Sérgio Armindo Lopes Crisóstomo

Information Dissemination in Random Networks

DISSERTATION

Submitted to the

Faculdade de Ciências, Universidade do Porto

and to the

Fakultät für Technische Wissenschaften, Alpen-Adria-Universität Klagenfurt

to obtain the joint academic degree of

Doutor em Ciência de Computadores

and

Doktor der Technischen Wissenschaften

April 2012

Sérgio Armindo Lopes Crisóstomo

Information Dissemination in Random Networks

Tese submetida à Faculdade de Ciências da Universidade do Porto para obtenção do grau de Doutor em Ciência de Computadores

Orientadores: Professor Doutor João Barros, Universidade do Porto, e Univ.-Prof. Dr.-Ing. Christian Bettstetter, Alpen-Adria-Universität Klagenfurt

Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
2012

Sérgio Armindo Lopes Crisóstomo

Information Dissemination in Random Networks

DISSERTATION

zur Erlangung des akademischen Grades
Doktor der Technischen Wissenschaften

Alpen-Adria-Universität Klagenfurt

Fakultät für Technische Wissenschaften

1. Begutachter:	Professor Doutor João Barros
Institut:	Faculdade de Engenharia
	Universidade do Porto
2. Begutachter:	Univ.-Prof. Dr.-Ing. Christian Bettstetter
Institut:	Institut für Vernetzte und Eingebettete Systeme
	Alpen-Adria-Universität Klagenfurt

April 2012

Word of honour

I hereby confirm on my honour that I personally prepared the present academic work and carried out myself the activities directly involved with it. I also confirm that I have used no resources other than those declared. All formulations and concepts adopted literally or in their essential content from printed, unprinted or Internet sources have been cited according to the rules for academic work and identified by means of footnotes or other precise indications of source.

The support provided during the work, including significant assistance from my supervisors has been indicated in full.

The academic work has not been submitted to any other examination authority. The work is submitted in printed and electronic form. I confirm that the content of the digital version is completely identical to that of the printed version.

I am aware that a false declaration will have legal consequences.

Signature: _____

Porto, Portugal, 25th of April 2012

To Lurdes, Armindo, and Paula.

Acknowledgments

I would like to start by thanking my supervisors Professor João Barros and Professor Christian Bettstetter for their mentoring, encouragement, and for giving me the opportunity to interact with exciting people in an international working environment.

A special thanks to Professor Christian Bettstetter for having hosted me at the University of Klagenfurt for a couple of years. At the University of Klagenfurt I found a fantastic working and social environment. I had the opportunity to make friendship and to collaborate with a wonderful team. A special thanks to Udo Schilcher, Helmut Adam, Evsen Yanmaz, Nikolaj Marchenko, Johannes Klinglmayr, István Fehérvári, Wilfried Elmenreich, Günther Brandner, Torsten Andre, Alexander Gogolev, Michael Gyarmati, Alexander Tyrrell, Andrea Krammer, and Kornelia Lienbacher.

I would like to express a special thanks to my colleagues Pedro Brandão and Rui Prior for their friendship, support, and all the insightful discussions. I also would like to express my gratitude to my colleagues from the Department of Computer Science, for their support and for the extra hours they had to lecture during my stay at the University of Klagenfurt. A special thanks to Alexandra Ferreira for helping me in many administrative issues.

At the Instituto de Telecomunicações I had the pleasure to work with a wonderful international team. My thanks to Paulo Falcão, Tiago Vinhoza, Rui Costa, João Vilela, Diogo Ferreira, João Patriarca, Gerhard Maierbacher, Luís Pinto, Pedro Santos, Traian Abrudan, Mate Boban, João Rodrigues, Mari Nistor, Saurabh Shintre, Fabio Hedayioglu, Luísa Lima, Fausto Vieira, Susana Cruz, Mariana Kaiseler, Ian Marsh, Rui Meireles, Hugo Pinto, Sílvia Ribeiro, and many others.

I express my gratitude to Susana Sargento, José Ruela, Manuel Ricardo, Jorge Mamede, and Eurico Inocêncio - former research collaborators with whom I learned a lot.

Many thanks to my friends from the undergraduate student times at the University of Porto who played a very special role in my live: Gabriel Falcão, Miguel Falcão, Nuno Pinto, and Gustavo Martins.

Last, but not least, I would like to thank all my family, especially my fiancé Paula and my mother Lurdes for all their support, patience, and love.

This work was supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science & Technology (FCT)) under grant SFRH/BD/28835/2006, the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant 20214/15935/23108, and the European Science Foundation (ESF) under the program MiNEMA.

Abstract

This dissertation focuses on the study of information dissemination in communication networks with a broadcast medium. The main problem we address is how to disseminate efficiently a message from a source node to all other network nodes. In terms of efficiency we target two goals: (1) to deliver a source message to all network nodes with high probability; and (2) to use as few transmissions as possible for a given target reachability.

In this context, our main focus is devoted to probabilistic dissemination algorithms. Modeling networks as random graphs, which are built from stochastic processes, and using methods from graph theory and stochastic geometry we address both replication based and network coded information dissemination approaches.
The first contribution is an analytical study of probabilistic flooding which answers the question of which is the minimum common network-wide forwarding probability each node should use such that a flooded message is obtained by all nodes with high probability.
Next, we address the question of which benefits can be expected from network coded based probabilistic flooding. We compare these benefits with the ones from the well established replication based MultiPoint Relay flooding. The study of their efficiency is performed both by analytical techniques and numerical methods.
Finally, we apply the insights gained from the study of information dissemination algorithms to the design of a sensor-actuator networked system for emergency response in indoor scenarios. The system guides people to the exits of a building via the shortest safe paths, computed autonomously by each node whenever a new measurement collected by a sensor is flooded throughout the network.

Resumo

Esta dissertação foca o estudo de disseminação de informação em redes de comunicação sobre um meio de difusão. O problema central estudado é como disseminar eficientemente uma mensagem gerada por um nó fonte para os restantes nós da rede. Em termos de eficiência procura-se satisfazer dois objectivos principais: (1) a entrega de mensagens geradas por nós fontes a todos nós da rede, com probabilidade elevada; (2) um número de transmissões tão baixo quanto possível, que permita, no entanto, a satisfação de um dado objectivo de entrega.

Neste contexto, este trabalho foca o estudo de algoritmos de disseminação probabilísticos. Modelizando redes como grafos aleatórios resultantes de processos estocásticos e utilizando ferramentas da teoria de grafos e da geometria estocástica, estudamos técnicas de disseminação baseadas em replicação de mensagens e técnicas que recorrem a codificação de rede.
A primeira contribuição deste trabalho traduz-se por uma análise de processos de inundação probabilísticos, em que se procura determinar qual será a probabilidade de reencaminhamento mínima que deverá ser utilizada por cada nó por forma a que uma mensagem gerada seja recebida por todos os nós da rede com uma probabilidade elevada.
Seguidamente, analisamos que benefícios se podem obter pela incorporação de técnicas de codificação de rede em algoritmos de disseminação probabilísticos. Com esse propósito confrontamos o desempenho de técnicas de inundação baseados em codificação de rede com técnicas baseadas em replicação de mensagens, tendo seleccionado para o segundo caso o algoritmo de inundação *MultiPoint Relaying*.
Finalmente, aplicamos o conhecimento adquirido pelo estudo de processos de disseminação de informação à especificação e implementação de um sistema baseado numa rede de sensores-actuadores para suporte à evacuação de edifícios em situações de emergência. O sistema guia as pessoas para fora do edifício, indicando-lhes o caminho mais curto e seguro a seguir. Os caminhos são calculados autonomamente por cada nó, sempre que informação relevante recolhida por um sensor é disseminada pela rede.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit der Dissemination von Information in einem Kommunikationsnetzwerk mit Broadcast-Kanal. Die zentrale Frage, welcher wir uns in dieser Arbeit widmen, ist wie man eine Nachricht ausgehend von einem Quellknoten effizient an alle anderen Knoten im Netzwerk verteilt. Hierbei verfolgen wir zwei Hauptziele: (1) Die Nachricht soll mit hoher Wahrscheinlichkeit alle Knoten im Netzwerk erreichen; (2) Es sollen so wenige Übertragungen wie möglich stattfinden.

In diesem Zusammenhang wenden wir uns hauptsächlich Algorithmen zur probabilistischen Dissemination von Information zu. Wir modellieren Kommunikationsnetzwerke als Zufallsgraphen, die auf stochastischen Prozessen beruhen. Wir verwenden Methoden der Graphentheorie sowie der stochastischen Geometrie um Disseminationsalgorithmen basierend sowohl auf Nachrichtenweiterleitung als auch auf Network Coding zu analysieren.
Unser erstes Resultat ist eine analytische Studie von probabilistischem Flooding. In dieser Studie zeigen wir, wie die netzwerkweite Weiterleitungswahrscheinlichkeit gewählt werden soll, sodass eine Nachricht mit hoher Wahrscheinlichkeit alle Knoten im Netzwerk erreicht.
Als nächstes widmen wir uns der Frage, welche Vorteile ein probabilistischer Flooding-Algorithmus basierend auf Network Coding gegenüber klassischen Methoden hat. Dabei wird die Network-Coding Methode mit dem weit verbreiteten MultiPoint Relay-Algorithmus verglichen. Der Vergleich erfolgt mittels analytischer und numerischer Methoden.
Schlussendlich verwenden wir die Erkenntnisse der oben beschriebenen Studien dazu, um ein vernetztes Sensor-Aktuator-System zu entwerfen, welches als Notfallschutzsystem innerhalb von Gebäuden zum Einsatz kommen soll. Es soll Personen den kürzesten sicheren Pfad zu den Notausgängen anzeigen. Das Auffinden dieser Pfade erfolgt dabei verteilt basierend auf den Messungen der einzelnen Knoten, die über das gesamte Netzwerk disseminiert werden.

"Across the Dark Continent sound the never-silent drums:
the base of all the music, the focus of every dance;
the talking drums, the wireless of the unmapped jungle."

Irma Wassall

Contents

List of Figures

List of Algorithms

"A problem well stated is a problem half solved."

Charles Franklin Kettering

1 Introduction

Information dissemination in communication networks is a key function whose effectiveness depends both on the chosen dissemination algorithm and on the underlying network topology. It is needed, for example, in route discovery, link state advertisements, autoconfiguration, and query propagation in ad-hoc networks and peer-to-peer systems.

The process of (re-)transmitting a message can occur as a point-to-point or point-to-multipoint transmission. Point-to-point communication means that a node sends a message to only one of its neighbors. The node may then retransmit the same message to distinct neighbors using multiple transmissions. This model is applied, for example, in peer-to-peer networks. Point-to-multipoint communication implies that all neighbors of a transmitting node will receive the transmitted message (assuming no errors). This model is usually used in studies of networks with a broadcast medium such as wireless networks.

The main focus of this work is to understand how to disseminate efficiently a message from a source node to all other network nodes using point-to-multipont communications. We target two goals: one is to deliver a source message to all network nodes with high probability; the other one is to use as few transmissions as possible for a given target reachability. We address mainly probabilistic dissemination algorithms.

1.1 – Information Dissemination in Communication Networks

Until recently, information dissemination resorted to replication based forwarding where nodes replicate and forward the information they receive. A naïve way of disseminating a message to all nodes in a network with point-to-multipoint transmissions is pure flooding. When a node receives a broadcast message for the first time it will always forward it. In a network with n nodes, the number of transmissions of a source message using pure flooding is n. This technique leads to a high number of redundant and unnecessary transmissions.

A natural optimization goal that arises is to minimize the number of transmissions while achieving global outreach of the message sent. Finding an optimum scheme for disseminating a message in a given network with minimum overhead — i.e., finding the minimum connected dominating set — is however NP-complete [vJPHE02].

Two main classes of approximation algorithms were proposed to improve the efficiency of pure flooding. The first class comprises deterministic algorithms approximating connected dominating sets of networks [AWF02, LW02, AQL02]. Reference [AQL02] proposes a deterministic algorithm, which approximates the connected dominating set within a two-hop neighborhood of each node, thus forming a backbone of forwarding nodes and limiting the number of transmissions. The idea of using such a sub-set of nodes, also called *Multipoint relays* (MPR), has been implemented successfully in the Optimized Link State Routing (OLSR) protocol [CJA+03] for mobile ad-hoc networks.

The second class comprises probabilistic algorithms that introduce a stochastic element to the message forwarding process; these class of algorithms are commonly denoted as Probabilistic Flooding (PF) or gossiping [SCS03, HHL06, SB07, BBB+08, SRS07]. [1]

The study of reachability in PF with point-to-point communications has received good attention from the research community both by analysis and simulation (see [OKS10, GS11] and references therein). PF with point-to-multipoint communications, however, has mainly been addressed by means of simulations [HHL06, SCS03, KWB01, YOKM+06, YOKP05, SRS07]. Some of these studies yield better insight into the behavior of PF — with inspiration from percolation theory — but most conclusions do not generalize beyond the particular setup [SCS03, HHL06, KWB01].

[1] Some authors use both terms to refer to the same concept. Others use them in a way that in PF a node forwards a message to all its neighbors (point-to-multipoint communications), while in gossiping a node forwards a message to only one neighbor (point-to-point communications).

The seminal paper of Ahlswede, Li, Cai, and Yeung [ACLY00], where they proved that the max-flow min-cut capacity of a general multicast network can only be achieved by allowing intermediate nodes to mix different data flows, landmardked the appearance of the new *Network Coding* paradigm. Network coding research suggests that further reductions in the number of transmissions required for flooding could be achieved due to the ability of intermediate nodes to mix multiple messages through algebraic operations. More specifically, reference [FWLB06] quantifies these gains for ring and square lattice topologies, and presents a heuristic algorithm which outperforms probabilistic flooding for a class of random geometric graphs. Related work on the benefits of network coding includes a proof that the minimum energy single-source multicast problem with network coding becomes solvable in polynomial-time [LMHK04] and in a distributed manner [LRK+05]. The problem of multiple multicasts, which is closer to flooding, remains however an open problem [LRM+06].

1.2 – Main Contributions

Modeling networks as random graphs built from stochastic processes and using methods from graph theory and stochastic geometry we address both replication and network coded information dissemination paradigms. We perform both analytical and numerical studies of "state-of-the-art" probabilistic dissemination algorithms, comparing their performance with deterministic algorithms — where performance is measured by the number of transmissions needed to disseminate a message and by the reachability of a dissemination process.

Moreover, we apply the insights gained from the analysis of the information dissemination algorithms in the design of a sensor-actuator networked system for emergency response in indoor scenarios.

The main contributions of this thesis are as follows:

- We present a generic approach to estimate the probability of achieving global outreach with PF with a network-wide common forwarding probability. We derive an exact expression for the global outreach probability over Erdős Rényi Random graphs and an asymptotic expression for the global outreach probability in Random Geometric graphs that constitutes a good approximation at high node density. We address both reliable and unreliable links.

- We characterize analytically the transmission cost of network coded flooding over

Erdős Rényi Random graphs, Random Geometric graphs, and Small-World Networks. Moreover, we present a numerical study for the number of transmissions, delivery ratio, and delay of network coded and MPR flooding. We also analyze the interplay between the network topology with replication and network coded based flooding algorithms.

- We propose graph theoretical abstractions of a sensor network for supporting building evacuation in disaster scenarios and algorithms for the computation of shortest safe path to exits. The wireless sensor network collects hazard information that is flooded throughout the network. This information is used as input for the computation of the shortest safe paths for leaving the building. We implement a prototype of the system in which we evaluate the proposed solutions.

1.3 – Thesis Outline

The remainder of this thesis is organized as follows. In Chapter 2 we introduce the fundamental concepts and mathematical techniques used throughout the dissertation. In Chapter 3 we take a mathematical approach to the analysis of PF. We determine analytically how PF with constant forwarding probability behaves over mathematically tractable abstractions of a network, namely random graph models. In Chapter 4 we analyze network coded based flooding techniques and compare them to replication based deterministic flooding. We seek to understand which benefits in terms of number of transmission per message, reachability, and delay may we obtain from performing algebraic mixing of messages in intermediate forwarding nodes during an information dissemination process. In Chapter 5 we address the problem of designing a sensor network system for the support of building evacuation in disaster scenarios. We characterize the problem with the help of graph models and we propose algorithms that use flooded hazard information to compute shortest paths to safely leave the building. Finally, Chapter 6 concludes this dissertation, including possible directions of future work.

Parts of the work presented in this thesis were previously published in [CBB08a, CBB08b, CSBB09, CSBB12]. New unpublished results are also presented.

"Facts do not cease to exist because they are ignored."

Aldous Huxley

"We don't live in a world of reality,
we live in a world of perceptions."

Gerald J. Simmons

2

Models and Tools

This chapter provides an overview of the analytical tools and modeling assumptions considered in the work presented in this thesis.

2.1 – Definitions from Graph Theory

Let $G = (V, E)$ be a *graph* with a set of nodes V and a set of edges $E \subseteq \{\{u, v\} : u, v \in V, u \neq v\}$. The number of nodes of G, called the *order* of G, is denoted by $n = |V|$.

A node v is called *neighbor* of u if there exists an edge $\{u, v\} \in E$. The *degree* $d(u)$ of a node u is the number of edges adjacent to u, i.e., the number of neighbors of u. A *path* in a graph is a sequence of nodes such that from each of its nodes there is an edge to the next node in the sequence. The *length* of a path is the number of edges traversed by the path. A *shortest path* between two nodes is a path such that its length is minimum. A graph G is called *connected* if there is a path between any two distinct nodes $u, v \in V$.

The *distance* $L_{u,v}$ between a pair of nodes $\{u, v\}$ in a graph G (also known as the *geodesic distance*) is the number of edges in a shortest path connecting them. The *diameter* of a

graph G is the longest geodesic distance in the graph. The *average distance* L for a whole graph G is the average of the distances between every distinct pair of nodes in G. A node with distance $l \in \mathbb{Z}^+$ to a node u is called *l-hop neighbor* of u. The *l-hop neighborhood* $N^l(u)$ of a node u is the set of l-hop neighbors of u. The *1-hop neighborhood* $N^1(u)$ (or $N(u)$) of a node u is also called the *neighborhood* of u.

The *clustering coefficient* C_u for a node u is the ratio between the number of edges connecting the nodes within its neighborhood and the maximum number of edges that could connect them. The clustering coefficient C for a whole graph G is the average of the clustering coefficients of each node in G.

A *subgraph* $G' = (V', E')$ of G is a graph with $V' \subseteq V$ and $E' \subseteq E$. An *induced subgraph* G' of G is a subgraph in which for any pair of nodes $u, v \in V'$, $\{u, v\}$ is an edge of E' whenever $\{u, v\}$ is an edge of E (i.e. $\forall u, v \in V' : \{u, v\} \in E \Rightarrow \{u, v\} \in E'$). A maximal connected subgraph $G' = (V', E')$ is an induced connected subgraph of $G = (V, E)$ that no longer satisfies the property of being connected when adding an additional node from $V \backslash V'$ and the corresponding edges. A maximal connected subgraph of G is called a *connected component* of G. V' is a *dominating set* if all nodes u that are not within V' have an edge to a node $v' \in V'$, i.e. $\forall u \in V \backslash V' \, \exists v' \in V' : \{u, v'\} \in E$. If additionally the induced subgraph $G' = (V', E')$ is connected, the node set V' is a *connected dominating set*.

2.2 – Stochastic Geometry and Point Processes

The spatial location of the nodes of a network can be modeled deterministically, or in a probabilistic manner. Typical deterministic models include grid networks, line networks, and triangular lattices, which are applicable when the devices share a structured distribution. When this is not the case, the uncertainty on the location of nodes can be captured by spatial stochastic models that may integrate several performance metrics.

Point processes

A Point Process (PP) can be thought of as a random set of points $\{x_1, x_2, \ldots, x_n\}$ in a plane. Mathematically, a point process, Π, is a measurable mapping from some probability space $[\Omega, \mathcal{A}, \mathbf{P}]$ into $[\mathbb{N}, \mathcal{N}]$, where $\mathbb{N}$ is the family of all sequences $\Pi = \{x_i\}$ of points in $\mathbb{R}^2$ that satisfy two conditions [SKM85, Chap. 4]: (1) the sequence Π is locally finite, i.e., each bounded set of $\mathbb{R}^2$ must contain only a finite number of points of Π; (2) the sequence Π is simple, meaning that there is no accumulation of points, i.e., $x_i \neq x_j$ if $i \neq j$. $\mathcal{N}$

is the smallest α-algebra on $\mathbb{N}$ that makes all mappings $\Pi \mapsto \Pi(B)$ measurable, where B is a bounded Borel set. A realization of a PP is a random choice of one of the sequences in $\mathbb{N}$. The total number of points of a PP falling in a given region of space is a random variable, and, therefore, the number of points within a certain region of $\mathbb{R}^2$ can be analyzed probabilistically.

Poisson Point Process

The Poisson Point Process (PPP) is the simplest and most important class of spatial processes, and accounts for both regular distribution of points (homogeneous PPP) as well as more irregular deployments (inhomogeneous PPP).

A stationary/homogeneous PPP models the deployment of a regular set of points, in which no preference is given to specific regions of the plane. The homogeneous PPP is characterized by a density parameter λ, and the expected number of points in a given region $\mathcal{R}$ follows a Poisson distribution with parameter $\lambda \cdot A(\mathcal{R})$, where $A(\mathcal{R})$ is the area of $\mathcal{R}$. The expected number of points within a given region then varies with the density λ and the area of that region. The probability of n nodes being inside a region $\mathcal{R}$ is given by ([Kin93])

$$\mathrm{P}(n \text{ nodes in } \mathcal{R}) = \frac{(\lambda \cdot A(\mathcal{R}))^n}{n!}\, e^{-\lambda \cdot A(\mathcal{R})}. \tag{2.1}$$

A realization of spatial locations according to a Poisson point process with density λ can be obtained by first drawing a random number N of points from a Poisson distribution with parameter $\lambda \cdot A(\mathcal{R})$, followed by scattering those N points uniformly at random inside $\mathcal{R}$. These points can represent locations of the nodes of a network and we will henceforth refer to them as nodes.

2.3 – FKG Inequality

The FKG inequality [FKG71] expresses positive correlations between increasing events (see Ch. 2.2, [Gri99]). Consider two realizations $\mathcal{P}_1$ and $\mathcal{P}_2$ of a Poisson point process. We define a partial ordering $\mathcal{P}_1 \preceq \mathcal{P}_2$ if and only if every point of $\mathcal{P}_1$ is also present in $\mathcal{P}_2$. An event A is an increasing event if for every $\mathcal{P}_1 \preceq \mathcal{P}_2$, the indicator function I_A of the event A respects the relation $I_A(\mathcal{P}_1) \leq I_A(\mathcal{P}_2)$. If A and B are increasing events in a Poisson point process, then $\mathrm{P}(A \cap B) \geq \mathrm{P}(A)\,\mathrm{P}(B)$.

2.4 – Poisson Approximation by the Chen-Stein Method

The Poisson distribution arises as the limiting distribution of the sum of n low probability independent Bernoulli random variables. The Chen-Stein method generalizes this result to dependent random variables, as long as these dependencies become negligible as n converges to infinity. References [Che75, AGG89, AGG90] present this method, showing how to calculate a bound for the error of this approximation.

The *total variation distance* between distributions of two integer-valued random variables X, Y is (Ch. 1.6, [Pen03])

$$d_{\mathrm{TV}}(X,Y) \triangleq \sup_{A \subseteq \mathbb{Z}} \left| \mathrm{P}\left(X \in A\right) - \mathrm{P}\left(Y \in A\right) \right|. \tag{2.2}$$

A sequence X_n of integer-valued random variables converges in distribution to X if

$$\lim_{n \to \infty} d_{\mathrm{TV}}(X_n, X) = 0. \tag{2.3}$$

Suppose X_i with $i \in \mathcal{I}$ are Bernoulli random variables with $\mathrm{E}(X_i) = p_i$, where $\mathcal{I}$ is an arbitrary index set. Assume $W \triangleq \sum_{i \in \mathcal{I}} X_i$ and $\mathrm{E}(W) = \sum_{i \in \mathcal{I}} p_i$ is finite. A subset $\mathcal{N}_i \subset \mathcal{I}$ is a *neighborhood of dependence* of $i \in \mathcal{I}$ if for each X_j dependent of X_i, it follows that $j \in \mathcal{N}_i$. Let

$$b_1 \triangleq \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i} \mathrm{E}(X_i)\, \mathrm{E}(X_j), \tag{2.4}$$

$$b_2 \triangleq \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i, j \neq i} \mathrm{E}(X_i X_j). \tag{2.5}$$

Then we have $d_{\mathrm{TV}}\left(W, \mathrm{Po}\left(\mathrm{E}(W)\right)\right) \leq 2\left(b_1 + b_2\right)$.

2.5 – Wireless Link Model

In a wireless electromagnetic channel, nodes communicate thorough signals conveyed via electromagnetic waves. With free space propagation, using a constant transmitting power and fixed signal wavelength, the received signal strength will only depend on the distance between the sender and the receiver.

Quite often, the energy radiated from the transmitter encounters multiple obstacles before reaching an intended receiver. These obstacles cause reflection, diffraction, and scattering of the electromagnetic waves. Moreover, during a transmission, the sender, the receiver, and/or the obstacles may move. The receiver will, therefore, be subject to the combined interference of multiple electromagnetic waves, which varies in time and space in a non deterministic way.

The difficulty in accurately characterize the propagation of wireless signals in environments with obstacles and mobility led to the proposal of different models for predicting the received power [Gol05, Sch05]. These models may be classified as deterministic or statistical. Deterministic models typically employ ray tracing techniques, or are derived from empirical measurements. Statistical models are typically employed when the complexity of the environment (e.g. obstacles, multipath propagation, mobility) makes it very difficult to apply deterministic models. Examples are the log-normal shadow fading model for long-term or large distance variations of the instantaneous signal strength around average power, and Rayleigh/Rician fading model for short-term or short distance multipath fading which captures power variations on a wavelength scale [Sch05].

Next we will present a brief overview of the Free-Space Path Loss Model, the Simplified Path Loss Model, the Combined Path Loss and Shadowing Model, and the Combined Path Loss and Shadowing Model. The first three models lead to the definition of purely deterministic geometric wireless links between any pair of nodes. I.e. the distance between two nodes determine wether they establish a wireless link or not. The Combined Path Loss and Shadowing Model leads to a wireless link characterized by a geometric component (determined by the distance between nodes) plus an associated erasure probability which is may be derived from the log-normal distributed random component of the path loss.

2.5.1 Free-Space Path Loss Model

With free space propagation, if a node u transmits with power $p_t(u)$, the power $p_r(v)$ of the signal received by a node v located at distance $d(u,v)$ from u is given by (see [Mad08] pg. 133):

$$p_r(v) = p_t(u)\, g(u)\, g(v) \left(\frac{v_c}{4\pi\, f_0\, d(u,v)} \right)^2 . \tag{2.6}$$

In the above expression, $g(u)$ and $g(v)$ are the antenna gains of transmitter and the receiver, respectively, v_c is the speed of light, and f_0 is the center frequency of the transmitted signal.

The transmitted signal is received properly if the power received signal $p_r(v)$ is larger than or equal to some threshold power $p_{thr}(v)$, denoted as *receiver sensitivity*. Hence, whenever $p_r(v) \geq p_{thr}(v)$ we say that there is a link between u and v.

