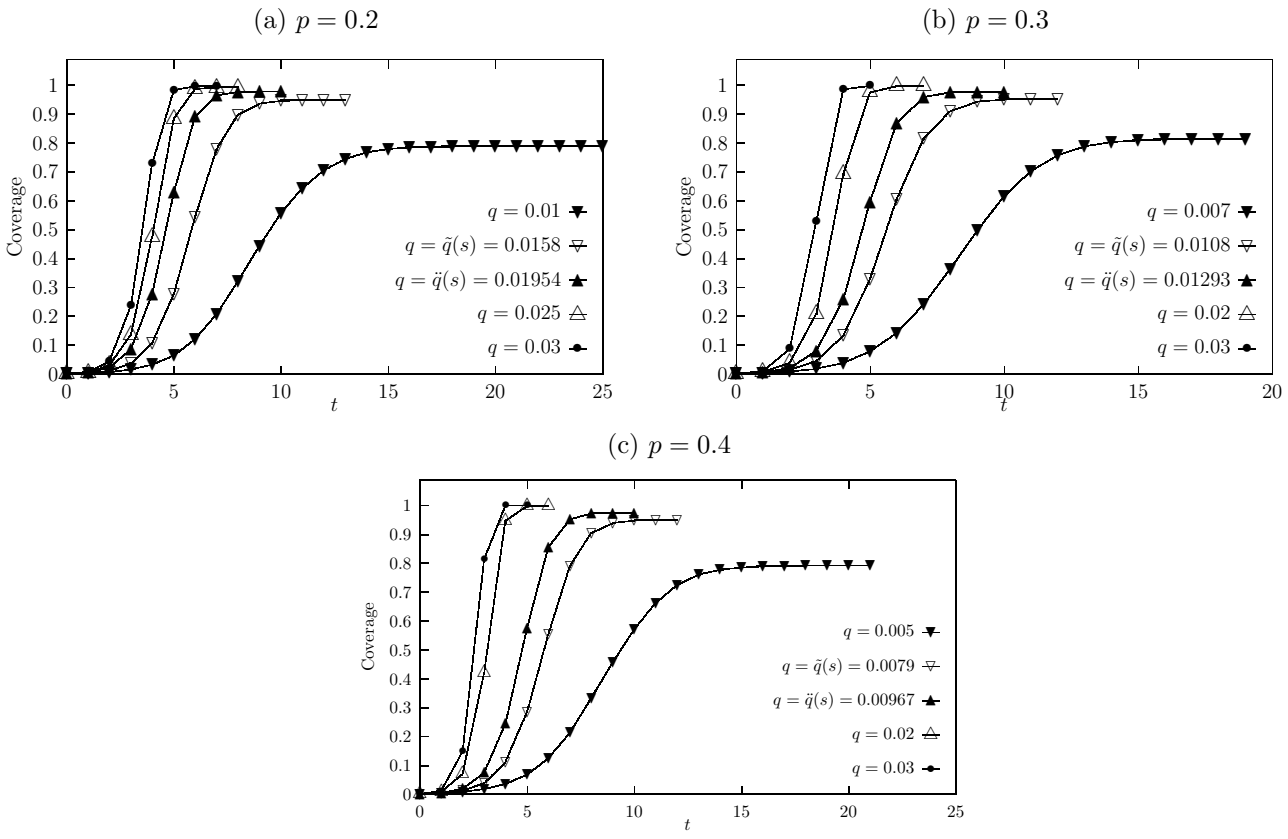


Figure D.2: *Coverage as a function of time for three ER topologies with connection probability* $p = 0.2, 0.3$ *and* 0.4 *and five different values of the forwarding probability.*

Figures D.5-{a,b,c} depict simulation results regarding coverage as a function of the forwarding probability q for three different GRG topologies with connectivity radius $r_c = 0.3$, $r_c = 0.4$ and $r_c = 0.5$ respectively. For each case, the analytical results (i.e., $\ddot{q}(s)$) along with previous works in the area (i.e., $\frac{1}{d}$, $\frac{1}{\lambda_1}$) are also depicted. It is clear that for any $q \leq \frac{1}{d}$ or $q \leq \frac{1}{\lambda_1}$, coverage is close to zero as expected from the corresponding literature [130], [129]. It is interesting to observe that for any $q \geq \ddot{q}(s)$, coverage is close to 1. The threshold value $\tilde{q}(s)$ obtained by simulations, is also shown to be close to $\ddot{q}(s)$. Eventually, $\ddot{q}(s)$ as given by the analysis is an efficient estimation of the threshold probability.

Table D.2 presents the parameters of each GRG topology along with the threshold probability, the analytical results (i.e., $\ddot{q}(s)$) and the previous works in the area

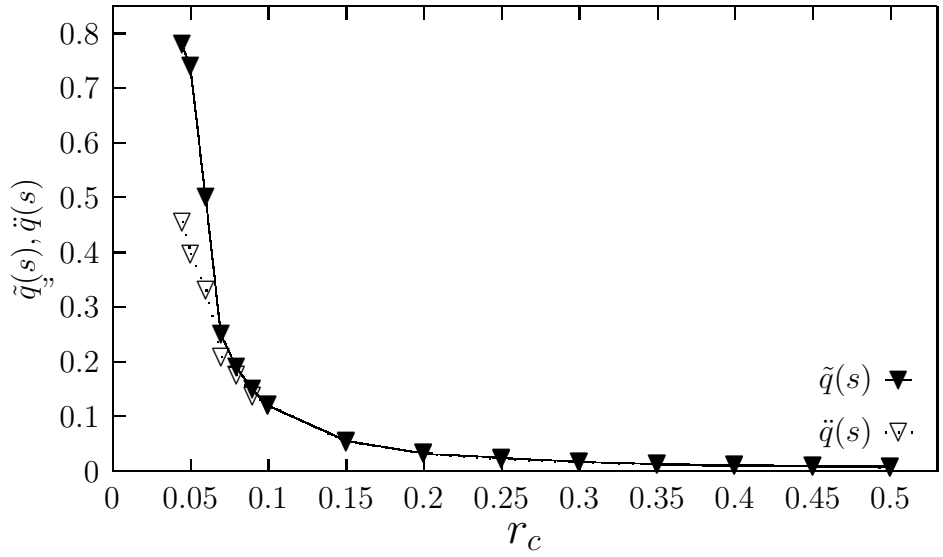


Figure D.3: *$\tilde{q}(s)$ and $\ddot{q}(s)$, as functions of the connectivity radius (r_c) for various GRG topologies.*

(i.e., $\frac{1}{d}$, $\frac{1}{\lambda_1}$).

A similar simulation setup is used for the model evaluation in ER topologies. Figures D.6-{a,b,c} depict simulation results regarding coverage as a function of the forwarding probability q for three different ER topologies with connection probability $p = 0.2$, $p = 0.3$ and $p = 0.4$ respectively. For each case, the analytical results (i.e., $\ddot{q}(s)$) along with previous works in the area (i.e., $\frac{1}{d}$, $\frac{1}{\lambda_1}$) are also depicted. It is also observed that $\ddot{q}(s)$ as given by the analysis is an efficient estimation of the threshold probability.

Table D.2: *GRG topology parameters*

n	r_c	$\frac{1}{\bar{d}}$	$\frac{1}{\lambda_1}$	$\tilde{q}(s)$	$\ddot{q}(s)$
1000	0.3	0.0047	0.0042	0.016	0.017
1000	0.4	0.0029	0.0026	0.010	0.011
1000	0.5	0.0020	0.0020	0.008	0.008

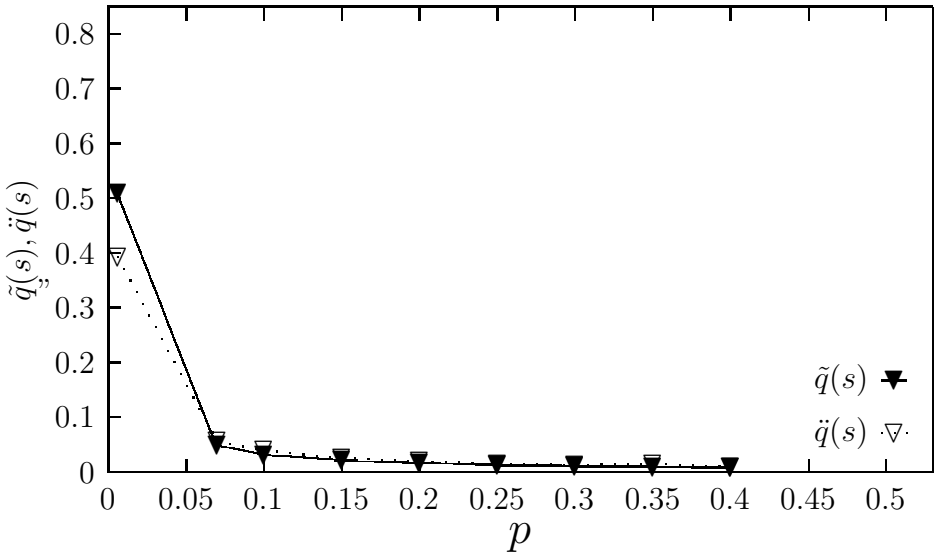


Figure D.4: *$\tilde{q}(s)$ and $\ddot{q}(s)$, as functions of the connection probability p for various ER topologies.*

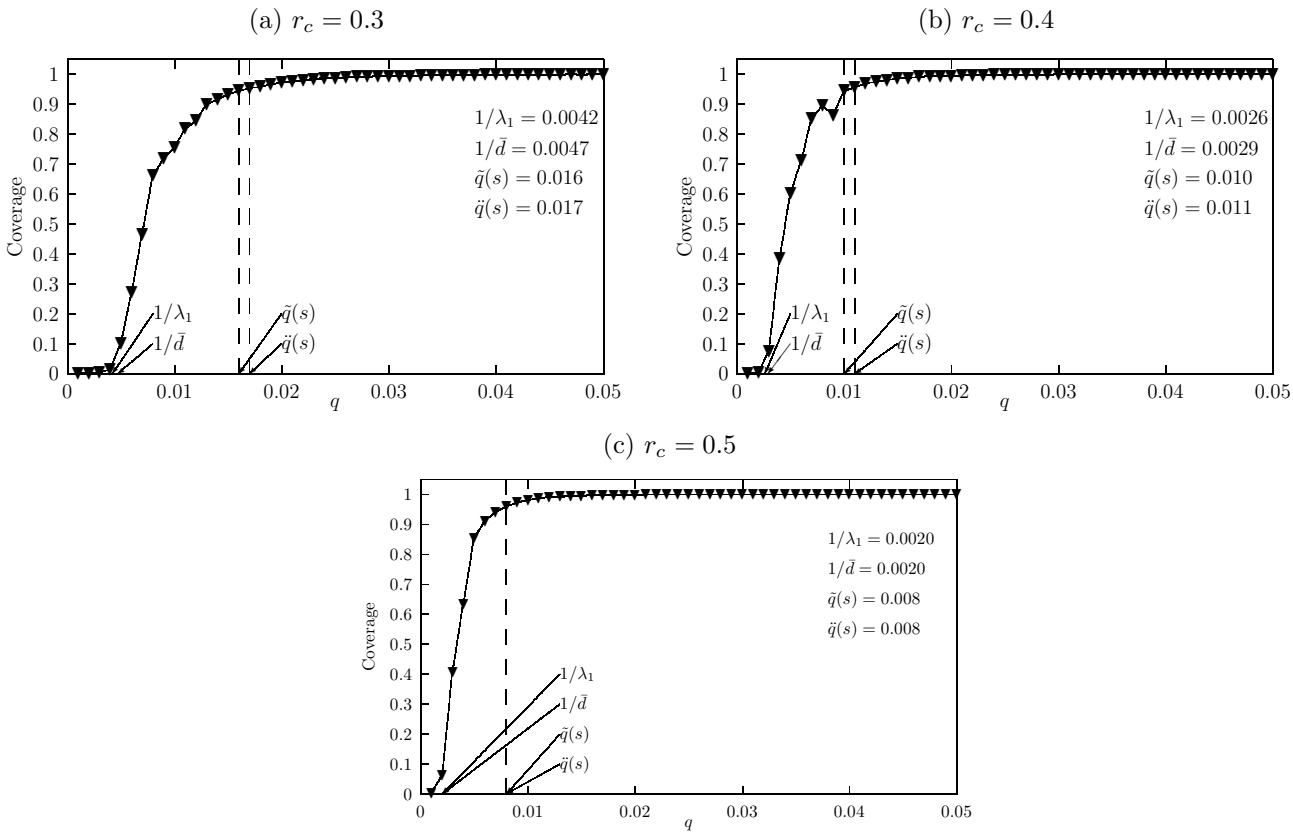


Figure D.5: *Coverage as a function of the forwarding probability for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5. *The values of* $\tilde{q}(s)$, $\ddot{q}(s)$, $\frac{1}{\bar{d}}$ *and* $\frac{1}{\lambda_1}$ *are also depicted. It is clear that for* $q \leq \frac{1}{\bar{d}}$ *or* $q \leq \frac{1}{\lambda_1}$, *coverage is close to zero while, for* $q \geq \ddot{q}(s)$ *almost full coverage is achieved.*

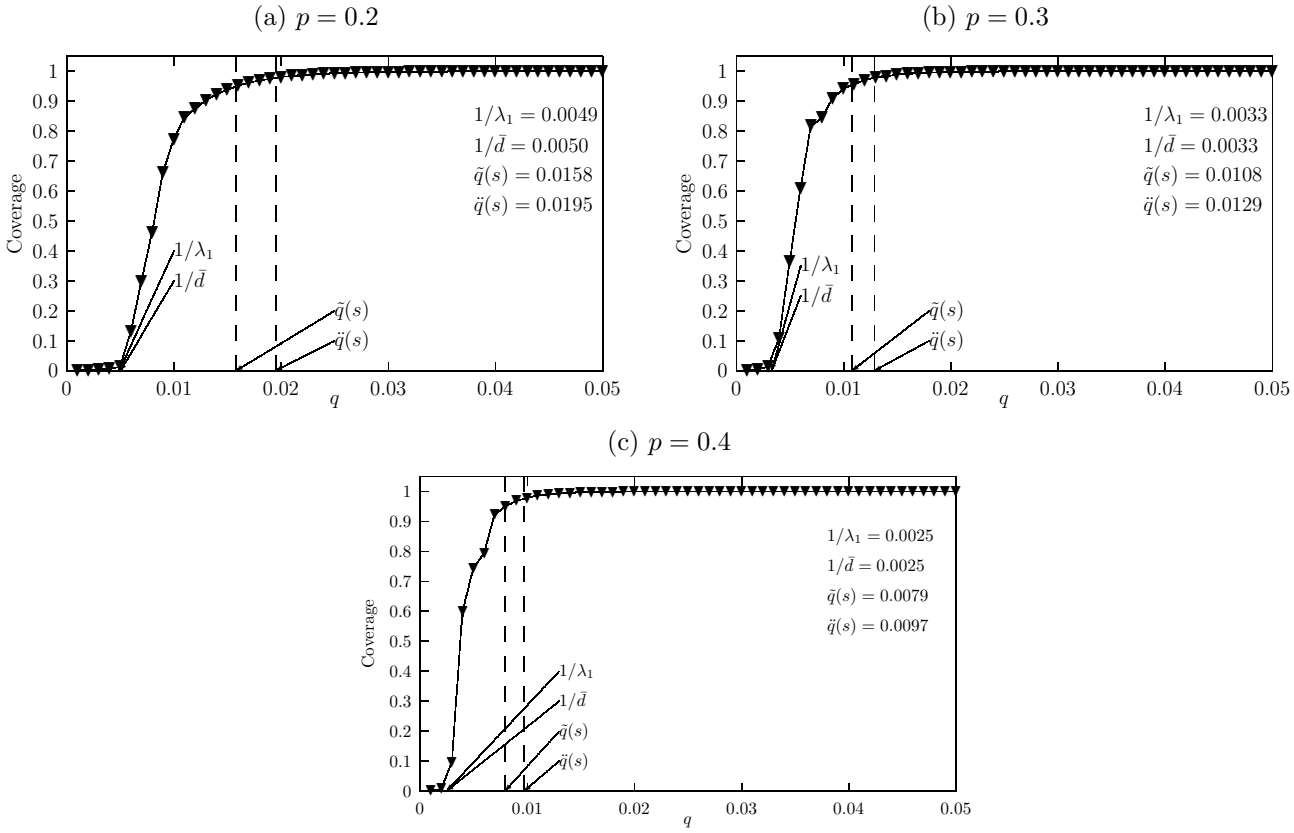


Figure D.6: *Coverage as a function of the forwarding probability for three ER topologies with connection probability $p = 0.2, 0.3$ and 0.4. The values of $\tilde{q}(s)$, $\ddot{q}(s)$, $\frac{1}{\bar{d}}$ and $\frac{1}{\lambda_1}$ are also depicted. It is clear that for $q \leq \frac{1}{\bar{d}}$ or $q \leq \frac{1}{\lambda_1}$, coverage is close to zero while, for $q \geq \ddot{q}(s)$ almost full coverage is achieved.*

Table D.3 presents the parameters of each ER topology along with the threshold probability, the analytical results (i.e., $\ddot{q}(s)$) and the previous works in the area (i.e., $\frac{1}{\bar{d}}$, $\frac{1}{\lambda_1}$).

Table D.3: *ER topology parameters*

n	p	$\frac{1}{\bar{d}}$	$\frac{1}{\lambda_1}$	$\tilde{q}(s)$	$\ddot{q}(s)$
1000	0.2	0.0050	0.0050	0.0160	0.0195
1000	0.3	0.0033	0.0033	0.0110	0.0129
1000	0.4	0.0025	0.0025	0.0080	0.0097

D.3 Discussion

In this chapter, the previously presented probabilistic flooding model is further analyzed in order to obtain an approximation of the threshold probability. It is shown that the threshold probability can be satisfactorily approximated by the higher roots of Equation (C.6). The particular roots (i.e., forwarding roots), are also used to confirm the existing results in the open literature. Subsequently, it is observed that the difference between two consecutive forwarding roots decreases as t increases. Based on this observation, a novel algorithm is introduced for the estimation of the threshold probability. The proposed algorithm, at each step, calculates the difference between two consecutive roots of the polynomial until the particular difference is smaller than a given threshold.

Simulation results confirm the analytical findings presented in this thesis and previous works in the area (i.e., [130], [129]). It is shown that the estimated threshold probability obtained by the proposed algorithm, results in global outreach under probabilistic flooding.

Chapter E

Spectral Properties of Coverage Polynomial

FUNDAMENTAL properties of probabilistic flooding not covered in any previous work are revealed in this chapter. In particular, the vector polynomial presented in Section C.3 (i.e., Equation (C.3)) is initially considered to come up with a new polynomial and the analysis focuses on certain properties of the adjacency matrix. These properties are exploited to derive analytical expressions regarding the behavior of probabilistic flooding. It is shown here that,

- the higher the *eigenvector centrality* of the node that initiates a probabilistic flooding process (i.e., *initiator* node), the higher the coverage;
- the probability of a node to be *covered* (i.e., to get the information message) is proportional to the eigenvector centrality of the particular node;
- the threshold probability that allows for global outreach depends entirely on the largest eigenvalue of the network's adjacency matrix;
- an analytical expression is derived regarding termination's time lower bound;
- a new probabilistic policy called m-Probabilistic Flooding is proposed and it is shown that the requirements for global outreach are independent of the underlying network's spectral properties.

E.1 Coverage Polynomial

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ denote the eigenvalues of the adjacency matrix $\mathcal{A}$ in decreasing order and $U_1, U_2, \ldots, U_n$ the corresponding normalized eigenvectors (i.e., $U_i^T U_i = 1$). Since graph G is connected, matrix $\mathcal{A}$ is irreducible [169]. According to the Perron-Frobenius theorem [145], an irreducible non-negative square matrix has a simple real positive largest eigenvalue λ_1 and the magnitude of any other eigenvalue does not exceed λ_1. Also, all eigenvectors apart from the *principal* one (i.e., the one corresponding to λ_1), have at least one negative element. Consequently,

- $\lambda_1 > \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_n$;
- $\lambda_1 \geq |\lambda_2| \geq |\lambda_3| \geq ... \geq |\lambda_n|$;
- $\min((U_k)_i) < 0\,, k > 1\,, 1 \leq i \leq n$;
- $\min((U_1)_i) \geq 0\,, 1 \leq i \leq n$.

The i-th element of the principal eigenvector U_1, corresponds to the *eigenvector centrality* of node i [170]. For notation simplicity, let $u_{1,i}$ denote the eigenvector centrality of node i. A useful property is the following: a graph is bipartite *iff* the adjacency matrix minimum eigenvalue magnitude equals the largest one, i.e., $\lambda_1 = |\lambda_n|$ [171]. Thus, for non-bipartite graphs, as the topologies considered here, $\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq ... \geq |\lambda_n|$.

The following lemma reveals the relationship between the initial conditions (i.e., P_0) and the topology (i.e., $\mathcal{A}$) to eigenvalue λ_1 and the principal eigenvector U_1, for large values of t.

Lemma 4 *For non-bipartite graphs, $\mathcal{A}^t P_0$ converges to $\lambda_1^t u_{1,s} U_1$, for large values of t.*

Proof 8 *The proof is included in Appendix A.1.*

Let $\Pi_t(q)$ be a polynomial as defined by the following recurrent expression,

$$\Pi_t(q) = q\lambda_1[\,\Pi_{t-1}(q) - \Pi_{t-2}(q)]\,, \tag{E.1}$$

where $\Pi_0(q) = 1$ and $\Pi_1(q) = q\lambda_1$. By analyzing Equation (E.1), similarly to [172], it is derived that,

$$\Pi_t(q) = \sum_{k=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{k+1} \binom{t-k+1}{t-2k+2} q^{t-k+1} \lambda_1^{t-k+1}. \tag{E.2}$$

Let $\hat{P}_t(q)$ denote $n \times 1$ vector polynomial such that for non-bipartite graphs,

$$\hat{P}_t(q) = \Pi_t(q) u_{1,s} U_1, \tag{E.3}$$

or,

$$\hat{P}_t(q) = \sum_{k=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{k+1} \binom{t-k+1}{t-2k+2} q^{t-k+1} \lambda_1^{t-k+1} u_{1,s} U_1. \tag{E.4}$$

For large values of t, vector polynomial $\mathcal{P}_t(q)$, given by Equation (C.4), can be approximated by $\hat{P}_t(q)$, given by Equation (E.4), in the sense that the values of the corresponding vector elements are close, as follows.

Lemma 5 *For non-bipartite graphs and large values of t, the elements of vector $\mathcal{P}_t(q)$ converge to those of vector $\hat{P}_t(q)$.*

Proof 9 *The proof is included in Appendix A.2.*

A direct consequence of Lemma 5 is that for non-bipartite graphs and large values of t, coverage can be approximated by a new polynomial, $\hat{C}_t(q) = 1/n \sum_{i=1}^{i=n} (\hat{P}_t(q))_i$ or,

$$\hat{C}_t(q) = \frac{1}{n} \Pi_t(q) u_{1,s} \sum_{i=1}^{i=n} (U_1)_i. \tag{E.5}$$

In order to find a more tractable form of $\Pi_t(q)$ than that of Equation (E.1) and Equation (E.2), the following Theorem is derived.

Theorem 3 *Polynomial $\Pi_t(q)$ is given by,*

$$\Pi_t(q) = \begin{cases} \frac{\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} & \\ \quad - \frac{\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}, & q > \frac{4}{\lambda_1}; \\ (t+1)2^t, & q = \frac{4}{\lambda_1}; \\ \frac{2(q\lambda_1)^{\frac{t+1}{2}} \sin[(t+1)\phi]}{\sqrt{4q\lambda_1 - (q\lambda_1)^2}}, & q < \frac{4}{\lambda_1}; \end{cases} \tag{E.6}$$

where $\phi = \arctan \frac{\sqrt{4q\lambda_1 - (q\lambda_1)^2}}{q\lambda_1}$.

Proof 10 *The proof is included in Appendix A.3.*

A worth mentioning outcome of Lemma 5 is that the probability for a node to get covered is proportional to its eigenvector centrality. Thus, node i_1 is more likely to get the information message than node i_2, if $u_{1,i_1} > u_{1,i_2}$. Another outcome is that since $\hat{P}_t$ is proportional to $u_{1,s}$, initiator nodes with higher eigenvector centrality result in higher coverage for the same timestep t.