2.5.2 Simplified Path Loss Model

For general tradeoff analysis of various system designs it is sometimes best to use a simple model that captures the essence of signal propagation. Thus, the following simplified model for path loss as a function of distance is commonly used for system design (see [Gol05] pg. 46):

$$p_r(v) = p_t(u)\, g_{ref} \left(\frac{d_{ref}}{d(u,v)} \right)^{\gamma}. \tag{2.7}$$

In the above expression, g_{ref} is a unitless constant which depends on the antenna characteristics and the average channel attenuation, d_{ref} is a reference distance for the antenna far-field. Parameter γ is the *path loss exponent* which depends on the propagation environment. Its value can be obtained via a minimum mean square error fit to empirical measurements. Typically it ranges from 2 in free space to 6 in environments with obstacles such as within office buildings with multiple floors (see [Gol05] pg. 47).

Due to scattering phenomena in the antenna near-field, Equation (2.7) is generally only valid at transmission distances $d(u,v) > d_{ref}$, where d_{ref} is typically assumed to be $1-10$ m indoors and $10-100$ m outdoors. The value of $g_{ref} < 1$ is sometimes set to the free space path loss at distance d_{ref}:

$$g_{ref} = \left(\frac{v_c}{4\pi\, f_0\, d_{ref}} \right)^2. \tag{2.8}$$

2.5.3 Combined Path Loss and Shadowing Model

In addition to path loss, a signal will typically experience random variation due to blockage from objects in the signal path, giving rise to a random variation about the path loss at a given distance. In addition, changes in reflecting surfaces and scattering objects can also cause random variation about the path loss. This effect is called *shadowing*. Since the location, size, and dielectric properties of the blocking objects as well as the changes in reflecting surfaces and scattering objects that cause the random attenuation are generally unknown, statistical models are widely used to characterize this attenuation. The most common model for this additional attenuation is log-normal shadowing. Therefore, in the combined path loss and shadowing model, the instantaneous reception power is a random variable with its mean value given in Equation (2.7). The power $p_r(v)$ of the signal received by a node v located at distance $d(u,v)$ from the transmitter u is

$$p_r(v) = p_t(u)\, g_{ref}\, g_s \left(\frac{d_{ref}}{d(u,v)} \right)^{\gamma}. \tag{2.9}$$

where g_s is a log-normal distributed random variable. In other words, the random variable $g_s^{(\mathrm{dB})} = 10\log(g_s)$ measured in dB is Gaussian distributed with mean zero and variance σ_s^2

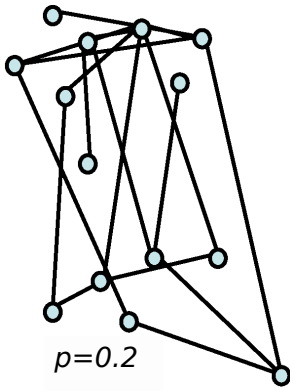


Figure 2.1: Erdős Rényi Random Graph with 14 nodes and edge probability $p = 0.2$

with a probability density function given by

$$f_{g_s^{(\mathrm{dB})}}(x) = \frac{1}{\sqrt{2\pi\sigma_s^2}}\, \mathrm{e}^{-\frac{x^2}{2\sigma_s^2}}\,, \tag{2.10}$$

Typical values of σ_s^2 range from 6 dB to 10 dB [Sch05].

2.6 – Network Models

In our study of dissemination algorithms we model networks as graphs generated by stochastic processes. We consider three graph models: (a) Erdős Rényi Random Graphs (ERGs) that are random graphs in which the existence of an edge between any pair of nodes is the outcome of a Bernoulli trial with parameter p; (b) Random Geometric Graphs (RGGs) for which the existence of an edge between a given pair of nodes is function of the geometric distance between them; (c) Small-World Networks (SWNs) that can be seen as an interpolation between regular graphs and random graphs.

2.6.1 Erdős Rényi Random Graphs

An ERG is a graph $G = (V, p)$ with node set V and edge set E built by independently sampling with probability p every element of the set $\{\{u, v\} : u, v \in V, u \neq v\}$ [Bol98] (see Fig. 2.1). The node degree is binomially distributed according to $\mathrm{Bin}(n-1, p)$ with an expected value $\mathbb{E}(d(v)) = (n-1)\,p$.

The reason we consider ERGs for network modeling is twofold: First, it is a simple model for networks that enables us to derive exact results. Secondly, wireless networks where

shadow fading is the dominant component of wireless propagation (e.g., many obstacles, indoor environments) may be well modelled by ERGs. The existence of a wireless link (edge in the graph model) between a given pair of nodes may be determined by the outcome of a Bernoulli trial in which the parameter p. This parameter is function of the variance of the log-normal distributed shadowing component g_s and the receiver sensitivity $p_{thr}(v)$, assuming that all nodes $v \in V$ have the same receiver sensitivity and use equal transmission power [GHH06, MAA08, FB08].

2.6.2 Random Geometric Graphs

RGGs constitute a class of graphs in which nodes are deployed over some region of an n-dimensional space, and for which the existence of an edge between any pair of nodes is fully determined by their geometric distance (see Fig. 2.2). RGGs are often used to model wireless outdoor scenarios without shadowing.

We restrict to square regions of some area A in a 2-dimensional space. Moreover, we use the following distance metrics:

- Euclidean distance metric $d_E(u,v) \triangleq \|u - v\|$;
- Toroidal distance metric $d_T(u,v) \triangleq \min_{z\in\mathbb{Z}^2} \|u - v + \sqrt{A}\, z\|$ (see [Pen03]).

The use of a toroidal distance metric avoids edge effects (Ch. 8, [Cre91]), thus simplifying the analysis for this network model. This approach is commonly used in the literature [MAA08, TMA09].

A set of distinct nodes $\{u, v\}$ will form an edge if $d_E(u,v)$ (or $d_T(u,v)$) is smaller than or equal to r, with $r > 0$. The parameter r models the node-independent transmission *range*.

In this thesis we consider RGGs in which node positions are generated by stochastic point processes.

Binomial Random Geometric Graphs

A *Binomial Random Geometric Graph* (B-RGG) [Pen03] is a random graph $G(V, r)$ in which a set of n nodes V is deployed, independently and uniformly at random, over a square $S_A \subset \mathbb{R}^2$ of area A. If we use the Euclidean distance, the edge set E of G contains all sets of nodes $\{u, v\}$ of $\{\{u, v\} : u, v \in V, u \neq v\}$, for which $d_T(u,v) \leq r$, where $r > 0$. Alternatively, if we use the toroidal distance, the edge set E of G contains all sets of nodes $\{u, v\}$ of $\{\{u, v\} : u, v \in V, u \neq v\}$, for which $d_T(u,v) \leq r$, where $r > 0$.

The number of nodes in a region $\mathcal{R} \subset S_A$ is binomially distributed according to $\mathrm{Bin}(n, \frac{A(\mathcal{R})}{A})$.

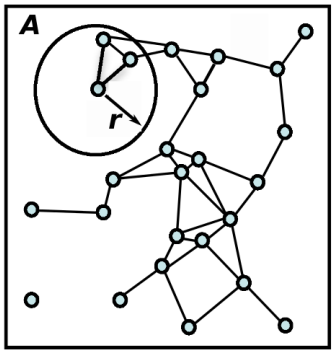


Figure 2.2: Random Geometric Graph in a square of area A, with 25 nodes and transmission range r.

Poisson Random Geometric Graphs

A *Poisson Random Geometric Graph* (P-RGG) $G(A, \lambda, r)$ is defined in the following way. Consider a Poisson point process Π of intensity $\lambda > 0$ on a $\sqrt{A} \times \sqrt{A}$ square $S_A \subset \mathbb{R}^2$, with a node located at each point generated by Π. We construct G by defining its set of nodes V as the nodes generated by Π. If we use the Euclidean distance, the edge set E of G contains all sets of nodes $\{u, v\}$ of $\{\{u, v\} : u, v \in V, u \neq v\}$, for which $d_T(u, v) \leq r$, where $r > 0$. Alternatively, if we use the toroidal distance, the edge set E of G contains all sets of nodes $\{u, v\}$ of $\{\{u, v\} : u, v \in V, u \neq v\}$, for which $d_T(u, v) \leq r$, where $r > 0$.

The number of nodes N of G is a Poisson random variable with mean $A\lambda$. Moreover, the number of nodes in a region $\mathcal{R} \subseteq S_A$ is Poisson distributed according to $\mathrm{Po}(\lambda)$.

2.6.3 Small-world Networks

SWNs [WS98] correspond to a class of random graphs which exhibit high clustering coefficients (i.e. neighboring nodes are likely to be connected) and small average path length — the diameter of a graph with n nodes is in fact bounded by a polynomial in $\log n$. The term "Small-world networks" itself was coined by Watts and Strogatz [WS98], who defined a class of models which interpolate between regular lattices and ERGs by introducing long-range shortcuts with a certain probability p. The most salient feature of these models is that for increasing values of p the average shortest-path length diminishes sharply, whereas the clustering coefficient remains practically constant during this transition.

SWN topologies are becoming potentially attractive in the context of communication networks, potentiating the spread of information. Resource discovery in wireless networks [Hel03], design of heterogeneous networks ([RKV04, DYT05]), and application to overlay networks for peer-to-peer communications ([MNW04, HLY04]) are just a few examples on which SWN topological properties are deemed to be particularly useful. Reference [New03] surveys com-

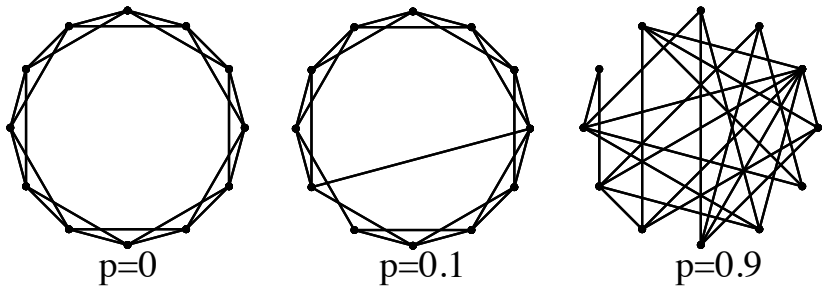


Figure 2.3: Small-World model with rewiring with 12 nodes and $k = 4$ for different values of the rewiring probability p.

munication processes in complex networks (including SWNs), such as search and navigation, network transmission, epidemics, and information dissemination processes.

In the original model of Watts and Strogatz (SWN with Rewiring [WS98]), shortcuts are introduced by rewiring the edges of the original ring lattice with a certain probability p. A typical variant was introduced by Newman and Watts (SWN with added Shortcuts [NW99]) where instead of reconnecting existing edges, new edges are added with probability p.

Kleinberg [Kle00] introduced an SWN model that has the property of navigability, where short paths not only exist, but can also be easily found using merely local information. The model consists of a grid to which shortcuts are added not uniformly but according to a harmonic distribution, such that the number of outgoing links per node is fixed and the link probability depends on the distance between the nodes. For this class of SWNs a greedy routing algorithm, in which a message is sent through the outgoing link that minimizes the distance to the destination, was shown to be effective.

Small-World Network with Rewiring A *Small-World Network with Rewiring* $G = (V, k, p)$ with set of nodes V, initial node degree k, and rewiring probability p is constructed as follows (see Fig. 2.3): the initial graph is a one-dimensional lattice of n nodes, with periodic boundary conditions (i.e. a ring), each node being connected to its k-hop neighborhood. The nodes are then visited one after the other; each edge connecting a node to one of its $k/2$ nearest neighbors in the clockwise sense is left in place with probability $1 - p$, and with probability p is reconnected to a randomly chosen other node.
From now on we will use the acronym SWN to specifically refer to *Small-World Networks with Rewiring*.

2.7 – Random Linear Network Coding

The multicast capacity of a network is only achieved through the use of Network Coding (NC), that is, algebraic mixing of packets in networks. This result was first proved in the seminal paper of Ahlswede, Li, Cai, and Yeung [ACLY00]. Following this, an algebraic framework was introduced by Koetter and Médard in [KM03]. This framework led to the introduction of Random Linear Network Coding (RLNC), a randomized and distributed scheme to perform network coding, with applications to a wide variety of settings. We now give a brief overview of RLNC including its properties and practical applications.

RLNC [HMK+06] is a distributed methodology for performing network coding, in which each node in the network independently and randomly selects a set of coefficients from a finite field and uses them to form linear combinations of the data packets it receives.

These linear combinations are then sent over the outgoing links of each node in the network. Each packet is sent along with the global encoding vector [KM03], which is the set of linear transformations that the original packet goes through on its path from the source to the destination. The global encoding vector enables the receivers to decode the original data using Gaussian elimination. The operations of RLNC for packetized networks are summarized in Algorithm 1 (from [LMKE05b]).

Now, if the coefficients are chosen at random from a large enough field, the resulting matrix is invertible with high probability, which explains why this approach is capable of achieving the multicast capacity of a network.

In [LMKE05a], RLNC is studied from the point of view of asynchronous packet networks. The results show that RLNC is capacity-achieving even on asynchronous lossy packet networks, provided that the packets received on a link arrive according to a process that admits an average rate. This statement holds both for lossy point-to-point (modeling wireline packet networks) and broadcast link models (suitable for wireless packet networks).

Practical Implementation and Applications

A framework for packetized network coding is presented in [CWJ03], which leverages the RLNC resilience against disruptions such as packet loss, congestion, and changes of topology, in order to guarantee robust communication over highly dynamic networks with minimal (or no) control information. The framework defines a packet format and a buffering model.

Algorithm 1 Random Linear Network Coding

Initialization (source node s): s forms the message packets $w_1, w_2, ..., w_r$ according to the same rules that the intermediate nodes use (shown below).
Operation at intermediate node v:
if packet received **then**
Gaussian elimination is performed with packets in buffer. Packet is stored in the buffer.
end if
for all outgoing edges **do**
Node v chooses all the packets $p_1, p_2, ..., p_L$ that are in his buffer.
Form packet $x := \sum_{l=1}^{L} \alpha_l p_l$, where α_l is chosen according to a uniform distribution over the elements of the finite field $\mathbb{F}_q$. The packet's global encoding vector γ, which satisfies $x := \sum_{k=1}^{r} \gamma_k w_k$, is placed in its header.
Send packet x.
end for
Decoding (sink nodes):
if packet received **then**
Gaussian elimination is performed with the packets already in the buffer.
if inverse of the matrix M^{-1} exists **then**
The sink node applies the inverse to the packets to obtain $w_1, w_2, ..., w_r$; otherwise, a decoding error occurs.
end if
end if

The packet format consists of the *global encoding vector* (kept in the header) and the payload, which is divided into vectors according to the field size (2^8 or 2^{16}, i.e. each symbol has 8 or 16 bits, respectively). Each of these symbols is then used as a building block for the linear operations performed by the nodes.

The buffering model divides the stream of packets into *generations* of size h, such that packets in the same generation are tagged with a common generation number. Each node sorts the incoming packets in a single buffer according to their generation number. When there is a transmission opportunity at an outgoing edge, the sending node generates a new packet, which contains a random linear combination of all packets in the buffer that belong to the *current* generation. If a packet is *non-innovative*, i.e. if it does not increase the rank of the decoding matrix available at the receiving node, then it is immediately discarded. As soon as the matrix of received packets has full rank, Gaussian elimination is performed at the receivers to recover the original packets.

RLNC seems particularly beneficial in dynamic and unstable networks — that is, networks where the structure or topology of the network varies within a short time, such as mobile ad-hoc networks and peer-to-peer content distribution networks. RLNC has been shown to extend naturally to packet networks with losses [LMKE05b], and, simultaneously, to provide increased resilience against failures in the network [HMK+06]. The inherent robustness properties of RLNC make it particularly suitable as a framework for dynamic and unstable networks, such as delay tolerant networks [WLB05] and content distribution

networks [GR05, DGWR07].

The benefits of RLNC in wireless environments with rare and limited connectivity, either due to mobility or battery scarcity, are highlighted in [WLB05, FWLB06], which propose an algorithm aimed at reducing the overhead of probabilistic flooding algorithms with applications in delay tolerant networks. Other RLNC based information dissemination schemes that increase reliability and robustness while reducing the incurred overhead can be found in [EFSC+06, DEH+05]. Since each node forwards a random linear combination independently of the information present at other nodes, its operation is completely decentralized. Moreover, when collecting a random combination of packets from a randomly chosen node, there is high probability of obtaining a linearly independent packet in each time. Thus, the problem of redundant transmissions, which is typical of traditional flooding approaches, is considerably reduced.

"Do not speak — unless it improves on silence."

Buddhist Sayings

"God does not play dice."

Albert Einstein

3

Probabilistic Flooding in Stochastic Networks

In Probabilistic Flooding (PF) with error-free point-to-multipoint communications, the transmission of a node is received by all its neighbors. The source node transmits a so called source message. Each of its neighbors then forwards the received message with some probability that may be common to all nodes, different for each node, or even adaptive.

In contrast to deterministic algorithms, PF algorithms do not guarantee that all nodes of a connected network will receive a flooded message even under ideal conditions (collision-free MAC and error-free propagation medium). The set of forwarding nodes of the communication subgraph generated by the PF process of a message needs to be a connected dominating set of the network in order to achieve global outreach. This in turn is only assured if the forwarding probability of all nodes equals one (pure flooding). Besides this special case, global information outreach can only be achieved with a probability smaller than one.

The study of reachability in PF with point-to-multipoint communications, however, has mainly been addressed by means of simulations. In this chapter, we take a mathematical approach to the analysis of PF. Our aim is to determine analytically how simple PF with constant forwarding probability behaves over mathematically tractable abstractions of a network, namely random graph models. We consider Erdős Rényi Random Graphs (ERGs)

and Poisson Random Geometric Graphs (P-RGGs) (or simply Random Geometric Graphs (RGGs) within this chapter; see Section 2.6). The results may be used as a reference for the study and development of more sophisticated algorithms.[1] In the following we refer to PF with point-to-multipoint communication only by PF, as our work addresses exclusively the reachability using this model.

Consider a flooding algorithm in which each node forwards a received message with a network-wide forwarding probability ω. We ask: How small can ω be while still achieving global outreach with high probability? We answer this question using methods from graph theory and stochastic geometry, and make the following main contributions:

- Presentation of a generic approach to estimate the probability of achieving global outreach;
- Derivation of an expression for the global outreach probability in ERGs;
- Derivation of an asymptotic expression for the global outreach probability in RGGs that constitutes a good approximation for dense RGGs;
- Detailed analysis of the ω required to achieve global outreach with high probability in ERGs and RGGs;
- Analysis of the global outreach probability of PF with unreliable transmission medium for ERGs and RGGs;
- Study of the border effects in RGGs and proposal of a PF heuristic that minimizes these effects.

The chapter is organized as follows. Section 3.1 describes the PF algorithm and gives the problem statement. Section 3.2 presents an analytical approach to compute the probability of global outreach. Section 3.3 employs this approach in ERGs, leading to an expression for the global outreach probability in such networks. In addition, we show (n, p, ω)-tuples leading to global outreach with high probability. Section 3.4 addresses RGGs, leading to an asymptotic expression for the global outreach probability. We also perform a numerical study, comparing analytical and simulation results evaluating the accuracy of the derived expressions. Again, we show (λ, A, r, ω)-tuples leading to global outreach with high probability. Section 3.5 studies border effects of RGGs on the global outreach probability and proposes a modification to the PF algorithm to address these effects. Section 3.6 presents analysis of global outreach probability in the presence of unreliable transmission medium. Section 3.7 discusses the achieved results in comparison to related work. Finally, Section 3.8 summarizes the main results of the chapter.

[1] E.g. PF with the forwarding probability given by some probability distribution; as function of local topology parameters; or as function of the dynamics of the dissemination process.

Main results of this chapter were published in [CSBB09, CSBB12] in collaboration with Christian Bettstetter, João Barros, and Udo Schilcher. Sections 3.1-3.8, excepting minor changes, are taken from [CSBB12], which were written with the corresponding co-authors.

3.1 – Probabilistic Flooding and Problem Statement

3.1.1 Probabilistic Flooding

A naïve way of disseminating a message to all nodes in a network is pure flooding. When receiving a broadcast message for the first time a node will always forward it. In a network with n nodes, the number of transmissions of a source message using pure flooding is n. This technique leads to a high number of redundant transmissions, which is commonly known as the broadcast storm problem [NTCS99].

Probabilistic flooding is a family of techniques that aim to reduce the number of redundant transmissions, in which the message forwarding is a probabilistic event [SCS03, HHL06, KWB01]. In general, each node v may have a distinct forwarding probability $\omega(v)$. We focus on the simple case where all nodes have the same forwarding probability. Only the source node u transmits the message always with probability 1. I.e. $\omega(v) = \omega \ \forall v \in V \setminus u$. The case $\omega = 1$ is equivalent to pure flooding. Algorithm 2 describes the flooding process.

We assume an error-free broadcast medium, i.e., a transmission from a node will be successfully received by all its neighbors. In this case, for an appropriate choice of ω leading to global information outreach, the expected number of transmissions is reduced from n to $(n-1)\,\omega + 1$.

3.1.2 Problem Statement

Let $G = (V, E)$ represent a network. A source node $u \in V$ intends to deliver a message m_u to all other nodes $v \in V$. The message m_u is disseminated through G using the flooding algorithm $A_{PF}(G, u, \omega)$.

Algorithm 2 Probabilistic flooding $A_{PF}(G, u, \omega)$

Let $G = (V, E)$ be a graph, $u \in V$ be a source node with a source message m_u to be disseminated, and $\omega \in [0, 1]$ be a forwarding probability common to all nodes $v \in V \setminus \{u\}$.

1. A source node u broadcasts its source message m_u.
2. Each node v that receives m_u for the first time re-broadcasts it with probability ω.

We are interested in the forwarding probability ω needed such that all nodes receive m_u with a given probability α. In a more formal way, let $V' \subseteq V$ denote the set of nodes that have received the message m_u after the completion of $A_{PF}(G, u, \omega)$. Our goal is to determine $\min\{\omega : \mathrm{P}(V' = V) \geq \alpha\}$. The term $\mathrm{P}(V = V')$ is the probability that all nodes of the network obtain the message. In the following, it is called *global outreach probability* $\Psi \triangleq \mathrm{P}(V' = V)$.

3.2 – Graph Sampling Approach

We present a generic approach to calculate the global outreach probability. First, using probabilistic flooding, we can construct a communication subgraph $G' = (V', E')$ of the network graph G in the following way: We start with a node set $V' = \{u\}$ containing only the source node and an empty edge set $E' = \{\}$. For each node v that forwards the message, we add all receiving nodes to V'. Additionally, we add edges $\{v, w\}$ between the forwarding node and the receiving nodes $w \in N(v)$ to E'.

Second, we construct an induced subgraph of G, called $G^* = (V^*, E^*)$, using *Graph sampling* (GS) explained in Algorithm 3. G^* helps us to analyze the probability of global outreach for given ω. We show how properties of a random graph G^* are related to those of a random graph G'. We study two properties:

- the event that G^* is connected, denoted as $C(G^*)$;
- the event that the nodes V^* are a dominating set of G, denoted as $D(V^*, G)$.

The event $C(G^*) \cap D(V^*, G)$ means that V^* is a connected dominating set of G.

Theorem 1 (Global outreach)**.** *The probability of global outreach using probabilistic flooding $A_{PF}(G, u, \omega)$ on a network $G = (V, E)$ is equal to the probability that the node set $V^* \subseteq V$ resulting from graph sampling $A_{GS}(G, u, \omega)$ is a connected dominating set of G:*

$$\Psi(G, \omega) = \mathrm{P}\big(C(G^*) \cap D(V^*, G)\big). \tag{3.1}$$

Algorithm 3 Graph sampling $A_{GS}(G, u, \omega)$

Let $G = (V, E)$ be a graph that represents the network and $u \in V$ a source node.

1. The node set V^* is obtained by uniformly sampling the node set $V \backslash \{u\}$ with probability ω and adding u.
2. The edge set E^* contains all edges of G that connect nodes within V^*, i.e. $E^* = \{\{u, v\} \in E : u, v \in V^*\}$.

Proof. A node can decide beforehand whether or not it will participate in the forwarding process if it receives a message. This is equivalent to the sampling process of algorithm A_{GS}. Hence, the set V^* can be associated with the set of nodes forwarding a message according to algorithm A_{PF} if and only if G^* is connected. If the set V^* is also a dominating set of G, all nodes in $V \backslash V^*$ are neighbors of at least one node in V^* and thus receive a message. □

3.3 – Probabilistic Flooding in Erdős Rényi Graphs

This section analyzes the probability of global outreach on an ERG G with n nodes and edge probability p. We derive the expression for the probability of global outreach and present (n, p, ω)-triples leading to a global outreach probability of 0.50, 0.80 and 0.95, respectively.

3.3.1 Derivation of the Outreach Probability

If G belongs to the class of ERGs, the probability of global outreach is given by the following theorem.

Theorem 2 (Global outreach in ERGs)**.** *The probability of global outreach using probabilistic flooding A_{PF} with forwarding probability ω in an ERG with n nodes and edge probability p is*

$$\begin{aligned} \Psi(n,p,\omega) \;=\; & \sum_{k=1}^{n} \mathrm{P_C}(k,\; p) \cdot \left(1-(1-p)^k\right)^{n-k} \\ & \cdot \binom{n-1}{k-1} \omega^{\,k-1} (1-\omega)^{n-k} \quad \textit{with} \end{aligned} \tag{3.2}$$

$$\mathrm{P_C}(m,\, p) = 1 - \sum_{j=1}^{m-1} \binom{m-1}{j-1} \mathrm{P_C}(j,\, p)\, (1-p)^{j\,(m-j)} \tag{3.3}$$

for $m \geq 1$ with starting value $\mathrm{P_C}(1,\, p) = 1$.

Fig. 3.1 plots Ψ over ω along with results from simulations of PF and GS. There is a critical interval of ω-values where Ψ increases from nearly zero to nearly one.

Proof. To prove Theorem 2, we show that connectivity of G^* and domination of G by V^* are mutually independent. Then, we characterize the connectivity and order of G^*, and the probability of domination of G by V^*. Based on this, we derive the global outreach probability.

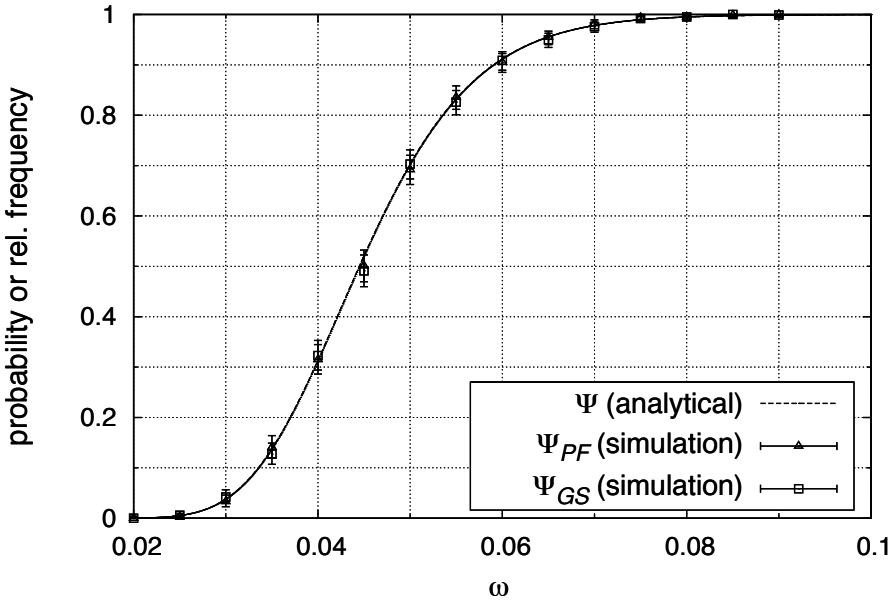


Figure 3.1: Global outreach probability Ψ for PF in ERGs. Parameters: $n = 1000$ nodes, edge probability $p = 0.15$ and message forwarding probability ω. Comparison of simulated PF and GS algorithms with analytical expression for Ψ. Simulation results are obtained from 1 000 experiments. Each experiment is run over a new ERG. Each data point (with its 95% confidence interval limits) represents the relative frequency of the events "achieving global outreach (PF)" or "achieving a connected dominating set (GS)".

Connectivity and domination are independent The probability that V^* of G^* is a connected dominating set of G is

$$\mathrm{P}\big(C(G^*) \cap D(V^*, G)\big) = \mathrm{P}\big(C(G^*)\big) \cdot \mathrm{P}\big(D(V^*, G)\big)\,. \tag{3.4}$$

The event $C(G^*)$ is equivalent to the existence of a path connecting any pair of nodes of G^*. Since a path in G^* is a sequence of consecutive edges of G^*, the sample space of $C(G^*)$ is the set of edges $\{\{u, v\} : u, v \in V^*, u \neq v\}$. The event $D(V^*, G)$ denotes the existence of edges connecting any node in $V \setminus V^*$ to the node set V^*. Thus, its sample space is the edge set $\{\{u, v\} : u \in V^*, v \in V \setminus V^*\}$. In conclusion, since the existence of an edge in an ERG is independent of the existence of any other edge and since the sample spaces of $C(G^*)$ and $D(V^*, G)$ are disjoint edge sets, the events $C(G^*)$ and $D(V^*, G)$ are independent.

Connectivity of G^* Gilbert [Gil59] derived the recurrence relation (3.3) for the probability that an ERG $G(V, p)$ with $m = |V|$ nodes and edge probability p is connected. Therefore, the probability of G^* being connected is

$$\mathrm{P}(C(G^*) \mid N^*) = \mathrm{P_C}(N^*, p). \tag{3.5}$$

Order of G^* Since $V^*\backslash\{u\}$ is a uniformly sampled subset of $V\backslash\{u\}$, the number of nodes N^* is a random variable. N^*-1 is binomially distributed according to $\mathrm{Bin}(n-1,\omega)$. The "-1" stems from the source node sending with probability 1. Thus, the probability mass function of N^* is

$$\mathrm{P}(N^*{=}\,k) = \begin{cases} 0 & \text{if } k=0, \\ \binom{n-1}{k-1}\omega^{k-1}(1-\omega)^{n-k} & \text{otherwise.} \end{cases} \tag{3.6}$$

Domination of G The probability that V^* is a dominating set of G is

$$\mathrm{P}(D(V^*,G)\mid N^*) = \left(1-(1-p)^{N^*}\right)^{n-N^*}. \tag{3.7}$$

For a given node $u\in V\backslash V^*$ the probability of having no edge to any of the nodes in V^* is $(1-p)^{N^*}$. Hence, the probability of an edge to at least one of them is $1-(1-p)^{N^*}$. Since this probability is independent for each node $u\in V\backslash V^*$, (3.7) gives the result.

Proof of Th. 2 Summing the conditional probabilities for connectivity and domination over all possible k of N^*, each of them multiplied with the probability $\mathrm{P}(N^*{=}k)$, gives

$$\Psi{=}\sum_{k=1}^{n}\mathrm{P}(C(G^*)\mid N^*)\cdot\mathrm{P}(D(V^*,G)\mid N^*)\cdot\mathrm{P}(N^*=k). \tag{3.8}$$

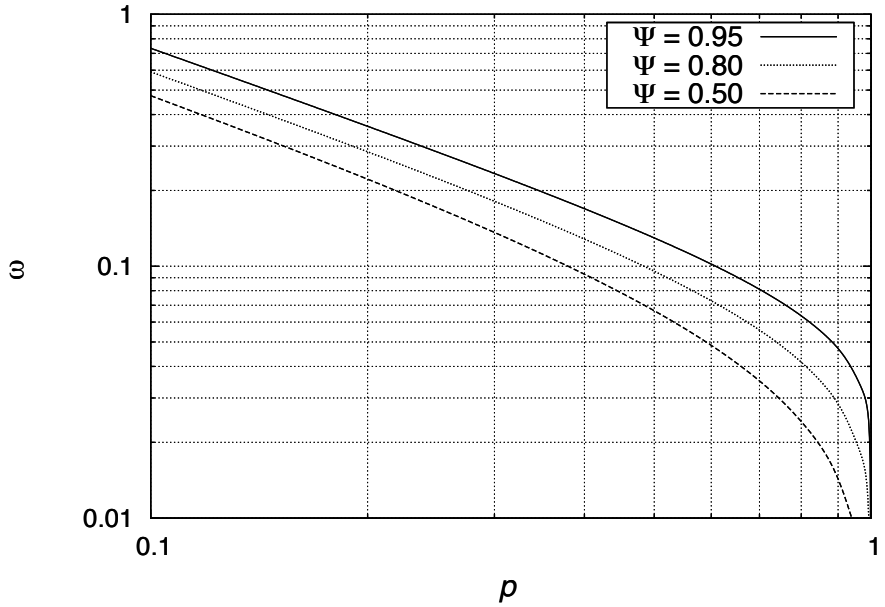
Substituting (3.5), (3.6), (3.7) into (3.8) yields (3.2). □

3.3.2 Parameters for Global Outreach

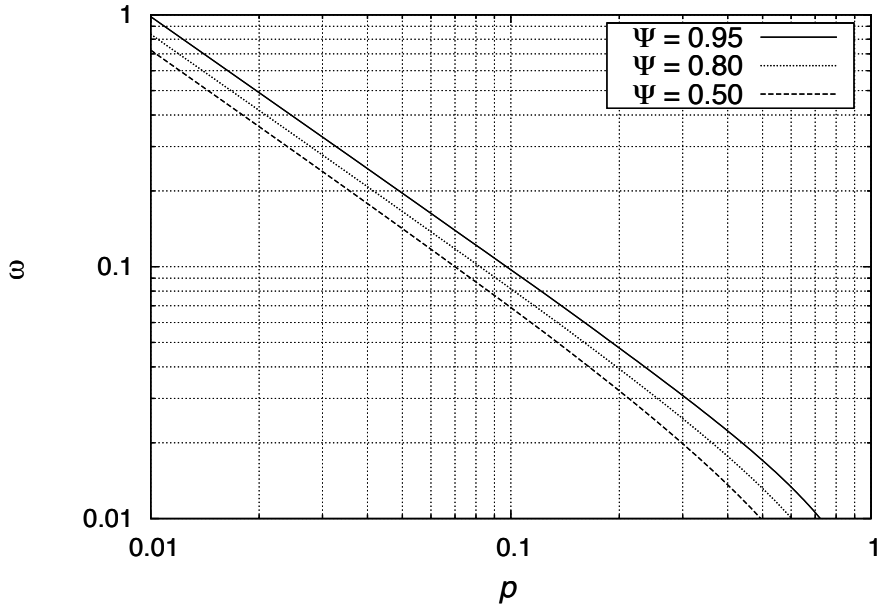
Let us illustrate how these results can be used for network design. The goal is to meet a target value for Ψ by creating or deploying networks and simultaneously tuning ω of the flooding algorithm. In practical applications, one is interested in high outreach probabilities — here we give design options for $\Psi{=}0.50$, 0.80 and 0.95.

If n is given, the parameters p and ω can be chosen. Fig. 3.2 plots the $(p,\,\omega)$-pairs required for achieving a high outreach probability Ψ with $n=100$ and 1000, respectively. The curves show a trade-off in the choice of the $(p,\,\omega)$-pairs. A sparse ERG (low p) requires higher ω-values. For well-connected ERGs (high p), small values of ω are sufficient to guarantee the desired Ψ. The plots also stress the non-linear dependence between these parameters.

If the ω is given, the parameters p and n can be determined. For $\omega=0.08$ and 0.5, Fig. 3.3 shows $(p,\,n)$-pairs ensuring $\Psi=0.50$, 0.80 and 0.95. The dependency between n and p for the same Ψ is non-linear, as expected from Theorem 2. For increasing p, with p close to 0, the required number of nodes experiences an expressive reduction. This trend is then smoothed, and this reduction becomes almost negligible as p approaches 1.

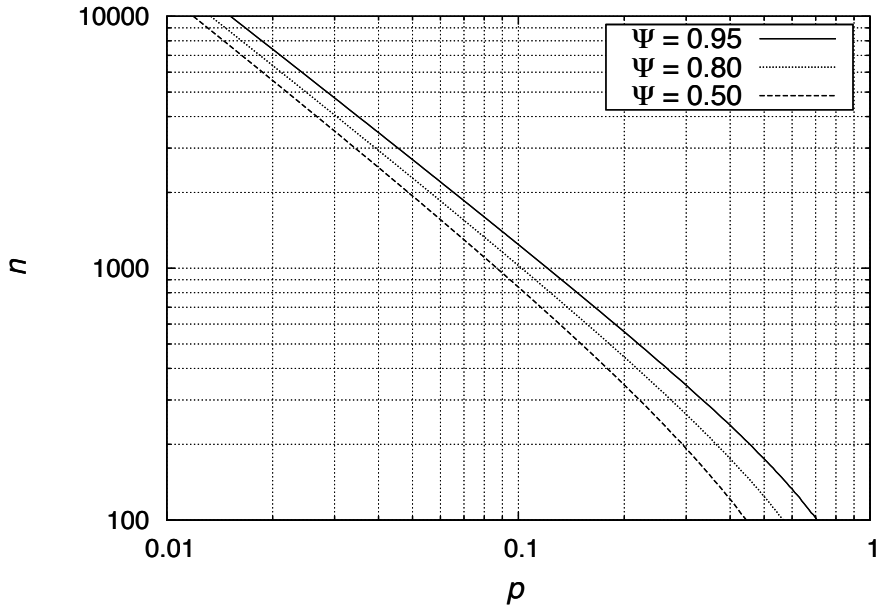


(a) Design options for $n = 100$ nodes.

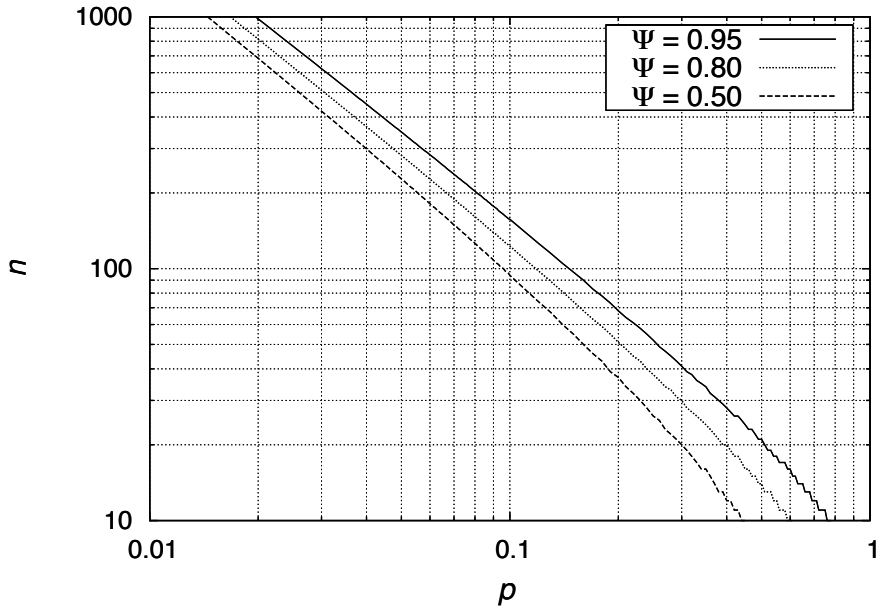


(b) Design options for $n = 1000$ nodes.

Figure 3.2: Probabilistic flooding in Erdős Rényi graphs. Plots show (p, ω)-pairs for global outreach probability $\Psi = 0.50$, 0.80, 0.95.



(a) Design options for $\omega = 0.08$.



(b) Design options for $\omega = 0.5$.

Figure 3.3: Probabilistic flooding in Erdős Rényi graphs. Plots show (p, n)-pairs for global outreach probability $\Psi = 0.50, 0.80, 0.95$.

3.4 – Probabilistic Flooding in Poisson Random Geometric Graphs without Border Effects

Let us now study the outreach probability in P-RGGs. We derive an asymptotic expression and complement it with simulations showing that this expression is a good approximation for finite node densities. Finally, we give system parameters leading to high outreach probability.

3.4.1 Derivation of the Outreach Probability

Theorem 1 gave us an expression for Ψ using A_{PF} in a general graph G. If G belongs to the class of RGGs, the outreach probability is given by the following theorem.

Theorem 3 (Global outreach probability in RGGs)**.** *Let* $\lambda^* \triangleq \lambda\omega + \frac{1}{A}$, $\alpha \triangleq \lambda^*\pi r^2 - \ln(A\,\lambda + 1)$, $\gamma \triangleq 1 - e^{-\lambda^*\pi r^2}$. *The probability of global outreach of algorithm* A_{PF} *with forwarding probability* ω *in an RGG* $G(A, \lambda, r)$ *is*

$$\Psi(A,\, \lambda,\, r,\, \omega) = \gamma\, e^{-e^{-\alpha}} + \varepsilon_\psi, \tag{3.9}$$

where $\lim_{\lambda\to\infty} \varepsilon_\psi = 0$.

If $\lambda \to \infty$, α may be kept constant by setting r to be a function of λ, under the conditions of Theorem 3.

To prove this theorem we (1) specialize the Graph sampling (GS) approach to the specifics of the underlying spatial Poisson point process of the definition of RGGs; (2) derive an asymptotic expression for the probability of absence of isolated nodes in G^* that is also a lower bound; (3) derive an asymptotic expression for the probability of connectivity of the subgraph G^*; (4) derive an asymptotic expression for the probability of V^* being a dominating set of G that is also a lower bound; (5) analyze the dependence between node isolation in G^* and domination of G by V^*, and derive an asymptotic expression for the probability of occurrence of both events that is also a lower bound; (6) analyze the dependence between the connectivity of G^* and the domination of G by V^*, and derive an asymptotic expression for the probability of occurrence of both events.

3.4.1.1 Graph Sampling and Poisson Processes

The GS approach can be applied to RGGs by performing the node sampling directly on the Poisson point process Π that generates G. Given the properties of Poisson processes, Π

can be decomposed into two independent thinned Poisson point processes Π^* and $\Pi^\diamond$. The process Π^* with intensity $\lambda\,\omega$ is the set of all forwarding nodes, and the process $\Pi^\diamond$ with intensity $\lambda\,(1-\omega)$ is the set of non-forwarding nodes.

Let Π_s^* denote a process Π^* on a square S_A of area A plus the addition of a source node placed uniformly at random in S_A. This leads to a subgraph of forwarding nodes $G^*(V^*, E^*)$ with node density $\lambda^* = \lambda\,\omega + \frac{1}{A}$. The process $\Pi^\diamond$ on S_A leads to a subgraph of non-forwarding nodes $G^\diamond(V^\diamond, E^\diamond)$ with node density $\lambda^\diamond = \lambda\,(1-\omega)$.

The subgraph G^* has N^*+1 nodes, where N^* is a Poisson random variable distributed according to $\mathrm{Po}(A\,\lambda^*)$. $G^\diamond$ has $N^\diamond$ nodes, where $N^\diamond$ follows $\mathrm{Po}(A\,\lambda^\diamond)$.

3.4.1.2 Node Isolation in G^*

To achieve global outreach, the graph of forwarding nodes G^* needs to be connected (Theorem 1). A necessary but not sufficient condition for connectivity is the absence of isolated nodes in G^*. We state the asymptotic expression for the probability of G^* having no isolated nodes and we show how this probability relates to the probability of G^* being connected. To do so, we use the *total variation distance* between the distributions of two integer-valued random variables X, Y, defined as ([Pen03], see 2.4):

$$d_{\mathrm{TV}}(X,Y) \triangleq \sup_{A\subseteq\mathbb{Z}} \left|\mathrm{P}\left(X\in A\right) - \mathrm{P}\left(Y\in A\right)\right|. \tag{3.10}$$

Lemma 1. *Let $\alpha^* \triangleq \lambda^*\pi\,r^2 - \ln\left(A\,\lambda^*\right)$. The total variation distance between the distribution of the number W^* of isolated nodes in G^* and the Poisson distribution with parameter $e^{-\alpha^*}$ converges to 0 when $\lambda\to\infty$. That is*

$$\lim_{\lambda\to\infty} d_{\mathrm{TV}}\left(W^*, Po\left(e^{-\alpha^*}\right)\right) = 0. \tag{3.11}$$

A proof can be found in A.1.

Proposition 1 (Isolated nodes in G^*)**.** *Let $I\left(G^*\right)$ denote the event that G^* has isolated nodes, and furthermore, let $\alpha^* \triangleq \lambda^*\pi\,r^2 - \ln\left(A\,\lambda^*\right)$ and $\gamma \triangleq 1 - e^{-\lambda^*\,\pi\,r^2}$. The probability of G^* having no isolated nodes is*

$$P\left(\neg I\left(G^*\right)\right) = \gamma\,e^{-e^{-\alpha^*}} + \varepsilon_I, \tag{3.12}$$

where $\varepsilon_I \geq 0$ and $\lim_{\lambda\to\infty}\varepsilon_I = 0$.

Proof. The nodes of G^* are distributed in S_A according to Π_s^*. The node density of G^* is therefore $\lambda^* = \lambda\,\omega + \frac{1}{A}$, and the probability that a node $v^*\in V^*$ is isolated is

$$\mathrm{P}\left(\mathrm{iso}\left(v^*\right)\right) = e^{-\lambda^*\,\pi\,r^2}. \tag{3.13}$$

The probability that there is no isolated node in G^* is

$$\mathrm{P}\left(\neg I\left(G^*\right)|N^*\right) = \mathrm{P}\left(\bigcap_{i=1}^{N^*+1} \neg \operatorname{iso}\left(v_i^*\right)\right). \tag{3.14}$$

The isolation events are not independent from node to node if the corresponding nodes are close enough to each other. Further, the events $\neg\mathrm{iso}\left(v_i^*\right)$ are increasing events with respect to ω and λ. Application of the FKG inequality ([FKG71], see 2.3) to (3.14) leads to

$$\mathrm{P}\left(\neg I\left(G^*\right)|N^*\right) \geq \prod_{i=1}^{N^*+1} \mathrm{P}\left(\neg \operatorname{iso}\left(v_i\right)\right) = \left[1 - e^{-\lambda^* \pi r^2}\right]^{N^*+1} \tag{3.15}$$

which can be re-written as

$$\mathrm{P}\left(\neg I\left(G^*\right)|N^*\right) = \left(1 - e^{-\lambda^* \pi r^2}\right)^{N^*+1} + \xi_I, \tag{3.16}$$

with $\xi_I \geq 0$. Applying the law of total probability yields

$$\begin{aligned}
&\mathrm{P}\left(\neg I\left(G^*\right)\right) = \mathrm{E}\left(\mathrm{P}\left(\neg I\left(G^*\right)|N^*\right)\right) \\
&= \mathrm{E}\left(\left(1 - e^{-\lambda^* \pi r^2}\right)^{N^*+1}\right) + \mathrm{E}\left(\xi_I\right) \\
&= \left(1 - e^{-\lambda^* \pi r^2}\right) \cdot \\
&\cdot \sum_{k=0}^{\infty} \left(1 - e^{-\lambda^* \pi r^2}\right)^k \frac{\left(A\,\lambda^*\right)^k e^{-A\,\lambda^*}}{k!} + \varepsilon_I \\
&= \left(1 - e^{-\lambda^* \pi r^2}\right) e^{-e^{-\left(\lambda^* \pi r^2 - \ln\left(A\,\lambda^*\right)\right)}} + \varepsilon_I \\
&= \gamma\, e^{-e^{-\alpha^*}} + \varepsilon_I,
\end{aligned} \tag{3.17}$$

with $\varepsilon_I = \mathrm{E}\left(\xi_I\right) \geq 0$ and $\gamma = 1 - e^{-\lambda^* \pi r^2}$, proving (3.12).

From Lemma 1, the probability of having no isolated nodes in G^* is upper bounded in the following way:

$$\begin{aligned}
\mathrm{P}\left(\neg I\left(G^*\right)\right) &= \mathrm{P}(W^* = 0) \\
&\leq e^{-e^{-\alpha^*}} + d_{\mathrm{TV}}\left(W^*, \mathrm{Po}\left(e^{-\alpha^*}\right)\right).
\end{aligned} \tag{3.18}$$

Combining (3.17) with (3.18), yields

$$\gamma\, e^{-e^{-\alpha^*}} + \varepsilon_I \leq e^{-e^{-\alpha^*}} + d_{\mathrm{TV}}\left(W^*, \mathrm{Po}\left(e^{-\alpha^*}\right)\right). \tag{3.19}$$

Moreover, by combining the last equation with (3.11) of Lemma 1, and since $\lim_{\lambda\to\infty} \gamma = 1$, we get $\lim_{\lambda\to\infty} \varepsilon_I = 0$. □

The probability $\mathrm{P}\left(\neg I\left(G^*\right)\right)$ converges to $\gamma\, e^{-e^{-\alpha^*}}$. This expression is a lower bound for the same probability.

3.4.1.3 Connectivity of G^*

We now determine an asymptotic expression for the probability that G^* is connected.

Proposition 2 (Connectivity of G^*)**.** *With* $\alpha^* \triangleq \lambda^* \pi r^2 - \ln(A\lambda^*)$ *and* $\gamma \triangleq 1 - e^{-\lambda^* \pi r^2}$, *the probability that* G^* *is connected is*

$$P(C(G^*)) = \gamma e^{-e^{-\alpha^*}} + \varepsilon_I - \varepsilon_C, \tag{3.20}$$

where $\varepsilon_I \geq 0$, $\varepsilon_C \geq 0$, $\lim_{\lambda\to\infty} \varepsilon_I = 0$ *and* $\lim_{\lambda\to\infty} \varepsilon_C = 0$.

Proof. The absence of isolated nodes in G^* is a necessary condition for its connectivity. Therefore, $\mathrm{P}(C(G^*)) \leq \mathrm{P}(\neg I(G^*))$, which can be rewritten as

$$\mathrm{P}(C(G^*)) = \mathrm{P}(\neg I(G^*)) - \varepsilon_C \tag{3.21}$$

where $\varepsilon_C \geq 0$. Applying (3.12) in (3.21) yields

$$\mathrm{P}(C(G^*)) = \gamma e^{-e^{-\alpha^*}} + \varepsilon_I - \varepsilon_C, \tag{3.22}$$

where $\lim_{\lambda\to\infty} \varepsilon_I = 0$, thus proving (3.20).

Penrose [Pen97] bridges the problem of the absence of isolated nodes in an RGG G, to the problems of (a) the longest of the nearest neighbor distances between the nodes of G and (b) the longest edge of the minimum spanning tree connecting the nodes of G. The probability of having no isolated node in an RGG G is asymptotically the same as the probability of G being connected. Hence,

$$\lim_{\lambda\to\infty} \mathrm{P}(C(G^*)) = \lim_{\lambda\to\infty} \mathrm{P}(\neg I(G^*)). \tag{3.23}$$

By combining (3.23) with (3.21) we get $\lim_{\lambda\to\infty} \varepsilon_C = 0$. □

Fig. 3.4 compares the asymptotic expression of $\mathrm{P}(C(G^*))$ with the relative frequencies of simulated graphs for which G^* has no isolated node or is connected, respectively. Algorithm A_{GS} is used for simulations.

3.4.1.4 Domination of G

We now determine the asymptotic expression for the probability that V^* is a dominating set of G, which is also a lower bound for the same probability.

Lemma 2. *Let* $\alpha^\diamond \triangleq \lambda^* \pi r^2 - \ln(A\lambda^\diamond)$. *The total variation distance between the distribution of the number* $W^\diamond$ *of non-dominated nodes of* G *and the Poisson distribution with parameter* $e^{-\alpha^\diamond}$ *converges to* 0 *when* $\lambda \to \infty$. *That is*

$$\lim_{\lambda\to\infty} d_{\mathrm{TV}}\left(W^\diamond, Po\left(e^{-\alpha^\diamond}\right)\right) = 0. \tag{3.24}$$

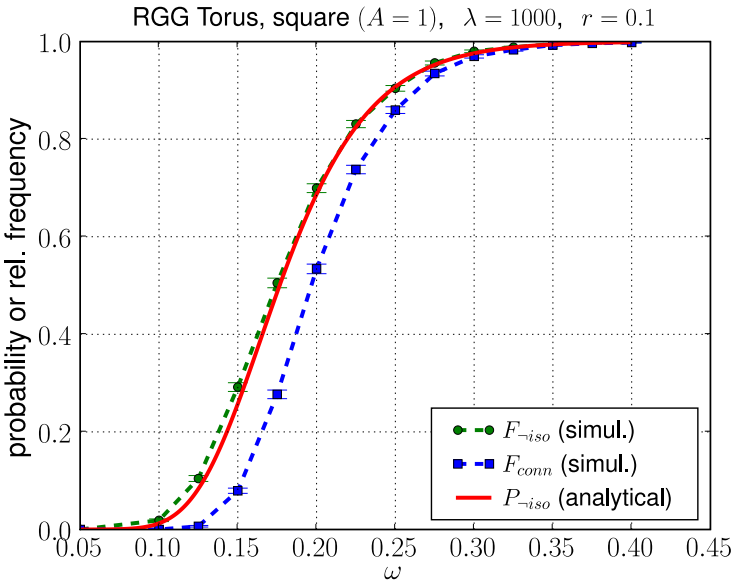


Figure 3.4: Probability of connectivity/no isolated node ($\varepsilon_C = \varepsilon_I = 0$); relative frequency of the experiments of A_{GS} yielding connected graphs; and relative frequency of experiments of A_{GS} yielding graphs with no isolated node. Simulation results are obtained from 10 000 experiments. Each experiment is run over a new RGG.

A proof can be found in A.2.

Proposition 3 (Domination of G)**.** *Let* $\alpha^\diamond \triangleq \lambda^* \pi r^2 - \ln A\lambda^\diamond$. *The probability that* V^* *is a dominating set of* G *is*

$$\mathrm{P}(D(V^*, G)) = e^{-e^{-\alpha^\diamond}} + \varepsilon_D, \tag{3.25}$$

where $\varepsilon_D \geq 0$ *and* $\lim_{\lambda\to\infty} \varepsilon_D = 0$.

Proof. The nodes of G^* are distributed in S_A according to Π_s^* with density $\lambda^* = \lambda\,\omega + \frac{1}{A}$. The set V^* dominates G if for every node $v^\diamond \in V^\diamond$ there is at least one edge $\{v^\diamond, w^*\}$ such that $w^* \in V^*$. In this case, we say that $v^\diamond$ is dominated by V^*. Hence, the probability that a node $v^\diamond \in V^\diamond$ is dominated by V^* is equal to the probability that there is at least one node from Π_s^* within the circle of radius r centered at the position of node $v^\diamond$. That is

$$\mathrm{P}\left(\mathrm{dom}\left(v^\diamond\right)\right) = 1 - e^{-\lambda^*\,\pi\,r^2}. \tag{3.26}$$

Let the random variable $N^\diamond$ be the number of nodes of $V^\diamond$. Then, the probability that G is dominated by V^* is

$$\mathrm{P}\left(D\left(V^*, G\right)|N^\diamond\right) = \mathrm{P}\left(\bigcap_{i=1}^{N^\diamond} \mathrm{dom}\left(v_i^\diamond\right)\right). \tag{3.27}$$

The domination events for nodes of $V^\diamond$ close to each other, i.e. within distance $d(v_i^\diamond, v_j^\diamond) \leq 2\,r$ from each other, are dependent. They are increasing events with respect to ω and λ.