E.2 Threshold Probability Analysis

The number of sent information messages under probabilistic flooding can be reduced by reducing the value of the forwarding probability q. However, small values of q may result in a large number of not covered nodes. Thus, in order to maximize the effectiveness of probabilistic flooding, the forwarding probability should be set to the minimum value that is capable of ensuring global outreach, i.e., the *threshold* probability. In particular, the objective here is the derivation of a lower bound regarding forwarding probability, denoted by $\hat{q}$, such that $\hat{C}_t(\hat{q}) \to 1$. The fact that coverage, given by Equation (E.5), is proportional to $\Pi_t(q)$, is exploited in the analysis next. The following corollary shows that $\Pi_t(q)$, given by Equation (E.6), is continuous at $q = \frac{4}{\lambda_1}$.

Corollary 3 $\Pi_t(q)$ *is continuous at* $q = \frac{4}{\lambda_1}$.

Proof 11 *The proof is included in Appendix A.4.*

The following corollary shows that for $q \geq \frac{4}{\lambda_1}$, as t increases, $\Pi_t(q)$ increases as well.

Corollary 4 *For* $q \geq \frac{4}{\lambda_1}$, *it holds that,* $\Pi_{t+1}(q) > \Pi_t(q)$, *for all timesteps* $t > 0$.

Proof 12 *The proof is included in Appendix A.5.*

The following theorem shows that for $q \geq \frac{4}{\lambda_1}$, polynomial $\Pi_t(q)$ is strictly positive and increasing for all timesteps $t > 0$.

Theorem 4 *The minimum value of q, such that $\Pi_{t+1}(q) > \Pi_t(q) > 0$, for all $t > 0$, is $\frac{4}{\lambda_1}$.*

Proof 13 *The proof is included in Appendix A.6.*

Since the coverage polynomial $\hat{C}_t(q)$, given by Equation (E.5), is proportional to $\Pi_t(q)$ for large values of t, the following corollary can be proved.

Corollary 5 *A lower bound regarding forwarding probability is given by $\hat{q} = \frac{4}{\lambda_1}$, for large values of timestep t.*

Proof 14 *The proof is included in Appendix A.7.*

E.3 Special Cases

The determination of the largest eigenvalue of a graph is not feasible in all cases. For example, in a large scale network, the resources required to determine λ_1 may overcome the benefits obtained by probabilistic flooding [174]. Additionally, in environments of high mobility, where the topology changes very fast, the determination of λ_1, may be impossible [175]. However, by using the properties of algebraic graph theory, a less tight and easier to compute than before, a lower bound regarding forwarding probability can be derived [176], as given by Corollary 6, based on the average number of neighbor nodes in the network, i.e., $\bar{d}$.

Corollary 6 *A forwarding probability of $\bar{q} = 4/\bar{d}$ is sufficient to provide global outreach, for large values of timestep t.*

Proof 15 *The proof is included in Appendix A.9.*

The effectiveness of Corollary 6 is demonstrated in the sequel using simulation results.

For the case of the Erdős-Rényi Graphs, the average vertex degree is closely related to the connection probability. Thus, Corollary 7 can be derived.

Corollary 7 *For Erdős-Rényi Graphs with n vertexes and connection probability p, the forwarding probability given by,*

$$q_e = \frac{4}{np}, \tag{E.7}$$

allows for global outreach.

Proof 16 *The average vertex degree of Erdős-Rényi Graphs is $\bar{d} = np$, where n is the number of vertices and p the connection probability. Using Corollary 6 it follows that, $q_e = \bar{q} = 4/\bar{d} = \frac{4}{np}$*

For the case of the Geometric Random Graphs, it is shown in [177] that the Cumulative Distribution Function (i.e., *CDF*) of the Euclidean distance in GRG topologies is given by,

$$F(r_c) = \begin{cases} \frac{1}{2}{r_c}^4 - \frac{8}{3}{r_c}^3 + \pi {r_c}^2, & r_c \in [0,1]; \\ 2\,{r_c}^2 \arcsin\left(\frac{1}{r_c}\right) - 2\,{r_c}^2 \arccos\left(\frac{1}{r_c}\right) & \\ \qquad + 4\sqrt{{r_c}^2 - 1}, & r_c \in (1, \sqrt{2}]. \end{cases}.$$

An alternative proof is also given in Appendix A.8. Consequently, Corollary 8 can be derived.

Corollary 8 *For Geometric Random Graphs with n vertexes and connectivity radius r_c the forwarding probability given by,*

$$q_g = \frac{4}{nF(r_c)}, \tag{E.8}$$

allows for global outreach.

Proof 17 *The proof is included in Appendix A.10.*

The largest eigenvalue of regular graphs is closely related to the graph degree d_R thus, the analytical expressions regarding coverage and threshold probability presented in Section E.1 and Section E.2 respectively, can be rewritten as it is shown in Corollary 9.

Corollary 9 *For $t \geq 0$, probabilistic flooding coverage is given by,*

$$C_{R,t}(q) = \begin{cases} \frac{1}{n}\frac{\left(qd_R + \sqrt{q^2 d_R^2 - 4qd_R}\right)^{t+1}}{2^{t+1}\sqrt{q^2 d_R^2 - 4qd_R}} & \\ \qquad - \frac{1}{n}\frac{\left(qd_R - \sqrt{q^2 d_R^2 - 4qd_R}\right)^{t+1}}{2^{t+1}\sqrt{q^2 d_R^2 - 4qd_R}}, & \textit{for } q > \frac{4}{d_R}; \\ \frac{1}{n}(t+1)2^t, & \textit{for } q = \frac{4}{d_R}; \\ \frac{1}{n}\frac{2(qd_R)^{\frac{t+1}{2}}\sin[(t+1)\phi]}{\sqrt{4qd_R - (qd_R)^2}}, & \textit{for } q < \frac{4}{d_R}; \end{cases}$$

and Threshold probability is given by,

$$\hat{q}_R = \frac{4}{d_R}.$$

Proof 18 *The proof is included in Appendix A.11.*

E.4 Termination Time

It is interesting to investigate termination time under probabilistic flooding for sufficient values of q such that global outreach is guaranteed with high probability (i.e., $C_t(q) \to 1$). Let $\mathcal{T}(q,s)$ denote the termination time under probabilistic flooding started from node s with forwarding probability q. Under blind flooding (i.e., $q=1$), it is $C_{\mathcal{T}(1,s)} \leq D$. However, it is not trivial to derive any boundaries regarding termination time for $q<1$, since the information message may not be forwarded through certain links that belong to the shortest paths towards the rest of the nodes. Thus, for the values of q that result in global outreach with high probability, termination time is expected to be increased (i.e., $\mathcal{T}(q,s) \geq \mathcal{T}(1,s)$, for $q<1$). Employing the analytical findings presented in this thesis, a lower bound regarding termination time for $q \geq \hat{q}$ (i.e., $\mathcal{T}_{LB}(q,s)$) is derived by theorem 5.

Theorem 5 *Termination time under probabilistic flooding is lower bounded by,*

$$\mathcal{T}_{LB}(q,s) = \begin{cases} \frac{\mathcal{W}_0\left(\frac{1.9nln2}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)}{ln2} - 1, & for\ q=\hat{q}; \\ \frac{ln\left(\frac{n\sqrt{q^2\lambda_1^2-4q\lambda_1}}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)}{ln\left(\frac{q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}}{2}\right)} - 1, & for\ q>\hat{q}; \end{cases} \tag{E.9}$$

where $\mathcal{W}_0$ is the 0-branch of Lambert-W function.

Proof 19 *The proof is included in Appendix A.12.*

A worth mentioning outcome of Theorem 5 is that the lower bound regarding termination time decreases as initiator's node eigenvector centrality (i.e., $u_{1,s}$) increases. Although there is no analytical expression regarding termination time under probabilistic flooding, it is shown in the simulations section that termination time also decreases as initiator's node eigenvector centrality increases.

For the case of the regular graphs, Corollary 10 derives an analytical expression regarding termination's time lower bound (i.e., $\mathcal{T}_{R,LB}$) as a function of the forwarding probability and the graph degree (i.e., d_R).

Corollary 10

$$\mathcal{T}_{R,LB}(q, d_R) = \begin{cases} \frac{\mathcal{W}_0(1.9nln2)}{ln2} - 1, & for\ q = \hat{q}_R; \\ \frac{ln\left(n\sqrt{q^2 d_R^2 - 4qd_R}\right)}{ln\left(\frac{qd_R + \sqrt{q^2 d_R^2 - 4qd_R}}{2}\right)} - 1, & for\ q > \hat{q}_R. \end{cases} \tag{E.10}$$

Proof 20 *The proof is included in Appendix A.13.*

E.5 m-Probabilistic Flooding

In order to achieve global outreach under probabilistic flooding, the forwarding probability should be set to a sufficiently large value. According to the analytical results presented in Section E.2, probabilistic flooding coverage and threshold probability are closely related to the principal eigenvector and the largest eigenvalue of the corresponding graph. However, Corollaries 6, 7, 8 and 9, presented in Section E.3, may be advantageous when the determination of the graph's spectral properties is impossible. However, the structural properties of the graph (i.e., the connection probability for the Erdős-Rényi graphs, the connectivity radius for the Geometric Random Graphs or the graph degree for the Regular graphs) should be known. In the case where neither spectral nor structural graph properties are known, the proposed policy called *m-Probabilistic Flooding* (mPF) can be applied in order to achieve global outreach. Under mPF, node i forwards the information message to its neighbors with probability $p_i = \frac{m}{d_i}$, where m is a constant value defined by the initiator node and d_i is the number of neighbors of node i. Thus, every node forwards the information message on an average of m of its neighbors. As per the analysis performed in [172] and [178], mPF's behavior can be captured by Equation (E.11),

$$\mathcal{P}_t = Q\left(\mathcal{P}_{t-1} - \mathcal{P}_{t-2}\right), \tag{E.11}$$

where $\mathcal{P}_t$ is the $n \times 1$ approximation vector and Q is an $n \times n$ matrix such that,

$$(Q)_{ij} = \begin{cases} \frac{m}{d_j}, & (i,j) \in E(G); \\ 0, & (i,j) \notin E(G). \end{cases}$$

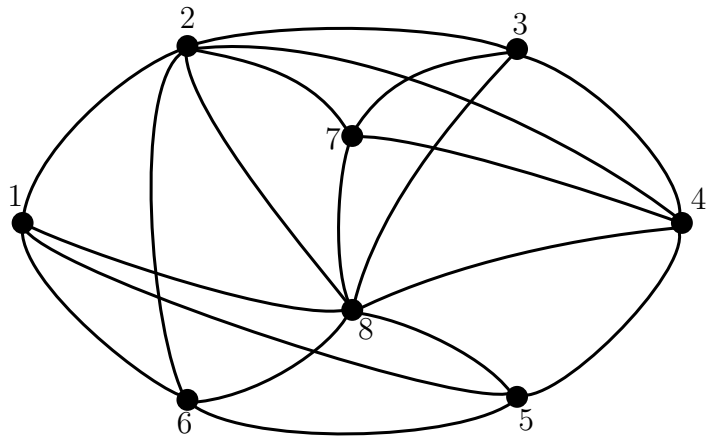


Figure E.1: *A connected network topology with eight nodes.*

For example, for the topology depicted in Figure E.1 the corresponding Q matrix is given by,

$$Q = \begin{bmatrix} 0 & \frac{m}{6} & 0 & 0 & \frac{m}{4} & \frac{m}{4} & 0 & \frac{m}{7} \\ \frac{m}{4} & 0 & \frac{m}{4} & 0 & 0 & \frac{m}{4} & \frac{m}{4} & \frac{m}{7} \\ 0 & \frac{m}{6} & 0 & \frac{m}{5} & 0 & 0 & \frac{m}{4} & \frac{m}{7} \\ 0 & \frac{m}{6} & \frac{m}{4} & 0 & \frac{m}{4} & 0 & \frac{m}{4} & \frac{m}{7} \\ \frac{m}{4} & 0 & 0 & \frac{m}{5} & 0 & \frac{m}{4} & 0 & \frac{m}{7} \\ \frac{m}{4} & 0 & 0 & 0 & \frac{m}{4} & 0 & 0 & \frac{m}{7} \\ 0 & \frac{m}{6} & \frac{m}{4} & \frac{m}{5} & 0 & 0 & 0 & \frac{m}{7} \\ \frac{m}{4} & \frac{m}{6} & \frac{m}{4} & \frac{m}{5} & \frac{m}{4} & \frac{m}{4} & \frac{m}{4} & 0 \end{bmatrix}.$$

It is observed that the sum of each column elements of table Q equals to m (i.e., $\sum_{j=1}^{j=n}(Q)_{ij} = m$). Consequently, the following Lemma regarding the spectral properties of matrix Q can be derived,

Lemma 6 *m is an eigenvalue of Q corresponding to the left eigenvector $\vec{1}^T$, where $\vec{1}^T$ is the transpose of all ones vector.*

Proof 21 *The proof is included in Appendix A.14.*

Lemma 7 derives coverage at timestep t under mPF as a function of m.

Lemma 7 *Coverage under mPF at timestep t is given by,*

$$C_{M,t}(m) = \frac{1}{n} \sum_{k=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{k+1} \binom{t-k+1}{t-2k+2} m^{t-k+1}$$

Proof 22 *The proof is included in Appendix A.15.*

The following Theorem shows that the minimum value of m that allows for global outreach is given by, $\hat{m} = 4$.

Theorem 6 *A lower bound regarding m which allows for global outreach under mPF is given by,*

$$\hat{m} = 4.$$

Proof 23 *The proof is included in in Appendix A.16.*

In accordance with Theorem 5, a lower bound regarding termination time under mPF is given by Equation (E.12).

$$\mathcal{T}_{M,LB}(m) = \begin{cases} \frac{\mathcal{W}_0(1.9nln2)}{ln2} - 1, & \text{for } m = \hat{m}; \\ \frac{ln\left(n\sqrt{m^2-4m}\right)}{ln\left(\frac{m+\sqrt{m^2-4m}}{2}\right)} - 1, & \text{for } m > \hat{m}. \end{cases} \tag{E.12}$$

E.6 Simulation Results and Evaluation

This section presents the simulation results for the evaluation of the analytical expressions regarding threshold probability, termination time and mPF. It is shown that

- the analytical expression of threshold probability is capable of covering over 95% of the network's nodes;
- starting a probabilistic flooding process from a node with higher eigenvector centrality results in higher network coverage;
- nodes with higher eigenvector centrality are more likely to get the information message;

- termination time decreases as the initiator's node eigenvector centrality increases;

- m-Probabilistic Flooding is capable of achieving global outreach in either ER or GRG topologies, regardless of the underlying graph's spectral properties.

The effects of the binomial approximation are also highlighted and it is observed that for the more dense topologies which require a certain small forwarding probability to get covered, the analytical findings are significantly close to the simulation results.

E.6.1 Coverage

Figures E.2-{a,b,c} depict simulation results regarding coverage as a function of time for two different initiator nodes (s_1, s_2) for three GRG topologies with connectivity radius $r_c = 0.3, 0.4$ and 0.5, respectively and forwarding probability $q = 0.033$. Let $C_t\,|_{s=s_1}$ and $C_t\,|_{s=s_2}$ correspond to the probabilistic flooding coverage initiated by node s_1 and s_2, respectively. The eigenvector centrality of s_1 is larger than this of s_2 (i.e., $u_{1,s_1} > u_{1,s_2}$). In all cases, it is observed that coverage is larger when probabilistic flooding is initiated from s_1 than that from s_2. This is expected since coverage is shown to be proportional to the eigenvector centrality $u_{1,s}$ of the initiator node s (i.e., Equation (E.5)).

Figures E.3-{a,b,c} depict simulation results regarding coverage as a function of time for two different initiator nodes (s_1, s_2) for various ER topologies with connection probability $p = 0.2, 0.3$ and 0.4, respectively and forwarding probability $q = 0.033$. Eigenvector centrality of s_1 is larger that this of s_2 (i.e., $u_{1,s_1} > u_{1,s_2}$). In all cases, it is also observed that coverage is larger when probabilistic flooding is initiated from s_1 than this of s_2.

As depicted, for the case of GRG topologies $u_{1,s_1} >> u_{1,s_2}$, while for the case of ER topologies u_{1,s_1} is close to u_{1,s_2} (even though $u_{1,s_1} > u_{1,s_2}$). For the former case, the difference $C_t\,|_{s=s_1} - C_t\,|_{s=s_2}$ is observed in Figures E.2-{a,b,c} to be larger than that depicted in Figures E.3-{a,b,c} for the latter case. This is another confirmation of the analytical findings given by Equation (E.5).

The probability of a node to be covered (hereinafter referred to as *coverage probability*) is shown to be proportional to its eigenvector centrality as given by Lemma 5. Figures E.4-{a,b,c} depict the average values of 1000 independent runs

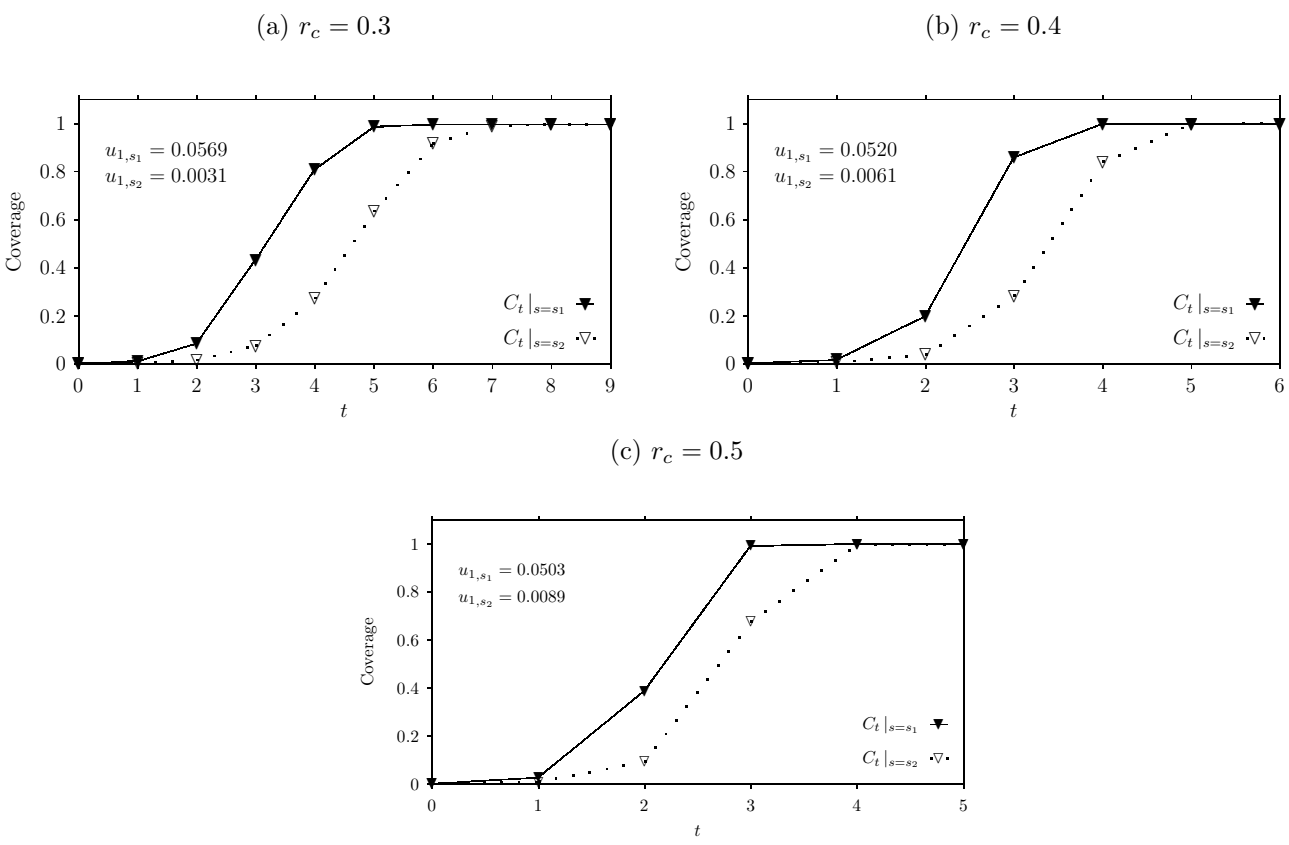


Figure E.2: *Coverage as a function of the time t for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5 *and two different initiator nodes* s_1, s_2 *such that* $u_{1,s_1} > u_{1,s_2}$.

for 15 arbitrarily selected nodes regarding coverage as a function of eigenvector centrality for a GRG topology with connectivity radius $r_c = 0.3$ and forwarding probability $q = 0.0167, 0.02$ and 0.025, respectively. The particular results support the analytical findings of Theorem 3 that coverage probability is proportional to eigenvector centrality for large values of t. For small values of t, coverage probability may not increase as eigenvector centrality increases, however, a trend of increment is observed.

Figures E.5-{a,b,c} depict the average values of 1000 independent runs for 15 arbitrarily selected nodes regarding coverage as a function of eigenvector centrality for an ER topology with connection probability $p = 0.2$ and forwarding probability

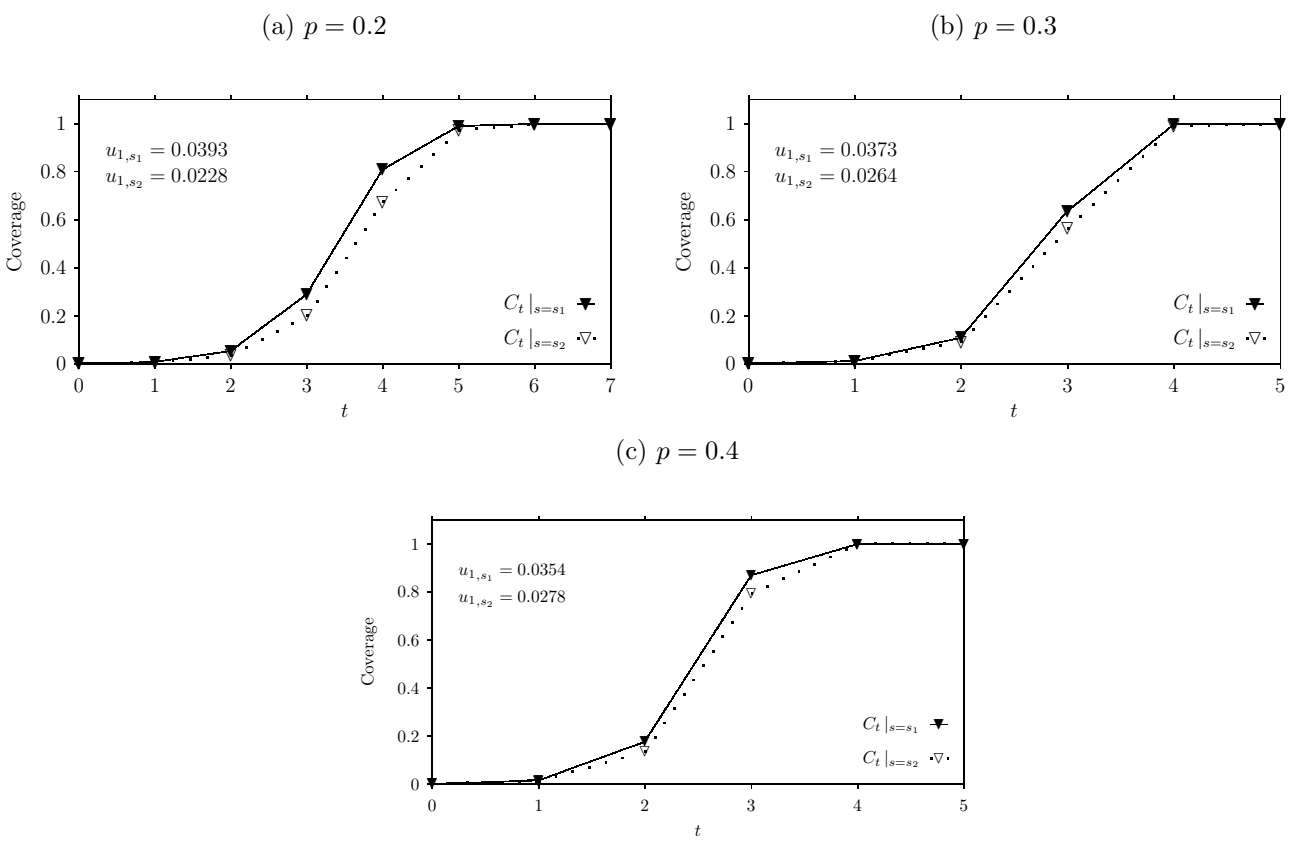
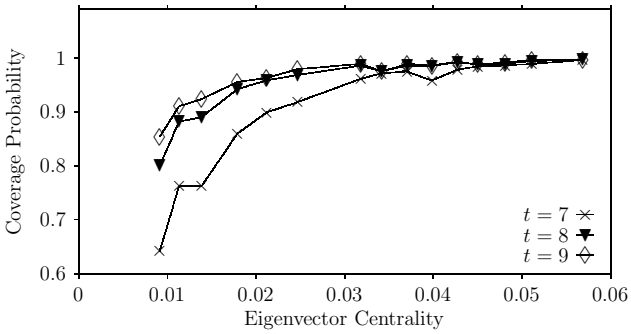


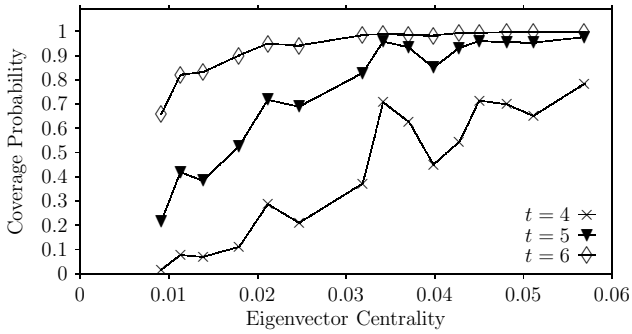
Figure E.3: *Coverage as a function of the time t for three ER topologies with connection probability* $p = 0.2, 0.3$ *and* 0.4 *and two different initiator nodes* s_1, s_2 *such that* $u_{1,s_1} > u_{1,s_2}$.