Application of the FKG inequality to (3.27) leads to

$$\mathrm{P}\left(D\left(V^*,G\right)|N^\diamond\right) \geq \prod_{i=1}^{N^\diamond} \mathrm{P}\left(\mathrm{dom}\left(v_i^\diamond\right)\right) = \left[1-e^{-\lambda^*\,\pi\,r^2}\right]^{N^\diamond},$$

$$\mathrm{P}\left(D\left(V^*,G\right)|N^\diamond\right) = \left(1-e^{-\lambda^*\,\pi\,r^2}\right)^{N^\diamond} + \xi_D, \tag{3.28}$$

with $\xi_D \geq 0$. Applying the law of total probability yields

$$\begin{aligned}\mathrm{P}\left(D\left(V^*,G\right)\right) &= \mathrm{E}\left(\mathrm{P}\left(D\left(V^*,G\right)|N^\diamond\right)\right)\\ &= \mathrm{E}\left(\left(1-e^{-\lambda^*\,\pi\,r^2}\right)^{N^\diamond}\right) + \mathrm{E}\left(\xi_D\right)\\ &= e^{-e^{-\left(\lambda^*\,\pi\,r^2-\ln(A\,\lambda^\diamond)\right)}} + \varepsilon_D \;=\; e^{-e^{-\alpha^\diamond}} + \varepsilon_D,\end{aligned} \tag{3.29}$$

with $\varepsilon_D = \mathrm{E}\left(\xi_D\right) \geq 0$, proving (3.25).

From Lemma 2, the probability that there is no non-dominated node in G is upper bounded as follows:

$$\begin{aligned}\mathrm{P}\left(D\left(V^*,G\right)\right) &= \mathrm{P}(W^\diamond = 0)\\ &\leq e^{-e^{-\alpha^\diamond}} + d_{\mathrm{TV}}\left(W^\diamond, \mathrm{Po}\left(e^{-\alpha^\diamond}\right)\right).\end{aligned} \tag{3.30}$$

Combining (3.29) with (3.30) yields

$$\varepsilon_D \leq d_{\mathrm{TV}}\left(W^\diamond, \mathrm{Po}\left(e^{-\alpha^\diamond}\right)\right). \tag{3.31}$$

Finally, combining (3.31) with (3.24), we get $\lim_{\lambda\to\infty}\varepsilon_D = 0$. □

This shows that the domination probability converges asymptotically to $e^{-e^{-\alpha^\diamond}}$, which is also a lower bound. Fig. 3.5 compares the analytical expression/lower bound with simulation results. As ω increases, the relative frequency of graphs where V^* is a dominating set approaches 1, and the difference between the analytical expression and simulation results becomes negligible.

3.4.1.5 Dependence between Node Isolation & Domination

The events $\neg I\left(G^*\right)$ and $D(V^*,G)$ depend on each other. Let us analyze this dependency and derive an asymptotic expression for $\mathrm{P}(\neg I(G^*) \cap D(V^*,G))$ which is also a lower bound for this probability. The event $\neg I\left(G^*\right) \cap D(V^*,G)$ is equivalent to the event that the sum of all non-dominated nodes and isolated forwarding nodes in G is 0. The following lemma helps us to derive the joint probability. A proof can be found in A.3.

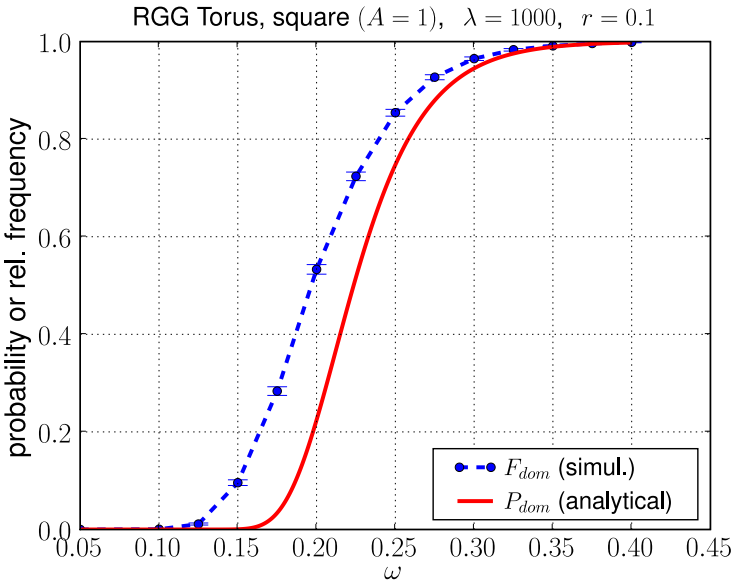


Figure 3.5: Probability of domination (assuming $\varepsilon_D = 0$) in comparison to the relative frequency of the experiments of the algorithm A_{GS} yielding graphs with no non-dominated nodes.

Lemma 3. *Let* $\alpha \triangleq \lambda^* \pi r^2 - \ln(A\lambda + 1)$ *and* $W^{\diamond *}$ *be the sum of all non-dominated nodes and isolated forwarding nodes in* G*. The total variation distance between the distribution of* $W^{\diamond *}$ *and a Poisson distribution* $Po(e^{-\alpha})$ *converges to* 0 *when* $\lambda \to \infty$*, i.e.*

$$\lim_{\lambda \to \infty} d_{\mathrm{TV}}\left(W^{\diamond *}, Po\left(e^{-\alpha}\right)\right) = 0. \tag{3.32}$$

Proposition 4. *The probability that* G^* *has no isolated node and its nodes* V^* *dominate* G *is*

$$\mathrm{P}\left(\neg I(G^*) \cap D(V^*, G)\right) = \mathrm{P}\left(\neg I(G^*)\right) \cdot \mathrm{P}\left(D(V^*, G)\right) + \varepsilon_{ID}, \tag{3.33}$$

where $\varepsilon_{ID} \geq 0$*, and*

$$\lim_{\lambda \to \infty} \varepsilon_{ID} = 0. \tag{3.34}$$

Proof. The events $\neg I(G^*)$ and $D(V^*, G)$ are increasing events with respect to ω and λ. The FKG inequality yields

$$\mathrm{P}(\neg I(G^*) \cap D(V^*, G)) = \mathrm{P}(\neg I(G^*)) \cdot \mathrm{P}(D(V^*, G)) + \varepsilon_{ID}$$

where $\varepsilon_{ID} \geq 0$, thus proving (3.33). Propositions 1, 3 yield

$$\begin{aligned} &\mathrm{P}(\neg I(G^*) \cap D(V^*, G)) \\ &= \left(\gamma\, e^{-e^{-\alpha^*}} + \varepsilon_I\right)\left(e^{-e^{-\alpha^\diamond}} + \varepsilon_D\right) + \varepsilon_{ID} \\ &= \gamma\, e^{-e^{-\alpha}} + \varepsilon_D\, \gamma\, e^{-e^{-\alpha^*}} + \varepsilon_I\left(e^{-e^{-\alpha^\diamond}} + \varepsilon_D\right) + \varepsilon_{ID}, \end{aligned} \tag{3.35}$$

with $\alpha \triangleq \lambda^* \pi\, r^2 - \ln(A\,\lambda + 1)$. From Lemma 3, we have

$$\begin{aligned} \mathrm{P}(\neg I(G^*) \cap D(V^*, G)) &= \mathrm{P}(W^{\diamond *} = 0) \\ &\leq e^{-e^{-\alpha}} + d_{\mathrm{TV}}\left(W^{\diamond *}, \mathrm{Po}\left(e^{-\alpha}\right)\right). \end{aligned} \tag{3.36}$$

Combining this expression with (3.35) yields

$$\begin{aligned} \gamma\, e^{-e^{-\alpha}} &+ \varepsilon_D\, \gamma\, e^{-e^{-\alpha^*}} + \varepsilon_I \left(e^{-e^{-\alpha^\diamond}} + \varepsilon_D\right) + \varepsilon_{ID} \\ &\leq e^{-e^{-\alpha}} + d_{\mathrm{TV}}\left(W^{\diamond *}, \mathrm{Po}\left(e^{-\alpha^{\diamond *}}\right)\right). \end{aligned} \tag{3.37}$$

Since $\lim_{\lambda\to\infty} \varepsilon_I = \lim_{\lambda\to\infty} \varepsilon_D = \lim_{\lambda\to\infty} \varepsilon_{ID} = 0$, and $\lim_{\lambda\to\infty} \gamma = 1$, combining (3.32) with (3.37) yields $\lim_{\lambda\to\infty} \varepsilon_{ID} = 0$. □

3.4.1.6 Dependence between Connectivity and Domination

Let us study the event $C(G^*) \cap D(V^*, G)$, inferring the asymptotic behavior of dependence between connectivity of G^* and domination of G by V^*.

Proposition 5. *The probability of G^* being connected and V^* dominating G is*

$$\mathrm{P}(C(G^*) \cap D(V^*, G)) = \mathrm{P}(C(G^*))\mathrm{P}(D(V^*, G)) + \varepsilon_{CD}, \tag{3.38}$$

where $\varepsilon_{CD} \geq 0$ and $\lim_{\lambda\to\infty} \varepsilon_{CD} = 0$.

Proof. The events $C(G^*)$ and $D(V^*, G)$ are increasing events with respect to ω and λ. The FKG inequality yields

$$\mathrm{P}(C(G^*) \cap D(V^*, G)) = \mathrm{P}(C(G^*))\mathrm{P}(D(V^*, G)) + \varepsilon_{CD}, \tag{3.39}$$

where $\varepsilon_{CD} \geq 0$, thus proving (3.38). From (3.23), the probability of no isolated node in an RGG G is asymptotically the same as the probability of G being connected. Thus,

$$\lim_{\lambda\to\infty} \mathrm{P}(C(G^*) \cap D(V^*, G)) = \lim_{\lambda\to\infty} \mathrm{P}(\neg I(G^*) \cap D(V^*, G)). \tag{3.40}$$

Conjugating (3.39) and (3.33) with (3.40), we get

$$\begin{aligned} &\lim_{\lambda\to\infty} \mathrm{P}(C(G^*)) \cdot \mathrm{P}(D(V^*, G)) + \lim_{\lambda\to\infty} \varepsilon_{CD} = \\ &\lim_{\lambda\to\infty} \mathrm{P}(\neg I(G^*)) \cdot \mathrm{P}(D(V^*, G)) + \lim_{\lambda\to\infty} \varepsilon_{ID}. \end{aligned} \tag{3.41}$$

Combining this equation with (3.40), we get

$$\begin{aligned} &\lim_{\lambda\to\infty} \mathrm{P}(\neg I(G^*)) \cdot \mathrm{P}(D(V^*, G)) + \lim_{\lambda\to\infty} \varepsilon_{CD} = \\ &\lim_{\lambda\to\infty} \mathrm{P}(\neg I(G^*)) \cdot \mathrm{P}(D(V^*, G)) + \lim_{\lambda\to\infty} \varepsilon_{ID}. \end{aligned} \tag{3.42}$$

Combining (3.34) with (3.42), we get $\lim_{\lambda\to\infty} \varepsilon_{CD} = 0$, thus proving the proposition.

□

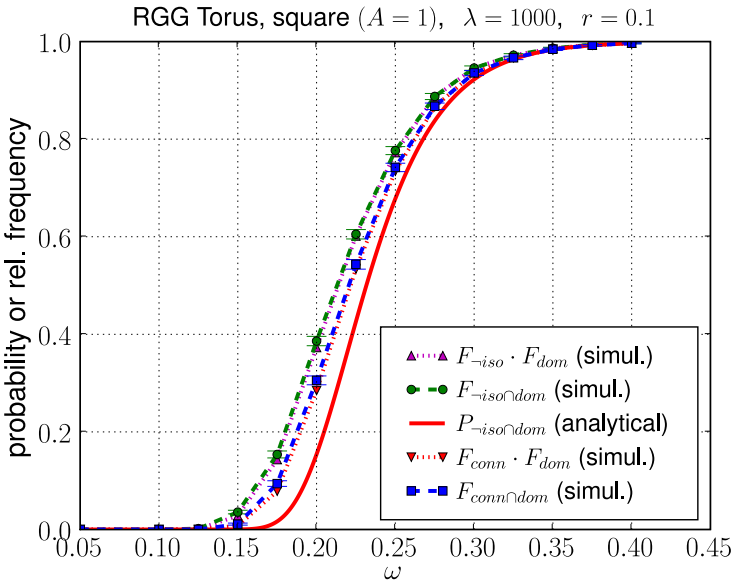


Figure 3.6: Dependence between graph domination and connectivity/node isolation (algorithm A_{GS}) in comparison to the analytical results (assuming $\varepsilon_I = 0$, $\varepsilon_D = 0$, and $\varepsilon_{ID} = 0$). The relative frequency $F_{\neg iso \cap dom}$ of simulated graphs with no isolated nodes in G^* and simultaneously having V^* dominating G is higher than the corresponding analytical expression, and is also slightly higher than $F_{\neg iso} F_{dom}$. Moreover, the relative frequency $F_{conn \cap dom}$ of graphs where G^* is connected and simultaneously G is dominated by V^* is slightly higher than $F_{conn} F_{dom}$.

In summary, the dependency between connectivity and domination becomes negligible as the node density increases. The term $\gamma\, e^{-e^{-\alpha}}$ is the asymptotic expression and lower bound for the joint probability of both events. This joint probability converges from above to the product of the individual probabilities. Fig. 3.6 shows analytical and simulation results that evidence these facts.

3.4.1.7 Proof of Theorem 3 (Global Outreach in RGGs)

Combining Propositions 2, 3 and 5, we get

$$\begin{aligned} \Psi(A, \lambda, r, \omega) &= \left[\gamma\, e^{-e^{-\alpha^*}} + \varepsilon_I - \varepsilon_C\right] \left[e^{-e^{-\alpha^\diamond}} + \varepsilon_D\right] + \varepsilon_{CD} \\ &= \gamma\, e^{-e^{-\alpha}} + \varepsilon_\psi, \end{aligned} \tag{3.43}$$

where

$$\varepsilon_\psi \triangleq \varepsilon_D\, \gamma\, e^{-e^{-\alpha^*}} + (\varepsilon_I - \varepsilon_C) \left(e^{-e^{-\alpha^\diamond}} + \varepsilon_D\right) + \varepsilon_{CD}. \tag{3.44}$$

As $\lim_{\lambda\to\infty} \varepsilon_I = \lim_{\lambda\to\infty} \varepsilon_C = \lim_{\lambda\to\infty} \varepsilon_D = \lim_{\lambda\to\infty} \varepsilon_{CD} = 0$, we get $\lim_{\lambda\to\infty} \varepsilon_\psi = 0$, thus proving the theorem.

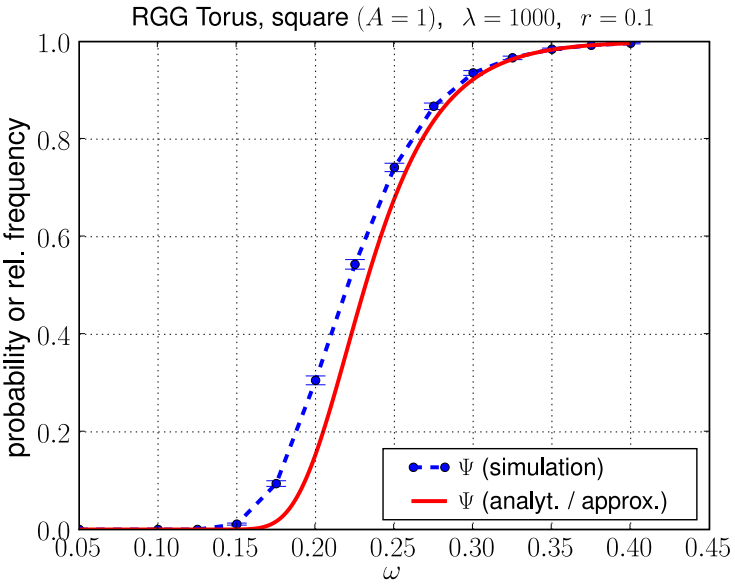
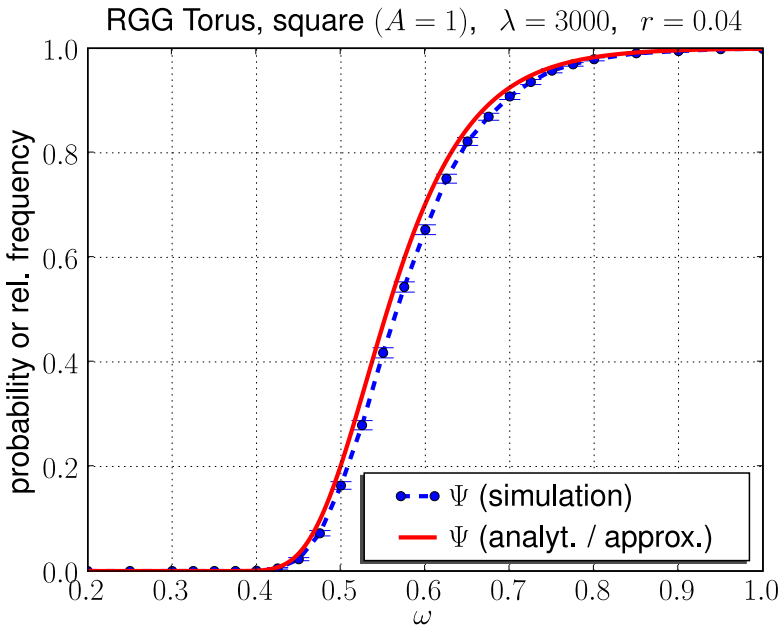


Figure 3.7: Probability of global outreach (assuming $\varepsilon_\Psi = 0$) in comparison to the relative frequency of the experiments of the algorithm A_{GS} yielding connected dominating sets.

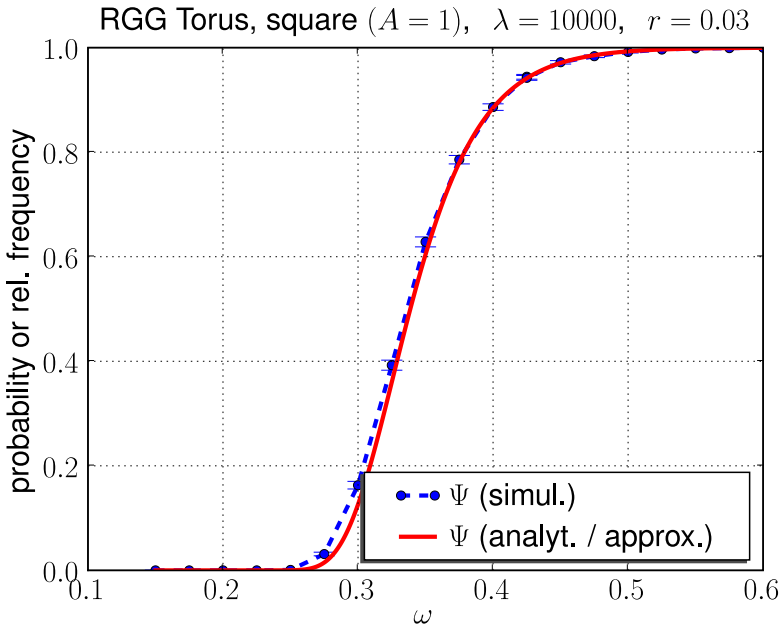
In summary, $\gamma\, e^{-e^{-\alpha}}$ is the asymptotic expression of the global outreach probability $\Psi(A, \lambda, r, \omega)$ for PF with forwarding probability ω on an RGG $G(A,\, \lambda,\, r)$. Fig. 3.7 compares analytical and simulation results. As ω increases, Ψ_{GS} converges to the analytical expression of Ψ. The difference becomes negligible for high values of Ψ.

3.4.2 Simulation of Outreach Probability in RGGs

Figs. 3.8(a) and 3.8(b) show the probability/relative frequency of floodings yielding global outreach in RGGs on $G(1,\, \lambda,\, r)$ over ω. The analytical curve Ψ represents the asymptotic expression ($\varepsilon_\psi = 0$) of $\Psi(A, \lambda, r, \omega)$ derived in Theorem 3. The curve Ψ_{GS} is obtained from the application of the algorithm A_{GS} to RGGs on a torus. Comparing results for different ω-values, there is a critical interval where Ψ changes from zero to one. The simulated value Ψ_{GS} may lie below or above the asymptotic curve of Ψ, depending on the graph parameters. Moreover, the difference between simulated and asymptotic values is relatively small and becomes negligible as ω increases. This behavior is due to the interplay of the components from which our expression for Ψ was derived.



(a) Parameters: $A = 1$; $\lambda = 3\,000$; $r = 0.04$, $\omega = 0.2...1.0$.



(b) Parameters: $A = 1$; $\lambda = 10\,000$; $r = 0.03$; $\omega = 0.1...0.6$.

Figure 3.8: Global outreach probability ($\varepsilon_\Psi = 0$) and relative frequency of experiments yielding connected dominating sets when applying A_{GS} to RGGs on a torus. Each simulated data point (with its respective 95% confidence interval limits) is obtained from 10 000 experiments. Each experiment is run over a new RGG.

3.4.3 Parameters for Global Outreach

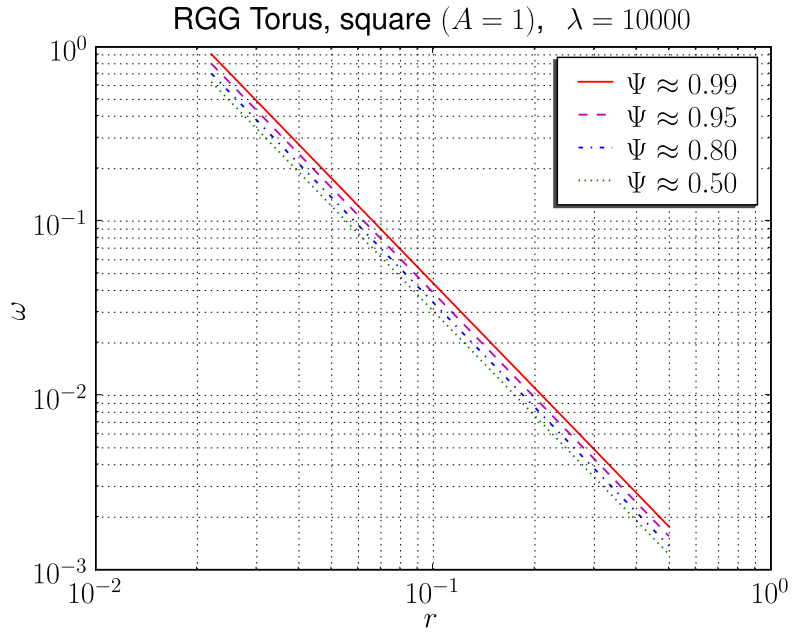
The goal is to meet a target value Ψ by tuning ω. If A and λ are given, the parameters ω and r can be chosen. Fig. 3.9(a) plots the (ω, r)-pairs that approximately achieve a required Ψ for RGGs with $A = 1$ and $\lambda = 1000$. The curves show a clear trade-off in the choice of the (ω, r)-pairs. Sparse RGGs (low values of r) require higher values of ω. For well-connected RGGs (high values of r), small values of ω are sufficient to approximately achieve the desired Ψ. Again, the plots stress the non-linear dependence between these two parameters.

If A and r are given, the parameters ω and λ may be chosen. Fig. 3.9(b) plots the (ω, λ)-pairs that approximately achieve a required outreach probability Ψ for RGGs with $A = 1$ and $r = 0.1$. The curves show the existence of a clear trade-off in the choice of the (ω, λ)-pairs. Sparse RGGs (low values of λ) require higher values of ω. For well-connected RGGs (high values of λ), small values of ω are sufficient to approximately achieve the desired Ψ.

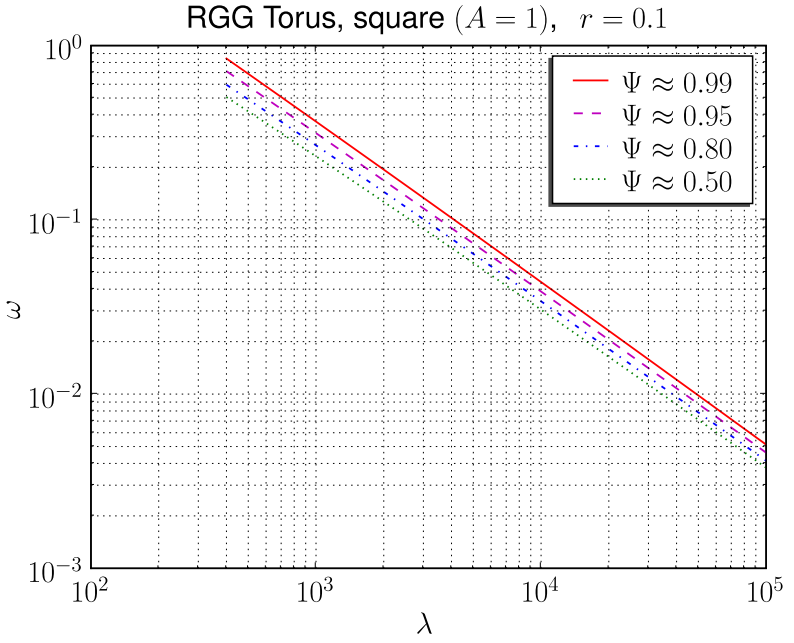
3.5 – Probabilistic Flooding in Poisson Random Geometric Graphs with Border Effects

The global outreach probability of PF with a network-wide forwarding probability degrades if we drop the assumption of a torus distance metric. This degradation in RGGs with Euclidean distance metric is due to border effects.

The probability of a node receiving a message is directly affected by the following parameters: the forwarding probability ω, the number of its neighbors, and by the probability of its neighbors having the message. The border nodes — located at distance smaller than the transmission radius from the border of the square — are expected to have a smaller number of neighbors when compared to central nodes. Therefore, the smaller neighborhood of the border nodes implies that these nodes are less likely to receive a source message when using a PF algorithm with constant ω. This fact has a significant impact in Ψ, since it is the product of the probabilities of each node receiving a message.



(a) Design options for $\lambda = 10\,000$.



(b) Design options for $r = 0.1$.

Figure 3.9: Probabilistic flooding in Random geometric graphs on a torus. PF and RGG parameter tuples for a global outreach probability $\Psi = \{0.50, 0.80, 0.95, 0.99\}$. Plots (a) shows (ω, r)-tuples and plots (b) shows (ω, λ)-tuples that approximately achieve the aforementioned global outreach probability.

We now show how a modification to the PF algorithm, that we denote as Border-Aware Probabilistic flooding (BAPF), minimizes the penalization incurred in Ψ due to these border effects. The main idea is that border nodes should receive a message with the same probability as non-border nodes receive it. To do so, border nodes use an increased forwarding probability ω' depending on their location:

$$\omega'(u) \triangleq \begin{cases} 1 - \sqrt[\phi(u)]{(1-\omega)^{\pi r^2}} & \text{if } u \text{ is a border node,} \\ \omega & \text{otherwise;} \end{cases}$$

where $\phi(u) \triangleq \min_{v \in N(u)} a(v)$ and $a(v)$ is the coverage area of the node v lying within the square of area A.

Figs. 3.10(a) and 3.10(b) show the probability/relative frequency of floodings yielding global outreach in RGGs on $G(1,\, \lambda,\, r)$ over ω. We can observe that the frequencies of global outreach of the BAPF algorithm over RGGs with Euclidean distance closely match the ones of the PF algorithm with constant forwarding probability over RGGs with toroidal distance.

In conclusion, the impact of border effects on global outreach can be minimized by using PF with increased forwarding probability for border nodes.