$q = 0.040, 0.042$ and 0.050, respectively. It is also observed that for large values of t, the higher the eigenvector centrality, the higher the coverage probability.

(a) $q = 0.0167$



(b) $q = 0.02$



(c) $q = 0.03$

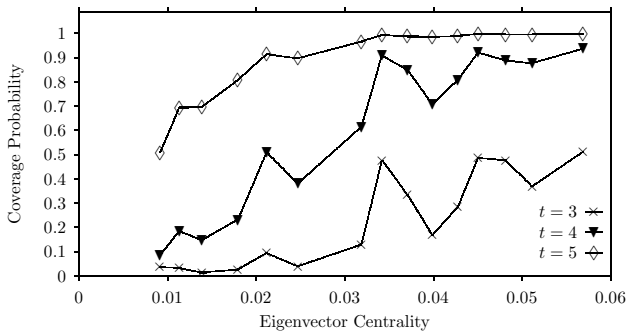
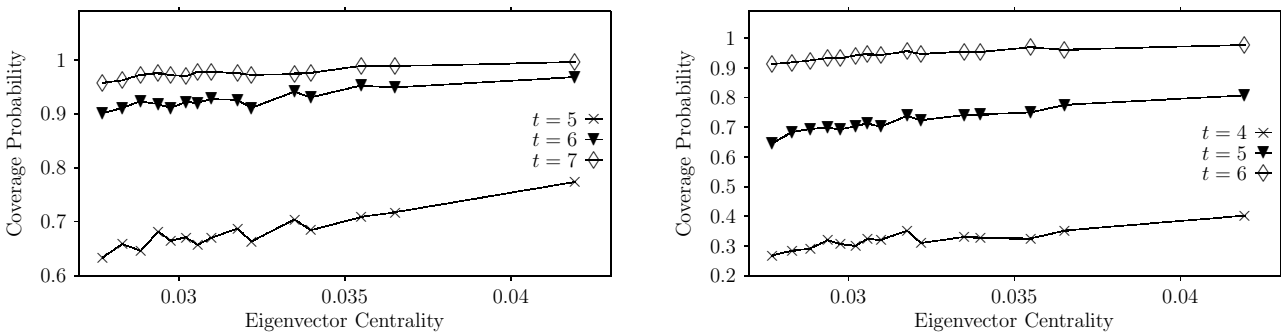


Figure E.4: *Node coverage probability as a function of eigenvector centrality for a GRG topology with connectivity radius $r_c = 0.3$, for three different values of the forwarding probability $q = 0.0167, 0.02$ and 0.025 and for three different timesteps.*

(a) $q = 0.040$

(b) $q = 0.042$



(c) $q = 0.050$

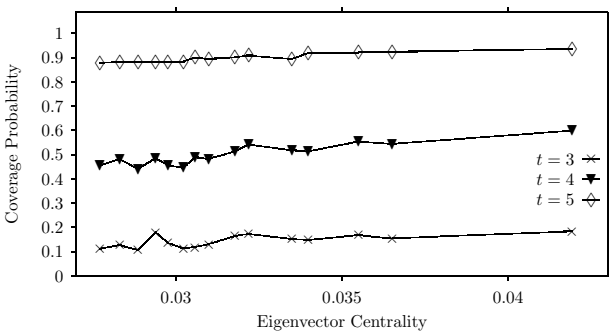


Figure E.5: *Node coverage probability as a function of eigenvector centrality for an ER topology with connection probability* $p = 0.2$*, for three different values of the forwarding probability* $q = 0.040, 0.042$ *and* 0.050 *and for three different timesteps.*

E.6.2 Threshold probability

Figures E.6-{a,b,c} depict simulation results regarding coverage as a function of the forwarding probability q for various GRG topologies with connectivity radius $r_c = 0.3, 0.4$ and 0.5 respectively. For each case, the analytical expressions of $\hat{q}$ (given by Corollary 5), $\bar{q}$ (given by Corollary 6) and q_g (given by Corollary 8) along with the threshold probability obtained by simulations ($\tilde{q}$) are also depicted. It is observed that $q_g = \bar{q} \geq \hat{q} \geq \tilde{q}$ is satisfied, which confirms that $\hat{q}$, $\bar{q}$ and q_g (the analytical results) can be employed under probabilistic flooding to achieve global network outreach.

Figures E.7-{a,b,c} depict simulation results regarding coverage as a function of the forwarding probability q for various ER topologies with connection probability $p = 0.2, 0.3$ and 0.4 respectively. For each case, the analytical expressions of $\hat{q}$, $\bar{q}$ and q_e along with the threshold probability obtained by simulations ($\tilde{q}$) are also depicted. It is observed that $q_e = \bar{q} \geq \hat{q} \geq \tilde{q}$ is satisfied, which confirms that $\hat{q}$, $\bar{q}$ and q_e (the analytical results) can be employed under probabilistic flooding to achieve global network outreach. Another observation is that as the topologies become denser (i.e., r_c increases), the difference $\hat{q} - \tilde{q}$ and the difference $\bar{q} - \tilde{q}$ decrease. This behavior is explained by the employment of the binomial approximation.

As it is mentioned in the analysis, the calculation of $\hat{q}$ (given by Corollary 5) may not be feasible for various reasons (e.g., need for global knowledge). Even though it is closer to $\tilde{q}$, as observed by the simulation results, $\bar{q}$, q_e and q_g are alternative approaches allowing for global outreach. Moreover, as network topologies become more dense, $\bar{q}$, q_e and q_g approach $\hat{q}$ and $\tilde{q}$.

Table E.1 and Table E.2 presents the parameters of each GRG and ER topology respectively, along with the threshold probability (i.e., $\tilde{q}$), the analytical results (i.e., $\ddot{q}, \hat{q}, \bar{q}, q_g$ and q_e) and the previous works in the area (i.e., $\frac{1}{d}$, $\frac{1}{\lambda_1}$). In all cases it is observed that $\hat{q}$ and $\ddot{q}$ are very close to $\tilde{q}$. For the case of sparse topologies, the analytical expressions $\bar{q}$, q_g and q_e are slightly larger than $\tilde{q}$ while, for the cases of dense topologies the particular expressions are very close to $\tilde{q}$.

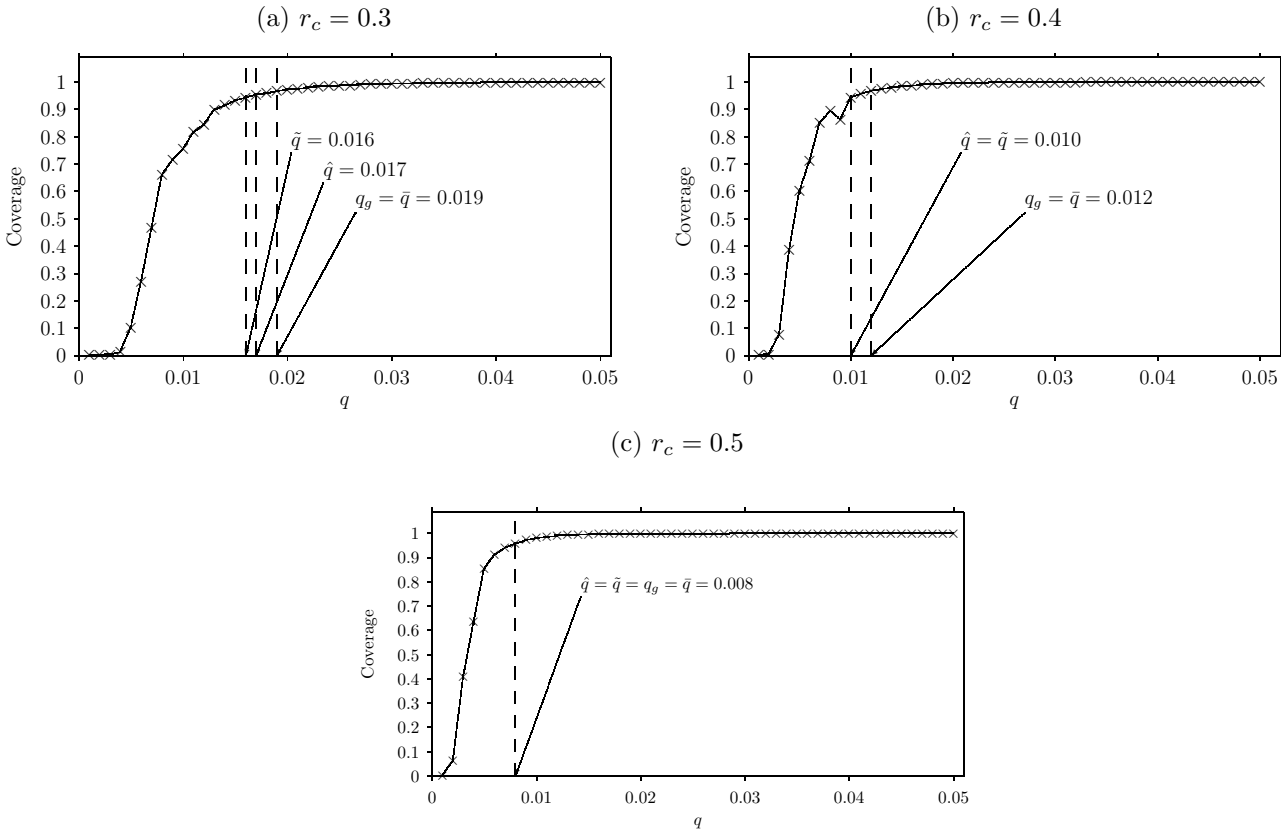


Figure E.6: *Network coverage as a function of the forwarding probability for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5.

Table E.1: *GRG topology parameters*

n	p	$\frac{1}{\bar{d}}$	$\frac{1}{\lambda_1}$	$\tilde{q}$	$\ddot{q}$	$\hat{q}$	$\bar{q}$	q_g
1000	0.3	0.0047	0.0042	0.016	0.017	0.017	0.019	0.019
1000	0.4	0.0029	0.0026	0.010	0.011	0.010	0.012	0.012
1000	0.5	0.0020	0.0020	0.008	0.008	0.008	0.008	0.008

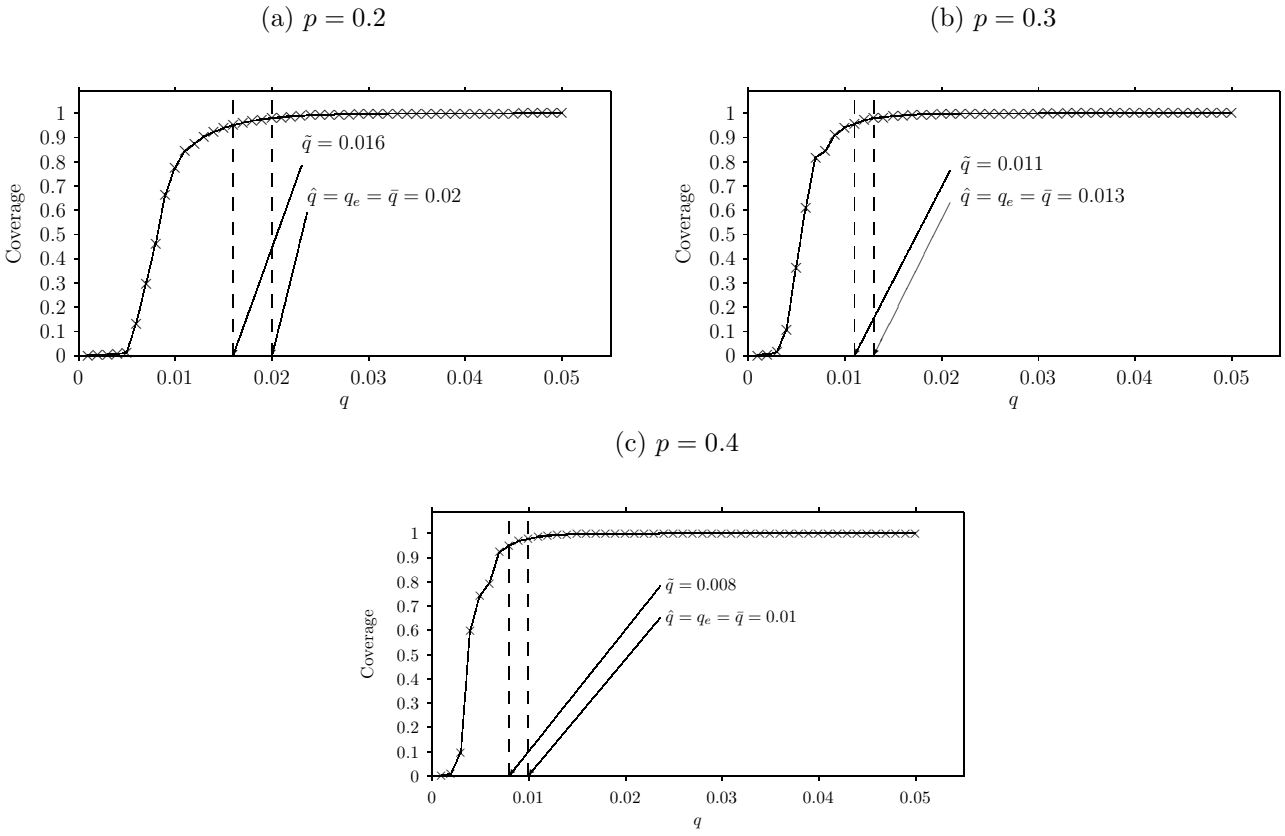


Figure E.7: *Network coverage as a function of the forwarding probability for three ER topologies with connection probability* $p = 0.2, 0.3$ *and* 0.4.

Table E.2: *ER topology parameters*

n	p	$\frac{1}{\bar{d}}$	$\frac{1}{\lambda_1}$	$\tilde{q}$	$\ddot{q}$	$\hat{q}$	$\bar{q}$	q_e
1000	0.2	0.0050	0.0050	0.0160	0.0195	0.020	0.020	0.020
1000	0.3	0.0033	0.0033	0.0110	0.0129	0.013	0.013	0.013
1000	0.4	0.0025	0.0025	0.0080	0.0097	0.010	0.010	0.010

E.6.3 Termination time

Figures E.8-{a,b,c} depict simulation results regarding termination time and its analytical lower bound (i.e., Theorem 5) as a function of the forwarding probability q for various GRG topologies with connectivity radius $r_c = 0.3, 0.4$ and 0.5 respectively. As it is expected, termination time (i.e., solid line) decreases as the forwarding probability increases. It is observed that the analytical expression of the termination's time lower bound (i.e., dotted line) is always smaller than the corresponding termination time obtained by the simulations. It is also observed that in all cases the analytical expression of the termination's time lower bound, follows the decrement rate of the termination time. For example in Figure E.8-{a}, termination time decreases more rapidly for $q \in [0.02, 0.03]$ than for $q \in [0.03, 0.04]$ and this behavior is also captured by the analytical expression of the termination's time lower bound. Figures E.9-{a,b,c} depict simulation results regarding termination time and its analytical lower bound (i.e., Theorem 5) as a function of the forwarding probability q for various ER topologies with connection probability $p = 0.2, 0.3$ and 0.4 respectively. It is also observed that in all cases the analytical expression of the termination's time lower bound, follows the decrement rate of the termination time. For example in Figure E.9-{a}, termination time decreases more rapidly for $q \in [0.025, 0.03]$ than for $q \in [0.04, 0.045]$ and this behavior is also captured by the analytical expression of the termination's time lower bound.

Figures E.10-{a,b,c} depict simulation results regarding termination time as a function of the forwarding probability q for two different initiator nodes (s_1, s_2) for various GRG topologies with connectivity radius $r_c = 0.3, 0.4$ and 0.5 respectively. It is observed that the eigenvector centralities of the selected initiator nodes are different such that $u_{1,s_1} > u_{1,s_2}$. These values are also depicted in the corresponding figures. Although there is no analytical expression regarding termination time under probabilistic flooding, in all cases, it is clearly observed that termination time for the initiator node s_1, is lower than this of the initiator node s_2.

Figures E.11-{a,b,c} depict simulation results regarding termination time as a function of the forwarding probability q for two different initiator nodes (s_1, s_2) for various ER topologies with connection probability $p = 0.2, 0.3$ and 0.4 respectively. It is observed that the eigenvector centralities of the selected initiator nodes are different such that $u_{1,s_1} > u_{1,s_2}$. It is also observed that termination time for the

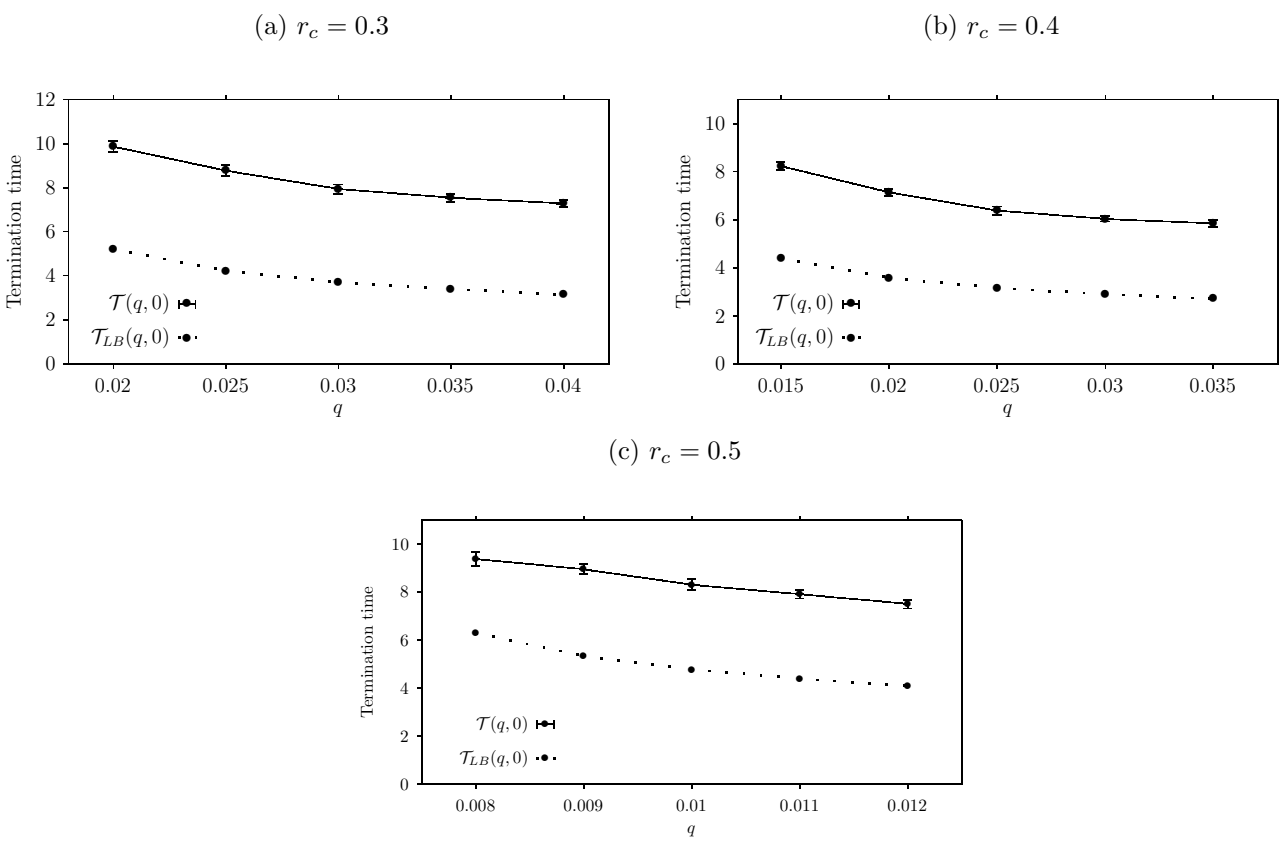
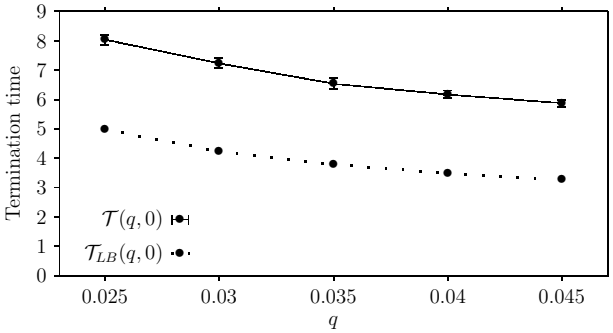


Figure E.8: *Termination time ($\mathcal{T}$) and its analytical lower bound ($\mathcal{T}_{LB}$) as a function of forwarding probability for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5. *95%-Confidence Intervals are also depicted.*

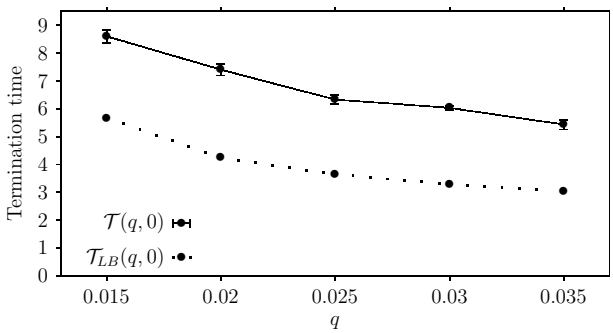
initiator node s_1, is lower than this of the initiator node s_2.

As depicted, for the case of GRG topologies $u_{1,s_1} >> u_{1,s_2}$, while for the case of ER topologies u_{1,s_1} is close to u_{1,s_2} (even though $u_{1,s_1} > u_{1,s_2}$). For the former case, the difference $\mathcal{T}(q, s_2) - \mathcal{T}(q, s_1)$ is observed to be larger than that depicted for the latter case.