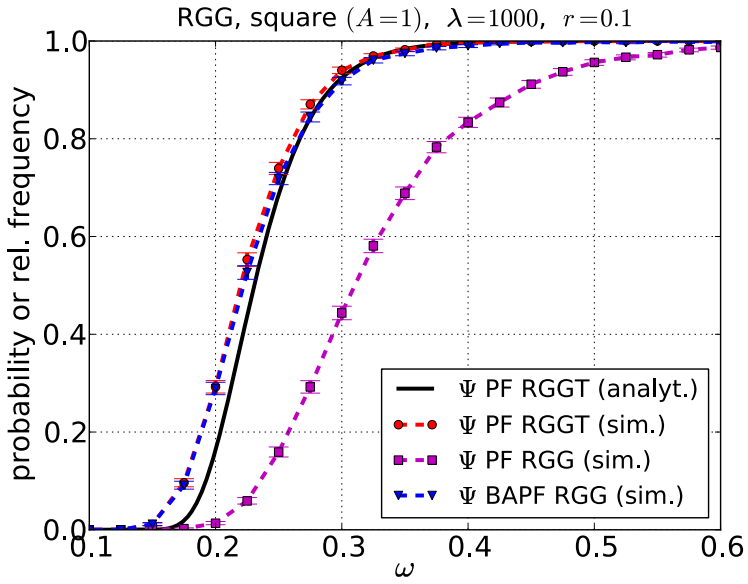
3.6 – Probabilistic Flooding with Unreliable Links

In this section we drop the assumption of an error-free broadcast medium, but consider networks with erasure channels (Ch. 7, [CT06]). A message may fail to be received by each neighbor of the transmitting node independently with probability ζ. To model this unreliable transmission medium, it suffices to sample uniformly at random the edge set E of the graph model $G(V, E)$ of the network with probability $(1 - \zeta)$. This yields a new graph $G_\zeta(V, E_\zeta)$.

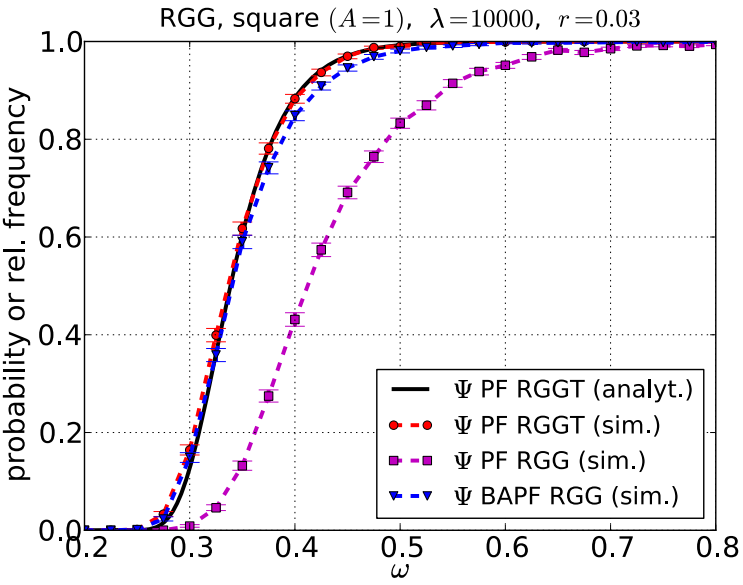
The PF algorithm is now analysed over this graph to take into account the effects of unreliable transmissions. Theorem 1 still holds for PF over networks with erasure channels after replacing G by G_ζ. Therefore, (3.1) becomes:

$$\Psi(G, \zeta, \omega) = \mathrm{P}\big(C(G^*_\zeta) \cap D(V^*, G_\zeta)\big). \tag{3.46}$$

We now specialize this expression for ERGs and RGGs.



(a) Parameters: $n = 1\,000$; $p = 0.15$; $\omega = 0.1...0.6$.



(b) Parameters: $A = 1$; $\lambda = 10\,000$; $r = 0.03$; $\omega = 0.2...0.8$.

Figure 3.10: Global outreach probability ($\varepsilon_\Psi = 0$). Relative frequency of experiments when applying: (1) A_{PF} to RGGs on a torus; (2) A_{PF} to RGGs with Euclidean distance metric (border effects); (3) BAPF to RGGs with euclidean distance metric, which minimizes the border effects. Each simulated data point (with its respective 95% confidence interval limits) is obtained from 5 000 experiments. Each experiment is run over a new network.

Erdős Rényi Graphs with Unreliable Links The outreach probability in ERGs with erasure channels is given by Theorem 2 if we replace p by $p(1-\zeta)$ in all expressions.

Proof. The proof is similar to the one of Theorem 2 after replacing G by G_ζ.

The graph G_ζ is also an ERG $G(n, p\,(1-\zeta))$. Hence, since G_ζ has an edge probability $p\,(1-\zeta)$, it is sufficient to replace p by $p\,(1-\zeta)$ in each expression within the proof of Theorem 2. □

Fig. 3.11(a) plots Ψ over ω for distinct values of ζ, along with results from simulations of PF. The simulation results match perfectly the analytical expression of Ψ.

Random Geometric Graphs with Unreliable Links The outreach probability in RGGs with erasure channels is given by Theorem 3 if we redefine λ^* as $(\lambda\,\omega + \frac{1}{A})\,(1-\zeta)$.

Proof. The proof is similar to the one of Theorem 3 after replacing G by G_ζ. Therefore, we give a brief sketch of the proof, highlighting the main differences.

The edge set E_ζ of the RGG G_ζ is built by sampling E with probability $(1-\zeta)$. Hence, the probability that a node $v^* \in V^*$ is isolated (Eq. (3.13)) changes to:

$$\mathrm{P}\left(\mathrm{iso}\left(v^*\right)\right) = e^{-\lambda^*\,(1-\zeta)\,\pi\,r^2}, \tag{3.47}$$

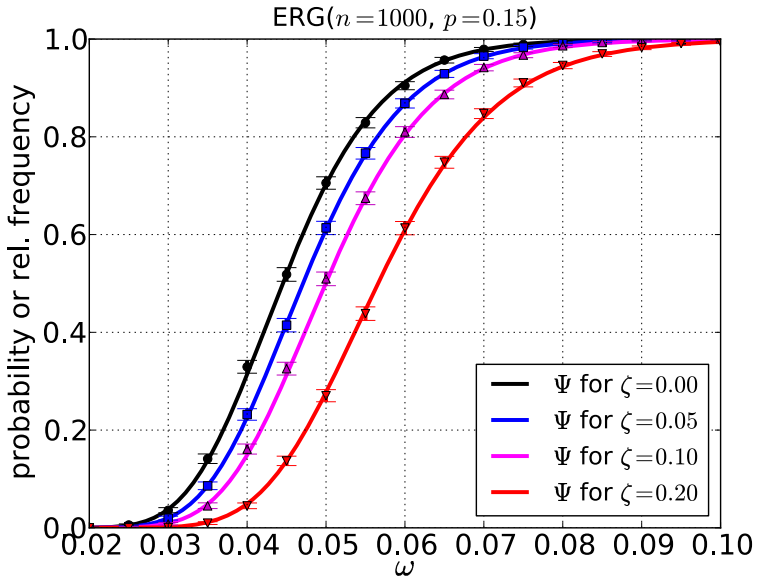
and the probability that a node $v^\diamond \in V^\diamond$ is dominated by V^* (Eq. (3.26)) becomes:

$$\mathrm{P}\left(\mathrm{dom}\left(v^\diamond\right)\right) = 1 - e^{-\lambda^*\,(1-\zeta)\,\pi\,r^2}. \tag{3.48}$$

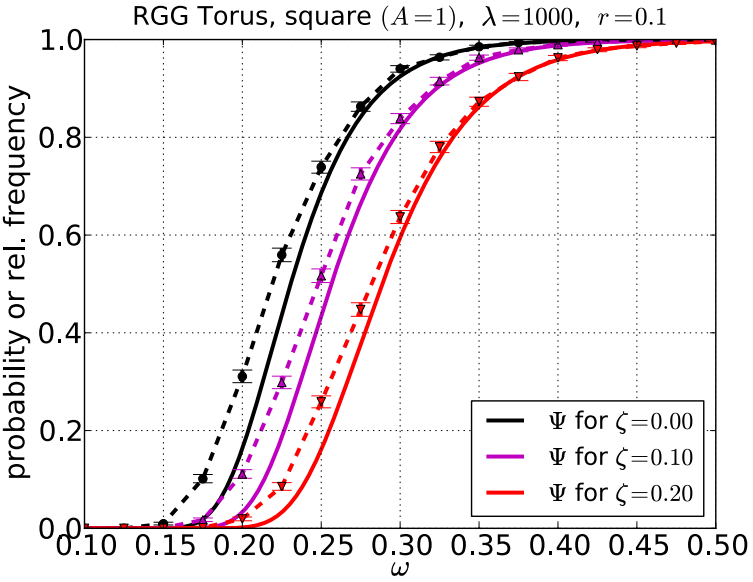
With these new node isolation and node domination probabilities we get the probabilities for the connectivity and domination events. Moreover, since G_ζ is an RGG, the dependency between these events is asymptotically negligible (Propositions 4 and 5). Therefore, it suffices to follow the steps of the proof of Theorem 3 to prove the above result.

□

Fig. 3.11(b) plots Ψ over ω for distinct values of ζ, along with results from simulations of PF. The difference between simulated and asymptotic values is relatively small and becomes negligible as ω increases.



(a) Parameters: $n = 1\,000$; $p = 0.15$; $\zeta \in \{0.0, 0.05, 0.1, 0.2\}$; $\omega = 0.2...1.0$.



(b) Parameters: $A = 1$; $\lambda = 1\,000$; $r = 0.1$; $\zeta \in \{0.0, 0.05, 0.1, 0.2\}$; $\omega = 0.1...0.5$.

Figure 3.11: Global outreach probability ($\varepsilon_\Psi = 0$) and relative frequency of experiments when applying A_{PF} to ERGs and RGGs on a torus with an unreliable transmission medium. Each simulated data point (with its respective 95% confidence interval limits) is obtained from 5 000 experiments. Each experiment is run over a new network.

3.7 – Discussion and Related Work

The problem of PF outreach has mainly been addressed by simulations [HHL06, SCS03, KWB01, YOKM+06, YOKP05]. Some of these papers suggest a connection between PF and percolation theory, which is, however, not deeply explored, as most conclusions do not generalize beyond the particular setup [SCS03, HHL06, KWB01].

This chapter addresses the following sub-problems: (a) characterization of the sets of forwarding and non-forwarding nodes; (b) connectivity of forwarding nodes; (c) domination of non-forwarding nodes by forwarding nodes; and (d) independence between connectivity and domination. We applied this approach to two random graph models.

For ERGs, we derive an exact expression for the global outreach probability. The analysis of RGGs brings challenges due to the local correlation among edges. Our approach to the connectivity sub-problem of the set of forwarding nodes is inspired by [FM08, Pen97, Bet04]. To cope with the dependencies among (a) node isolation/connectivity events, (b) node domination events, and (c) the dependency between isolation/connectivity and domination events, we derive inequalities involving the probabilities of the mentioned events by resorting to the FKG inequality [FKG71]. In the asymptotic case, these inequalities become equalities.

Hence, we derive three lemmas regarding the asymptotic distribution of the total number of isolated nodes, non-dominated nodes, and isolated plus non-dominated nodes. In the proof of these lemmas, we follow a strategy based on the application of the Poisson approximation by the Chen-Stein method [Che75, AGG89, AGG90]. Reference [YWLF06] addresses the problem of the number of isolated nodes in wireless ad-hoc networks with Bernoulli nodes. Its method represents an alternative approach to derive a proof of the lemma concerning the asymptotic distribution of the total number of isolated plus non-dominated nodes (Lemma 3). Although that paper considers boundary effects, the derived distribution is still the same asymptotically as the one for the toroidal model. Nevertheless, for the proof of this lemma we opted to use the same approach as for Lemmas 1 and 2. Contrary to ERGs, where the global outreach probability is known exactly, in RGGs we only derive an approximate expression whose error becomes negligible asymptotically.

3.8 – Concluding Remarks

We analyzed how to set a system-wide forwarding probability ω of probabilistic flooding, such that all network nodes ultimately receive a message with high probability. For this purpose, we proposed a graph sampling method, which can be applied in arbitrary networks. This method yields an induced subgraph, whose node set is obtained by sampling the total node set uniformly at random with probability ω. We proved that the events "all nodes receive a flooded message" and "the induced subgraph is connected and its nodes dominate the network graph" have the same probability, and thus, the analysis of global outreach in probabilistic flooding can be performed by analyzing the properties of the induced subgraph.

In networks modeled as Erdős Rényi graphs, we derived the exact expression for the probability of global outreach. In random geometric graphs — as often used in modeling wireless ad-hoc networks — the local correlation among edges results in stochastic dependencies, but, in our model, these dependencies become asymptotically negligible with increasing node density. We derived an asymptotic expression for the global outreach probability, which is also a good approximation for high node density.

"Eggs cannot be unscrambled."

American Proverb

4

Network Coded Information Dissemination

In the previous chapter we analyzed the reachability of Probabilistic Flooding with a network-wide common forwarding probability. We derived analytical expressions for the global outreach probability in networks with a broadcast medium, both with reliable and unreliable links.

In this chapter we devote our attention to the study of the trade-offs between distinct networking paradigms — replication based forwarding and network coded forwarding — in the dissemination of information.

When nodes communicate over the wireless medium, the broadcast property of the channel enables us to optimize the flooding process with respect to the number of transmissions, with obvious repercussions on the overall energy expenditure and bandwidth consumption. Typically, flooding resorts to replication based forwarding where nodes replicate and forward the information they receive.

Since the basic problem of finding the minimum energy transmission scheme for broadcasting a set of messages in a given network is known to be NP-complete, flooding optimization often relies on approximation algorithms. In the class of probabilistic flooding algorithms, messages are forwarded according to a set of predefined probabilistic rules, whereas in the

class of deterministic algorithms, the forwarding decision is always the same for each set of input parameters.

Multipoint relay (MPR) flooding is a deterministic algorithm, which approximates the connected dominating set within a two-hop neighborhood of each node, thus forming a backbone of forwarding nodes and limiting the number of transmissions. This algorithm plays a key role in the Optimized Link State Routing (OLSR) protocol for mobile ad-hoc networks.

The spectra of dissemination algorithms was recently enlarged by the advent of the Network Coding (NC) paradigm ([ACLY00, FLBW06]), in which intermediate nodes are allowed to mix information flows through algebraic operations. Research suggests that NC based flooding algorithms yields further reductions in the number of transmissions required for flooding a message in a network. More specifically, Reference [FWLB06] quantifies these gains for ring and square lattice topologies, and presents a heuristic algorithm which outperforms probabilistic flooding [1] for a class of random geometric graphs. Related work on the benefits of network coding includes a proof that the minimum energy single-source multicast problem with network coding becomes solvable in polynomial-time [LMHK04] and in a distributed manner [LRK+05]. The problem of multiple multicasts, which is closer to flooding, remains however an open problem [LRM+06]. Early results on improvements in terms of throughput, security and energy efficiency are surveyed in [FLBW06].

Seeking to understand how information dissemination techniques compete over network topologies with broadcast medium, we compare replication and network coded based flooding techniques with respect to the number of transmissions, delivery ratio, and the end-to-end delay. More specifically, we base our analysis on Erdős Rényi Random Graphs (ERGs), Binomial Random Geometric Graphs (B-RGGs) (or simply Random Geometric Graphs (RGGs) within this chapter), and Small-World Networks (SWNs) (see Section 2.6); and shed some light on the impact of the network topology on the behavior of two main representatives: the NC flooding algorithm of [FWLB06] and the MPR flooding algorithm of [CJA+03, AQL02].

We present the following main contributions:

- An analytical characterization of the transmission cost of network coded flooding;
- A set of simulation results for the number of transmissions, delivery ratio, and delay trade-offs between network coding and MPR flooding;
- A critical discussion of the interplay between network topology and replication and network coded based flooding algorithms.

[1] Reference [FWLB06] denotes *probabilistic flooding* as *probabilistic routing*.

This chapter is organized as follows. Section 4.1 presents the algorithms under study. Section 4.2 gives an asymptotic analysis of the number of transmissions required by NC flooding algorithm. Section 4.3 presents a simulation study that compares delay, delivery ratio, and number of transmissions of both NC and MPR flooding algorithms. Moreover, the impact of the topology in the performance of both algorithms is also inferred. Finally, Section 4.4 summarizes the main results presenting some concluding remarks.

Main results of this chapter were published in [CBB08a, CBB08b] in collaboration with Christian Bettstetter and João Barros. Sections 4.1-4.4 were adapted from [CBB08a] and [CBB08b], which were written with the corresponding co-authors.

4.1 – Flooding Algorithms

4.1.1 Multipoint Relaying

In its simplest form, *pure flooding* means that all nodes retransmit the received messages. In a network with n nodes, the number of retransmissions of a source message using pure flooding is $n-1$.

Multipoint relaying ([AQL02, JLMV01]) is deemed to reduce the number of duplicate retransmissions while forwarding a broadcast message. This technique reduces the set of nodes retransmitting the message in such a away that a message forwarded by a node is guaranteed to reach (assuming lossless transmissions) all the two-hop neighbors of that node. For this purpose, each node selects a subset of its neighbors ("multipoint relays") that ensure connectivity to every two-hop neighbor. Although finding the optimal MPR set is an NP-complete problem, efficient heuristics are available for its calculation [Vie98].

In this study we resort to the heuristic described in Algorithm $MPR_{Selection}$ for the MPR set computation, and Algorithm MPR_{Flood} for MPR-based flooding. Asymptotic analysis of these two MPR algorithms can be found in [JLMV01].

4.1.2 Random Linear Network Coding based Flooding

Random linear network coding can be viewed as a distributed method for combining different data flows ([HMK+06], [CWJ03]). The basic principle is that each node in the network selects independently and randomly a set of coefficients and uses them to form linear combinations of the messages it receives. These linear combinations are then sent over the outgoing links. The global encoding vector, i.e. the matrix of coefficients corresponding

Algorithm 4 $MPR_{Selection}$ [AQL02]

Let $N(u)$ denote the set of one-hop neighbors of u, and $N^2(u)$ denote the set of two-hop neighbors of u.

1. Start with an empty multipoint relay set $MPR(u)$.
2. Select those one-hop neighbor nodes in $N(u)$ as multipoint relays which are the only neighbor of some node in $N^2(u)$, and add these one-hop neighbor nodes to the multipoint relay set $MPR(u)$.
3. While there still exist some nodes in $N^2(u)$ which are not covered by the multipoint relay set $MPR(u)$:
   (a) For each node in $N(u)$ not in $MPR(u)$ compute the number of nodes that it covers among the uncovered nodes in the set $N^2(u)$.
   (b) Add that node of $N(u)$ in $MPR(u)$ for which this number is maximum.

Algorithm 5 MPR_{Flood} [JLMV01]

1. A source node u broadcasts its source message m_u.
2. Each node v that receives m_u re-broadcasts it only if:
   (a) v is a multipoint relay of the previous hop of the message, and
   (b) the message was not previously forwarded by v.

to the operations performed on the messages, is sent along in the packet header to ensure that the end receivers are capable of decoding the original data. Specifically, it was shown that if the coefficients are chosen at random from a large enough finite field, Gaussian elimination succeeds with high probability [HMK+06].

The NC algorithm used in our study ([FWLB06]). combines random linear network coding with a probabilistic forwarding algorithm. The proposed algorithm (Algorithm NC_{FWB}), resorts to a heuristic that assigns to each node v a probabilistic forwarding factor $f(v)$. This forwarding factor is set to be inversely proportional to the degree $d(v)$, i.e., $f(v) = \frac{\gamma}{d(v)}$, where $\gamma \geq 0$ is a scaling factor whose value depends on the topology [FWLB06].

A node that receives a linearly independent combination of messages will form and broadcast new random linear combinations of the current and previously received messages depending on this forwarding factor.

Algorithm 6 NC_{FWB} [FWLB06]

1. Associate with each node v a forwarding factor $f(v)$.
2. Node v transmits its source message $\max\{1, \lfloor f(v) \rfloor\}$ times, and an additional time with probability $p = f(v) - \max\{1, \lfloor f(v) \rfloor\}$ if $p > 0$.
3. When a node v receives linearly independent messages, it broadcasts a linear combination over the span of the received coding vectors $\lfloor f(v) \rfloor$ times, and an additional time with probability $p = f(v) - \lfloor f(v) \rfloor$.

4.2 – Asymptotic Analysis of Network Coded Flooding

4.2.1 Problem Statement

Let $G = (V, E)$ be a connected graph and furthermore let $M = \{m_u : u \in V\}$ be a set of messages. Assume that every node $u \in V$ acts as a source node intending to deliver a source message m_u to every other node. In the NC flooding process, one transmission of a node refers to broadcasting a message or a linear combination of messages to all neighbors of the node.

We are interested in the number T_{NC} of required transmissions per source message, such that all nodes can decode all messages $m_u \in M$. Our goal is to characterize the expected value $\mathrm{E}(T_{NC})$ in ERGs, RGGs, and SWNs.

4.2.2 General Bounds

Let D be a random variable representing the degree of an arbitrary node in G. Furthermore, let $\mathrm{E}_D(\mathrm{g}(D))$ denote the expected value of some function $\mathrm{g}(D)$ of the random variable D, and let $\xi_D = \mathrm{E}_D(D^{-1})$ be the first negative moment of D.

Theorem 4. *For a transmission scheme defined by Algorithm NC_{FWB}, with γ chosen to ensure that all nodes can decode all messages, the expected value $\mathrm{E}_D(T_{NC})$ is bounded as follows:*

$$(n-1)\ \gamma\ \xi_D + 1 \leq \mathrm{E}_D(T_{NC}) \leq (n-1)\ \gamma\ \xi_D + \max(1, \gamma). \tag{4.1}$$

For $\gamma \leq 1$ the bounds are tight.

Proof. Let S_t be the total number of transmissions performed by all source nodes for the

transmission of their source messages, and I_t be the total number of transmissions performed by intermediate nodes due to reception of linearly independent combination of messages. As there are n source messages, the expected number of transmissions per source message is

$$\mathrm{E}_D(T_{NC}) = \frac{\mathrm{E}_D(S_t) + \mathrm{E}_D(I_t)}{n}. \tag{4.2}$$

To characterize S_t we define S as the random variable representing the number of transmissions performed by a source node to broadcast its source message. Since G has n sources,

$$\mathrm{E}_D(S) = \frac{1}{n}\,\mathrm{E}_D(S_t). \tag{4.3}$$

According to step 2 of Algorithm NC_{FWB}:

$$S = \begin{cases} 1, & \text{for } D \geq \gamma \quad (4.4a) \\ \left\lfloor \frac{\gamma}{D} \right\rfloor + S', & \text{for } D < \gamma \quad (4.4b) \end{cases}$$

where S' is a Bernoulli random variable representing the outcome of a potential additional transmission, with $\mathrm{P}(S' = 1) = B = \frac{\gamma}{D} - \left\lfloor \frac{\gamma}{D} \right\rfloor$. The conditioned expected value of S' is

$$\begin{aligned} \mathrm{E}_D(S'|D<\gamma) &= \mathrm{E}_D(B|D<\gamma) \\ &= \mathrm{E}_D\left(\frac{\gamma}{D} - \left\lfloor \frac{\gamma}{D} \right\rfloor \middle| D<\gamma\right). \end{aligned} \tag{4.5}$$

The conditioned expected value of S given $D < \gamma$ is

$$\begin{aligned} \mathrm{E}_D(S|D<\gamma) = & \\ &= \mathrm{E}_D\left(\left\lfloor \frac{\gamma}{D} \right\rfloor \middle| D<\gamma\right) + \mathrm{E}_D\left(S'|D<\gamma\right) \\ &= \mathrm{E}_D\left(\left\lfloor \frac{\gamma}{D} \right\rfloor \middle| D<\gamma\right) \\ &+ \mathrm{E}_D\left(\frac{\gamma}{D}\middle|D<\gamma\right) - \mathrm{E}_D\left(\left\lfloor \frac{\gamma}{D} \right\rfloor \middle| D<\gamma\right) \\ &= \mathrm{E}_D\left(\frac{\gamma}{D}\middle|D<\gamma\right) \\ &\leq \gamma, \end{aligned} \tag{4.6}$$

because $D \geq 1$.
Conjugating (4.6) with the fact that $S \geq 1$, we get:

$$1 \leq \mathrm{E}_D(S) \leq \max(1, \gamma). \tag{4.7}$$

With (4.3), we obtain

$$n \leq \mathrm{E}_D(S_t) \leq n\ \max(1, \gamma). \tag{4.8}$$

To characterize I_t we define I as the random variable representing the number of transmissions performed by an intermediate node due to the reception of a linearly independent combination of messages. According to step 3 of Algorithm NC_{FWB}:

$$I = \left\lfloor \frac{\gamma}{D} \right\rfloor + I', \tag{4.9}$$

where I' is a Bernoulli random variable representing the outcome of a potential additional transmission, with $\mathrm{P}(I' = 1) = B = \frac{\gamma}{D} - \left\lfloor \frac{\gamma}{D} \right\rfloor$. The expected value of I' is:

$$\mathrm{E}_D(I') = \mathrm{E}_D(B) = \mathrm{E}_D\left(\frac{\gamma}{D}\right) - \mathrm{E}_D\left(\left\lfloor \frac{\gamma}{D} \right\rfloor\right). \tag{4.10}$$

The expected value of I is:

$$\begin{aligned} \mathrm{E}_D(I) &= \mathrm{E}_D\left(\left\lfloor \frac{\gamma}{D} \right\rfloor\right) + \mathrm{E}_D(I') \\ &= \mathrm{E}_D\left(\frac{\gamma}{D}\right) = \gamma\, \mathrm{E}_D\left(\frac{1}{D}\right) = \gamma\, \xi_D. \end{aligned} \tag{4.11}$$

Since after the completion of the transmission process of all n messages, the rank increase of the decoding matrix of each node is $n-1$ and since G has n nodes, we have

$$\begin{aligned} \mathrm{E}_D(I_t) &= n\,(n-1)\,\mathrm{E}_D(I) \\ &= n\,(n-1)\,\gamma\,\xi_D. \end{aligned} \tag{4.12}$$

Finally, from (4.2), (4.8) and (4.12), we get (4.1). □

4.2.3 Bounds for Erdős Rényi Random Graphs

Corollary 1. *Let $G = (V, p)$ be a connected ERG, $\epsilon_1 = O\left(\frac{1}{(n-1)\,p}\right)$, and $\epsilon_2 = (1-p)^{n-1}$. For a transmission scheme defined by Algorithm NC_{FWB}, with γ chosen to ensure that all nodes can decode all messages, we have*

$$\frac{\gamma}{p} + 1 \leq \mathrm{E}_D(T_{NC}) \leq \frac{\gamma}{p}\,\frac{1+\epsilon_1}{1-\epsilon_2} + \max(1, \gamma). \tag{4.13}$$

Proof. ERGs have a Binomial degree distribution $\mathrm{B}(n-1,\, p)$. As we consider connected graphs, however, we must use a conditioned degree distribution. We know that each node has at least one neighbor, i.e., $d(u) > 0 \;\; \forall u \in V$. For this reason, we assume a positive Binomial distribution, which can be obtained by normalizing the Binomial distribution with the factor $1 - \mathrm{P}(D = 0)$. This yields

$$\mathrm{P}(D = d) = \frac{1}{1 - q^{n-1}} \binom{n-1}{d} p^d\, q^{n-1-d}, \tag{4.14}$$

with $q = 1 - p$ and $d \in \mathbb{Z}^+$.
The first negative moment of the degree is thus:

$$\begin{aligned} \xi_D = \mathrm{E}_D\left(\frac{1}{D}\right) &= \sum_{d=1}^{n-1} \frac{1}{d}\,\mathrm{P}(D = d) \\ &= \frac{1}{1 - q^{n-1}} \sum_{d=1}^{n-1} \frac{1}{d} \binom{n-1}{d} p^d\, q^{n-1-d}. \end{aligned} \tag{4.15}$$

This function can be developed into the following series [Rem04]:

$$\xi_D = \frac{1}{1-q^{n-1}} \sum_{i=0}^{r-1} \frac{i!\, q^i}{p^{i+1}\,(n-1)^{[i+1]}} + o\left(\frac{1}{(n-1)^{[r]}}\right),$$

for any $r \in \mathbb{Z}^+$, with $s^{[j]} = \frac{s!}{(s-j)!}$. Moreover, it can be rewritten as:

$$\xi_D = \frac{1}{1-q^{n-1}} \left(\frac{1}{(n-1)\,p} + O\left(\frac{1}{((n-1)\ p)^2}\right)\right). \tag{4.16}$$

Hence, we can compute

$$\begin{aligned}
(n-1)\ \xi_D = & \\
&= \frac{n-1}{1-q^{n-1}} \left(\frac{1}{(n-1)\,p} + O\left(\frac{1}{((n-1)\,p)^2}\right)\right) \\
&= \frac{1}{p\ (1-q^{n-1})} \left(1 + O\left(\frac{1}{(n-1)\,p}\right)\right) \\
&= \frac{1}{p}\,\frac{1+\epsilon_1}{1-\epsilon_2} && (4.17) \\
&\geq \frac{1}{p} && (4.18)
\end{aligned}$$

with $\epsilon_1 = O\left(\frac{1}{(n-1)\,p}\right)$ and $\epsilon_2 = q^{n-1} = (1-p)^{n-1}$.
Replacing (4.17) and (4.18) in (4.1), we get (4.13). □

Fig. 4.1 plots the analytical and simulation results in ERGs, showing that the simulated average value of T_{NC} lies within the analytical bounds of $\mathrm{E}_D(T_{NC})$ assuming $\epsilon_1 = \epsilon_2 = 0$. Section 4.3.1 explains the used simulation method.