(a) $p = 0.2$



(b) $p = 0.3$



(c) $p = 0.4$

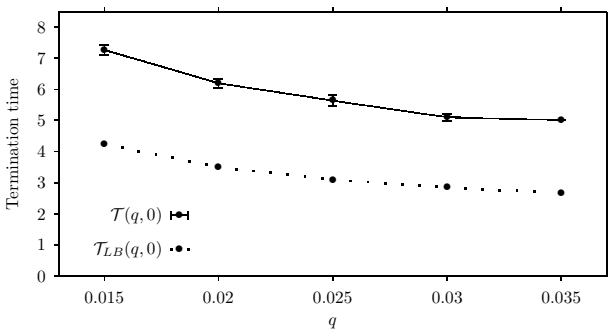


Figure E.9: *Termination time ($\mathcal{T}$) and its analytical lower bound ($\mathcal{T}_{LB}$) as a function of forwarding probability for three ER topologies with connection probability $p = 0.2, 0.3$ and 0.4. 95%-Confidence Intervals are also depicted.*

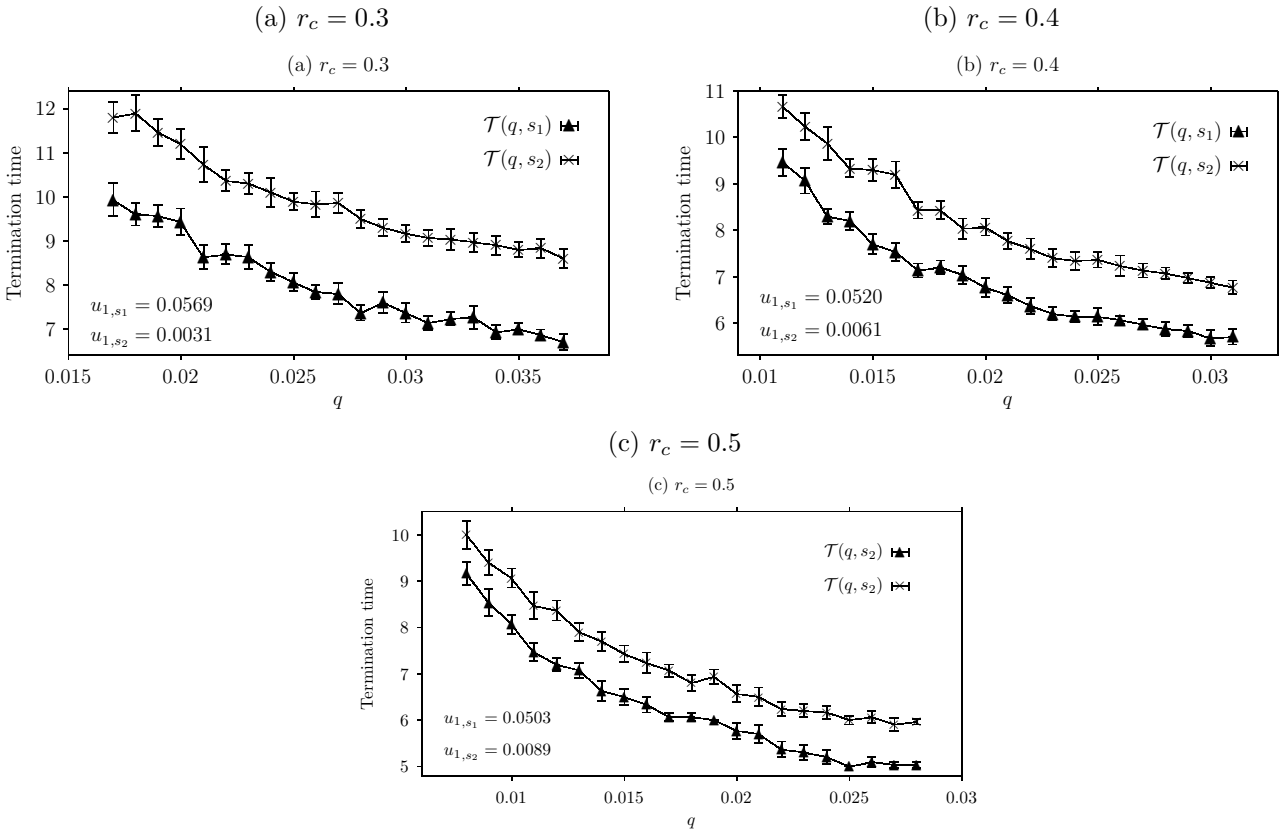


Figure E.10: *Termination time ($\mathcal{T}$) as a function of forwarding probability for three GRG topologies with connectivity radius $r_c = 0.3, 0.4$ and 0.5, for two different initiator nodes s_1, s_2 such that $u_{1,s_1} > u_{1,s_2}$. 95%-Confidence Intervals are also depicted.*

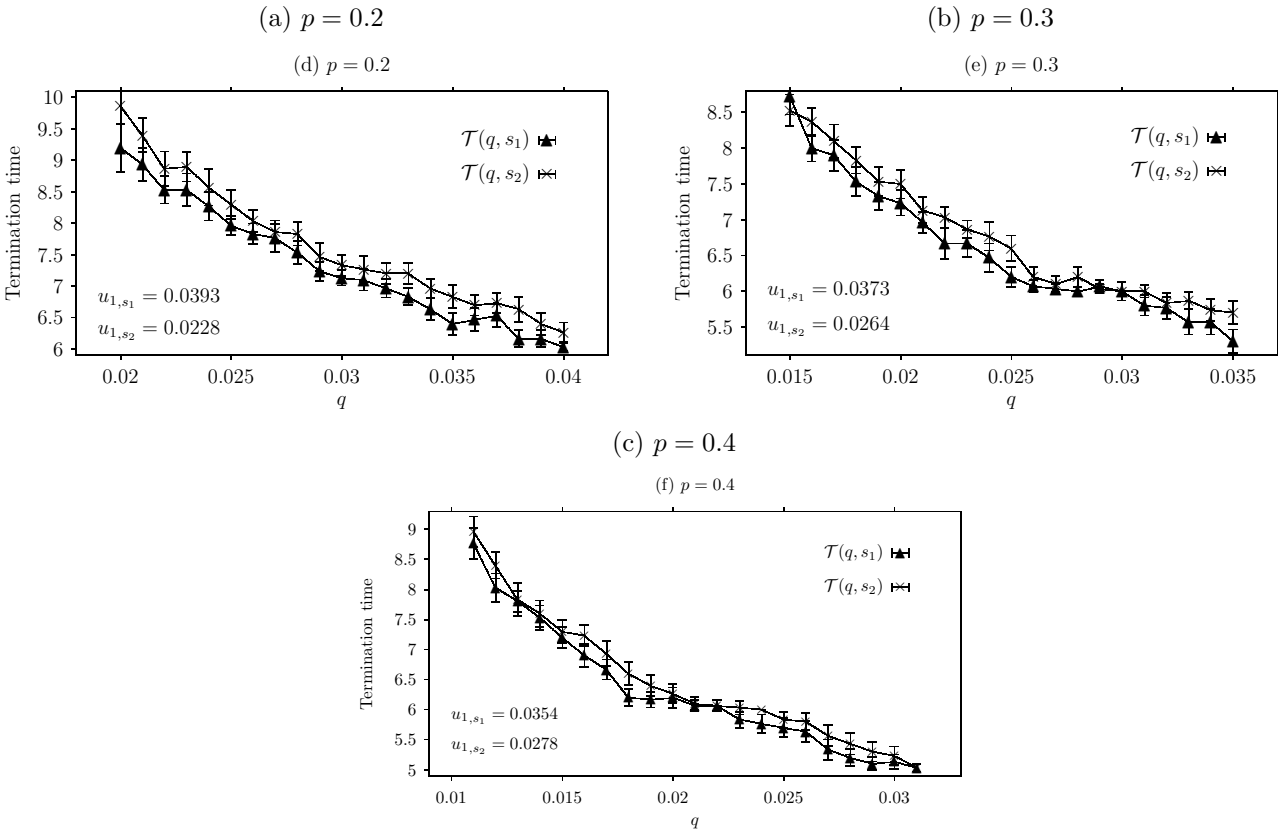


Figure E.11: *Termination time ($\mathcal{T}$) as a function of forwarding probability for three ER topologies with connection probability* $p = 0.2, 0.3$ *and* 0.4*, for two different initiator nodes* s_1*,* s_2 *such that* $u_{1,s_1} > u_{1,s_2}$*. 95%-Confidence Intervals are also depicted.*

E.6.4 m-Probabilistic Flooding

Figures E.12-{a,b,c} depict coverage under mPF as a function of m for various GRG topologies with connectivity radius $r_c = 0.3, 0.4$ and 0.5 respectively. The horizontal and the vertical dotted lines correspond to the 95% of global outreach and $m = \hat{m}$ respectively. It is observed that in all cases that coverage is over 0.95 for $m = \hat{m} = 4$, while for $m < \hat{m}$ coverage is below 0.95. Also, the coverage increment obtained by increasing m over $\hat{m}$ is limited, which supports the analytical finding of Theorem 6. For example, for the case of GRG with connectivity radius $r_c = 0.4$ (i.e., Figure E.12-{b}) the coverage is 0.926, 0.973 and 0.989 for $m = 3$, $m = 4$ and $m = 5$ respectively. Thus, increasing m from three to four, coverage is increased by about 5%, while increasing m from four to five, coverage is increased by about 1%.

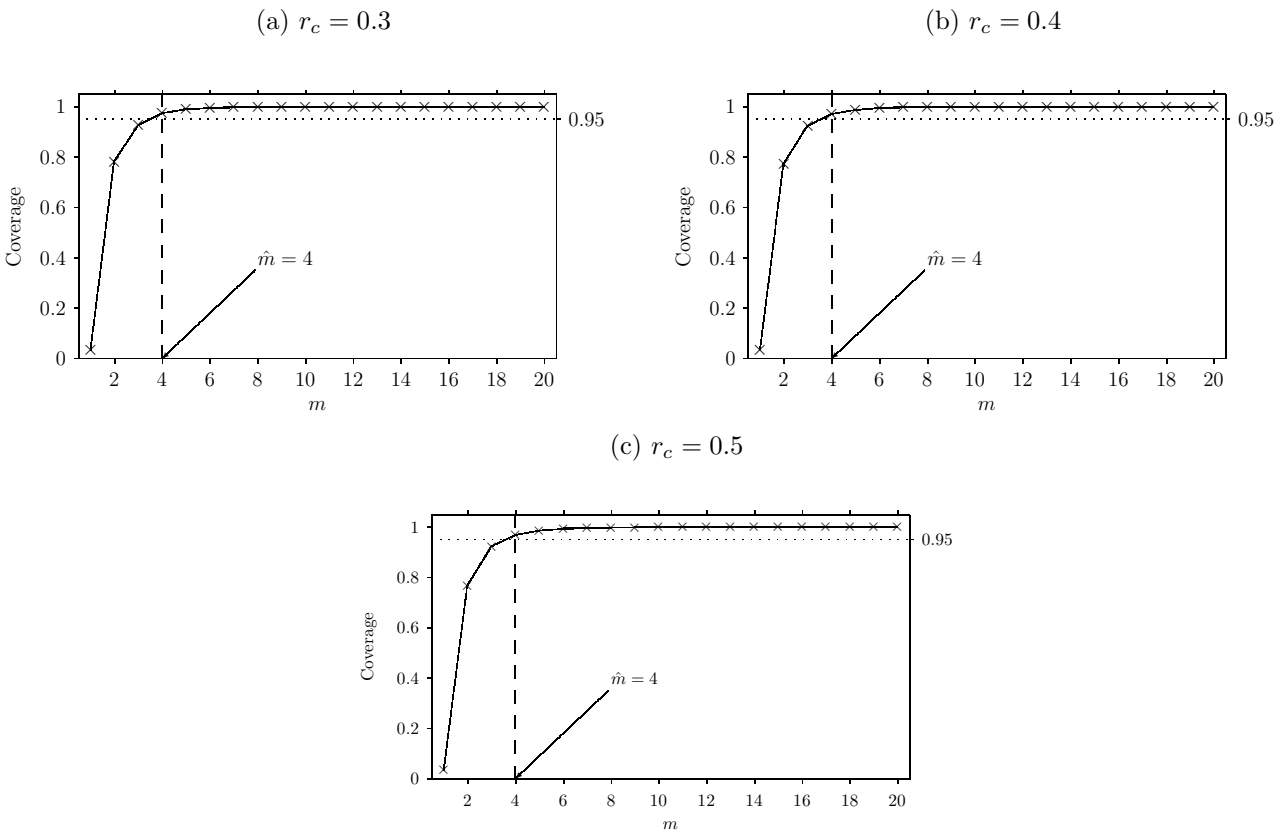
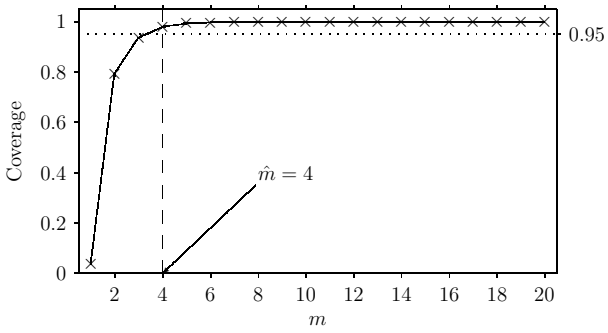


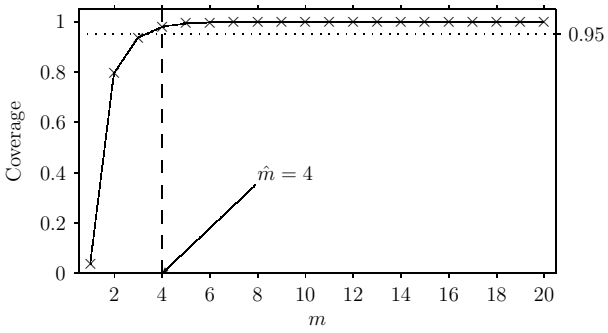
Figure E.12: *Network coverage under mPF as a function of m for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5.

Figures E.13-{a,b,c} depict coverage under mPF as a function of m for various ER topologies with connection probability $p = 0.2, 0.3$ and 0.4 respectively. It is also observed that that in all cases that coverage is over 0.95 for $m = \hat{m} = 4$, while for $m < \hat{m}$ coverage is bellow 0.95.

(a) $p = 0.2$



(b) $p = 0.3$



(c) $p = 0.4$

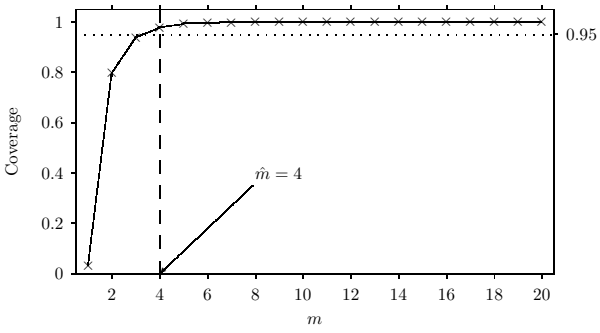
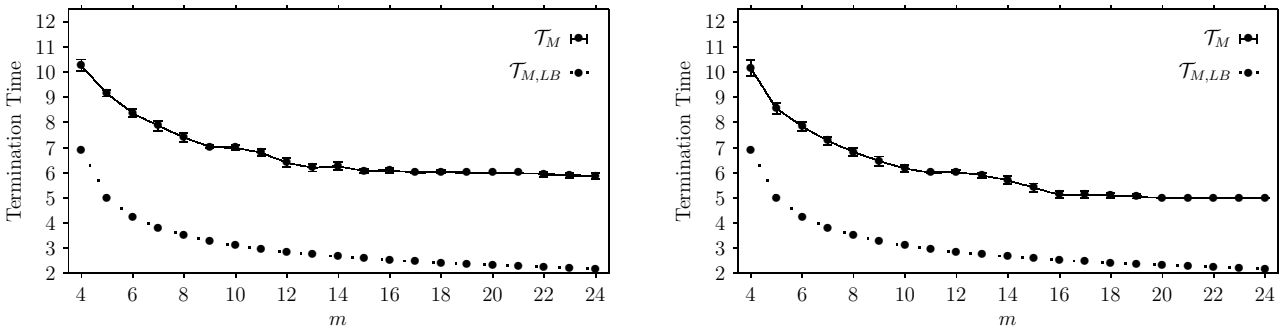


Figure E.13: *Network coverage under mPF as a function of m for three ER topologies with connection probability* $p = 0.2, 0.3$ *and* 0.4.

Figures E.14-{a,b,c} depict simulation results regarding termination time under mPF (i.e., solid line) and its analytical lower bound (i.e., dotted line) obtained by Equation E.12, as a function of m for various GRG topologies with connectivity radius $r_c = 0.3, 0.4$ and 0.5 respectively. As it is expected, termination time (i.e., solid line) decreases as m increases. It is observed that the analytical expression of the termination's time lower bound (i.e., dotted line) is always smaller than the corresponding termination time obtained by the simulations. It is also observed in

all cases that the analytical expression of termination's time lower bound, follows the decrement rate of the termination time. For example, termination time decreases more rapidly for $m \in [4, 8]$ than for $m \in [16, 24]$ and this behavior is also captured by the analytical expression of termination's time lower bound.

(a) $r_c = 0.3$ (b) $r_c = 0.4$



(c) $r_c = 0.5$

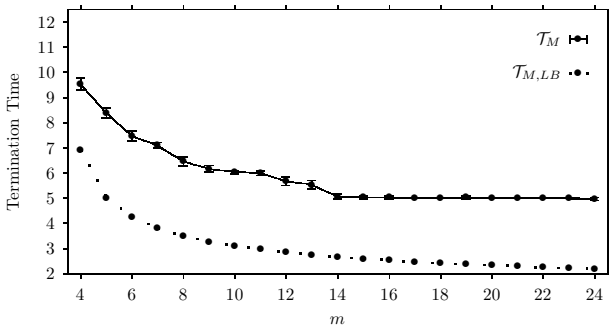
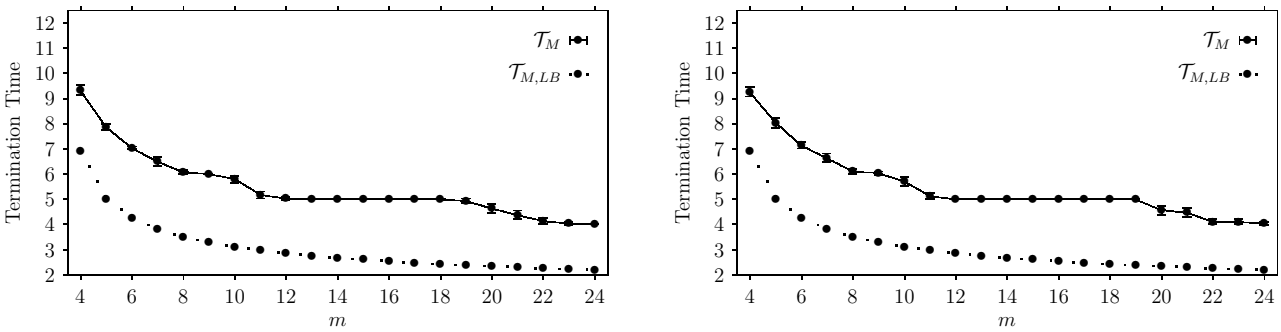


Figure E.14: *Termination time ($\mathcal{T}_M$) under mPF and its analytical lower bound ($\mathcal{T}_{M,LB}$) as a function of m for three GRG topologies with connectivity radius* $r_c = 0.3, 0.4$ *and* 0.5. *95%-Confidence Intervals are also depicted.*

Figures E.15-{a,b,c} depict simulation results regarding termination time under mPF (i.e., solid line) and its analytical lower bound (i.e., dotted line) obtained by Equation E.12, as a function of m for various ER topologies with connection probability $p = 0.2, 0.3$ and 0.4 respectively. It is also observed that the analytical expression of the termination's time lower bound (i.e., dotted line) is always smaller than the corresponding termination time obtained by the simulations and in all cases

the analytical expression of termination's time lower bound, follows the decrement rate of the termination time.

(a) $p = 0.2$ (b) $p = 0.3$



(c) $p = 0.4$

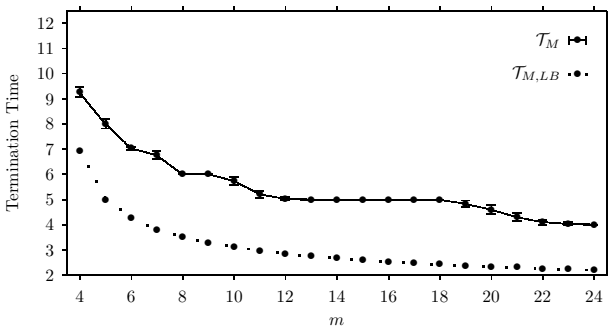


Figure E.15: *Termination time ($\mathcal{T}_M$) under mPF and its analytical lower bound ($\mathcal{T}_{M,LB}$) as a function of m for three ER topologies with connection probability* $p = 0.2, 0.3$ *and* 0.4. *95%-Confidence Intervals are also depicted*

E.7 Discussion

Probabilistic flooding has been studied in this chapter using algebraic graph theory's elements, in order to derive analytical results regarding coverage. In particular, a lower bound of the forwarding probability allowing for global network outreach and a lower bound of the termination time are derived. It is shown that threshold probability is inversely proportional to the largest eigenvalue of the network's adjacency

matrix and the probability of a node to get the information message is proportional to the eigenvector centrality of the initiator node and itself for large values of t. Analytical findings reveal that,

- starting probabilistic flooding from a node with higher eigenvector centrality results in higher network coverage;
- nodes with higher eigenvector centrality are more probable to get covered; and
- termination time lower bound decreases as the initiator's node eigenvector centrality increases.

For the cases of GRG, ER and Regular topologies, lower bounds regarding forwarding probability which are independent of the spectral properties of the particular topologies are derived. Also, a new probabilistic flooding policy called m-Probabilistic Flooding is proposed and it is shown that the requirements for global outreach are independent of the underlying network's spectral properties. The simulation results support the claims of the analysis, giving further insight into the particulars of probabilistic flooding.

Chapter F

Conclusions and Future Work

THE increasing number of WSN and IoT installations and the proliferating demand for high data rates lead to the introduction of the fifth generation mobile phone system (5G). A wide variety of communication characteristics with different requirements regarding cost, data rate, latency, mobility and reliability emerge from the expected increment in human-type and machine-type communications. In such environments, several applications and services that are available in different network locations, regularly need to disseminate information for various purposes like data collection or to reveal useful features (e.g., sink node location, routing information etc). In order to meet the 5G requirements regarding latency, reliability and high network coverage, information dissemination mechanisms should be implemented efficiently in a way that low-latency in guaranteed while not wasting valuable network resources. Blind flooding, is a common mechanism for reliable information dissemination capable of reaching all network nodes in a deterministic manner (i.e., global outreach). Although, it adopts a significant amount of redundant transmissions, leading to the broadcast storm problem thus, undermining the network resources. An increasing volume of research attention has been observed recently regarding probabilistic flooding. Probabilistic flooding can be seen as a suitable alternative for pruning transmissions by employing some fixed probability for message forwarding among neighbor nodes. The basic idea is to employ small values for the forwarding probability in order to exclude any redundant links that would not result in coverage increment.

F.1 Summary and Conclusions

In this thesis, probabilistic flooding is modeled and analyzed using algebraic graph theory elements and an $n \times 1$ vector is introduced such that the element i corresponds to the probability that the node i has received the information message. Subsequently, employing the binomial approximation, the particular vector is further analyzed and an analytical expression regarding coverage is derived in the form of a polynomial. The higher roots of the particular polynomial are shown to be a satisfactory approximation to the forwarding probability. Moreover, the polynomial's roots, are also used to confirm the existing results in the open literature and based on certain observations, a novel algorithm is proposed for the estimation of the threshold probability. The simulation results presented in Section C.4 confirm the analytical findings and it is shown that the analytical expression regarding coverage, captures the probabilistic flooding behavior satisfactorily for small values of the forwarding probability. However, for large values of the forwarding probability, there may be a significant difference between the analytical and the simulation results due to the employment of the binomial approximation. It is also demonstrated that the number of messages under probabilistic flooding may be significantly smaller than in blind flooding, at the expense of increased termination time.

In the sequel, the coverage polynomial is further analyzed and fundamental properties of probabilistic flooding not covered in any previous work are revealed. In particular, the previously mentioned polynomial is initially considered and a new polynomial that depends on the largest eigenvalue and the principal eigenvector of the network's adjacency matrix is introduced. Also, analytical expressions regarding coverage, a lower bound of the forwarding probability allowing for global network outreach, and a lower bound of the termination time are derived. It is shown that threshold probability is inversely proportional to the largest eigenvalue of the network's adjacency matrix and the probability of a node to get the information message is proportional to the eigenvector centrality of the initiator node and itself for large values of time t. Analytical findings reveal that,

- starting probabilistic flooding from a node with higher eigenvector centrality results in higher network coverage;

- nodes with higher eigenvector centrality are more probable to get covered and

- termination's time lower bound decrease as the initiator's node eigenvector centrality increase.

Finally, a new probabilistic flooding policy called m-Probabilistic Flooding is proposed and it is shown that the requirements for global outreach are independent of the underlying network's spectral properties.

F.2 Future Work Directions

Future work directions are related to (i) studying probabilistic flooding for the cases of power law topologies and (ii) applying probabilistic flooding in real WSNs and examine the effectiveness of the analytical findings. As it is mentioned in Section C.3, the employment of the binomial approximation results in the requirement for small values of the forwarding probability for the validation of the analytical model. Note that the benefits obtained by probabilistic flooding are limited when the forwarding probability is set to a large value. However, power law topologies require large values of the forwarding probability for global outreach [179] thus, the analytical model presented in this thesis may need complementary adjustments to capture satisfactorily probabilistic flooding in the particular topologies. Moreover, a possible direction is the implementation of a real WSN consisting of several nodes with embedded processors and wireless interfaces in order to evaluate the effectiveness of the analytical model in a realistic environment.