4.2.4 Bounds for Binomial Random Geometric Graphs

Corollary 2. *Let $G = (V, r)$ be a connected RGG in a square with toroidal distance metric and area $A \gg \pi\ r^2$, and let $\beta = \frac{\pi\, r^2}{A}$, and $\epsilon_1 = O\left(\frac{1}{(n-1)\,\beta}\right)$, and $\epsilon_2 = (1-\beta)^{n-1}$ For a transmission scheme defined by Algorithm NC_{FWB}, with γ chosen to ensure that all nodes can decode all messages,*

$$\frac{\gamma}{\beta} + 1 \leq \mathrm{E}_D(T_{NC}) \leq \frac{\gamma}{\beta}\,\frac{1+\epsilon_1}{1-\epsilon_2} + \max(1, \gamma). \tag{4.19}$$

Proof. RGGs have a Binomial degree distribution $\mathrm{B}(n-1,\ \frac{\pi r^2}{A})$. Similar to Section 4.2.3, we derive:

$$\begin{aligned}
(n-1)\,\xi_D &= \frac{1}{\beta}\,\frac{1+\epsilon_1}{1-\epsilon_2} && (4.20) \\
&\geq \frac{1}{\beta} && (4.21)
\end{aligned}$$

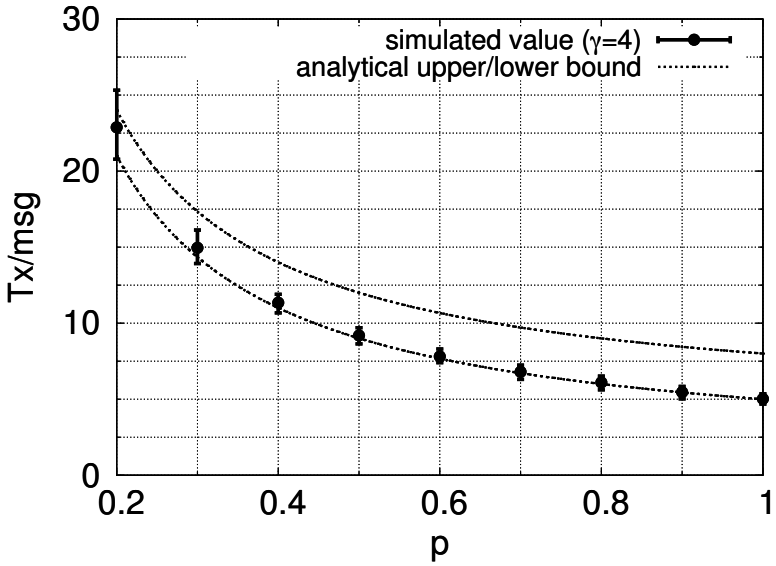


Figure 4.1: Number of Transmissions per Message using Network Coded Flooding in Erdős Rényi Random Graphs with 50 nodes

with $\beta = \frac{\pi r^2}{A}$, and $\epsilon_1 = O\left(\frac{1}{(n-1)\,\beta}\right)$, and $\epsilon_2 = (1-\beta)^{n-1}$.

Replacing (4.20) and (4.21) in (4.1), the above result follows. □

Fig. 4.2 plots the analytical and simulation results in RGGs, showing that the simulated average value of T_{NC} lies within the analytical bounds of $\mathrm{E}_D(T_{NC})$ assuming $\epsilon_1 = \epsilon_2 = 0$.

Corollaries 1 and 2 show that in ERGs and RGGs, the expected number of transmissions required to flood a message is asymptotically independent of the number of nodes n. It depends on other topological parameters and on the scaling factor γ of Algorithm NC_{FWB} which, according to the authors of [FWLB06], is independent of n.

4.2.5 Bounds for Small-World Networks

Corollary 3. *Let $G = (V, k, p)$ be a connected SWN, and let D be a random variable representing the degree of an arbitrary node in G. For a transmission scheme defined by the NC algorithm (Algorithm NC_{FWB}) with γ chosen to ensure that all nodes can decode all messages, we have*

$$(n-1)\,\frac{\gamma}{k} + 1 \leq \mathrm{E}_D(T_{NC}) \leq 2\,(n-1)\,\frac{\gamma}{k} + \max(1,\gamma). \tag{4.22}$$

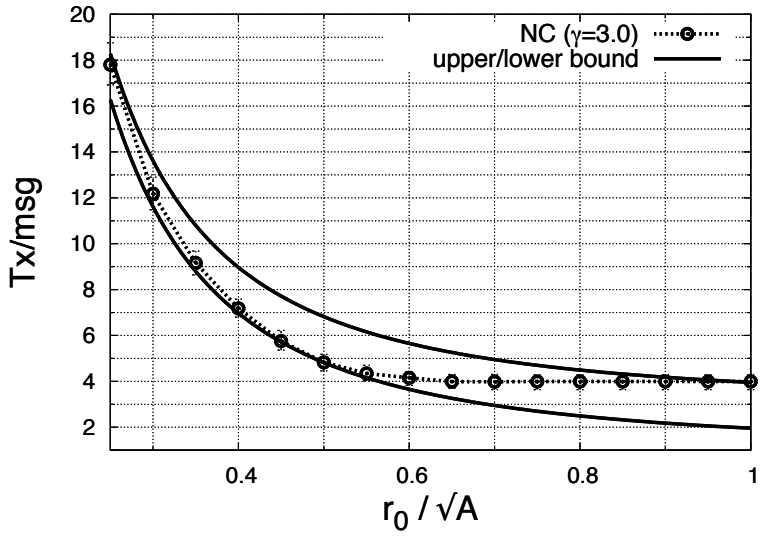


Figure 4.2: Number of Transmissions per Message using Network Coded Flooding in Random Geometric Graphs (toroidal space) with 50 nodes

Proof. The specialization of $\mathbb{E}_D(T_{NC})$ for SWNs is achieved by calculating

$$\xi_D = \sum_{d=1}^{n-1} \frac{1}{d} \mathrm{P}(D = d). \tag{4.23}$$

The probability mass function of the degree in SWN for large n is ([BW00]):

$$\begin{aligned} \mathrm{P}(D = d) \quad \approx \quad & \sum_{i=0}^{\min\left(d-\frac{k}{2},\frac{k}{2}\right)} \binom{\frac{k}{2}}{i} (1-p)^i \, p^{\frac{k}{2}-i} \\ & \cdot \frac{(\frac{k}{2}\, p)^{d-\frac{k}{2}-i}}{(d-\frac{k}{2}-i)!} \, e^{-p\frac{k}{2}}, \qquad d \geq \frac{k}{2}. \end{aligned} \tag{4.24}$$

The degree distribution $\mathrm{P}(D)$ presents two useful properties to derive an upper and lower bound for ξ_D. One is the expected degree $\mathrm{E}_D(D) = k$. The other is the minimum degree $\min(D) \geq k/2$, which arises from the lower cutoff of $\mathrm{P}(D) = d$, at degree $d = k/2$. Conjugating these results with the following inequality that holds for any discrete random variable $D > 0$ (see [Man69] and references therein):

$$\min(D) \leq \frac{1}{\mathrm{E}_D(D^{-1})} \leq \mathrm{E}_D(D), \tag{4.25}$$

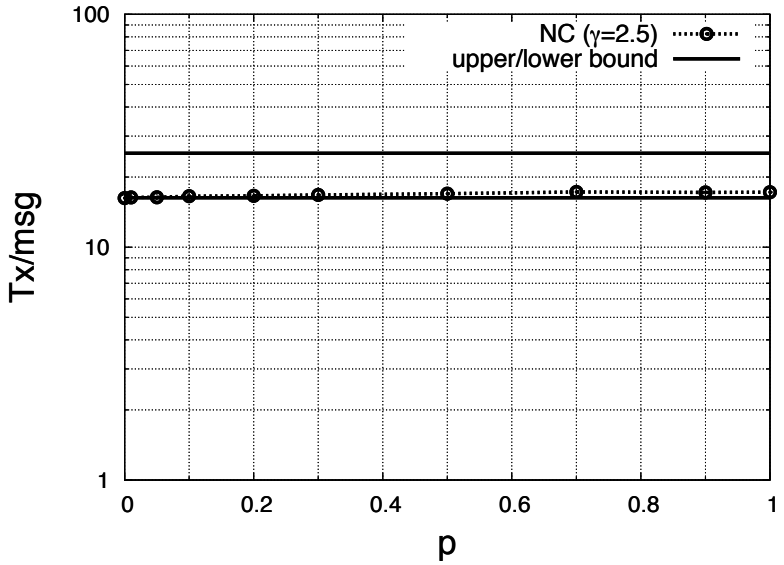
we get:

$$\frac{1}{k} \leq \mathrm{E}_D(D^{-1}) \leq \frac{2}{k}. \tag{4.26}$$

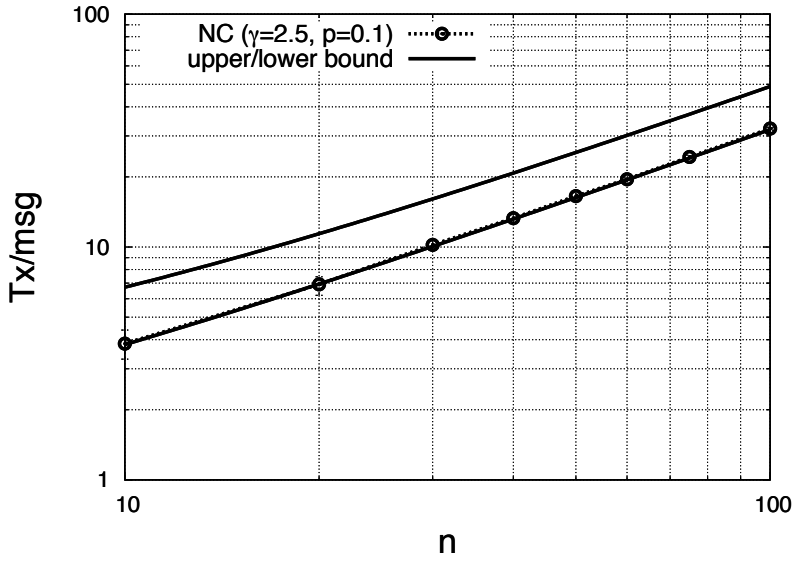
Replacing (4.26) in (4.1) we prove the above result.

□

Corollary 3 suggests that using the NC algorithm over SWNs, $\mathrm{E}_D(T_{NC})$ is independent of the rewiring probability p. Moreover, it shows that $\mathrm{E}_D(T_{NC})$ lies between two linear functions of n that differ by a factor of 2. Fig. 4.3(a) and Fig. 4.3(b) plot the analytical and simulation results of T_{NC} ($\gamma = 2.5$) in SWN graphs. They show that the average value of T_{NC} lies within the analytical bounds of $\mathrm{E}_D(T_{NC})$, and in particular that it is very close to the lower bound.



(a) As function of the rewiring probability



(b) As function of the number of nodes

Figure 4.3: Number of Transmissions per Message in Small-World Networks with Rewiring

4.3 – Simulation based Analysis

In this section we present a simulation study comparing NC and MPR flooding on ERGs, RGGs, and in SWN topologies.

4.3.1 Description of the Simulator and Simulation Setup

To conduct numeric analysis we developed a simulator written in C++. We implemented the RLNC based flooding, the MPR, and probabilistic flooding algorithms. The simulator enables the performance evaluation of these algorithms in networks modelled by ERGs, RGGs and SWNs. Moreover, it supports the analysis of topological properties of these graph models.

The MAC layer works in an idealized manner with perfect collision avoidance. The simulation time is divided in discrete rounds (time units), and each transmission/reception lasts one simulation round. In each round, the order of the node transmissions is randomly chosen, and each idle node is scheduled to transmit if and only if all its neighbors are idle (not in a receiving or transmitting state).

In our simulation setup each node has a message to be sent to all other network nodes. Therefore, each node acts simultaneously as a source, a relay, and as a sink.

For the implementation of random linear NC, we followed the framework described in [CWJ03] with coding operations over the $\mathbb{F}_{2^8}$ finite field (see Section 2.7). This field size is sufficient for practical networking scenarios ([CWJ03, FWLB06]) and has the advantage of allowing each field symbol to be stored in one byte. Packets are composed by a packet header and a data payload. The packet header contains a vector with the coding coefficients, while the data payload is meant for the coded message. A packet to transmit is generated by randomly combining the previously received packets through algebraic operations over $\mathbb{F}_{2^8}$. Decoding uses Gaussian-Jordan elimination allowing progressive decoding while coded packets are being received.

The complexity of decoding via Gaussian-Jordan elimination ($O(h^3)$, where h is the number of messages) is a limiting factor to the number of messages that can be combined (i.e. the generation size).

In the simulation of NC flooding all messages are allowed to be combined together if the opportunity arises. Thus the generation size is equal to the number of nodes n. Therefore the aforementioned decoding complexity constrains the number of nodes that can be used in the simulations.

4.3.2 Topology and Performance Metrics

Aiming at a reasonable comparison of NC and MPR, we consider the following metrics:

- Number of transmissions per source message T_{NC} required by the NC algorithm such that all nodes can decode all sent messages;
- Number of transmissions per source message T_{MPR} required by the MPR algorithm such that all nodes receive all sent messages;
- Delay: rounds elapsed between the transmission of a message by a source node and the reception (with MPR), or successful decoding (with NC) at a node;
- Delivery ratio (DR): ratio between number of messages successively received or decoded at a node and the number of sent messages.

For SWNs we further evaluate:

- Clustering coefficient C: defined in Section 2.1;
- Average distance L: defined in Section 2.1;
- Normalized Rank (NR): ratio between the rank of the decoding matrix of a node and the total number of source messages, averaged over all nodes;
- MPR set size: the cardinality of the MPR set of each node averaged over all nodes.

Each data point (mean, 10% and 90% quantile) in the simulation results is obtained from 100 repetitions of a simulation using different seeds for the random number generator.

4.3.3 Analysis of Erdős Rényi Random Graphs

In this set of simulations we compare MPR and NC flooding in ERGs with the edge probability $p \in [0.2, 1]$ and $n = 50$ nodes. For a fair comparison, Algorithm NC_{FWB} is simulated with scaling factors $\gamma \in \{0.5, 1.0, 2.0, 4.0\}$, chosen via simulation on an iterative trial-and-error approach to guarantee the existence of (γ, p) tuples that achieve 100% DR.

Fig. 4.4(a) shows that MPR flooding always guarantees 100% DR. NC requires sufficiently large γ for the same DR. Note that γ decreases with increasing p, and for $p = 1$ (fully connected graph) γ should be 0, since one transmission from a source reaches all other nodes (see Algorithm NC_{FWB}).

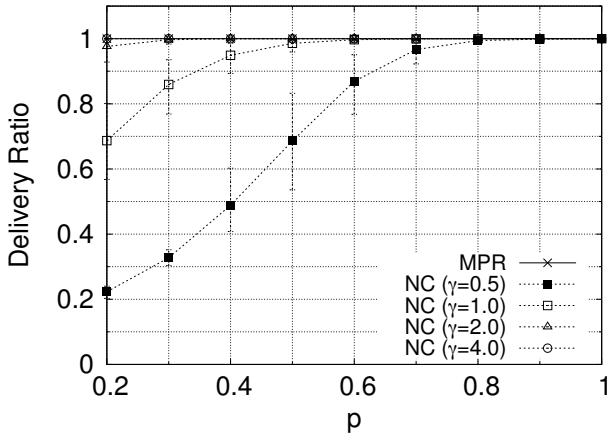
The "delay gain" obtained by NC is substantial for small p and sufficiently large γ (Fig. 4.4(b)). As $p \to 1$, the NC delay converges to the delay value of MPR (with $\gamma = 0$, not shown in the graph).

In Fig. 4.4(c), we observe that NC flooding (with sufficiently large γ and small p) outperforms MPR flooding in terms of number of transmissions. The fraction T_{NC}/T_{MPR} ranges from 0.6 ($\gamma = 4$) to 1 ($\gamma = 0$, not shown in the graph) when p increases from 0.2 to 1. This is an expected result, since with p converging to 1 the diameter of the graph reduces to 1 and consequently both NC and MPR schemes are able to broadcast a message with only one transmission.

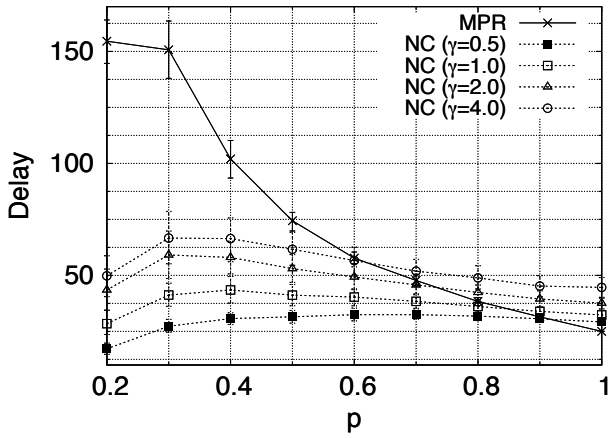
4.3.4 Analysis of Binomial Random Geometric Graphs

In this set of simulations, we compare both flooding algorithms in RGGs in a square area of size A, $n = 50$ nodes and the radio range r. We set $\frac{r}{\sqrt{A}} \in [0.4, 1]$, to ensure with high probability that all graph realizations are connected [Bet02]. The parametrization of the simulation results as a function of $\frac{r}{\sqrt{A}}$ enables the generalization of the results to different parameters.

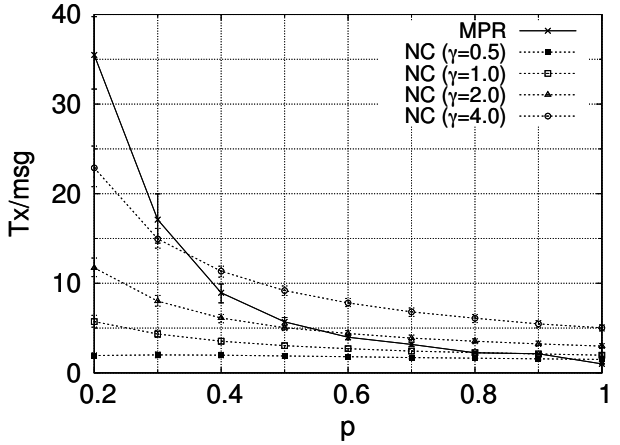
Fig. 4.5(a) presents the DR with different values of the scaling factor γ. Fig. 4.5(b) illustrates that NC, with sufficiently large γ (which can be inferred from Fig. 4.5(a)), presents a substantial "delay gain" (half the delay of MPR for $\frac{r}{\sqrt{A}} = 0.4$). This advantage vanishes as the network diameter converges to 1 ($r/\sqrt{A} \to 1$). From Fig. 4.5(c) we conclude that NC (Algorithm NC_{FWB}, with sufficiently large γ for 100% DR) presents no gain in terms of number of transmissions when compared to MPR. Since RGGs are often used to model wireless ad-hoc networks, this is a discouraging result for Algorithm NC_{FWB} [FWLB06], which cannot however be generalized to other NC algorithms.



(a) Delivery Ratio

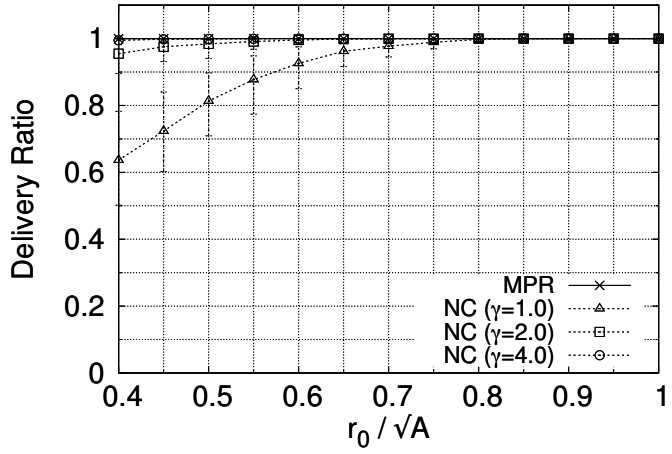


(b) Delay

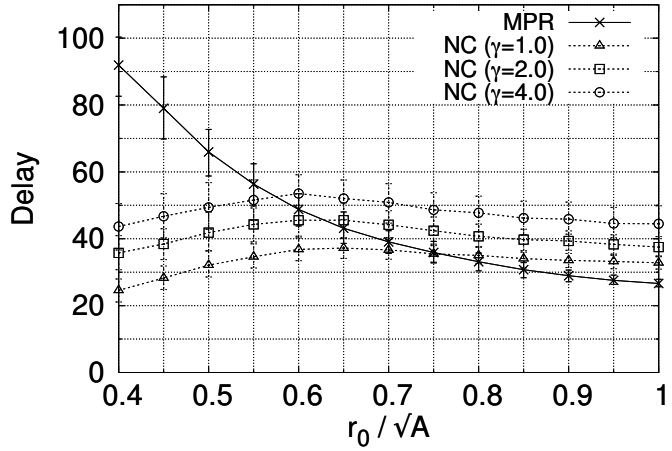


(c) Number of Transmissions per Message

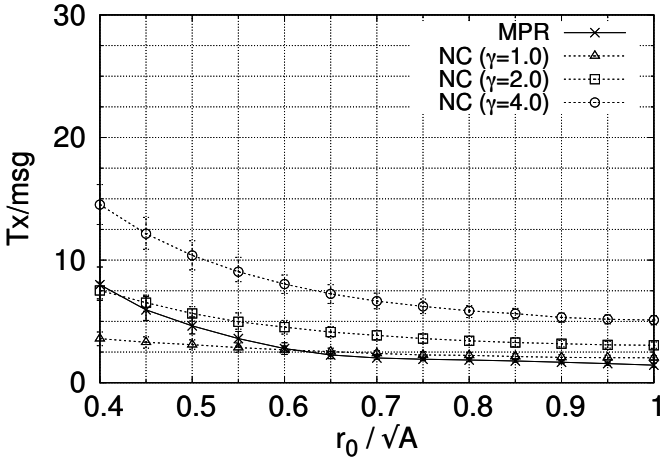
Figure 4.4: Analysis in Erdős Rényi Random Graphs



(a) Delivery Ratio



(b) Delay



(c) Number of Transmissions per Message

Figure 4.5: Analysis in Random Geometric Graphs (no torus)

We repeat the same simulations for an RGG with the nodes placed on a torus to avoid edge effects [Bet02]. To ensure connected graph realizations and a broad diameter range we set $\frac{r}{\sqrt{A}} \in [0.25, 1]$, recalling that it differs from the above non-toroidal case. Fig. 4.6(a) presents the DR for this case. Fig. 4.6(b) shows that NC with sufficiently large γ and small $\frac{r}{\sqrt{A}}$ still presents a substantial "delay gain" (1/3 the delay of MPR for $\frac{r}{\sqrt{A}} = 0.25$). From Fig. 4.6(c) we observe that in an RGG with torus geometry, with $r \ll \sqrt{A}$, NC again outperforms MPR in terms of the number of transmissions. The fraction T_{NC}/T_{MPR} ranges from 0.7 ($\gamma = 3$) to 1 (for $\gamma = 0$, not shown in the figure), as the diameter converges to 1. This behavior suggests that, as the diameter of the network falls, there is little or no benefit in using network coding. The distinct behaviors of T_{NC} with and without border effects suggest that Algorithm NC_{FWB} is affected negatively by the existence of border nodes in RGGs with average node degree smaller than the average degree of nodes near the center of the square.

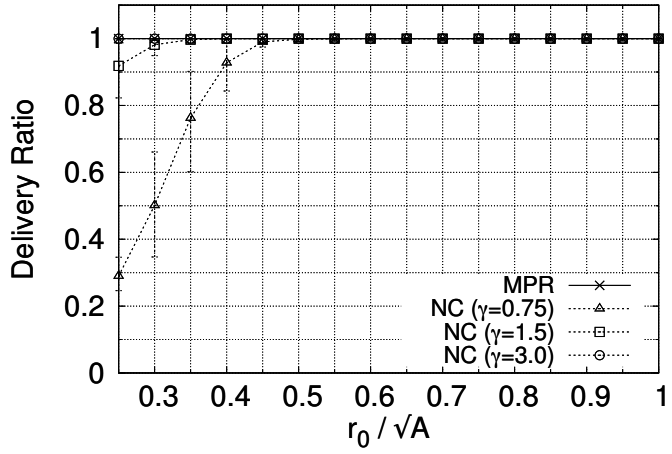
4.3.5 Analysis of Small-World Networks

We compare MPR and NC flooding in SWNs with $n = 50$ nodes, mean degree $k = 8$, and edge rewiring probability $p \in [0, 1]$.

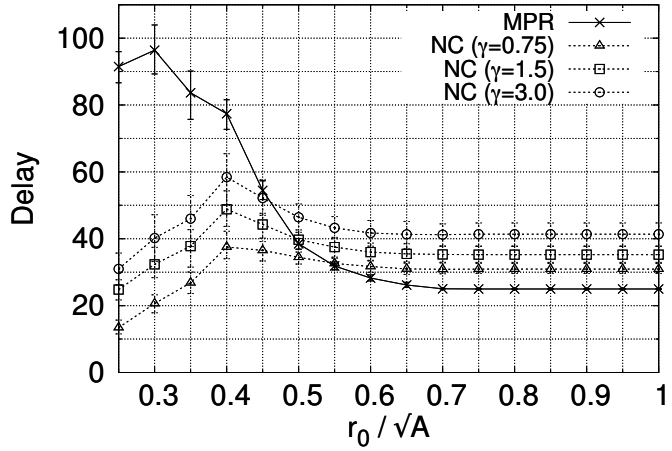
The NC algorithm (Algorithm NC_{FWB}) is simulated with scaling factors $\gamma \in \{0.5, 1.5, 2.5\}$, chosen via simulation on an iterative trial-and-error approach to guarantee the existence of (γ, p) tuples that achieve 100% DR.

Fig. 4.7(a) presents the normalized values of clustering coefficient $C(p)/C(0)$ and the average path length $L(p)/L(0)$, with $C(0) \simeq 0.64$ and $L(0) \simeq 3.57$. These curves follow the typical behavior of the topological properties of SWNs ([Wat99, New03]).

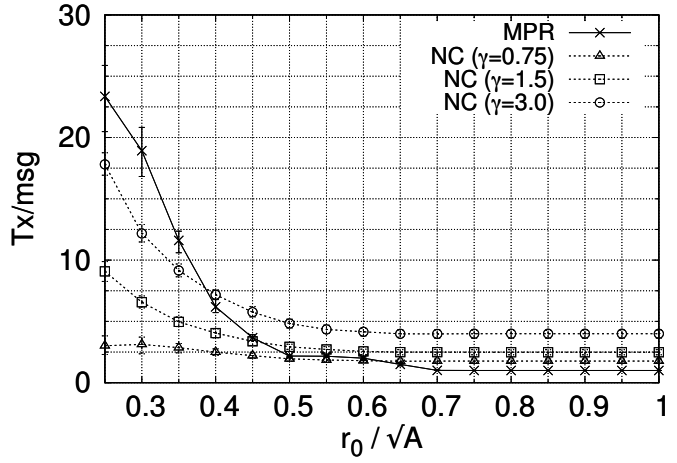
Fig. 4.7(b) presents the average MPR set size. This metric increases sharply for small p, stabilizing thereafter. Moreover, comparing Fig. 4.7(b) with Fig. 4.7(a) we find a correlation between the mean MPR set size and the reciprocal of C. This correlation can be interpreted as follows: since C_v is roughly equivalent to the probability of two neighbors of v being also neighbors of each other, a higher C_v implies a more 'cliquish' neighborhood. Therefore, the number of 1-hop neighbors necessary to reach all the 2-hop neighbors of v (MPR set) is expected to increase when C_v decreases. This is the observed case in our simulations when p converges from 0 to 1.



(a) Delivery Ratio



(b) Delay



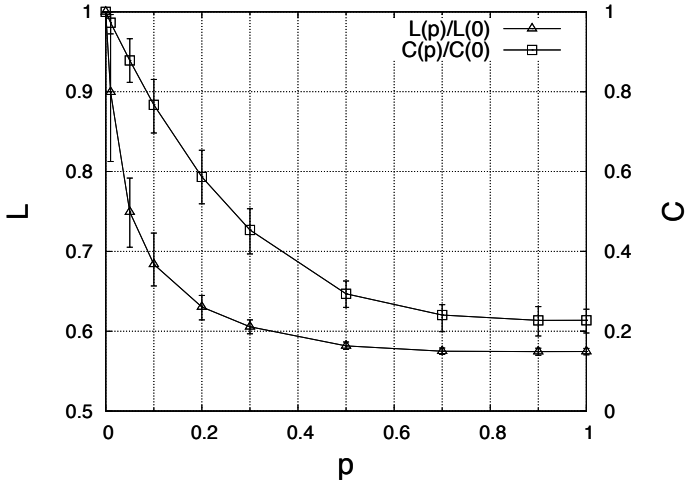
(c) Number of Transmissions per Message

Figure 4.6: Analysis in Random Geometric Graphs (torus)

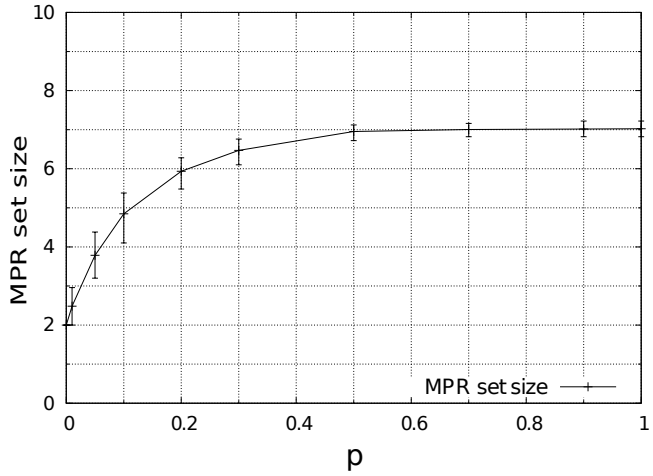
For successful decoding with random linear NC, the number of linearly independent combinations of coded messages received by a node needs to be equal to the number of source messages. Otherwise, a node may still decode a fraction of the source messages. This is illustrated in Fig. 4.7(c) and Fig. 4.8(a) which plot the NR and the DR respectively. For $\gamma = 0.5$ the NR is around 0.35 while the DR only reaches 0.2 (20%). For $\gamma = 1.5$ the NR is almost 1 and the DR is slightly smaller. With $\gamma = 2.5$ the NR is 1, yielding a DR of 100%. We also notice that for the same γ (e.g. $\gamma = 0.5$) both the NR and the DR keep fairly constant with p. This suggests that in SWNs the rewiring probability does not significantly affect the performance of the NC algorithm. This behavior can be interpreted as follows. Given that our NC algorithm is probabilistic, we might expect that the reduction of the diameter would contribute to an increase in the NR. On the other hand, since a larger C implies a higher number of redundant paths between nodes, we would expect the decrease of C to cause a decrease in NR. We argue that the combined reduction of L and C cancel one another yielding an almost constant NR (and DR) regardless of p.

Fig. 4.8(b) presents the number of transmissions per message for NC, MPR, and pure flooding. We observe that T_{MPR} degrades significantly with p, converging to the number of transmissions per message attained with pure flooding. As expected, T_{MPR} increases with the MPR set size (Fig. 4.7(b)). In contrast, T_{NC} is almost constant with the rewiring probability p, presenting a fairly low transmission cost when compared to pure flooding or MPR flooding. The fraction T_{NC}/T_{MPR} ranges from 0.77 ($p = 0$, with $\gamma = 1.5$) to 0.4 ($p = 1$, with $\gamma = 2.5$).

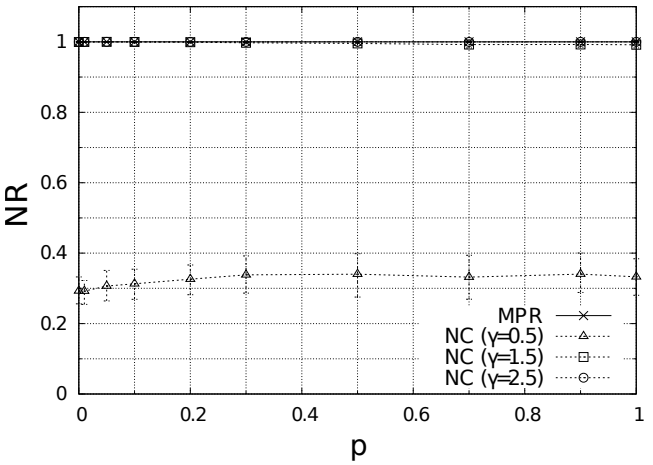
The delay behavior (Fig. 4.8(c)) presents the same trend as the number of transmissions. For sufficiently large γ, the delay ratio between NC and MPR ranges from around 0.5 ($p = 0$, with $\gamma = 1.5$) to 0.3 ($p = 1$, with $\gamma = 2.5$).



(a) Normalized Clustering Coefficient (C) and Distance (L)

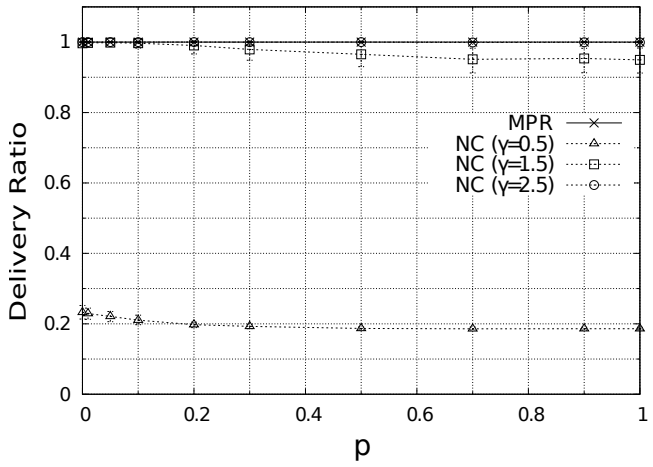


(b) MPR set size

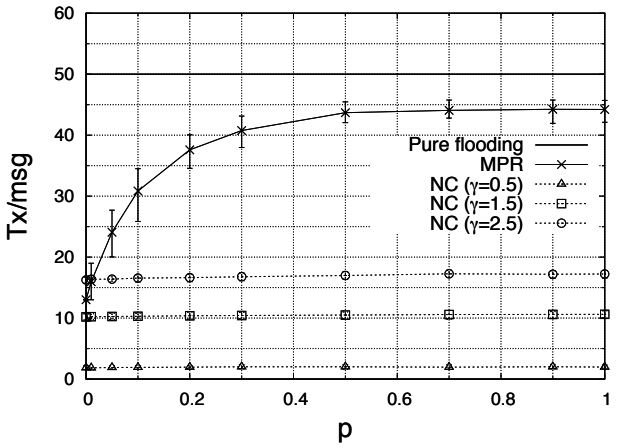


(c) Normalized Rank

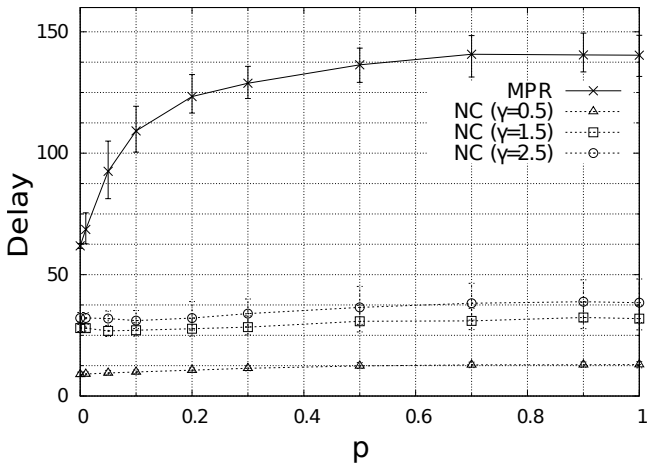
Figure 4.7: Analysis in Small-World Networks: (a) Normalized Clustering Coefficient and Distance; (b) MPR set size; and (c) Normalized Rank



(a) Delivery Ratio



(b) Number of Transmissions per Message



(c) Delay

Figure 4.8: Analysis in Small-World Networks: (a) Delivery Ratio; (b) Number of Transmissions per Message; and (c) Delay

4.4 – Concluding Remarks

Aiming at understanding how distinct forwarding paradigms influence the dissemination of information over communication networks, we selected one representative of each paradigm for our study: the network coding algorithm of [FWLB06] and replication based flooding MPR algorithm of [AQL02]. We evaluated (a) the number of transmissions per source message and (b) the incurred delay, and (c) the delivery ratio, under three relevant classes of random graph models.

Somewhat unintuitively, the analytical part of our work shows that over ERGs and RGGs, the number of transmissions required to flood a message with the NC flooding algorithm under consideration is asymptotically independent of the number of nodes. This observation becomes less surprising in retrospect, if we consider that in these classes of graphs the average node degree increases linearly with the number of nodes. Therefore, a higher number of nodes corresponds to a higher number of neighbors that can be reached by a single broadcast transmission. Since random linear network coding mixes multiple messages in a single transmission, it is very effective at exploiting the benefits of increased node density. With multipoint relays, however, the number of transmissions per message is not independent of the number of nodes.

In contrast, for SWNs, the analysis shows that the number of transmissions per message of the network coding algorithm scales linearly with the number of nodes. The reason for these distinct results can be understood by the fact that in SWNs with rewiring the average node degree remains fixed irrespective of the number of nodes.

Naturally, the number of transmissions depends on other features of the network topology, as evidenced both by Corollaries 1, 2, and 3, and by our simulation results. Consequently, the question as to which scheme should be preferred requires a nuanced answer.

In ERGs, NC flooding outperforms MPR flooding in terms of number of transmissions per source message; the extent of this gain is however deeply influenced by the diameter of the network. Reducing the diameter decreases both the number of transmissions and the delay gains. A unit diameter implies no gain at all.

In contrast, in general RGGs (non-toroidal distance metric) the considered NC flooding algorithm does not bring any benefits in terms of number of transmissions per message, when compared to MPR flooding. This appears to be in contradiction with the observation in [FWLB06]. However, it is worth noting that [FWLB06] focuses on RGGs on a torus and

compares NC with probabilistic flooding. Our results thus indicate that the existence of border effects in general RGG topologies has a negative effect on the performance of the considered NC flooding technique.

In SWNs, the analytical expression for the number of transmissions per message of network coding shows no dependency on the rewiring probability of the SWN model. This result is corroborated by the simulation results. In fact, the simulations highlight the stability of the NC performance metrics (delivery ratio, number of transmissions per message and delay) within all the rewiring range of the SWN model (i.e. with distinct clustering coefficients and average geodesic distance values). NC flooding demonstrates to be relatively immune to changes in local connectivity parameters (i.e. presence or absence of strong local connectivity), suggesting that the strengths of the studied network coding algorithm in SWN topologies stems from its network-wide coding/decoding operation paradigm. In turn, the MPR flooding algorithm does not produce significant overhead reduction in terms of number of transmissions per source message, in poorly clustered SWN topologies. Its *Achilles' heel* resides on the scoped and limiting view of the topological properties centered in the neighborhood of each node.

"The unavoidable price of reliability is simplicity."

C.A.R. Hoare

5

Applications in Dynamic Sensor Networks

In the previous chapters we analyzed main representatives of replication and network coded based dissemination algorithms. We studied their reachability and transmission cost by means of mathematical analysis and numerical experiments. Moreover we studied the interplay between the network topology and the performance of dissemination algorithms. To do so, we modelled networks as random graphs and we used tools from graph theory and stochastic geometry.

In this chapter we apply graph theoretical tools to model a sensor-actuator networked system to support building evacuation in disaster scenarios such as fires or earthquakes. Hazard information, continuously gathered by sensor nodes, is disseminated throughout the network. This information is used to infer, in a distributed way, secure exit paths to be followed by evacuees.

Long-established static building evacuation planning comprehends an a-priori identification of default exit paths. Static signaling panels, deployed at key points all over the building, indicate the default exit direction to follow. This approach presents, however, a major drawback. Due to the dynamics of the hazard affecting the building (e.g. the propagation of fire or smoke, damages in the structure of the building blocking pathways),

the predefined exit paths may be or may become unsafe.

Henceforth, we design a sensor-actuator networked system for emergency response in indoor scenarios that tackles the dynamics of hazard propagation over the building. The purpose of the system is to guide people to the exits of the building throughout secure paths. These paths are computed autonomously by each node, having as input sensor measurements flooded by all network nodes.

The main contributions of this work are:

- Modeling of a sensor-actuator network for the support of building evacuation in emergency scenarios by using graph abstractions of the building topology, associated sensor network, and radio connectivity among sensor nodes;
- Definition of path safety metrics;
- Algorithms for the computation of the shortest safest paths to exits;
- Implementation of a prototype of the system;
- Evaluation of sensor data dissemination using the prototype.

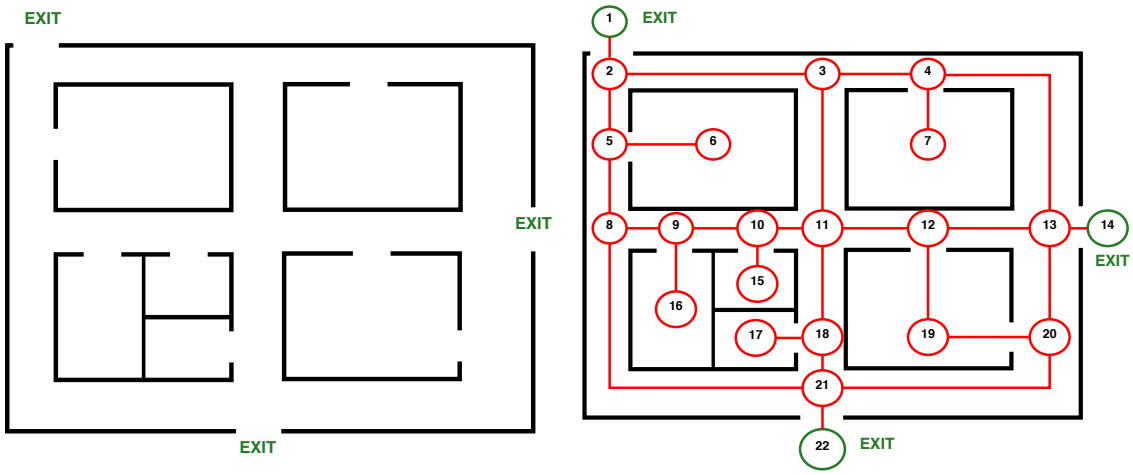
The chapter is organized as follows. Section 5.1 presents modeling options and assumptions. Section 5.2 explore alternative security metrics and presents algorithms to be used in a Wireless Sensor Network Network (WSN) to support building evacuation. Section 5.3 presents an overview of the functional architecture of the software of the wireless sensor nodes. Moreover it presents an evaluation of sensor data dissemination using the developed wireless sensor network prototype. Finally, Section 5.4 summarizes the main results of the chapter.

Parts of the results of this chapter were achieved in collaboration with Luís Pinto, Pedro Santos, Traian Abrudan, Fausto Vieira, and João Barros.

5.1 – Modeling Assumptions

5.1.1 Building Topology

A building layout or topology may be represented by an undirected graph $G_B(V_B, E_B)$, which we denote as *building graph* or B-graph:



(a) Building layout (b) B-graph

Figure 5.1: Building layout and corresponding B-graph

- The set of nodes (or B-nodes) V_B, represents rooms, halls, and other building regions where people usually stand.[1] Moreover, B-nodes also represent crossings (e.g. convergence between a room exit and a pathway along a corridor) and other regions where more than two pathways converge. B-nodes that have immediate access to the outside of the building are called *exit nodes*, or *exit B-nodes*. They form the set $V_{\mathrm{x}} \in V_B$ of exit nodes.

- The set of edges (or B-edges) E_B, connects B-nodes. They represent corridors, stairs and other pathways of the building layout thorough which people can walk.
  A B-edge $\{u, v\}$ has an associated length $d(u, v) > 0$ which is the walkable distance between its end nodes.

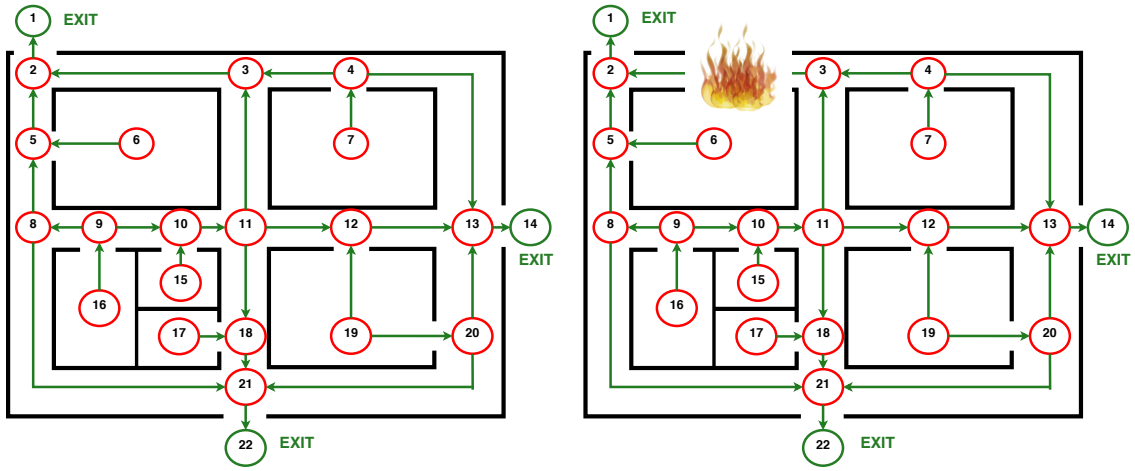
Therefore, B-nodes represent key region points where people may choose to follow one of the incident B-edges to move to an adjacent region (i.e. an adjacent B-node). Fig. 5.1 shows an example of a building layout represented by a B-graph.

5.1.2 Emergency Navigation Graph

A building evacuation plan can be represented by a directed acyclic graph $G_N(V_N, E_N)$, derived from the B-graph, that we denote as *emergency navigation graph* or N-graph:

- Its set of nodes V_N may be equal to the set of nodes V_B of the B-graph;

[1] We define a *building region* as a continuos space within a building in which people can stand or walk through (e.g. a room, a hall, a line-of-sight corridor, a line-of-sight staircase).



(a) Static navigation graph

(b) Static navigation graph fails to safely guide evacuees

Figure 5.2: Static navigation graph

- Its set of edges E_N may be built from the set of edges E_B by transforming each undirected edge $\{u, v\} \in E_B$ in a directed edge (u, v) or (v, u) in E_N.

This transformation must ensure that G_N is acyclic, and that from each node $u \in V_N$ there is a directed path to at least one exit node (usually the shortest path).

The major disadvantage of a static building evacuation plan is that the corresponding emergency navigation graph does not adapt to the existence or to the propagation of hazards. Some of the default exit paths may misguide evacuees throughout hazard regions endangering their lives. To overcome this limitation, the emergency navigation graph G_N needs to be dynamic (i.e a time-variant navigation graph). Having as input the information about the hazard propagation, it must adapt to ensure whenever possible, the existence of safe exit paths from any point in the building. This can be accomplished by deploying a network of sensor and decision nodes that continuously monitor the hazard state, updating accordingly the emergency navigation graph. Moreover, static signalization panels must be replaced by more versatile panels that display real-time information about the safe exit direction to follow, according to the evolution of the dynamic emergency navigation graph.

Fig. 5.2 presents a possible navigation graph for the building layout of Fig. 5.1, exemplifying the limitations of static navigation graphs.

5.1.3 Wireless Sensor Network

The hazard detection and propagation monitoring and the consequent computation of the emergency navigation graph is to be performed by a network of wireless sensor nodes

deployed all over a building. The physical deployement of the sensor nodes has to guarantee a convenient coverage of the building to ensure effective hazard detection. Moreover, it has to guarantee multi-hop radio connectivity between all sensor nodes to allow global dissemination of hazard information.

The WSN performs the following functions:

- Hazard sensing - nodes must be equipped with sensing units to collect relevant hazard information. The deployement of sensor nodes over the building must ensure that all its regions are conveniently monitored.
- Evacuation path computation - nodes have to compute the direction to be followed by evacuees to leave the building safely (i.e. to compute the emergency navigation graph). This function is performed having as input the sensed data mapped over the building graph.
- Actuation - nodes have to actuate on signaling devices (such as signalization panels) to inform evacuees about the safe exit path to follow. The actuation function has as input the emergency navigation graph.
- Communication - the data sensed by each node needs to be conveyed to possibly all other nodes to allow them to compute the emergency navigation graph.

We advocate a sensor network functional design in which each node is responsible for computing autonomously the exit paths. Such design is in principle more robust, and therefore, more appropriate to disaster scenarios than a design encompassing centralized decision services. In case of sensor node destruction due to hazard propagation, the wireless communication network may become disconnected. Even so, each functional node is still able to compute a subset of the exit paths resorting to its partial knowledge of the building hazard state. Hence, we opt for a WSN where all nodes have similar capabilities and functional roles. Each node is responsible for computing autonomously the emergency navigation graph from the sensed hazard information collected and disseminated by all nodes.

5.1.4 Spatial and Radio Graphs

The WSN has two associated graphs: A *spatial graph* also denoted as S-graph, and a *radio connectivity graph* denoted as R-graph.

The S-graph $G_S(V_S, E_S)$ is a specialization of the B-graph that maps the deployed sensor nodes into the building topology. The node set V_S is a superset of V_B (i.e. $V_B \subseteq V_S$) containing all deployed sensor nodes. Each B-node is mapped to one corresponding S-node in the S-graph. The remaining nodes (i.e. nodes of $V_S \setminus V_B$) are mapped into "intermediate

points" over B-edges. The "interpolated edges" of the B-graph are therefore mapped into 1-dimensional chains of S-edges (i.e. simple paths) in the S-graph. The computation of the emergency navigation graph G_N will be performed having as input the sensor measurements in the edges of the S-graph.

The R-Graph $G_R(V_R, E_R)$ represents the wireless connectivity between the sensor nodes. It constitutes the communication graph over which the S-graph mapped hazard information shall be disseminated. The node set V_R is equal to the node set V_S of the S-graph. The edge set E_R contains all node pairs which are in radio communication range of each other.

From a modeling perspective, we consider that each S-node has a sensing unit and an actuation unit per incident S-edge. Each sensing unit is in charge of collecting the data from the sensing probes of the corresponding S-edge. The emergency navigation graph G_N shall be computed by each sensor node having as input the S-graph and the sensor measurements associated to each of its S-edges. Each actuation unit is responsible for sending command instructions for the signaling devices associated to the corresponding S-edge, according to the computed emergency navigation graph G_N.

5.2 – Emergency Navigation Graph Computation

In this section we address the problem of finding shortest safest exit paths from a building suffering a disaster event. We present a generic solution in which we abstract a building topology by its B-graph, having hazard information associated to its edges. This generic method is, in practice, easily transposed to the WSN solution by using the S-graph (instead of the B-graph) as the representation of the building topology.

We first present a formal description of the problem of determining the emergency navigation graph G_N. Then we propose unambiguous quantifiable definitions for path safety and we propose a discrete and a continuous hazard metric to quantify path security. Finally, we propose algorithms to compute the shortest safest exit paths from nodes and from edges (i.e. the emergency navigation graph G_N).

5.2.1 Problem Statement

Let $G_B(V_B, E_B)$ be a graph that represents the topology of a building, with a set of exit nodes $V_{\mathrm{x}} \subseteq V_B$. Each edge $e = \{u, v\} \in E_B$ has a length $\mathrm{d}(e) > 0$. Moreover, let H be a totally ordered set of states whose elements represent the severity of a hazard, and let h be a mapping $E_B \to \mathrm{H}$ that assigns to each edge $e \in E_B$ a hazard state $\mathrm{h}(e)$, which is time-variant.

The mapping h is a function that takes as input a vector of sensor measurements of an edge e, and gives as output a hazard state $\mathrm{h}(e)$.

The goal is to find the shortest safest directed exit path from each point in the building (represented by G_B) to one of the exit nodes V_x, each time there is a change in the state $\mathrm{h}(e)$ of an edge e of E_B. Obtaining a solution to this problem encompasses solving two sub-problems:

1. Finding the shortest safest directed exit paths from the points of the building represented by nodes in V_B to one of the exit nodes in V_x. The solution to this problem yields a directed forest $G_F = (V_F, E_F)$ where $V_F = V_B$. The roots of the trees of G_F are the exit nodes $v_x \in V_\mathrm{x}$. Note that G_F is a navigation graph in which some of the edges of E_T may be absent.

2. Finding a shortest safest exit direction from any point in the building abstracted by an edge of E_B. This is particularly important for points mapped to edges of E_B which are absent from E_F.

5.2.2 Security Metrics

To be able to solve this optimization problem we must (a) agree on possible unambiguous definitions for path safety that are quantifiable, and, based on that, we need (b) to find a cost metric $\mathrm{c}(e)$ for an edge e that encompasses both its length $\mathrm{d}(e)$ and its hazard state $\mathrm{h}(e)$, that is in agreement with the path safety definition.