Bibliography

[1] C. S. Carr, S. D. Crocker, and V. G. Cerf, "Host-host communication protocol in the arpa network," in *Proceedings of the May 5-7, 1970, spring joint computer conference*, pp. 589–597, ACM, 1970.

[2] V. Cerf and R. Kahn, "A protocol for packet network intercommunication," *IEEE Transactions on communications*, vol. 22, no. 5, pp. 637–648, 1974.

[3] K. R. Fall and W. R. Stevens, *TCP/IP illustrated, volume 1: The protocols*. addison-Wesley, 2011.

[4] F. Wilder, *A guide to the TCP/IP Protocol Suite*. Artech House, Inc., 1998.

[5] P. D. Weadon, "The sigsaly story," 2002.

[6] N. Abramson, "Development of the alohanet," *IEEE transactions on Information Theory*, vol. 31, no. 2, pp. 119–123, 1985.

[7] N. Abramson, "The aloha system: another alternative for computer communications," in *Proceedings of the November 17-19, 1970, fall joint computer conference*, pp. 281–285, ACM, 1970.

[8] B. Stavenow, "Throughput-delay characteristics and stability considerations of the access channel in a mobile telephone system," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 12, pp. 105–112, ACM, 1984.

[9] C. Namislo, "Analysis of mobile radio slotted aloha networks," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 4, pp. 583–588, 1984.

[10] W. Crowther, R. Rettberg, D. Walden, S. Ornstein, and F. Heart, "A system for broadcast communication: Reservation-aloha," in *Proc. 6th Hawaii Int. Conf. Syst. Sci*, pp. 596–603, 1973.

[11] B. O'Hara and A. Petrick, *The IEEE 802.11 Handbook: A Designer's Companion*. Standards Information Network IEEE Press, 1999.

[12] J. Khan and A. Khwaja, *Building secure wireless networks with 802.11*. John Wiley & Sons, 2003.

[13] S. Haykin, *Communication systems*. John Wiley & Sons, 2008.

[14] M. Oliver and A. Escudero, "Study of different csma/ca ieee 802.11-based implementations," *EUNICE contribution*, 1999.

[15] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11 b," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2, pp. 836–843, IEEE, 2003.

[16] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell System Technical Journal*, vol. 45, no. 10, pp. 1775–1796, 1966.

[17] S. Weinstein and P. Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE transactions on Communication Technology*, vol. 19, no. 5, pp. 628–634, 1971.

[18] J. Jun, P. Peddabachagari, and M. Sichitiu, "Theoretical maximum throughput of ieee 802.11 and its applications," in *Network Computing and Applications, 2003. NCA 2003. Second IEEE International Symposium on*, pp. 249–256, IEEE, 2003.

[19] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation mimo-ofdm broadband wireless system: design, performance, and field trial results," *IEEE Communications Magazine*, vol. 40, no. 9, pp. 143–149, 2002.

[20] D. Skordoulis, Q. Ni, H.-H. Chen, A. P. Stephens, C. Liu, and A. Jamalipour, "Ieee 802.11 n mac frame aggregation mechanisms for next-generation high-throughput wlans," *IEEE Wireless Communications*, vol. 15, no. 1, 2008.

[21] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE assp magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[22] G. Prakash and S. Pal, "Wimax technology and its applications," *International Journal of Engineering Research and Applications*, vol. 1, no. 2, pp. 327–336, 2012.

[23] C. Weinschenk, "Speeding up wimax," *IT Business Edge*, 2010.

[24] H. Yin and S. Alamouti, "Ofdma: A broadband wireless access technology," in *Sarnoff Symposium, 2006 IEEE*, pp. 1–4, IEEE, 2006.

[25] S. Bluetooth, "How bluetooth technology works," 2006.

[26] R. C. Dixon, *Spread spectrum systems: with commercial applications*, vol. 994. Wiley New York, 1994.

[27] F. Bennett, D. Clarke, J. B. Evans, A. Hopper, A. Jones, and D. Leask, "Piconet: Embedded mobile networking," *IEEE Personal Communications*, vol. 4, no. 5, pp. 8–15, 1997.

[28] P. McDermott-Wells, "Bluetooth scatternet models," *IEEE potentials*, vol. 23, no. 5, pp. 36–39, 2004.

[29] "Traditional profile specifications — bluetooth technology website." URL https://www.bluetooth.com/specifications/profiles-overview, Accessed: 2018-11-10.

[30] E. T. Tiirola, K. P. Pajukoski, and E. Lahetkangas, "Time-division duplexing," Dec. 10 2015. US Patent App. 14/763,060.

[31] J. Haartsen, M. Naghshineh, J. Inouye, O. J. Joeressen, and W. Allen, "Bluetooth: Vision, goals, and architecture," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 2, no. 4, pp. 38–45, 1998.

[32] Q. Zhang, X.-l. Yang, Y.-m. Zhou, L.-r. Wang, and X.-s. Guo, "A wireless solution for greenhouse monitoring and control system based on zigbee technology," *Journal of Zhejiang University-Science A*, vol. 8, no. 10, pp. 1584–1587, 2007.

[33] A. Pascale, M. Nicoli, F. Deflorio, B. Dalla Chiara, and U. Spagnolini, "Wireless sensor networks for traffic management and road safety," *IET Intelligent Transport Systems*, vol. 6, no. 1, pp. 67–77, 2012.

[34] N. A. Somani and Y. Patel, "Zigbee: A low power wireless technology for industrial applications," *International Journal of Control Theory and Computer Modelling (IJCTCM) Vol*, vol. 2, pp. 27–33, 2012.

[35] P. Kinney *et al.*, "Zigbee technology: Wireless control that simply works," in *Communications design conference*, vol. 2, pp. 1–7, 2003.

[36] Z. Specification, "Zigbee alliance," *ZigBee Document 053474r06, Version*, vol. 1, 2006.

[37] H. B. Snderford Jr, "Binary phase shift keying modulation system," June 2 1992. US Patent 5,119,396.

[38] D. Saha and T. G. Birdsall, "Quadrature-quadrature phase-shift keying," *IEEE Transactions on Communications*, vol. 37, no. 5, pp. 437–448, 1989.

[39] S. Faruque, "Introduction to radio frequency," in *Radio Frequency Propagation Made Easy*, pp. 1–18, Springer, 2015.

[40] L. Ashiho, "Mobile technology: Evolution from 1g to 4g," *Electronics for you*, pp. 94–98, 2003.

[41] R. M. Gagliardi, "Frequency-division multiple access," in *Satellite Communications*, pp. 215–250, Springer, 1991.

[42] W. Stallings, *Wireless communications & networks*. Pearson Education India, 2009.

[43] M. Mouly, M.-B. Pautet, and T. Foreword By-Haug, *The GSM system for mobile communications*. Telecom publishing, 1992.

[44] A. Leppisaari and J. Hämäläinen, "General packet radio service," Mar. 11 2003. US Patent 6,532,227.

[45] A. Furuskar, S. Mazur, F. Muller, and H. Olofsson, "Edge: Enhanced data rates for gsm and tdma/136 evolution," *IEEE personal communications*, vol. 6, no. 3, pp. 56–66, 1999.

[46] T. S. Chan, "Time-division multiple access," *Handbook of Computer Networks: LANs, MANs, WANs, the Internet, and Global, Cellular, and Wireless Networks*, vol. 2, pp. 769–778, 2007.

[47] A. J. Viterbi and A. J. Viterbi, *CDMA: principles of spread spectrum communication*, vol. 122. Addison-Wesley Reading, MA, 1995.

[48] S. Alamouti, E. F. Casas, M. Hirano, E. Hoole, M. Jesse, D. G. Michelson, P. Poon, G. J. Veintimilla, and H. Zhang, "Method for frequency division duplex communications," Aug. 3 1999. US Patent 5,933,421.

[49] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *arXiv preprint arXiv:1706.05347*, 2017.

[50] K. Ashton *et al.*, "That 'internet of things' thing," *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009.

[51] R. Minerva, A. Biru, and D. Rotondi, "Towards a definition of the internet of things (iot)," *IEEE Internet Initiative*, vol. 1, pp. 1–86, 2015.

[52] R. Van Kranenburg, *The Internet of Things: A critique of ambient technology and the all-seeing network of RFID*. Institute of Network Cultures, 2008.

[53] G. Intelligence, "Definitive data and analysis for the mobile industry," *GSMAintelligence. com*, 2016.

[54] S. M. Babu, A. J. Lakshmi, and B. T. Rao, "A study on cloud based internet of things: Cloudiot," in *Communication Technologies (GCCT), 2015 Global Conference on*, pp. 60–65, IEEE, 2015.

[55] A. Noronha, R. Moriarty, K. O'Connell, and N. Villa, "Attaining iot value: How to move from connecting things to capturing insights," *Cisco, Tech. Rep.*, 2014.

[56] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[57] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE communications magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[58] V. Jindal, "History and architecture of wireless sensor networks for ubiquitous computing," *History*, vol. 7, no. 2, 2018.

[59] W. K. Seah, Z. A. Eu, and H.-P. Tan, "Wireless sensor networks powered by ambient energy harvesting (wsn-heap)-survey and challenges," in *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on*, pp. 1–5, Ieee, 2009.

[60] G. Koufoudakis, K. Oikonomou, and G. Tsoumanis, "Adapting Probabilistic Flooding in Energy Harvesting Wireless Sensor Networks," *Journal of Sensor and Actuator Networks*, vol. 7, no. 3, p. 39, 2018.

[61] K. Sohraby, D. Minoli, and T. Znati, *Wireless sensor networks: technology, protocols, and applications*. John Wiley & Sons, 2007.

[62] H. M. Ammari, A. Shaout, and F. Mustapha, "Sensing coverage in three-dimensional space: A survey," in *Handbook of Research on Advanced Wireless Sensor Network Applications, Protocols, and Architectures*, pp. 1–28, IGI Global, 2017.

[63] N. McKeown, "Software-defined networking," *INFOCOM keynote talk*, vol. 17, no. 2, pp. 30–32, 2009.

[64] H. Mostafaei and M. Menth, "Software-defined wireless sensor networks: A survey," *Journal of Network and Computer Applications*, 2018.

[65] O. N. Fundation, "Software-defined networking: The new norm for networks," *ONF White Paper*, vol. 2, pp. 2–6, 2012.

[66] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[67] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G mobile and wireless communications technology*. Cambridge University Press, 2016.

[68] R. Tang, J. Zhao, H. Qu, and Z. Zhang, "User-centric joint admission control and resource allocation for 5g d2d extreme mobile broadband: A sequential convex programming approach," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1641–1644, 2017.

[69] E. Dutkiewicz, X. Costa-Perez, I. Z. Kovacs, and M. Mueck, "Massive machine-type communications," *IEEE Network*, vol. 31, no. 6, pp. 6–7, 2017.

[70] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: current status and future perspectives toward 5g systems," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 10–17, 2015.

[71] 3GPP, *Overall description; Stage 2 (Release 12); Technical Specification TS 36.300 V11.7.0, Technical Specification Group Radio Access Network*, 9 2013.

[72] M. METIS, "wireless communications enablers for the twenty-twenty information society, eu 7th framework programme project," tech. rep., ICT-317669-METIS.

[73] Ericsson, "Ericsson mobility report, no. eab-15:010920," tech. rep.

[74] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE wireless communications*, vol. 11, no. 6, pp. 6–28, 2004.

[75] C. N. Ververidis and G. C. Polyzos, "Service discovery for mobile ad hoc networks: a survey of issues and techniques," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 3, 2008.

[76] L. B. Ruiz, J. M. Nogueira, and A. A. Loureiro, "Manna: A management architecture for wireless sensor networks," *IEEE communications Magazine*, vol. 41, no. 2, pp. 116–125, 2003.

[77] X.-L. Zheng and M. Wan, "A survey on data dissemination in wireless sensor networks," *Journal of Computer Science and Technology*, vol. 29, no. 3, pp. 470–486, 2014.

[78] M. U. Arumugam, "Infuse: a tdma based reprogramming service for sensor networks," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pp. 281–282, ACM, 2004.

[79] M. D. Krasniewski, R. K. Panta, S. Bagchi, C.-L. Yang, and W. J. Chappell, "Energy-efficient on-demand reprogramming of large-scale sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 4, no. 1, p. 2, 2008.

[80] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, and J.-P. Sheu, "The broadcast storm problem in a mobile ad hoc network," *Wireless networks*, vol. 8, no. 2-3, pp. 153–167, 2002.

[81] C. R. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE Journal on Selected areas in Communications*, vol. 15, no. 7, pp. 1265–1275, 1997.

[82] R. Ramanathan and M. Steenstrup, "Hierarchically-organized, multihop mobile wireless networks for quality-of-service support," *Mobile networks and applications*, vol. 3, no. 1, pp. 101–119, 1998.

[83] K. Wang, S. A. Ayyash, T. D. Little, and P. Basu, "Attribute-based clustering for information dissemination in wireless sensor networks," in *Proceeding of 2nd annual IEEE communications society conference on sensor and ad hoc communications and networks (SECON'05), Santa Clara, CA*, 2005.

[84] D. Tian, Y. Wang, H. Xia, and F. Cai, "Clustering multi-hop information dissemination method in vehicular ad hoc networks," *IET Intelligent Transport Systems*, vol. 7, no. 4, pp. 464–472, 2013.

[85] M. Jiang, "Cluster based routing protocol (cbrp) functional specification," *Internet draft*, 1998.

[86] X. Zheng, J. Wang, W. Dong, Y. He, and Y. Liu, "Bulk data dissemination in wireless sensor networks: analysis, implications and improvement," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1428–1439, 2016.

[87] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless networks*, vol. 8, no. 2/3, pp. 169–185, 2002.

[88] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, pp. 174–185, ACM, 1999.

[89] A. Segall, "Distributed network protocols," *IEEE transactions on Information Theory*, vol. 29, no. 1, pp. 23–35, 1983.

[90] A. O. Stauffer and V. C. Barbosa, "Probabilistic heuristics for disseminating information in networks," *IEEE/ACM Transactions on Networking*, vol. 15, no. 2, pp. 425–435, 2007.

[91] Y. Mylonas, M. Lestas, A. Pitsillides, P. Ioannou, and V. Papadopoulou, "Speed adaptive probabilistic flooding for vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 1973–1990, 2015.

[92] Q. Li, H. Rong, W. Sun, J. Wang, and J. Li, "A correlation-based energy balanced probabilistic flooding algorithm in wireless sensor network," in *Vehicular Technology Conference (VTC Spring), 2016 IEEE 83rd*, pp. 1–5, IEEE, 2016.

[93] M. Koseoglu, A. Bereketli, I. Yazgi, and B. Yeni, "Probabilistic broadcast for dense auv networks," in *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–5, IEEE, 2016.

[94] S. V. Margariti and V. V. Dimakopoulos, "On probabilistic flooding search over unstructured peer-to-peer networks," *Peer-to-Peer Networking and Applications*, vol. 8, no. 3, pp. 447–458, 2015.

[95] T. Saeed, M. Lestas, and A. Pitsillides, "Adaptive probabilistic flooding for nanonetworks employing molecular communication," in *Telecommunications (ICT), 2016 23rd International Conference on*, pp. 1–5, IEEE, 2016.

[96] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A new communication paradigm," *Computer Networks*, vol. 52, no. 12, pp. 2260–2279, 2008.

[97] S. Crisóstomo, U. Schilcher, C. Bettstetter, and J. Barros, "Probabilistic flooding in stochastic networks: Analysis of global information outreach," *Computer Networks*, vol. 56, no. 1, pp. 142–156, 2012.

[98] C. Betoule, T. Bonald, R. Clavier, D. Rossi, G. Rossini, and G. Thouenon, "Adaptive probabilistic flooding for multipath routing," in *New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on*, pp. 1–6, IEEE, 2012.

[99] K. Oikonomou, D. Kogias, and I. Stavrakakis, "Probabilistic flooding for efficient information dissemination in random graph topologies," *Computer Networks*, vol. 54, no. 10, pp. 1615 – 1629, 2010.

[100] B. Bollobás, *Random Graphs*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2001.

[101] B. Lichtblau and A. Dittrich, "Probabilistic breadth-first search-a method for evaluation of network-wide broadcast protocols," in *New Technologies, Mobility and Security (NTMS), 2014 6th International Conference on*, pp. 1–6, IEEE, 2014.

[102] W. Badreddine and M. Potop-Butucaru, "Peak transmission rate resilient crosslayer broadcast for body area networks," *arXiv preprint arXiv:1702.05031*, 2017.

[103] O. A. Al Tameemi, M. Chatterjee, K. Kwiat, and C. Kamhoua, "Napf: Percolation driven probabilistic flooding for interference limited cognitive radio networks."

[104] R. Gaeta and M. Sereno, "Generalized probabilistic flooding in unstructured peer-to-peer networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 12, pp. 2055–2062, 2011.

[105] F. Banaei-Kashani and C. Shahabi, "Criticality-based analysis and design of unstructured peer-to-peer networks as" complex systems"," in *Cluster Computing and the Grid, 2003. Proceedings. CCGrid 2003. 3rd IEEE/ACM International Symposium on*, pp. 351–358, IEEE, 2003.

[106] D. Tsoumakos and N. Roussopoulos, "Adaptive probabilistic search for peer-to-peer networks," in *Peer-to-Peer Computing, 2003.(P2P 2003). Proceedings. Third International Conference on*, pp. 102–109, IEEE, 2003.

[107] J. Han and Y. Liu, "Mutual anonymity for mobile p2p systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 8, pp. 1009–1019, 2008.

[108] F. Banaei-Kashani, C.-C. Chen, and C. Shahabi, "Wspds: Web services peer-to-peer discovery service," in *Proceedings of the International Conference on Internet Computing*, pp. 733–743, 2004.

[109] W. Jianping, R. Huihui, S. Wei, and L. Qiyue, "A correlation-based coverage-aware and energy-balanced probabilistic flooding algorithm," *International Journal of Sensor Networks*, vol. 25, no. 4, pp. 207–217, 2017.

[110] V. Drabkin, R. Friedman, G. Kliot, and M. Segal, "On reliable dissemination in wireless ad hoc networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 6, pp. 866–882, 2011.

[111] D. Reina, S. Toral, P. Jonhson, and F. Barrero, "Hybrid flooding scheme for mobile ad hoc networks," *IEEE communications letters*, vol. 17, no. 3, pp. 592–595, 2013.

[112] F. Palmieri, "A wave propagation-based adaptive probabilistic broadcast containment strategy for reactive manet routing protocols," *Pervasive and Mobile Computing*, vol. 40, pp. 628–638, 2017.

[113] M. Agarwal, M. Govil, M. Sinha, and A. Jhankal, "Energy conservation by improving flooding mechanism in manet," *Journal of Scientific & Industrial Research*, vol. 76, pp. 408–414, 2017.

[114] A. Xeros, M. Lestas, M. Andreou, and A. Pitsillides, "Adaptive probabilistic flooding for information hovering in vanets," in *Vehicular Networking Conference (VNC), 2010 IEEE*, pp. 239–246, IEEE, 2010.

[115] T. Saeed, Y. Mylonas, A. Pitsillides, V. Papadopoulou, and M. Lestas, "Modeling probabilistic flooding in vanets for optimal rebroadcast probabilities," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2018.

[116] Z. J. Haas, J. Y. Halpern, and L. Li, "Gossip-based ad hoc routing," *IEEE/ACM Transactions on Networking (ToN)*, vol. 14, no. 3, pp. 479–491, 2006.

[117] E. Zanaj, M. Baldi, and F. Chiaraluce, "Efficiency of the gossip algorithm for wireless sensor networks," in *Software, Telecommunications and Computer Networks, 2007. SoftCOM 2007. 15th International Conference on*, pp. 1–5, IEEE, 2007.

[118] D. Chang, K. Cho, N. Choi, Y. Choi, *et al.*, "A probabilistic and opportunistic flooding algorithm in wireless sensor networks," *Computer communications*, vol. 35, no. 4, pp. 500–506, 2012.

[119] J. Cartigny and D. Simplot, "Border node retransmission based probabilistic broadcast protocols in ad-hoc networks," in *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, pp. 10–pp, IEEE, 2003.

[120] Q. Zhang and D. P. Agrawal, "Dynamic probabilistic broadcasting in manets," *Journal of Parallel and Distributed Computing*, vol. 65, no. 2, pp. 220 – 233, 2005.

[121] I. S. Lysiuk and Z. J. Haas, "Controlled gossiping in ad hoc networks," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, pp. 1–6, IEEE, 2010.

[122] P. Kyasanur, R. R. Choudhury, and I. Gupta, "Smart gossip: An adaptive gossip-based broadcasting service for sensor networks," in *Mobile Adhoc and Sensor Systems (MASS), 2006 IEEE International Conference on*, pp. 91–100, IEEE, 2006.

[123] Z. Shi and H. Shen, "Adaptive gossip-based routing algorithm.," in *IPCCC*, pp. 323–324, Citeseer, 2004.

[124] A. Khelil, P. J. Marrón, C. Becker, and K. Rothermel, "Hypergossiping: A generalized broadcast strategy for mobile ad hoc networks," *Ad Hoc Networks*, vol. 5, no. 5, pp. 531–546, 2007.

[125] J.-P. Ryu, M.-S. Kim, S.-H. Hwang, and K.-J. Han, "An adaptive probabilistic broadcast scheme for ad-hoc networks," in *IEEE International Conference on High Speed Networks and Multimedia Communications*, pp. 646–654, Springer, 2004.

[126] R. Beraldi, "The polarized gossip protocol for path discovery in manets," *Ad Hoc Networks*, vol. 6, no. 1, pp. 79 – 91, 2008.

[127] R. Hu, "Efficient probabilistic information broadcast algorithm over random geometric topologies," in *Global Communications Conference (GLOBECOM), 2015 IEEE*, pp. 1–6, IEEE, 2015.

[128] Y. Sasson, D. Cavin, and A. Schiper, "Probabilistic broadcast for flooding in wireless mobile ad hoc networks," in *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, vol. 2, pp. 1124–1130, IEEE, 2003.

[129] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: An eigenvalue viewpoint," in *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pp. 25–34, IEEE, 2003.

[130] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *Computation: the micro and the macro view*, pp. 71–102, World Scientific, 1992.

[131] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.

[132] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 1, pp. 1–14, 2009.

[133] A. Socievole, F. De Rango, C. Scoglio, and P. Van Mieghem, "Assessing network robustness under sis epidemics: The relationship between epidemic threshold and viral conductance," *Computer Networks*, vol. 103, pp. 196–206, 2016.

[134] R. E. Kooij, P. Schumm, C. Scoglio, and M. Youssef, "A new metric for robustness with respect to virus spread," in *International Conference on Research in Networking*, pp. 562–572, Springer, 2009.

[135] M. Penrose, *Random Geometric Graphs*. Oxford studies in probability, Oxford University Press, 2003.

[136] G. Zhou, T. He, S. Krishnamurthy, and J. A. Stankovic, "Impact of radio irregularity on wireless sensor networks," in *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pp. 125–138, ACM, 2004.

[137] J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring, *GNU Octave version 4.2.1 manual: a high-level interactive language for numerical computations*, 2017.

[138] G. Bounova, "Octave networks toolbox," Aug. 2015.

[139] A. E. Brouwer and W. H. Haemers, *Spectra of graphs*. Springer Science & Business Media, 2011.

[140] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[141] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[142] M. E. Newman, "The structure of scientific collaboration networks," *Proceedings of the national academy of sciences*, vol. 98, no. 2, pp. 404–409, 2001.

[143] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England journal of medicine*, vol. 357, no. 4, pp. 370–379, 2007.

[144] M. Ripeanu, "Peer-to-peer architecture case study: Gnutella network," in *Peer-to-Peer Computing, 2001. Proceedings. First International Conference on*, pp. 99–100, IEEE, 2001.

[145] P. Van Mieghem, *Graph spectra for complex networks*. Cambridge University Press, 2010.

[146] R. Balakrishnan and K. Ranganathan, *A textbook of graph theory*. Springer Science & Business Media, 2012.

[147] P. R. Halmos, "What does the spectral theorem say?," *The American Mathematical Monthly*, vol. 70, no. 3, pp. 241–247, 1963.

[148] D. Cvetković, P. Rowlinson, and S. Simić, *An Introduction to the Theory of Graph Spectra*. London Mathematical Society Student Texts, Cambridge University Press, 2009.

[149] R. Grassi, S. Stefani, and A. Torriero, "Some new results on the eigenvector centrality," *Mathematical Sociology*, vol. 31, no. 3, pp. 237–248, 2007.

[150] J. A. Rodríguez, E. Estrada, and A. Gutiérrez, "Functional centrality in graphs," *Linear and Multilinear Algebra*, vol. 55, no. 3, pp. 293–302, 2007.

[151] E. Estrada and J. A. Rodriguez-Velazquez, "Subgraph centrality in complex networks," *Physical Review E*, vol. 71, no. 5, p. 056103, 2005.

[152] U. Brandes and D. Fleischer, "Centrality measures based on current flow," in *Annual symposium on theoretical aspects of computer science*, pp. 533–544, Springer, 2005.

[153] A. Goldman, "Optimal locations for centers in a network," *Transportation Science*, vol. 3, no. 4, pp. 352–360, 1969.

[154] S. L. Hakimi and S. Maheshwari, "Optimum locations of centers in networks," *Operations Research*, vol. 20, no. 5, pp. 967–973, 1972.

[155] S. L. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *Operations research*, vol. 12, no. 3, pp. 450–459, 1964.

[156] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[157] U. Brandes, P. Kenis, and D. Wagner, "Centrality in policy network drawings," in *International Symposium on Graph Drawing*, pp. 250–258, Springer, 1999.

[158] A. Bavelas, "A mathematical model for group structures," *Applied anthropology*, vol. 7, no. 3, pp. 16–30, 1948.

[159] F. Harary, "Status and contrastatus," *Sociometry*, vol. 22, no. 1, pp. 23–43, 1959.

[160] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[161] L. S. Shapley, "A value for n-person games," *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 307–317, 1953.

[162] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein, "The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics," *PLoS computational biology*, vol. 3, no. 4, p. e59, 2007.

[163] J. Yoon, A. Blumer, and K. Lee, "An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality," *Bioinformatics*, vol. 22, no. 24, pp. 3106–3108, 2006.

[164] G. Tsoumanis, K. Oikonomou, G. Koufoudakis, and S. Aïssa, "Energy-Efficient Sink Placement in Wireless Sensor Networks," *Computer Networks*, vol. 141, pp. 166 – 178, 2018.

[165] G. Tsoumanis, K. Oikonomou, S. Aïssa, and I. Stavrakakis, "A Recharging Distance Analysis for Wireless Sensor Networks," *Ad Hoc Networks*, vol. 75-76, pp. 80 – 86, 2018.

[166] E. Kavvadia, S. Sagiadinos, K. Oikonomou, G. Tsioutsiouliklis, and S. Aïssa, "Elastic Virtual Machine Placement in Cloud Computing Network Environments," *Computer Networks*, vol. 93, no. Part 3, pp. 435–447, 2015. Cloud Networking and Communications II.

[167] H. Liu, Q. Deng, S. Tian, X. Peng, and T. Pei, "Recharging schedule for mitigating data loss in wireless rechargeable sensor network," *Sensors*, vol. 18, no. 7, p. 2223, 2018.

[168] J. Roselin, P. Latha, and S. Benitta, "Maximizing the wireless sensor networks lifetime through energy efficient connected coverage," *Ad Hoc Networks*, vol. 62, pp. 1–10, 2017.

[169] D. L. Powers, "Graph partitioning by eigenvectors," *Linear Algebra and its Applications*, vol. 101, pp. 121–133, 1988.

[170] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Journal of mathematical sociology*, vol. 2, no. 1, pp. 113–120, 1972.

[171] L. Lovász and K. Vesztergombi, "Geometric representations of graphs," *Paul Erdos and his Mathematics*, 1999.

[172] K. Oikonomou, G. Koufoudakis, and S. Aïssa, "Probabilistic Flooding Coverage Analysis in Large Scale Wireless Networks," in *2012 19th International Conference on Telecommunications (ICT)*, pp. 1–6, April 2012.

[173] A. Brousseau, *Linear Recursion and Fibonacci Sequences*. Fibonacci Assoc., 1971.

[174] U. Kang, S. Papadimitriou, J. Sun, and H. Tong, "Centralities in large networks: Algorithms and observations," in *Proceedings of the 2011 SIAM International Conference on Data Mining*, pp. 119–130, SIAM, 2011.

[175] M. Nekovee, "Sensor networks on the road: the promises and challenges of vehicular ad hoc networks and grids," in *Workshop on ubiquitous computing and e-Research*, vol. 47, 2005.

[176] S. T. Roweis, "Em algorithms for pca and spca," in *Advances in neural information processing systems*, pp. 626–632, 1998.

[177] Z. Gao, D. Wen, D. Zuo, H. Liu, and X. Yang, "Probability density function of the euclidean distance between node pairs in rectangular random graphs and its applications in manets," *CHINESE JOURNAL OF ELECTRONICS*, vol. 15, no. 3, p. 521, 2006.

[178] G. Koufoudakis, K. Oikonomou, K. Giannakis, and S. Aïssa, "Probabilistic Flooding Coverage Analysis for Efficient Information Dissemination in Wireless Networks," *Computer Networks*, vol. 140, pp. 51 – 61, 2018.

[179] N. Makino, S. Arakawa, and M. Murata, "A flooding method for exchanging routing information in power–law networks," in *Communications, 2005 Asia-Pacific Conference on*, pp. 812–816, IEEE, 2005.

[180] K. Oikonomou, G. Koufoudakis, E. Kavvadia, and V. Chrissikopoulos, "A Wireless Sensor Network Innovative Architecture for Ambient Vibrations Structural Monitoring," *Key Engineering Materials*, vol. 628, 2014.

[181] G. Koufoudakis, N. Skiadopoulos, E. Magkos, and K. Oikonomou, "Synchronization Issues in an Innovative Wireless Sensor Network Architecture Monitoring Ambient Vibrations in Historical Buildings," *Key Engineering Materials*, vol. 628, 2014.

[182] K. Skiadopoulos, A. Tsipis, K. Giannakis, G. Koufoudakis, E. Christopoulou, K. Oikonomou, G. Kormentzas, and I. Stavrakakis, "Synchronization of data measurements in wireless sensor networks for iot applications," *Ad hoc networks*, Under Review.

[183] G. Koufoudakis, K. Oikonomou, S. Aïssa, and I. Stavrakakis, "Analysis of Spectral Properties for Efficient Coverage Under Probabilistic Flooding," in *2018 IEEE 19th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM) (IEEE WoWMoM 2018)*, (Chania, Crete, Greece), Jun 2018.

[184] S. Fanarioti, A. Tsipis, K. Giannakis, G. Koufoudakis, E. Christopoulou, K. Oikonomou, and I. Stavrakakis, "A Proposed Algorithm for Data Measurements Synchronization in Wireless Sensor Networks," in *Second International Balkan Conference on Communications and Networking 2018 (BalkanCom'18)*, (Podgorica, Montenegro), Jun 2018.

[185] G. Tsoumanis, G. Koufoudakis, K. Oikonomou, M. Avlonits, and N. Varotsis, "A Low-Cost Surface Wireless Sensor Network for Pollution Monitoring in the Ionian Sea," in *12th Panhellenic Symposium of Oceanography & Fisheries*, May 2018. Abstract.

[186] E. Kavvadia, G. Koufoudakis, and K. Oikonomou, "Robust Probabilistic Information Dissemination in Energy Harvesting Wireless Sensor Networks," in *2014 13th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)*, pp. 63–70, June 2014.

Appendices

Appendix A

Various Proofs

A.1 Proof of Lemma 4

Since the eigenvectors of $\mathcal{A}$ form a basis on $\mathbb{R}^n$ [145], P_0 can be written as a linear combination of $U_1, U_2, ..., U_n$,

$$P_0 = c_1U_1 + c_2U_2 + ... + c_nU_n. \tag{A.1}$$

By multiplying Equation (A.1) with $\mathcal{A}^t$,

$$\mathcal{A}^tP_0 = c_1\mathcal{A}^tU_1 + c_2\mathcal{A}^tU_2 + ... + c_n\mathcal{A}^tU_n = c_1\lambda_1^tU_1 + c_2\lambda_2^tU_2 + ... + c_n\lambda_n^tU_n,$$

or

$$\mathcal{A}^tP_0 = \lambda_1^t\left(c_1U_1 + c_2(\frac{\lambda_2}{\lambda_1})^tU_2 + ... + c_n(\frac{\lambda_n}{\lambda_1})^tU_n\right).$$

The magnitude for each of the eigenvalues of non-bipartite graphs, is strictly smaller than the largest one, i.e., $\lambda_1 > |\lambda_i|$, or $\frac{|\lambda_i|}{\lambda_1} < 1$, for $2 \leq i \leq n$. Consequently, for large values of t, $(\frac{\lambda_i}{\lambda_1})^t \to 0$ and thus,

$$\mathcal{A}^tP_0 \to \lambda_1^tc_1U_1. \tag{A.2}$$

Given that $\mathcal{A}^tP_0$ is not the zero vector for any $t > 0$, it is concluded that $c_1 \neq 0$. Since $\mathcal{A}$ is a real symmetric matrix, its left eigenvectors are the transpose of the right ones [145], thus

$$U_1^T\mathcal{A} = \lambda_1U_1^T$$

and also

$$U_1^T \mathcal{A}^t = \lambda_1^t U_1^T.$$

Note that U_1 is a normalized eigenvector of $\mathcal{A}$ (i.e., $U_1^T U_1 = 1$). Consequently, by multiplying Equation (A.2) with U_1^T it follows that,

$$U_1^T \mathcal{A}^t P_0 \to U_1^T \lambda_1^t c_1 U_1,$$

or,

$$U_1^T \lambda_1^t P_0 \to U_1^T \lambda_1^t c_1 U_1$$

or,

$$U_1^T \lambda_1^t P_0 \to \lambda_1^t c_1$$

or,

$$\frac{1}{c_1} \lambda_1^t U_1^T P_0 \to \lambda_1^t$$

or,

$$\frac{u_{1,s}}{c_1} \to 1$$

or,

$$c_1 \to u_{1,s}$$

and the proof is completed.

A.2 Proof of Lemma 5

It is observed from Equation (C.4) that the maximum degree of the polynomial is t (for $k = 1$) and the minimum $\lceil \frac{t}{2} \rceil$ (for $k = \lfloor \frac{t}{2} \rfloor + 1$). Given large values of t (such that $\lceil \frac{t}{2} \rceil$ also takes large values), Lemma 4 can be applied to Equation (C.4). Therefore,

$$\mathcal{P}_t(q) = \sum_{k=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{k+1} \binom{t-k+1}{t-2k+2} q^{t-k+1} \mathcal{A}^{t-k+1} P_0 \to$$
$$\to \sum_{k=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{k+1} \binom{t-k+1}{t-2k+2} q^{t-k+1} \lambda_1^{t-k+1} u_{1,s} U_1 = \hat{P}_t(q)$$

and the lemma is proved.

A.3 Proof of Theorem 3

Based on Equation (E.1) it follows that,

$$\Pi_t(q) - q\lambda_1\Pi_{t-1}(q) + q\lambda_1\Pi_{t-2}(q) = 0, \tag{A.3}$$

where $\Pi_0(q) = 1$ and $\Pi_1(q) = q\lambda_1$.
The characteristic equation [173] of Equation (A.3) is

$$\chi^2 - q\lambda_1\chi + q\lambda_1 = 0, \tag{A.4}$$

its roots being,

$$\chi_1 = \frac{q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2} \text{ and } \chi_2 = \frac{q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2}.$$

For $q = \frac{4}{\lambda_1}$, Equation (A.4) has a double root $\chi_1 = \chi_2 = 2$ and the solution of Equation (A.3) is given by,

$$\Pi_t(\frac{4}{\lambda_1}) = k_1 2^t + k_2 t 2^t. \tag{A.5}$$

Given the initial conditions it follows that,

$$\Pi_0(\frac{4}{\lambda_1}) = 1$$

or,

$$k_1 2^0 + k_2 0 2^0 = 1$$

or,

$$k_1 = 1$$

and

$$\Pi_1(\frac{4}{\lambda_1}) = \frac{4}{\lambda_1}\lambda_1$$

or,

$$k_1 2^1 + k_2 2^1 = 4$$

or,

$$k_1 + k_2 = 2$$

or,

$$k_2 = 1,$$

thus Equation (A.5) can be written as,

$$\Pi_t(\frac{4}{\lambda_1}) = (t+1)2^t.$$

For $q > \frac{4}{\lambda_1}$, Equation (A.4) has two distinct roots and the solution of Equation (A.3) is given by,

$$\Pi_t(q) = k_1 \left(\frac{q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2} \right)^t + k_2 \left(\frac{q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2} \right)^t, \qquad (A.6)$$

where k_1 and k_2 can be determined from the initial conditions. Thus for $t = 0$ it is satisfied that, $\Pi_0(q) = 1$, or $k_1 + k_2 = 1$, or $k_2 = 1 - k_1$ and for $t = 1$,

$$\Pi_1(q) = q\lambda_1$$

or ,

$$q\lambda_1 = k_1 \frac{q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2} + (1 - k_1) \frac{q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2}$$

or,

$$q\lambda_1 = \frac{k_1 q\lambda_1 + k_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} + q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1} - k_1 q\lambda_1 + k_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2}$$

or,

$$q\lambda_1 = \frac{2k_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} + q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2}.$$

Consequently,

$$k_1 = \frac{q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}$$

and

$$k_2 = \frac{\sqrt{q^2\lambda_1^2 - 4q\lambda_1} - q\lambda_1}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}$$

By replacing k_1 and k_2 to Equation (A.6), it follows that,

$$\Pi_t(q) = \frac{q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} \left(\frac{q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2} \right)^t + \\ + \frac{\sqrt{q^2\lambda_1^2 - 4q\lambda_1} - q\lambda_1}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} \left(\frac{q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}}{2} \right)^t$$

or,

$$\Pi_t(q) = \frac{\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} - \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}.$$

For $q < \frac{4}{\lambda_1}$, Equation (A.4) has two distinct complex roots and the solution of Equation (A.3) is given by,

$$\Pi_t(q) = \frac{\left(q\lambda_1 + i\sqrt{4q\lambda_1 - q^2\lambda_1^2}\right)^{t+1} - \left(q\lambda_1 - i\sqrt{4q\lambda_1 - q^2\lambda_1^2}\right)^{t+1}}{2^{t+1}i\sqrt{4q\lambda_1 - q^2\lambda_1^2}}. \tag{A.7}$$

Let ϕ such that, $\tan\phi = \frac{\sqrt{4q\lambda_1 - q^2\lambda_1^2}}{q\lambda_1}$ (polar coordinates). Consequently,

$$\Pi_t(q) = \frac{\left[2\sqrt{q\lambda_1}(cos\phi + isin\phi)\right]^{t+1} - \left[4q\lambda_1(cos\phi - isin\phi)\right]^{t+1}}{2^{t+1}i\sqrt{4q\lambda_1 - q^2\lambda_1^2}}$$

or,

$$\Pi_t(q) = \frac{\left[2\sqrt{q\lambda_1}(cos\phi + isin\phi)\right]^{t+1} - \left[2\sqrt{q\lambda_1}\lambda_1(cos\phi - isin\phi)\right]^{t+1}}{2^{t+1}i\sqrt{4q\lambda_1 - q^2\lambda_1^2}}$$

or,

$$\Pi_t(q) = \frac{(2\sqrt{q\lambda_1})^{t+1}\left[cos\left[(t+1)\phi\right] + isin\left[(t+1)\phi\right]\right]^{t+1}}{2^{t+1}i\sqrt{4q\lambda_1 - q^2\lambda_1^2}} -$$
$$\frac{(2\sqrt{q\lambda_1})^{t+1}\left[cos\left[(t+1)\phi\right] - isin\left[(t+1)\phi\right]\right]^{t+1}}{2^{t+1}i\sqrt{4q\lambda_1 - q^2\lambda_1^2}} =$$

or,

$$\Pi_t(q) = \frac{2^{t+1}(q\lambda_1)^{\frac{t+1}{2}}\left[2isin\left[(t+1)\phi\right]\right]}{2^{t+1}i\sqrt{4q\lambda_1 - q^2\lambda_1^2}}$$

or,

$$\Pi_t(q) = \frac{2(q\lambda_1)^{\frac{t+1}{2}}sin\left[(t+1)\phi\right]}{\sqrt{4q\lambda_1 - q^2\lambda_1^2}}$$

and the theorem is proved.

A.4 Proof of Corollary 3

For $q \to \frac{4}{\lambda_1}^+$, L'Hôpital's rule implies that,

$$\lim_{q\to\frac{4}{\lambda_1}^+} \Pi_t(q) = \lim_{q\to\frac{4}{\lambda_1}^+} \frac{\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} - \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^+} \Pi_t(q) = \lim_{q\to\frac{4}{\lambda_1}^+} \frac{\left[\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} - \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}\right]'}{\left[2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right]'}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^+} \Pi_t(q) = \lim_{q\to\frac{4}{\lambda_1}^+} \frac{\left[\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}\right]' - \left[\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}\right]'}{\left[2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right]'} \tag{A.8}$$

The derivative of the left part of the numerator is,

$$\left[(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = (t+1)\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)'$$

or,

$$\left[(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = (t+1)\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(\lambda_1 + \frac{2q\lambda_1^2 - 4\lambda_1}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}\right)$$

or,

$$\left[(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = (t+1)\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(\frac{2\lambda_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} + 2q\lambda_1^2 - 4\lambda_1}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}\right)$$

or,

$$\left[(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = \frac{(t+1)\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(2\lambda_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} + 2q\lambda_1^2 - 4\lambda_1\right)}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} \tag{A.9}$$

The derivative of the right part of the numerator is,

$$\left[(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = (t+1)\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)'$$

or,

$$\left[(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = (t+1)\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(\lambda_1 - \frac{2q\lambda_1^2 - 4\lambda_1}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}\right)$$

$$\left[(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = (t+1)\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(\frac{2\lambda_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} - 2q\lambda_1^2 + 4\lambda_1}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}}\right)$$

or,

$$\left[(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1})^{t+1}\right]' = \frac{(t+1)\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(2\lambda_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} - 2q\lambda_1^2 + 4\lambda_1\right)}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} \tag{A.10}$$

The derivative of the denominator is,

$$\left[2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right]' = 2^{t+1}\frac{2q\lambda_1^2 - 4\lambda_1}{2\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} \tag{A.11}$$

It is observed that Equation (A.9), Equation (A.10) and Equation (A.11) have common denominator, thus after replacement to Equation (A.8) it follows that,

$$\lim_{q\to\frac{4}{\lambda_1}^+}\Pi_t(q) = \lim_{q\to\frac{4}{\lambda_1}^+}\left[\frac{(t+1)\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(2\lambda_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} + 2q\lambda_1^2 - 4\lambda_1\right)}{2^{t+1}(2q\lambda_1^2 - 4\lambda_1)} - \frac{(t+1)\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^t \left(2\lambda_1\sqrt{q^2\lambda_1^2 - 4q\lambda_1} - 2q\lambda_1^2 + 4\lambda_1\right)}{2^{t+1}(2q\lambda_1^2 - 4\lambda_1)}\right]$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^+}\Pi_t(q) = \frac{(t+1)\,(4+0)^t\,(0+8\lambda_1 - 4\lambda_1)}{2^{t+1}(8\lambda_1 - 4\lambda_1)} - \frac{(t+1)\,(4-0)^t\,(0-8\lambda_1 + 4\lambda_1)}{2^{t+1}(8\lambda_1 - 4\lambda_1)}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^+}\Pi_t(q) = \frac{(t+1)4^t 4\lambda_1}{2^{t+1}4\lambda_1} - \frac{(t+1)4^t(-4\lambda_1)}{2^{t+1}4\lambda_1}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^+}\Pi_t(q) = 2\frac{(t+1)4^t 4\lambda_1}{2^{t+1}4\lambda_1}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^+}\Pi_t(q)=\frac{(t+1)4^t4\lambda_1}{2^t4\lambda_1}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^+}\Pi_t(q)=(t+1)2^t=\Pi_t(\frac{4}{\lambda_1})$$

For $q\to\frac{4}{\lambda_1}^-$ it is,

$$\Pi_t(q)=\frac{2(q\lambda_1)^{\frac{t+1}{2}}sin\left[(t+1)\phi\right]}{\sqrt{4q\lambda_1-(q\lambda_1)^2}}$$

where,

$$\phi=\arctan\frac{\sqrt{4q\lambda_1-(q\lambda_1)^2}}{q\lambda_1}$$

or,

$$\sqrt{4q\lambda_1-(q\lambda_1)^2}=q\lambda_1\tan\phi.$$

Thus,

$$\Pi_t(q)=\frac{2(q\lambda_1)^{\frac{t+1}{2}}sin\left[(t+1)\phi\right]}{q\lambda_1\tan\phi}$$

or,

$$\Pi_t(q)=2(q\lambda_1)^{\frac{t-1}{2}}\frac{sin\left[(t+1)\phi\right]}{\tan\phi}$$

Consequently,

$$\lim_{q\to\frac{4}{\lambda_1}^-}\Pi_t(q)=\lim_{q\to\frac{4}{\lambda_1}^-}2(q\lambda_1)^{\frac{t-1}{2}}\frac{sin\left[(t+1)\phi\right]}{\tan\phi}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^-}\Pi_t(q)=\lim_{q\to\frac{4}{\lambda_1}^-}2\cdot4^{\frac{t-1}{2}}\frac{sin\left[(t+1)\phi\right]}{\tan\phi}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^-}\Pi_t(q)=\lim_{q\to\frac{4}{\lambda_1}^-}2\cdot2^{t-1}\frac{sin\left[(t+1)\phi\right]}{\tan\phi}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^-}\Pi_t(q)=2^t\lim_{q\to\frac{4}{\lambda_1}^-}\frac{sin\left[(t+1)\phi\right]}{\tan\phi}.$$

Note that for $q \to \frac{4}{\lambda_1}^-$ it is $\frac{\sqrt{4q\lambda_1-(q\lambda_1)^2}}{q\lambda_1} \to 0^+$ and since arctan is an increasing function it follows that,

$$\arctan \frac{\sqrt{4q\lambda_1-(q\lambda_1)^2}}{q\lambda_1} \to 0^+ \text{ or,} \phi \to 0^+$$

thus,

$$\lim_{q\to\frac{4}{\lambda_1}^-} \Pi_t(q) = 2^t \lim_{\phi\to 0^+} \frac{sin\left[(t+1)\phi\right]}{\tan\phi}.$$

L'Hôpital's rule implies that,

$$\lim_{q\to\frac{4}{\lambda_1}^-} \Pi_t(q) = 2^t \lim_{\phi\to 0^+} \frac{\left[sin\left[(t+1)\phi\right]\right]'}{\left[\tan\phi\right]'}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^-} \Pi_t(q) = 2^t \lim_{\phi\to 0^+} \frac{(t+1)\cos\left[(t+1)\phi\right]}{1/\cos^2\phi}$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^-} \Pi_t(q) = 2^t(t+1) \lim_{\phi\to 0^+} \left[\cos^2\phi\cos\left[(t+1)\phi\right]\right]$$

or,

$$\lim_{q\to\frac{4}{\lambda_1}^-} \Pi_t(q) = 2^t(t+1) = \Pi_t(\frac{4}{\lambda_1})$$

which completes the proof.

A.5 Proof of Corollary 4

For $q = \frac{4}{\lambda_1}$, it is satisfied that,

$$\Pi_t(q) = (t+1)2^t \text{ and } \Pi_{t+1}(q) = (t+2)2^{t+1}.$$

Obviously,

$$t+2 > t+1 \text{ and } 2^{t+1} > 2^t$$

thus,

$$\Pi_{t+1}(q) > \Pi_t(q)$$

For $q > \frac{4}{\lambda_1}$, it is satisfied that,

$$q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1} > q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1} > 0$$

Consequently,

$$\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} > \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} > 0 \tag{A.12}$$

Since $q\lambda_1 > 4$, it follows that,

$$q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1} > 4.$$

Thus,

$$\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right) - 2 > 0. \tag{A.13}$$

Obviously the following inequality also holds,

$$\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right) - 2 > \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right) - 2. \tag{A.14}$$

By multiplying inequality (A.12) with inequality (A.14) and with respect to inequality (A.13) it follows that,

$$\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}\left[\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right) - 2\right] > \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}\left[\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right) - 2\right]$$

or,

$$\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+2} - 2\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} > \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+2} - 2\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}$$

or,

$$\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+2} - \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+2} > 2\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} - 2\left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}$$

and by multiplying with $\frac{1}{2^{t+2}\sqrt{q^2\lambda_1^2-4q\lambda_1}}$ it follows that,

$$\frac{\left(q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+2}-\left(q\lambda_1-\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+2}}{2^{t+2}\sqrt{q^2\lambda_1^2-4q\lambda_1}} > \frac{2\left(q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1}-2\left(q\lambda_1-\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1}}{2^{t+2}\sqrt{q^2\lambda_1^2-4q\lambda_1}}$$

or,

$$\frac{\left(q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+2}-\left(q\lambda_1-\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+2}}{2^{t+2}\sqrt{q^2\lambda_1^2-4q\lambda_1}} > \frac{\left(q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1}-\left(q\lambda_1-\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2-4q\lambda_1}}$$

or,

$$\Pi_{t+1}(q) > \Pi_t(q),$$

which completes the proof.

A.6 Proof of Theorem 4

For $q > \frac{4}{\lambda_1}$, it is satisfied that,

$$\Pi_t(q) = \frac{\left(q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1}-\left(q\lambda_1-\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2-4q\lambda_1}},$$

which is strictly positive for any $t > 0$ since,

$$\sqrt{q^2\lambda_1^2-4q\lambda_1} > 0 \text{ and } \left(q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1} > \left(q\lambda_1-\sqrt{q^2\lambda_1^2-4q\lambda_1}\right)^{t+1}.$$

For $q = \frac{4}{\lambda_1}$, it is satisfied that

$$\Pi_t(\frac{4}{\lambda_1}) = (t+1)2^t$$

which is strictly positive for any $t > 0$.
For $q < \frac{4}{\lambda_1}$, it is satisfied that

$$\Pi_t(q) = \frac{2(q\lambda_1)^{\frac{t+1}{2}} sin\left[(t+1)\phi\right]}{\sqrt{4q\lambda_1 - q^2\lambda_1^2}},$$

which is negative for all positive values of t that satisfy

$$(2m-1)\pi < (t+1)\phi < 2m\pi$$

or,

$$\frac{(2m-1)\pi}{\phi} - 1 < t < \frac{2m\pi}{\phi} - 1, m \in \mathbb{N}_{>0}.$$

In order to complete the proof, it is enough to shown that there is at least one integer value in the range $\left(\frac{(2m-1)\pi}{\phi} - 1, \frac{2m\pi}{\phi} - 1\right)$ or,

$$\left(\frac{2m\pi}{\phi} - 1\right) - \left(\frac{(2m-1)\pi}{\phi} - 1\right) > 1.$$

Note that,

$$\phi = \arctan\frac{\sqrt{4q\lambda_1 - (q\lambda_1)^2}}{q\lambda_1} < \frac{\pi}{2}.$$

Consequently,

$$\left(\frac{2m\pi}{\phi} - 1\right) - \left(\frac{(2m-1)\pi}{\phi} - 1\right) = \frac{\pi}{\phi} > \frac{\pi}{\frac{\pi}{2}} = 2.$$

Thus, based on Corollary 4, $q = \dfrac{4}{\lambda_1}$ is the minimum forwarding probability such that $\Pi_{t+1}(q) > \Pi_t(q) > 0$ for all $t > 0$.

A.7 Proof of Corollary 5

For any value q' of the forwarding probability, a necessary condition for global outreach is that $\hat{C}_{t+1,s}(q') > \hat{C}_t(q') > 0$ to be satisfied, for all $t > 0$. Consequently, the lower bound regarding the forwarding probability $\hat{q}$, is the minimum value of q such that, $\hat{C}_{t+1}(\hat{q}) > \hat{C}_t(\hat{q}) > 0$. Since $\hat{C}_t(q)$ is proportional to $\Pi_t(q)$ for large values of t, based on Theorem 4, it is concluded that $\hat{q} = \dfrac{4}{\lambda_1}$.

A.8 The CDF of the Euclidian distance in GRG topologies

Let u, v be two arbitrary nodes in a GRG topology in $[0,1] \times [0,1]$ plane with connectivity radius r_c. Also let (x_u, y_u) and (x_v, y_v) be the coordinates of u and v respectively. x_u, y_u, x_v and y_v are random variables independent and identically distributed (iid) in $[0,1]$. The probability density function (PDF) of x_u and x_v is given by,

$$f_{X_u}(x_u) = f_{X_v}(x_v) = f_X(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \in (-\infty, 0) \cup (1, +\infty) \end{cases} \tag{A.15}$$

The PDF of $-x_v$ is given by :

$$f_{-X_v}(x_v) = \begin{cases} 1 & -x_v \in [0,1] \\ 0 & -x_v \in (-\infty, 0) \cup (1, +\infty) \end{cases} = \begin{cases} 1 & x_v \in [-1,0] \\ 0 & x_v \in (-\infty, -1) \cup (0, +\infty) \end{cases} \tag{A.16}$$

Let $\Delta_x = x_u - x_v = x_u + (-x_v)$ and $\Delta_y = y_u - y_v = y_u + (-y_v)$. Consequently, the Euclidian distance of u and v is given by, $\sqrt{\Delta_x^2 + \Delta_y^2}$. The PDF of Δ_x (i.e., $f_{\Delta_x}(t)$) is obtained by the convolution of f_{X_u} and f_{-X_v} as following,

$$f_{\Delta_x}(t) = (f_{X_u} * f_{-X_v})(t) = \int_{\xi=-\infty}^{\xi=+\infty} f_{X_u}(\xi) f_{-X_v}(t-\xi) \, \mathrm{d}\xi$$

or,

$$f_{\Delta_x}(t) = \int_{\xi=-\infty}^{\xi=0} f_{X_u}(\xi) f_{-X_v}(t-\xi) \, \mathrm{d}\xi + \int_{\xi=0}^{\xi=1} f_{X_u}(\xi) f_{-X_v}(t-\xi) \, \mathrm{d}\xi + \int_{\xi=1}^{\xi=+\infty} f_{X_u}(\xi) f_{-X_v}(t-\xi) \, \mathrm{d}\xi \tag{A.17}$$

Since $f_{X_u}(\xi) = 0$ for $\xi \in (-\infty, 0) \cup (1, +\infty)$ it follows that,

$$f_{\Delta_x}(t) = \int_{\xi=0}^{\xi=1} f_{X_u}(\xi) f_{-X_v}(t-\xi) \, \mathrm{d}\xi \tag{A.18}$$

For $t < -1 : \; t - \xi < -1 - \xi$ consequently, $t - \xi < -1$ (since $0 \leq \xi \leq 1$) or, $f_{-X_v}(t-\xi) = 0$ or, $f_{\Delta_x}(t) = 0$.

For $t > 1 : t-\xi > 1-\xi$ consequently, $t-\xi > 0$ (since $0 \leq \xi \leq 1$) or, $f_{-X_v}(t-\xi) = 0$ or, $f_{\Delta_x}(t) = 0$.
For $-1 \leq t \leq 0$ equation A.18 becomes

$$f_{\Delta_x}(t) = \int_{\xi=0}^{\xi=t+1} f_{X_u}(\xi) f_{-X_v}(t-\xi)\,\mathrm{d}\xi + \int_{\xi=t+1}^{\xi=1} f_{X_u}(\xi) f_{-X_v}(t-\xi)\,\mathrm{d}\xi$$

Since $f_{-X_v}(t-\xi) = 0$ for $\xi \in [t+1, 1]$ it follows that,

$$f_{\Delta_x}(t) = \int_{\xi=0}^{\xi=t+1} f_{X_u}(\xi) f_{-X_v}(t-\xi)\,\mathrm{d}\xi$$

thus,

$$f_{\Delta_x}(t) = \int_{\xi=0}^{\xi=t+1} \mathrm{d}\xi \Rightarrow f_{\Delta_x}(t) = t+1$$

For $0 < t \leq 1$ equation A.18 becomes

$$f_{\Delta_x}(t) = \int_{\xi=0}^{\xi=t} f_{X_u}(\xi) f_{-X_v}(t-\xi)\,\mathrm{d}\xi + \int_{\xi=t}^{\xi=1} f_{X_u}(\xi) f_{-X_v}(t-\xi)\,\mathrm{d}\xi$$

Since $f_{-X_v}(t-\xi) = 0$ for $\xi \in [0, t]$ it follows that,

$$f_{\Delta_x}(t) = \int_{\xi=t}^{\xi=1} f_{X_u}(\xi) f_{-X_v}(t-\xi)\,\mathrm{d}\xi$$

Thus,

$$f_{\Delta_x}(t) = \int_{\xi=t}^{\xi=1} \mathrm{d}\xi \Rightarrow f_{\Delta_x(t)} = 1-t$$

Consequently,

$$f_{\Delta_x}(t) = \begin{cases} 0 & t \in (-\infty, -1) \cup (1, +\infty) \\ 1+t & t \in [-1, 0] \\ 1-t & t \in (0, 1] \end{cases} \tag{A.19}$$

Let $f_{\Delta_x^2}(t)$ and $F_{\Delta_x^2}(t)$ be the PDF and the CDF of Δ_x^2 respectively. Since $\Delta_x^2 \geq 0$, it is clear that

$$F_{\Delta_x^2}(t) = 0 \text{ for } t \leq 0$$

For $t > 0$,

$$F_{\Delta_x^2}(t) = \Pr[\Delta_x^2 \leq t] = \Pr[-\sqrt{t} \leq \Delta_x \leq \sqrt{t}]$$

or,

$$F_{\Delta_x^2}(t) = F_{\Delta_x}\left(\sqrt{t}\right) - F_{\Delta_x}\left(-\sqrt{t}\right)$$

or,

$$\left(F_{\Delta_x^2}(t)\right)' = \left(F_{\Delta_x}\left(\sqrt{t}\right)\right)' - \left(F_{\Delta_x}\left(-\sqrt{t}\right)\right)'$$

or,

$$f_{\Delta_x^2}(t) = \frac{1}{2\sqrt{t}}\left[f_{\Delta_x}\left(\sqrt{t}\right) + f_{\Delta_x}\left(-\sqrt{t}\right)\right] \tag{A.20}$$

and by replacing f_{Δ_x} from Equation (A.19) it follows that,

$$f_{\Delta_x^2}(t) = \begin{cases} \frac{1}{2\sqrt{t}}\left(1 - \sqrt{t} + 1 - \sqrt{t}\right) & \sqrt{t} \in (0, 1] \\ 0 & \sqrt{t} \in (1, +\infty) \end{cases}$$

or,

$$f_{\Delta_x^2}(t) = \begin{cases} \frac{1}{\sqrt{t}} - 1 & t \in (0, 1] \\ 0 & t \in (1, +\infty) \end{cases} \tag{A.21}$$

Let $\Delta_y = y_u - y_v = y_u + (-y_v)$ and also let $f_{\Delta_y^2}(t)$ and $F_{\Delta_y^2}(t)$ denote the PDF and the CDF of Δ_y^2 respectively. Similarly,

$$f_{\Delta_y^2}(t) = \begin{cases} \frac{1}{\sqrt{t}} - 1 & t \in (0, 1] \\ 0 & t \in (1, +\infty) \end{cases} \tag{A.22}$$

Let $f_{\Delta_x^2+\Delta_y^2}(t)$ denote the PDF of $\Delta_x^2 + \Delta_y^2$ or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \left(f_{\Delta_x^2} * f_{\Delta_y^2}\right)(t) = \int_{\xi=-\infty}^{\xi=+\infty} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\, \mathrm{d}\xi$$

Since $f_{\Delta_x^2}(\xi) = 0$ for $\xi \in (-\infty, 0] \cup (1, +\infty)$ it follows that,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=0}^{\xi=1} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\,\mathrm{d}\xi$$

For $t<0$: It is $t-\xi<0$ consequently,

$$f_{\Delta_x^2+\Delta_y^2}(t) = 0$$

For $t>2$: It is $t-\xi>1$ consequently,

$$f_{\Delta_x^2+\Delta_y^2}(t) = 0$$

For $0 \le t \le 1$

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=0}^{\xi=t} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\,\mathrm{d}\xi + \int_{\xi=t}^{\xi=1} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\,\mathrm{d}\xi$$

Since $t-\xi<0$ for $\xi \in [t,1]$ it follows that,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=0}^{\xi=t} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\,\mathrm{d}\xi$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=0}^{\xi=t} \left(\frac{1}{\sqrt{\xi}}-1\right)\left(\frac{1}{\sqrt{t-\xi}}-1\right)\mathrm{d}\xi =$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=0}^{\xi=t} \frac{1}{\sqrt{\xi\,(t-\xi)}}\mathrm{d}\xi - \int_{\xi=0}^{\xi=t} \frac{1}{\sqrt{\xi}}\mathrm{d}\xi - \int_{\xi=0}^{\xi=t} \frac{1}{\sqrt{t-\xi}}\mathrm{d}\xi + \int_{\xi=0}^{\xi=t} \mathrm{d}\xi$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=0}^{\xi=t} \left(\arcsin\left(\frac{2\xi-t}{t}\right)\right)' \mathrm{d}\xi - \int_{\xi=0}^{\xi=t} \left(2\sqrt{\xi}\right)' \mathrm{d}\xi - \int_{\xi=0}^{\xi=t} \left(-2\sqrt{t-\xi}\right)' \mathrm{d}\xi + \int_{\xi=0}^{\xi=t} (\xi)'\,\mathrm{d}\xi$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \arcsin\left(\frac{2\xi-t}{t}\right)\Bigg|_{\xi=0}^{\xi=t} - 2\sqrt{\xi}\Big|_{\xi=0}^{\xi=t} + 2\sqrt{t-\xi}\Big|_{\xi=0}^{\xi=t} + \xi|_{\xi=0}^{\xi=t}$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) - 2\sqrt{t} - 2\sqrt{t} + t$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \pi - 4\sqrt{t} + t$$

For $1 < t \leq 2$,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=0}^{\xi=t-1} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\,\mathrm{d}\xi + \int_{\xi=t-1}^{\xi=1} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\,\mathrm{d}\xi$$

Since $f_{\Delta_y^2}(t-\xi) = 0$ for $\xi \leq t-1$ it follows that,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=t-1}^{\xi=1} f_{\Delta_x^2}(\xi)\, f_{\Delta_y^2}(t-\xi)\,\mathrm{d}\xi$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \int_{\xi=t-1}^{\xi=1} \left(\frac{1}{\sqrt{\xi}} - 1\right)\left(\frac{1}{\sqrt{t-\xi}} - 1\right)\mathrm{d}\xi$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \arcsin\left(\frac{2\xi - t}{t}\right)\Big|_{\xi=t-1}^{\xi=1} - 2\sqrt{\xi}\Big|_{\xi=t-1}^{\xi=1} + 2\sqrt{t-\xi}\Big|_{\xi=t-1}^{\xi=1} + \xi|_{\xi=t-1}^{\xi=1}$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \arcsin\left(\frac{2-t}{t}\right) - \arcsin\left(\frac{t-2}{t}\right) - 2 + 2\sqrt{t-1} + 2\sqrt{t-1} - 2 + 1 - t + 1$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \arcsin\left(\frac{2-t}{t}\right) - \arcsin\left(\frac{t-2}{t}\right) + 4\sqrt{t-1} - t - 2$$

Consequently, the PDF of $\Delta_x^2 + \Delta_y^2$ is given by,

$$f_{\Delta_x^2+\Delta_y^2}(t) = \begin{cases} 0 & t \in (-\infty, 0) \cup (2, +\infty) \\ \pi - 4\sqrt{t} + t & t \in [0, 1] \\ \arcsin\left(\frac{2-t}{t}\right) - \arcsin\left(\frac{t-2}{t}\right) + 4\sqrt{t-1} - t - 2 & t \in (1, 2] \end{cases} \tag{A.23}$$

Let $F_{\Delta_x^2+\Delta_y^2}(t)$ denote the CDF of $\Delta_x^2+\Delta_y^2$. Consequently,

$$F_{\Delta_x^2+\Delta_y^2}(t)=\Pr\left[\Delta_x^2+\Delta_y^2\leq t\right]$$

or,

$$F_{\Delta_x^2+\Delta_y^2}(t)=\Pr\left[\sqrt{\Delta_x^2+\Delta_y^2}\leq\sqrt{t}\right]$$

or,

$$F_{\Delta_x^2+\Delta_y^2}(t)=F_{\sqrt{\Delta_x^2+\Delta_y^2}}\left(\sqrt{t}\right)$$

By differentiating this equation it follows that,

$$\left(F_{\Delta_x^2+\Delta_y^2}(t)\right)'=\left(F_{\sqrt{\Delta_x^2+\Delta_y^2}}\left(\sqrt{t}\right)\right)'$$

or,

$$f_{\Delta_x^2+\Delta_y^2}(t)=\frac{1}{2\sqrt{t}}f_{\sqrt{\Delta_x^2+\Delta_y^2}}\left(\sqrt{t}\right)$$

or,

$$f_{\sqrt{\Delta_x^2+\Delta_y^2}}\left(\sqrt{t}\right)=2\sqrt{t}f_{\Delta_x^2+\Delta_y^2}(t)$$

or,

$$f_{\sqrt{\Delta_x^2+\Delta_y^2}}(t)=2tf_{\Delta_x^2+\Delta_y^2}\left(t^2\right)$$

or,

$$f_{\sqrt{\Delta_x^2+\Delta_y^2}}(t)=\begin{cases}0 & t^2\in(-\infty,0)\cup(2,+\infty)\\ 2\pi t-8t^2+2t^3 & t^2\in[0,1]\\ 2t\arcsin\left(\frac{2}{t^2}-1\right)-2t\arcsin\left(1-\frac{2}{t^2}\right) & \\ \quad+8t\sqrt{t^2-1}-2t^3-4t & t^2\in(1,2]\end{cases}$$

Consequently,

$$f_{\sqrt{\Delta_x^2+\Delta_y^2}}(t)=\begin{cases}0 & t\in(-\infty,0)\cup(\sqrt{2},+\infty)\\ 2\pi t-8t^2+2t^3 & t\in[0,1]\\ 2t\arcsin\left(\frac{2}{t^2}-1\right)-2t\arcsin\left(1-\frac{2}{t^2}\right) & \\ \quad+8t\sqrt{t^2-1}-2t^3-4t & t\in(1,\sqrt{2}]\end{cases}\tag{A.24}$$

In order to derive the analytical expression regarding Euclidean distance's CDF, the following property is employed.

$$-\frac{2}{t^2\sqrt{1-\frac{1}{t^2}}} = -\frac{4}{\frac{2t^3}{t}\sqrt{1-\frac{1}{t^2}}}$$

or,

$$-\frac{2}{t^2\sqrt{1-\frac{1}{t^2}}} = -\frac{4}{t^3\sqrt{\frac{4}{t^2}-\frac{4}{t^4}}}$$

or,

$$-\frac{2}{t^2\sqrt{1-\frac{1}{t^2}}} = -\frac{4}{t^3\sqrt{1-\left(\frac{2}{t^2}-1\right)^2}}$$

or,

$$-\frac{1}{t^2\sqrt{1-\frac{1}{t^2}}} - \frac{1}{t^2\sqrt{1-\frac{1}{t^2}}} = -\frac{4}{t^3\sqrt{1-\left(\frac{2}{t^2}-1\right)^2}}$$

or,

$$-\int \frac{1}{t^2\sqrt{1-\frac{1}{t^2}}}\mathrm{d}t - \int \frac{1}{t^2\sqrt{1-\frac{1}{t^2}}}\mathrm{d}t = -\int \frac{4}{t^3\sqrt{1-\left(\frac{2}{t^2}-1\right)^2}}\mathrm{d}t + \mathrm{C}$$

or,

$$\arcsin\left(\frac{1}{t}\right) - \arccos\left(\frac{1}{t}\right) = \arcsin\left(\frac{2}{t^2}-1\right) + \mathrm{C}$$

Setting $t = \sqrt{2}$ it follows that, $\mathrm{C} = 0$. Consequently,

$$\arcsin\left(\frac{1}{t}\right) - \arccos\left(\frac{1}{t}\right) = \arcsin\left(\frac{2}{t^2}-1\right)$$

or,

$$2\left(\arcsin\left(\frac{1}{t}\right) - \arccos\left(\frac{1}{t}\right)\right) = 2\arcsin\left(\frac{2}{t^2}-1\right)$$

or,

$$2\left(\arcsin\left(\frac{1}{t}\right) - \arccos\left(\frac{1}{t}\right)\right) = \arcsin\left(\frac{2}{t^2}-1\right) + \arcsin\left(\frac{2}{t^2}-1\right)$$

or,

$$2\left(\arcsin\left(\frac{1}{t}\right) - \arccos\left(\frac{1}{t}\right)\right) = \arcsin\left(\frac{2}{t^2}-1\right) - \arcsin\left(1-\frac{2}{t^2}\right) \qquad \text{(A.25)}$$

Substituting Equation (A.25) into Equation (A.24) it follows that,

$$f_{\sqrt{\Delta_x^2+\Delta_y^2}}(t) = \begin{cases} 0 & t \in (-\infty, 0) \cup (\sqrt{2}, +\infty) \\ 2\pi t - 8t^2 + 2t^3 & t \in [0, 1] \\ 4t \arcsin\left(\frac{1}{t}\right) - 4t \arccos\left(\frac{1}{t}\right) & \\ \quad + 8t\sqrt{t^2-1} - 2t^3 - 4t & t \in (1, \sqrt{2}] \end{cases} \tag{A.26}$$

Consequently, the CDF of the Euclidean distance (i.e., $F_{\sqrt{\Delta_x^2+\Delta_y^2}}(t)$) is obtained by the integral of Equation (A.26),

$$F_{\sqrt{\Delta_x^2+\Delta_y^2}}(t) = \begin{cases} 0 & t \in (-\infty, 0) \cup (\sqrt{2}, +\infty) \\ \pi t^2 - \frac{8}{3}t^3 + \frac{1}{2}t^4 + \mathrm{C}_1 & t \in [0, 1] \\ 2\,t^2 \arcsin\left(\frac{1}{t}\right) - 2\,t^2 \arccos\left(\frac{1}{t}\right) + 4\sqrt{t^2-1} & \\ \quad + \frac{8\left(t^2-1\right)^{\frac{3}{2}}}{3} - 2\,t^2 - \frac{t^4}{2} + \mathrm{C}_2 & t \in (1, \sqrt{2}] \end{cases}$$

Constants C_1 and C_2 can be obtained by setting $t = 0$ and $t = \sqrt{(}2)$ respectively. Thus, $\mathrm{C}_1 = 0$ and $\mathrm{C}_2 = \frac{1}{3}$. Consequently,

$$F_{\sqrt{\Delta_x^2+\Delta_y^2}}(t) = \begin{cases} 0 & t \in (-\infty, 0) \cup (\sqrt{2}, +\infty) \\ \pi t^2 - \frac{8}{3}t^3 + \frac{1}{2}t^4 & t \in [0, 1] \\ 2\,t^2 \arcsin\left(\frac{1}{t}\right) - 2\,t^2 \arccos\left(\frac{1}{t}\right) + 4\sqrt{t^2-1} & \\ \quad + \frac{8\left(t^2-1\right)^{\frac{3}{2}}}{3} - 2\,t^2 - \frac{t^4}{2} + \frac{1}{3} & t \in (1, \sqrt{2}] \end{cases} \tag{A.27}$$

A.9 Proof of Corollary 6

Let Y denote the all ones $1 \times n$ vector such that $Y = [1, 1, 1...., 1]$ and d_i the degree of vertex i. The sum of the node degrees in any graph equals twice the number of edges i.e., $\sum_{i=1}^{n}(d_i) = 2|E(G)|$. Consequently,

$$Y^T \mathcal{A} Y = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}(a_{ij})\right) = \sum_{i=1}^{n}(d_i) = 2|E(G)|.$$

Rayleigh quotient is upper bounded by λ_1 i.e., $\lambda_1 \geq \frac{X^T \mathcal{A} X}{X^T X}$, for any non-zero vector $X \in \mathbb{R}^n$. The equality holds *iff* X is the principal eigenvector of the matrix $\mathcal{A}$. By

setting $X = Y$ it follows that,

$$\lambda_1 \geq \frac{Y^T \mathcal{A} Y}{Y^T Y} = \frac{2|E(G)|}{N} = \bar{d}.$$

Therefore,

$$4/\bar{d} \geq 4/\lambda_1$$

or,

$$\bar{q} \geq \hat{q}.$$

Since $\hat{q}$ is a lower bound with respect to the forwarding probability, for large values of timestep t it follows that,

$$\hat{C}_t(\bar{q}) \geq \hat{C}_t(\hat{q}) > 0.$$

A.10 Proof of Corollary 8

From Appendix A.8, the CDF of the Euclidean distance between node pairs in Geometric Random Graphs in a $[0,1] \times [0,1]$ plain is given by,

$$F(r_c) = \begin{cases} \frac{1}{2}{r_c}^4 - \frac{8}{3}{r_c}^3 + \pi {r_c}^2 & ,r_c \in [0,1] \\ 2\,{r_c}^2 \arcsin\left(\frac{1}{r_c}\right) - 2\,{r_c}^2 \arccos\left(\frac{1}{r_c}\right) + 4\sqrt{{r_c}^2 - 1} & ,r_c \in (1, \sqrt{2}] \end{cases}.$$

Consequently, the average node degree is $nF(r_c)$ and using Corollary 6 it is satisfied that, $q_g = \frac{4}{nF(r_c)}$.

A.11 Proof of Corollary 9

According to [172], coverage (C_t) is a polynomial of q and degree t that is,

$$C_t(q) = \frac{1}{n} \sum_{i=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{i+1} \binom{t-i+1}{t-2i+2} w_{t-i+1,s} q^{t-i+1}. \tag{A.28}$$

where $w_{t,s}$ denotes the summation of the number of walks of size t starting from node s. For a regular graph of degree d_R the number of walks of size t starting from

node s is given by, $w_{t,s} = d_R^t$. By substituting the latter into Equation (A.28) it follows that,

$$C_{R,t}(q) = \frac{1}{n}\sum_{i=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{i+1} \binom{t-i+1}{t-2i+2} d_R^{t-i+1} q^{t-i+1}. \tag{A.29}$$

Let for $t \geq 0$, $\Pi_{R,t}(q)$ be a polynomial of q such that,

$$\Pi_{R,t}(q) = qd_R[\,(\Pi_{R,t-1}(q) - \Pi_{R,t-2}(q)] \tag{A.30}$$

with initial conditions, $\Pi_0(q) = 1$ and $\Pi_1(q) = qd_R$. Similarly to Equation (E.1) and Equation (E.2) it follows that,

$$\Pi_{R,t}(q) = \sum_{k=1}^{\lfloor \frac{t}{2} \rfloor + 1} (-1)^{k+1} \binom{t-k+1}{t-2k+2} d_R^{t-k+1} q^{t-k+1} \tag{A.31}$$

Consequently,

$$C_{R,t}(q) = \frac{1}{n}\Pi_{R,t}(q). \tag{A.32}$$

Solving A.30 similarly as in Appendix A.3 it follows that,

$$C_{R,t}(q) = \begin{cases} \frac{1}{n}\frac{\left(qd_R + \sqrt{q^2 d_R^2 - 4qd_R}\right)^{t+1} - \left(qd_R - \sqrt{q^2 d_R^2 - 4qd_R}\right)^{t+1}}{2^{t+1}\sqrt{q^2 d_R^2 - 4qd_R}} & \text{for } q > \frac{4}{d_R} \\ \frac{1}{n}(t+1)2^t & \text{for } q = \frac{4}{d_R} \\ \frac{1}{n}\frac{2(qd_R)^{\frac{t+1}{2}}\sin[(t+1)\phi]}{\sqrt{4qd_R - (qd_R)^2}} & \text{for } q < \frac{4}{d_R} \end{cases}$$

Analyzing the above equation similarly to Appendix A.6 it follows that,

$$\hat{q}_R = \frac{4}{d_R}$$

A.12 Proof of Theorem 5

For $q = \hat{q}$ it is

$$C_t(\hat{q}) = \frac{1}{n}(t+1)2^t u_{1,s} \sum_{i=1}^{n} (U_1)_i$$

Since global outreach in this work is considered as $C_t \geq 0.95$ it follows that,

$$\frac{1}{n}(t+1)2^t u_{1,s} \sum_{i=1}^{n} (U_1)_i \geq 0.95$$

or,

$$(t+1)2^t \geq \frac{0.95n}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}$$

or,

$$ln2(t+1)2^{t+1} \geq \frac{1.9nln2}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}$$

or,

$$ln2(t+1)e^{(t+1)ln2} \geq \frac{1.9nln2}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}.$$

Let $\mathcal{W}_0$ denote the 0-branch of Lambert's W function. Since $\mathcal{W}_0$ is a strictly increasing function it follows that,

$$\mathcal{W}_0\left(ln2(t+1)e^{(t+1)ln2}\right) \geq \mathcal{W}_0\left(\frac{1.9nln2}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)$$

or,

$$(t+1)ln2 \geq \mathcal{W}_0\left(\frac{1.9nln2}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)$$

or,

$$t \geq \frac{\mathcal{W}_0\left(\frac{1.9nln2}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)}{ln2} - 1$$

thus,

$$\mathcal{T}(\hat{q},s) > \frac{\mathcal{W}_0\left(\frac{1.9nln2}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)}{ln2} - 1$$

For $q > \hat{q}$ it is,

$$C_t(q) = \frac{1}{n}\frac{\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1} - \left(q\lambda_1 - \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} u_{1,s}\sum_{i=1}^{n}(U_1)_i\,.$$

Consequently,

$$C_t(q) < \frac{1}{n}\frac{\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} u_{1,s}\sum_{i=1}^{n}(U_1)_i$$

Therefore, when

$$\frac{1}{n}\frac{\left(q\lambda_1 + \sqrt{q^2\lambda_1^2 - 4q\lambda_1}\right)^{t+1}}{2^{t+1}\sqrt{q^2\lambda_1^2 - 4q\lambda_1}} u_{1,s}\sum_{i=1}^{n}(U_1)_i = 1$$

it follows that, $C_t(q) < 1$. Solving this equation we get,

$$\frac{u_{1,s}\sum_{i=1}^{n}(U_1)_i}{n\sqrt{q^2\lambda_1^2-4q\lambda_1}}\left(\frac{q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}}{2}\right)^{t+1}=1$$

or,

$$\left(\frac{q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}}{2}\right)^{t+1}=\frac{n\sqrt{q^2\lambda_1^2-4q\lambda_1}}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}$$

or,

$$t=\frac{ln\left(\frac{n\sqrt{q^2\lambda_1^2-4q\lambda_1}}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)}{ln\left(\frac{q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}}{2}\right)}-1$$

However, for this value of t it is $C_t(q) < 1$ consequently,

$$\mathcal{T}(q,s)>\frac{ln\left(\frac{n\sqrt{q^2\lambda_1^2-4q\lambda_1}}{u_{1,s}\sum_{i=1}^{n}(U_1)_i}\right)}{ln\left(\frac{q\lambda_1+\sqrt{q^2\lambda_1^2-4q\lambda_1}}{2}\right)}-1$$

A.13 Proof of Corollary 10

From Corollary 9 it is,

$$C_{R,t}(q)=\frac{1}{n}(t+1)2^t \text{ for } q=\hat{q}_{d_R}$$

and

$$C_{R,t}(q)=\frac{1}{n}\frac{\left(qd_R+\sqrt{q^2d_R^2-4qd_R}\right)^{t+1}-\left(qd_R-\sqrt{q^2d_R^2-4qd_R}\right)^{t+1}}{2^{t+1}\sqrt{q^2d_R^2-4qd_R}} \text{ for } q>\hat{q}_{d_R}$$

Analyzing similarly to Appendix A.12 it follows that,

$$\mathcal{T}_R(\hat{q})>\frac{\mathcal{W}_0\left(1.9nln2\right)}{ln2}-1$$

and

$$\mathcal{T}_R(q>\hat{q})>\frac{ln\left(n\sqrt{q^2d_R^2-4qd_R}\right)}{ln\left(\frac{qd_R+\sqrt{q^2d_R^2-4qd_R}}{2}\right)}-1$$

respectively.

A.14 Proof of Lemma 6

Multiplying the all ones vector ($\vec{1}^T$) with matrix Q it follows that,

$$\vec{1}^T Q = [111\ldots1]Q = [mmm\ldots m] = m\vec{1}^T$$

Consequently, m is an eigenvalue of Q corresponding to eigenvector $\vec{1}^T$.

A.15 Proof of Lemma 7

Analyzing Equation (E.11) similarly to the works presented in [172] and [178] vector $\mathcal{P}_t$ is given by,

$$\mathcal{P}_t = \left[\sum_{k=1}^{\lfloor\frac{t}{2}\rfloor+1} (-1)^{k+1}\binom{t-k+1}{t-2k+2}Q^{t-k+1}\right] P_0 \tag{A.33}$$

Coverage at timestep t is given by,

$$C_{M,t} = \frac{1}{n}\sum_{j=1}^{j=n}(\mathcal{P}_{t,j})$$

or,

$$C_{M,t} = \frac{1}{n}\vec{1}^T\mathcal{P}_t$$

and by substituting vector $\mathcal{P}_t$ from (A.33) it follows that,

$$C_{M,t} = \frac{1}{n}\vec{1}^T\left[\sum_{k=1}^{\lfloor\frac{t}{2}\rfloor+1} (-1)^{k+1}\binom{t-k+1}{t-2k+2}Q^{t-k+1}\right] P_0$$

or,

$$C_{M,t} = \frac{1}{n}\left[\sum_{k=1}^{\lfloor\frac{t}{2}\rfloor+1} (-1)^{k+1}\binom{t-k+1}{t-2k+2}\vec{1}^T Q^{t-k+1}\right] P_0.$$

Since $\vec{1}^T$ is a left eigenvector of Q it follows that,

$$C_{M,t} = \frac{1}{n}\left[\sum_{k=1}^{\lfloor\frac{t}{2}\rfloor+1} (-1)^{k+1}\binom{t-k+1}{t-2k+2}\vec{1}^T m^{t-k+1}\right] P_0$$

or,

$$C_{M,t} = \frac{1}{n}\left[\sum_{k=1}^{\lfloor\frac{t}{2}\rfloor+1}(-1)^{k+1}\binom{t-k+1}{t-2k+2}m^{t-k+1}\right]\vec{1}^T P_0$$

or,

$$C_{M,t} = \frac{1}{n}\sum_{k=1}^{\lfloor\frac{t}{2}\rfloor+1}(-1)^{k+1}\binom{t-k+1}{t-2k+2}m^{t-k+1}$$

A.16 Proof of Theorem 6

Analyzing m-Probabilistic Flooding coverage for $t > 0$ similarly to Appendix A.3 it follows that,

$$C_{M,t}(m) = \begin{cases} \frac{1}{n}\frac{\left(m+\sqrt{m^2-4m}\right)^{t+1}-\left(m-\sqrt{m^2-4m}\right)^{t+1}}{2^{t+1}\sqrt{m^2-4m}} & \text{for } m > 4 \\ \frac{1}{n}(t+1)2^t & \text{for } m = 4 \\ \frac{1}{n}\frac{2(m)^{\frac{t+1}{2}}\sin[(t+1)\phi]}{\sqrt{4m-m^2}} & \text{for } m < 4 \end{cases} \tag{A.34}$$

and similarly to in Corollary 3, Corollary 4 and Theorem 4 it follows that, $\hat{m} = 4$ is the minimum value of m that allows for global outreach under mPF.

Appendix B

Simulation Details

For the generation of network topologies and the evaluation of the analytical findings, a proprietary object oriented simulator was developed in C++ programming language. The UML class diagram of the simulator is depicted in Figure B.1. Each node within a topology is an object of the "Node" class and has an "ID" property that represents the node. The topology is an object that contains nodes and the corresponding links among them. This class includes methods for adding, removing and connecting nodes and methods for saving the adjacency matrix of the topology in GNU Octave compatible file format. For the graph generation, the "baseGraph" class was developed that includes an object of the "Topology" class. Furthermore, this class is inherited by "ER" and "GRG" classes for the generation of ER and GRG topologies respectively. For the simulation of probabilistic flooding and mPF,

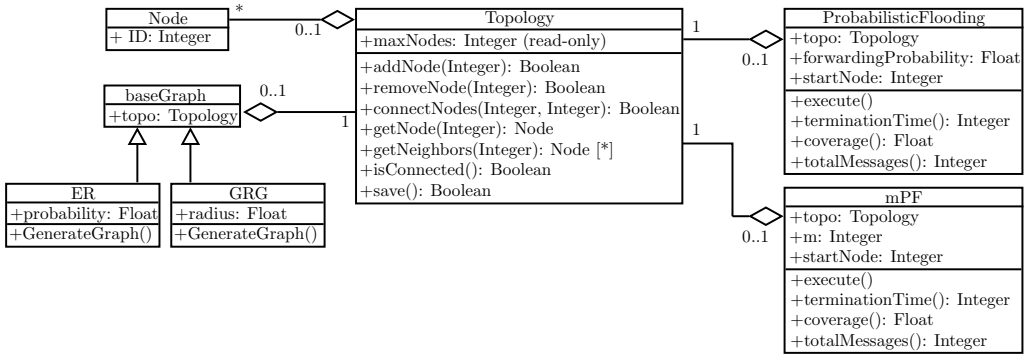


Figure B.1: *The UML class diagram of the simulator.*

"ProbabilisticFlooding" and "mPF" classes respectively were developed. Each class

includes an object of the "Topology" class and various parameters regarding their operation such as the forwarding probability and the initiator node. Moreover, methods for the execution of probabilistic flooding and mPF are also included along with methods for obtaining the simulation results regarding coverage, termination time and total number of sent messages.

The simulator maintains two lists of nodes that should transmit at current and at next timestep respectively. Moreover a boolean variable that depicts whether a node is covered, is also included. For the case of probabilistic flooding, each node that receives the information message for the first time, generates for each neighbor (excluding the node that received the information message from) a random value between zero and one, and if this value is smaller than the forwarding probability, transmits the message to the neighbor node. If this node is uncovered, then it is added to the list of nodes that should transmit the information message at next timestep as depicted in Figure B.2. For the case of mPF, each node that receive the information message for the first time, selects randomly "m" of its neighbors (excluding the node that received the information message from) and transmits the information message towards them. The information dissemination mechanism, terminates when the list of nodes that should transmit the information message at next timestep, is empty. Finally, several methods for obtaining the simulation results (e.g., number of covered nodes, termination time etc) are executed.

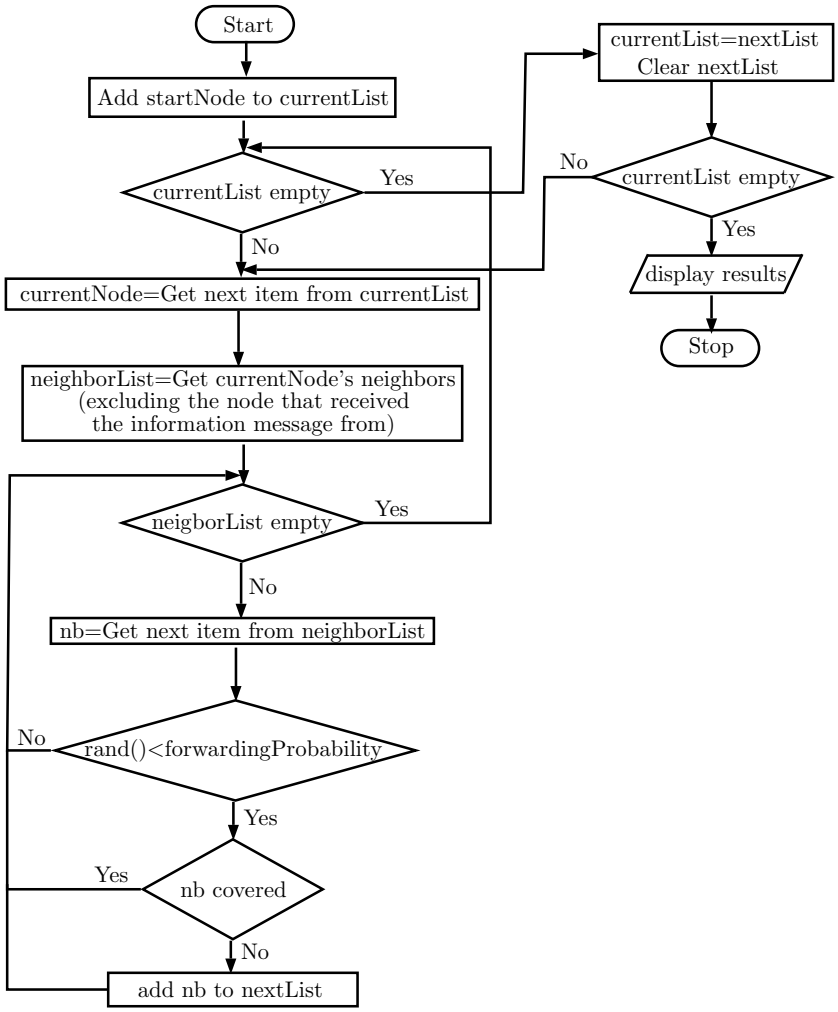


Figure B.2: *Probabilistic flooding flowchart of the simulator.*

Publication Table

Refereed Journals

- G. Koufoudakis, K. Oikonomou, K. Giannakis, and S. Aïssa, "Probabilistic Flooding Coverage Analysis for Efficient Information Dissemination in Wireless Networks," *Computer Networks*, vol. 140, pp. 51 – 61, 2018
- G. Koufoudakis, K. Oikonomou, and G. Tsoumanis, "Adapting Probabilistic Flooding in Energy Harvesting Wireless Sensor Networks," *Journal of Sensor and Actuator Networks*, vol. 7, no. 3, p. 39, 2018
- G. Tsoumanis, K. Oikonomou, G. Koufoudakis, and S. Aïssa, "Energy-Efficient Sink Placement in Wireless Sensor Networks," *Computer Networks*, vol. 141, pp. 166 – 178, 2018
- K. Oikonomou, G. Koufoudakis, E. Kavvadia, and V. Chrissikopoulos, "A Wireless Sensor Network Innovative Architecture for Ambient Vibrations Structural Monitoring," *Key Engineering Materials*, vol. 628, 2014
- G. Koufoudakis, N. Skiadopoulos, E. Magkos, and K. Oikonomou, "Synchronization Issues in an Innovative Wireless Sensor Network Architecture Monitoring Ambient Vibrations in Historical Buildings," *Key Engineering Materials*, vol. 628, 2014
- K. Skiadopoulos, A. Tsipis, K. Giannakis, G. Koufoudakis, E. Christopoulou, K. Oikonomou, G. Kormentzas, and I. Stavrakakis, "Synchronization of data measurements in wireless sensor networks for iot applications," *Ad hoc networks*, Under Review

Refereed Conferences

- G. Koufoudakis, K. Oikonomou, S. Aïssa, and I. Stavrakakis, "Analysis of Spectral Properties for Efficient Coverage Under Probabilistic Flooding," in *2018 IEEE 19th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM) (IEEE WoWMoM 2018)*, (Chania, Crete, Greece), Jun 2018
- S. Fanarioti, A. Tsipis, K. Giannakis, G. Koufoudakis, E. Christopoulou, K. Oikonomou, and I. Stavrakakis, "A Proposed Algorithm for Data Measurements Synchronization in Wireless Sensor Networks," in *Second International Balkan Conference on Communications and Networking 2018 (BalkanCom'18)*, (Podgorica, Montenegro), Jun 2018
- G. Tsoumanis, G. Koufoudakis, K. Oikonomou, M. Avlonits, and N. Varotsis, "A Low-Cost Surface Wireless Sensor Network for Pollution Monitoring in the Ionian Sea," in *12th Panhellenic Symposium of Oceanography & Fisheries*, May 2018. Abstract
- E. Kavvadia, G. Koufoudakis, and K. Oikonomou, "Robust Probabilistic Information Dissemination in Energy Harvesting Wireless Sensor Networks," in *2014 13th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)*, pp. 63–70, June 2014
- K. Oikonomou, G. Koufoudakis, and S. Aïssa, "Probabilistic Flooding Coverage Analysis in Large Scale Wireless Networks," in *2012 19th International Conference on Telecommunications (ICT)*, pp. 1–6, April 2012

Glossary

`Adjacency matrix`	An $n \times n$ symmetric matrix, such that the value of element at row i and column j indicates whether the nodes i and j are connected (i.e.,$(\mathcal{A})_{i,j} = 1$) or not (i.e.,$(\mathcal{A})_{i,j} = 0$)
`Initiator node`	The node of a network that initiates probabilistic flooding.
`Forwarding probability`	The probability of a node to forward a particular information message to its neighbors under probabilistic flooding.
`Threshold probability`	The minimum value of the forwarding probability that results in global outreach under probabilistic flooding.
`Coverage probability`	The probability of a node to get covered.
`Eigenvector`	Eigenvector of a linear transformation is a non-zero vector that its direction is not changed when the particular linear transformation is applied to it.
`Eigenvalue`	Eigenvalue of a linear transformation $\mathcal{A}$ is a is a scalar λ for which there exists a non-zero vector U such that $\mathcal{A}\lambda = \lambda U$.
`Principal Eigenvector`	The eigenvector that corresponds to the largest eigenvalue

Notations

G	The corresponding graph of a network
V	The set of the graph's vertices
E	The set of the graph's edges
n	Number of vertices/nodes in the graph/network
d_i	The degree of node i
d_R	The degree of the regular graph/network
$\bar{d}$	The average degree of the graph/network
$w_{l,i}$	The number of walks of length l starting from vertex i
D	Graph/Network diameter
$\mathcal{A}$	The adjacency matrix of the graph
$\det(\mathcal{A})$	The determinant of matrix $\mathcal{A}$
$tr(\mathcal{A})$	The trace of matrix $\mathcal{A}$
$\Pi_{\mathcal{A}}$	The characteristic polynomial of matrix $\mathcal{A}$
I	The identity matrix
λ_i	The ith largest eigenvalue of the adjacency matrix
U_i	The normalized eigenvector of the adjacency matrix corresponding to eigenvalue λ_i
q	Forwarding probability
$\check{q}(s)$	Threshold Probability
$\tilde{q}(s)$	Threshold Probability obtained by the simulations
s	Initiator node
P_t	$n \times 1$ vector, such that the value of i's element indicates whether node i is covered (i.e. $(P_t)_i = 1$) or not (i.e. $(P_t)_i = 0$) until timestep t
$\check{P}_t$	$n \times 1$ vector, such that the value of i's element indicates whether node i is covered (i.e. $(\check{P}_t)_i = 1$) or not (i.e. $(\check{P}_t)_i = 0$) at timestep t

$\mathcal{P}_t$	$n \times 1$ vector, such that the value of i's element indicates the probability that node i has received the information message until timestep t
C_t	Coverage
$\mathcal{T}$	Termination time
$\mathcal{Z}_t$	Number of messages sent until time t
r_c	The connectivity radius of a Geometric Random Graph
$\rho_{s,t}$	The forwarding root of polynomial $\mathtt{P}_t(q)$
$\Delta(s,t)$	The difference between two consecutive forwarding roots (i.e., $\rho_{s,t} - \rho_{s,t-1}$)
$\delta(s,t)$	The forwarding probability threshold for Algorithm 1

Abbreviations

IP	Internet Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
Modem	Modulator - Demodulator
IEEE	Institute of Electrical and Electronics Engineers
UHF	Ultra High Frequency
GPRS	General Packet Radio Service
DSSS	Direct Sequence Spread Spectrum
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
MAC	Media Access Control
OFDM	Orthogonal Frequency Division Multiplexing
WiGig	Wireless Gigabit Alliance
WiMAX	Worldwide Interoperability for Microwave Access

SOFDMA	Scalable Orthogonal Frequency Division Multiple Access
QoS	Quality of Service
SS	Subscriber Station
BS	Base Station
MAC service data units	MSDUs
MPDUs	MAC protocol data units
AES	Advanced Encryption Standard
3DES	Triple Data Encryption Standard
IETF	Internet Engineering Task Force
EAP	Extensible Authentication Protocol
WPAN	Wireless Personal Area Network
FHSS	Frequency Hopping Spread Spectrum
BPSK	Binary Phase Shift Keying
OQPSK	Offset Quadrature Phase Shift Keying
FFD	Full Function Device
RFD	Reduced Function Device
FDMA	Frequency Division Multiple Access

GSM	Global System for Mobile communications
GPRS	General Packet Radio Services
EDGE	Enhanced Data rates for GSM Evolution
SPIN	Sensor Protocols for Information via Negotiation
GPS	Global Positioning System
ADSL	Asymmetric Digital Subscriber Line
TDMA	Time Division Multiple Access
CDMA	Code Division Multiple Access
FDD	Frequency Division Duplex
TDD	Time Division Duplex
NOMA	Non-Orthogonal Multiple Access
IoT	Internet of Things
IIoT	Industrial Internet of Things
SD-WSNs	Software Defined Wireless Sensor Networks
MBB	Mobile BroadBand
xMBB	Extreme Mobile BroadBand
mMTC	Massive Machine-Type Communication

uMTC	Ultra-reliable Machine-Type Communication
SDN	Software Defined Networking
API	Application Programming Interface
NFV	Network Function Virtualization
VNF	Virtualized Network Functions
NFVI	Network Function Virtualization Infrastructure
WSN	Wireless Sensor Network
GRG	Geometric Random Graphs
ER	Erdős-Rényi
AUV	Autonomous Underwater Vehicles
ISP	Internet Service Provider
AODV	Ad hoc On-demand Distance Vector
MANET	Mobile Ad-hoc Networks
VANET	Vehicular Ad-hoc Networks
mPF	m-Probabilistic Flooding

Index