The cost $\mathrm{c}(P)$ of a path P is a function of the costs $\mathrm{c}(e)$ of its edges, according to the following expression:

$$\mathrm{c}(P) \triangleq \sum_{e \in P} \mathrm{c}(e). \tag{5.1}$$

If a path P_1 is safer than a path P_2, denoted as $P_1 \prec P_2$, then $\mathrm{c}(P_1) < \mathrm{c}(P_2)$.

In the following we present two alternative definitions of path safety encompassing a discrete and a continuous hazard state set H.

5.2.2.1 Discrete Hazard Metric (DHM)

Let $\mathrm{H} \in \mathbb{Z}$ be a totally ordered finite countable set of states that represent the severity of a hazard. With no loss of generality, assume $\mathrm{H} = \{0, 1, 2, 3\}$. The set is ordered by increasing level of danger. Table 5.1 assigns a meaning to each element of H.

A path P_1 is safer than a path P_2 (i.e. $P_1 \prec P_2$) if one of the following conditions holds:

(a) P_2 has at least one edge whose hazard state is higher than the maximum of the hazard

Table 5.1: Semantics of the elements of H

State	Alias	Semantic
0	GREEN	Safe
1	YELLOW	Low danger
2	ORANGE	Medium danger
3	RED	High danger

states of the edges of P_1;

(b) The maximum of the hazard states (h_{max}) of the edges of P_1 is equal to the maximum of P_2, and P_1 has a smaller sum of the edge lengths with a hazard state equal to the maximum;

(c) In case the above stated maximum hazard states and edge length sums are equal, P_1 is safer than P_2 if, only considering the edges with hazard state smaller than h_{max}, one of the conditions (a) or (b) hold.

We now present an edge and a path safety cost function that is in agreement to the above path safety definition. The strategy we follow in the definition of the edge cost function is to weight the edge length by a factor that grows exponentially with its hazard state.
A safety cost $\mathrm{c}(e)$ of an edge e with length $\mathrm{d}(e)$ and hazard state $\mathrm{h}(e)$ is given by:

$$\mathrm{c}(e) \triangleq \mathrm{d}(e) \cdot \alpha^{\mathrm{h}(e)}, \tag{5.2}$$

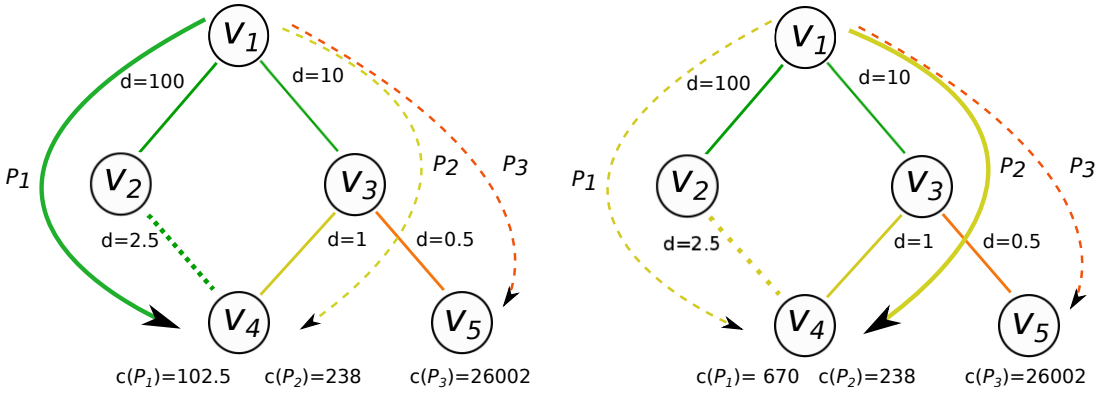
and the safety cost of a path P is:

$$\mathrm{c}(P) \triangleq \sum_{e \in P} \mathrm{c}(e) = \sum_{e \in P} \mathrm{d}(e) \cdot \alpha^{\mathrm{h}(e)}. \tag{5.3}$$

A choice of the parameter α ensuring that the path safety cost function is in agreement with the above path safety definition, is:

$$\alpha = \frac{\sum_{e \in E_B} \mathrm{d}(e)}{\min_{e \in E_B}\{\mathrm{d}(e)\}}. \tag{5.4}$$

Proof. According to the path safety definition, a path P_2 with a total edge length ϵ in hazard state $n+1$ is more dangerous than any other path P_1 with maximum edge hazard state n. With no loss of generality, consider a particular case in which P_2 is composed by just one edge with the smallest edge length ϵ of the graph G_B, and that it has a hazard state of $n+1$. Moreover, assume that a hypothetical path P_1 composed by the remaining edges, has a total



(a) $\mathrm{h}(\{v_2, v_4\})$ is GREEN. P_1 is the safest path

(b) $\mathrm{h}(\{v_2, v_4\})$ changed to YELLOW, P_2 is now the safest path

Figure 5.3: Comparison of exit path safety costs from node v_1 to exit nodes in $V_x = \{v_4, v_5\}$, using the DHM model. The hazard state of the edges is represented by its colors (see Table 5.1). Three alternative paths are considered: $P_1 = (v_1, v_2, v_4)$, $P_2 = (v_1, v_3, v_4)$, and $P_3 = (v_1, v_3, v_5)$.

edge length $(\sum_{e \in E_B} \mathrm{d}(e)) - \epsilon$ with hazard state n. Hence,

$$\begin{aligned} \mathrm{c}(P_2) &> \mathrm{c}(P_1) \\ \epsilon \cdot \alpha^{n+1} &> \left(\left(\sum_{e \in E_B} \mathrm{d}(e) \right) - \epsilon \right) \cdot \alpha^n \\ \alpha &> \frac{\left(\sum_{e \in E_B} \mathrm{d}(e) \right) - \epsilon}{\epsilon} \\ \alpha &> \frac{\sum_{e \in E_B} \mathrm{d}(e)}{\epsilon} - 1 \end{aligned} \tag{5.5}$$

Conjugating the above inequality with the fact that $\epsilon = \min_{e \in E_B}\{\mathrm{d}(e)\}$, an appropriate choice for α is the one given by (5.4). □

Fig. 5.3 shows alternative escape paths together with its costs from node v_1 to exit nodes v_4 and v_5. The hazard state of each edge is represented by its color, according to Table 5.1. The value of α (calculated by equation (5.4)) used to determine the path costs is 228. Fig. 5.3(a) shows that, although paths P_2 and P_3 are shorter than P_1, they have edges with higher hazard state (*YELLOW* and *ORANGE*, respectively). Therefore, P_1 is, according to the above safety definition, the safest of the three paths. This is corroborated by its path safety cost which is the smallest. In Fig 5.3(b), the hazard state of edge $\{v_2, v_4\}$ changed from *GREEN* to *YELLOW*. P_2 is now the safest path, since it has the smallest total edge length with hazard state *YELLOW*, and P_3 has an edge with higher hazard state (*ORANGE*). Again, this is in agreement with the chosen cost function, which now yields for P_2 the smallest safety cost.

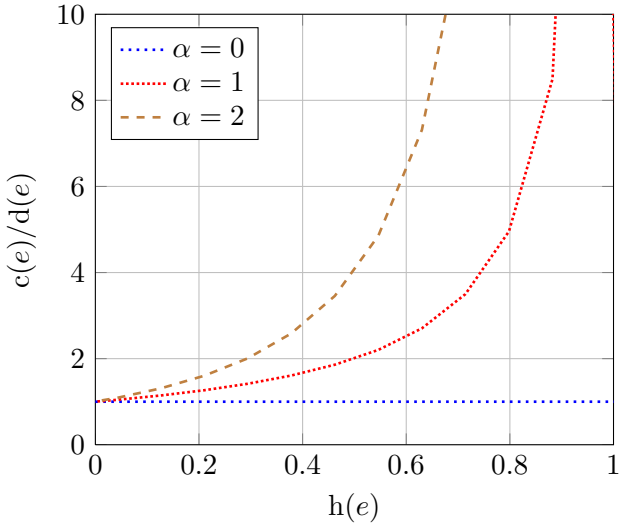


Figure 5.4: Normalized edge safety cost $\frac{\mathrm{c}(e)}{\mathrm{d}(e)}$ in the CHM model as function of the edge hazard state $\mathrm{h}(e)$ for $\alpha \in \{0, 1, 2\}$.

5.2.2.2 Continuous Hazard Metric (CHM)

Let $\mathrm{H} \subset \mathbb{R}$ be a totally ordered continuous set of states that represent the severity of a hazard. With no loss of generality, assume $\mathrm{H} = [0, 1)$. The set is ordered by increasing level of danger. The value $\mathrm{h}(e) = 0$ corresponds to a safe state while $\mathrm{h}(e) \to 1$ corresponds to a very dangerous state, meaning that edge e should not be used for building evacuation purposes.

We define the safety cost $\mathrm{c}(e)$ of an edge e to be a function of its hazard state $\mathrm{h}(e)$ and its length $\mathrm{d}(e)$ according the following expression:

$$\mathrm{c}(e) \triangleq \frac{\mathrm{d}(e)}{(1 - \mathrm{h}(e))^{\alpha}}, \tag{5.6}$$

where $\alpha \geq 0$ is a parameter that controls the grow rate of $\mathrm{c}(e)$.

The strategy used in the definition of the above edge safety cost function is to cause the edge length to grow hyperbolically (for $\alpha > 0$) with its hazard state. A secure edge ($\mathrm{h}(e) = 0$) has safety cost equal to its length $\mathrm{d}(e)$. Furthermore, as the hazard state of an edge converges to one, the corresponding safety cost converges asymptotically to infinity, reducing sharply the likelihood of its use in a escape path.

Fig. 5.4 plots the edge safety cost $\mathrm{c}(e)$ normalized over $\mathrm{d}(e)$ for distinct values of the exponent α. The value $\alpha = 0$, is a special case where the safety cost is not influenced by the hazard state $\mathrm{h}(e)$, assuming a value $\mathrm{c}(e) = \mathrm{d}(e)$.

The safety cost of a path P is given by:

$$\mathrm{c}(P) \triangleq \sum_{e \in P} \frac{\mathrm{d}(e)}{(1 - \mathrm{h}(e))^{\alpha}}. \tag{5.7}$$

5.2.3 Safest Exit Paths from Nodes

The problem of finding the minimum cost exit paths from each of the nodes in V_B to one of the exit nodes $v_x \in V_{\mathrm{x}}$ can be classified as a *Multiple-Destination Shortest Paths* (MDSP) problem. A generic strategy to solve this problem is the following:

A) Transform the MDSP problem into a *Single-Destination Shortest Paths* (SDSP) problem by:

   (a) Adding a virtual destination node v_d to V_B;

   (b) Adding a virtual edge $\{v_d, v_x\}$ with cost $\mathrm{c}(\{v_d, v_x\}) = 0$ to E_B for each node $v_x \in V_{\mathrm{x}}$.

B) Solve the SDSP problem for G_B with v_d as destination node using a "standard" SDSP algorithm (e.g. Bellman-Ford or Dijkstra's algorithm [CSRL01, Chapter 24]). The outcome should be a minimum cost tree G_T rooted at v_d.

C) Derive the solution to the MDSP problem from G_T. It should be the induced forest G_F that arises by removing v_d from G_T.

Data Structures in Dijkstra's Algorithm

The SDSP Dijkstra's algorithm resorts to two key data structures (see [CSRL01, Chap. 24]):

- A predecessor array $\pi[v]$ indexed by the nodes v of V_B, having as values, either a node of V_B or the symbol NIL. The symbol NIL means "no object at all", indicating that a node v has no predecessor whenever $\pi[v] = \mathrm{NIL}$.

- A cost estimate array $\hat{\mathrm{c}}[v]$ indexed by the nodes of V_B. Each $\hat{\mathrm{c}}[v]$ value is an upper bound for the cost of a minimum cost path from the destination v_d to a node $v \in V_B$.

These arrays are updated in such a way that at the end of the execution of the Dijkstra's algorithm, each value $\hat{\mathrm{c}}[v]$ represents the cost of a minimum cost directed path between the destination node v_d and the node v. Moreover, each $\pi[v]$ value represents the predecessor node of v in that path.

MDSP Method

The solution, denoted as *MDSP method*, that we propose to the MDSP problem is an adaptation of the SDSP Dijkstra's algorithm. It has the advantage over the generic approach described above of not needing a virtual destination node.

Let us first define the minimum cost $\delta(u, V)$ of the paths between a node u and the nodes of a node set V as:

$$\delta(u,V) \triangleq \begin{cases} \min\left(\mathrm{c}(P) : u \overset{P}{\rightsquigarrow} v, v \in V\right) & \text{if there is a path from } u \text{ to} \\ & \quad \text{at least one node } v \in V. \\ \infty & \text{otherwise.} \end{cases}$$

The meaning of the cost estimate array $\hat{\mathrm{c}}[v]$ in the *MDSP method* has a central difference from the one assumed in the SDSP Dijkstra's algorithm. For each node $v \in V_B \setminus V_\mathrm{x}$, the cost $\hat{\mathrm{c}}[v]$ is an upper bound for $\delta(v, V_\mathrm{x})$. I.e., it is an upper bound for the minimum of the costs of all the paths between v and the destination nodes $v_x \in V_\mathrm{x}$ (and not with a specific destination node).

The MDSP method is composed by Algorithm 7 and 8. Algorithm 7 initializes the predecessor and cost estimate arrays $\pi[v]$ and $\hat{\mathrm{c}}[v]$. In particular, $\hat{\mathrm{c}}[v_x]$ is initialized with a cost equal to zero for all destination nodes $v_x \in V_\mathrm{x}$ (being the main difference to the Algorithm *INITIALIZE-SINGLE-SOURCE* in [CSRL01, Chap. 24]).

Algorithm 8 differs from Algorithm *DIJKSTRA* in [CSRL01, Chap. 24] by replacing the call to *INITIALIZE-SINGLE-SOURCE* by a call to *MDSP_INITIALIZE_ARRAYS* (Algorithm 7). S is a set of nodes for which the minimums of the costs of the paths to nodes $v_x \in V_\mathrm{x}$ has already been determined (i.e. $\hat{\mathrm{c}}[v] = \delta(u, V_\mathrm{x}), \forall v \in S$). Q is a min-priority queue of nodes $v \in V_B$ keyed by the cost estimate values $\hat{\mathrm{c}}[v]$. The algorithm repeatedly selects the node $u \in V_B \setminus S$ with the minimum cost estimate, adds u to S, and updates the minimum cost estimate $\hat{\mathrm{c}}[v]$ and the predecessor node $\pi[v]$ for all neighbors v of u.

Algorithm 7 MDSP_INITIALIZE_ARRAYS

MDSP_INITIALIZE_ARRAYS (G_B, V_x)

1: **for** each node $v \in V_B$ **do**
2: $\quad\hat{\mathrm{c}}[v] \leftarrow \infty$
3: $\quad\pi[v] \leftarrow NIL$
4: **end for**
5: **for** each node $v_x \in V_\mathrm{x}$ **do**
6: $\quad\hat{\mathrm{c}}[v_x] \leftarrow 0$
7: **end for**

After the completion of Algorithm 8, the predecessor subgraph $G_F = (V_F, E_F)$ induced by the predecessor array $\pi[v]$ constitutes a minimum cost directed forest having as roots the nodes in the destination node set V_x:

Algorithm 8 MDSP_MAIN

MDSP_MAIN $(G_B, V_\mathrm{x}, \mathrm{c}())$

```
1: MDSP_INITIALIZE_ARRAYS (G_B, V_x)
2: S ← ∅
3: Q ← V_B
4: while Q ≠ ∅ do
5:    u ← EXTRACT_MIN(Q)
6:    S ← S ∪ {u}
7:    for each node v ∈ N(u) do
8:       if ĉ[v] > ĉ[u] + c({u, v}) then
9:          ĉ[v] ← ĉ[u] + c({u, v})
10:         π[v] ← u
11:      end if
12:   end for
13: end while
```

- V_F corresponds to the set of vertices of G_B with non-NIL predecessors plus the destination node set V_x (i.e. $V_F = \{v \in V_B : \pi[v] \neq \mathrm{NIL}\} \cup V_\mathrm{x}$);

- E_F is a directed edge set induced by $\pi[v]$ for nodes in V_F (i.e. $E_F = \{(v, \pi[v]) \in E_B : v \in V_F \setminus V_\mathrm{x}\}$).

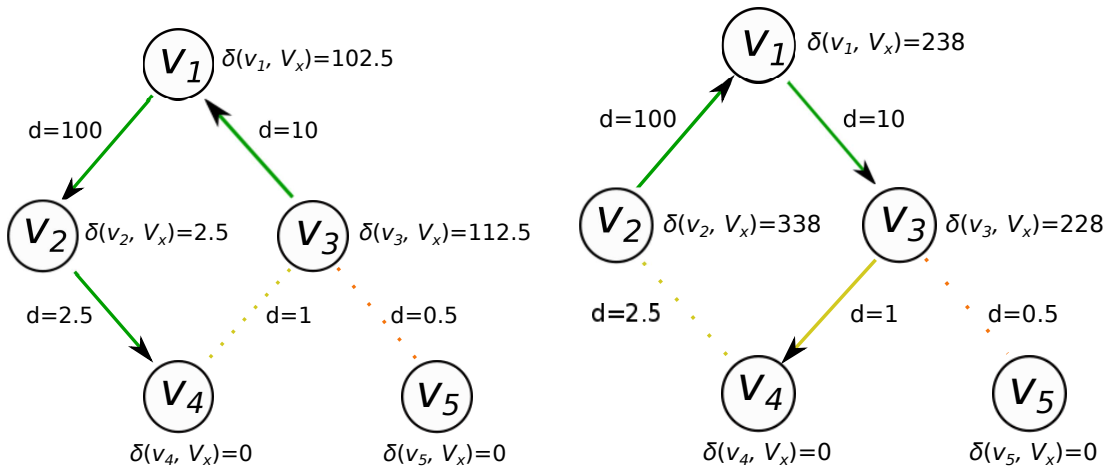
Fig. 5.5 presents the minimum cost exit forest determined by applying MDSP method to the graph of Fig. 5.3. Edge costs are determined using the DHM model. In Fig. 5.5(a) the exit forest is composed by two trees. One is a directed path (v_3, v_1, v_2, v_4). The other is composed by the single node v_5 (only a root node with no edges). In Fig. 5.5(b) the hazard state of edge $\{v_2, v_4\}$ has changed to YELLOW. The exit forest is now composed by a tree that is the directed path (v_1, v_2, v_3, v_4), and another tree with the single root node v_5.

5.2.4 Safest Exit Paths from Edges

After applying the MDSP method to G_B the exit paths/directions are defined only for spatial points mapped into nodes of V_B or into edges of E_F. Therefore, the exit paths/directions remain undefined for points mapped into edges of $E_B \setminus E_F$. We now show how to determine the exit paths/directions from a given point at any edge of E_B.

Consider an edge $\{u, v\} \in E_B$ with length $\mathrm{d}(\{u, v\})$ and cost $\mathrm{c}(\{u, v\})$. Moreover, consider a point p mapped over an edge $\{u, v\}$ at "mapped distance" $\beta \cdot \mathrm{d}(\{u, v\})$ from u, where $\beta \in (0, 1)$. Departing from p, the exit path in G_F to be followed is the one with the starting node

$$s = \begin{cases} u & \text{if } (u, v) \in E_F \\ & \quad \vee\, (v, u) \notin E_F \wedge \delta(u, V_\mathrm{x}) + \beta \cdot \mathrm{c}(\{u, v\}) < \delta(v, V_\mathrm{x}) + (1 - \beta) \cdot \mathrm{c}(\{u, v\}), \\ v & \text{otherwise.} \end{cases}$$



(a) h($\{v_2, v_4\}$) is GREEN (b) h($\{v_2, v_4\}$) changed to YELLOW

Figure 5.5: Minimum cost exit forest determined by the MDSP method. Exit node set (forest roots) is $V_x = \{v_4, v_5\}$. Edge costs follow the DHM model. The hazard state of the edges is represented by its colors (see Table 5.1).

Note that the values $\delta(u, V_{\mathrm{x}})$ and $\delta(v, V_{\mathrm{x}})$ are equal to the values of $\hat{\mathrm{c}}[u]$ and $\hat{\mathrm{c}}[v]$ respectively, at the completion of Algorithm 8.

5.3 – System Design

In this section we present an overview of the functional architecture of the wireless sensor nodes software. We focus on the presentation of general design guidelines, avoiding entering into platform specific details.

The software architecture is divided in two main functional planes, as shown in Fig. 5.6. The radio communication plane, associated to the R-graph, encompasses all the functions to support the communication between nodes (e.g. neighborhood inference, information dissemination). The spatial topology plane, associated to the S-graph, encompasses all the functions concerning the calculation of the emergency navigation graph, from hazard sensing to actuation decisions on signaling devices.

5.3.1 Sensing and Navigation Plane

This logical plane encompasses all the S-graph related functions of a node necessary to compute the emergency navigation graph. These functions are: (1) the classification

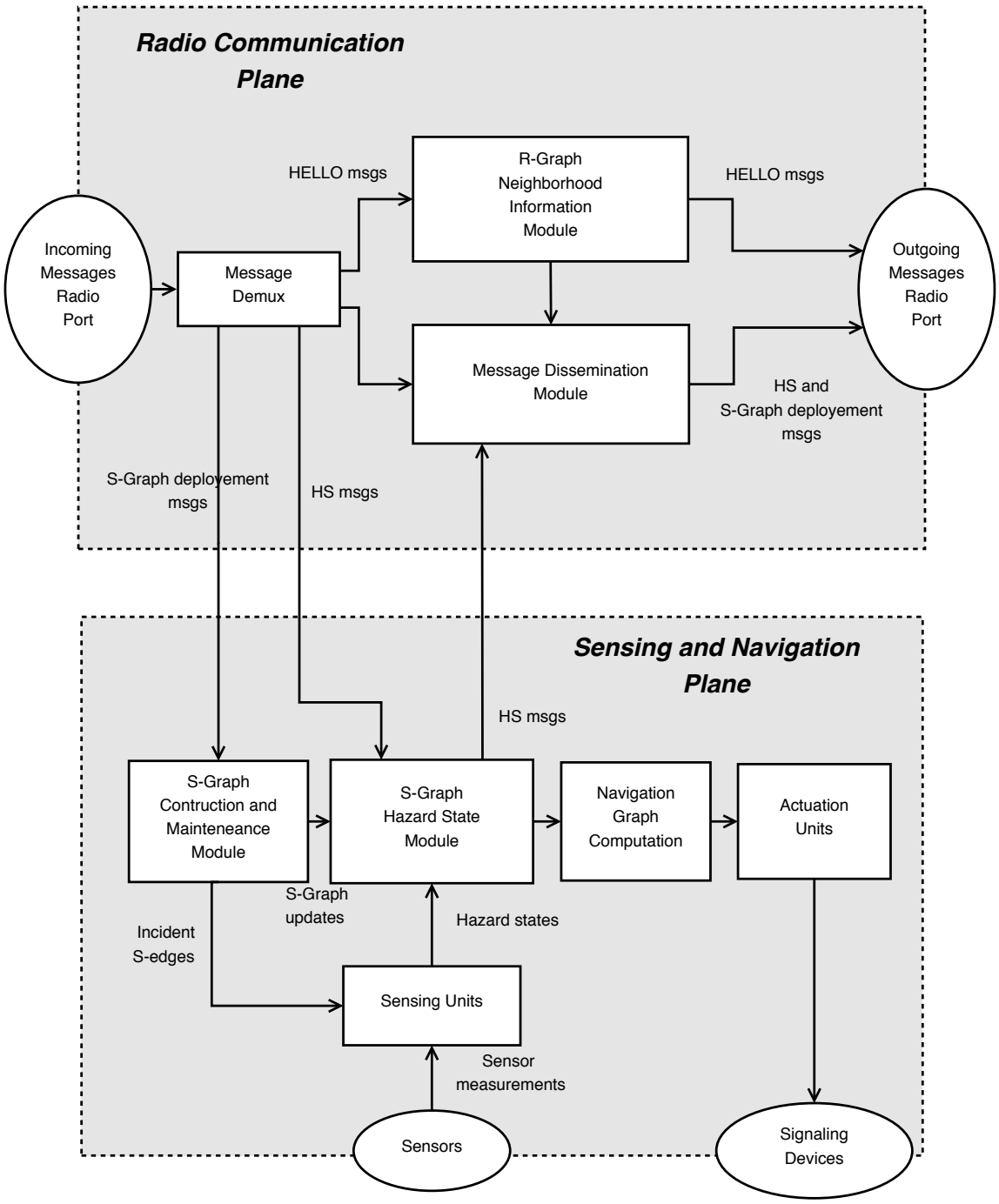


Figure 5.6: Functional architecture of a wireless sensor node for emergency evacuation support

of the sensed data collected by the attached sensors; (2) the reception, processing and generation of *S-graph deployement* messages and *S-graph Hazard State (SGHS)* messages; (3) the computation of the emergency navigation graph, and (4) the actuation on signaling devices, according to the emergency navigation graph, whenever a severe hazard situation arises.

Next we present a brief functional description of the modules of this logical plane.

S-graph Construction and Maintenance Module: This module is responsible for acquiring and maintaining the information about the spatial topology (represented by an S-graph G_S). The S-graph information is to be sent during the system deployment phase. This information is conveyed by *S-graph deployement* messages that carry information about the S-edges of the WSN. These messages are generated by a configuration node, being delivered to all sensor nodes using a dissemination algorithm.

Sensing Units: Each node has a sensing unit associated to each of its adjacent S-edges. Each of these units is responsible for collecting the sensor data associated to the corresponding S-edge. The sensor data is classified according to one of the security metrics chosen for the system (see Section 5.2.2) before being conveyed to the *S-graph Hazard State Module*.

S-graph Hazard State Module: The role of this module in each node is to collect and store information about the hazard state of each region of the building (mapped into the edges of the S-graph). This module receives hazard state information from (1) the S-edges incident to the node (via the *Sensing Units*), and (2) from the S-edges adjacent to the other S-nodes, via SGHS messages disseminated throughout the network. It is also responsible for generating the SGHS messages with the hazard state information of the S-edges of the node, which shall be disseminated throughout the network. These messages are generated either periodically or upon the occurrence of a substantial change in the hazard state of an S-edge.

Navigation Graph Computation: This module is responsible for computing the emergency navigation graph G_N from the S-graph hazard state information provided by the *S-graph Hazard State Module*. This computation is performed using the algorithms defined in Sections 5.2.3 and 5.2.4.

Actuation Units: Each node may have an actuation unit per incident S-edge. These units gather the exit direction for the associated S-edges, which are obtained from the emergency navigation graph G_N. The exit direction information is used to actuate accordingly on the attached signaling device.

5.3.2 Radio Communication Plane

This logical plane encompasses all the R-graph related functions of a node necessary to disseminate throughout the network the hazard state of each edge of the S-graph. The dissemination algorithm to be used in this emergency evacuation system should give guarantees that each disseminated hazard state message reaches all network nodes. This a major requirement in order to ensure that all network nodes compute an equivalent emergency navigation graph.

Next we present a brief functional description of the modules of this logical plane.

Message Demux: This module demultiplexes the messages received at the *Incoming Messages Radio Port* according to its type. For instance, *S-graph deployement messages* and SGHS messages are sent to the appropriate module in the Sensing and Navigation Plane. Moreover, they are also forwarded to the *Message Dissemination Module* for further dissemination of these messages through the network.

R-graph Neighborhood Information Module: This module is responsible for keeping track of the radio (R-graph) neighborhood of each node. It may feed this information to the *Message Dissemination Module*, depending on the needs of the chosen dissemination algorithm.

Message Dissemination Module: This module is responsible for applying the forwarding rules of a chosen message dissemination algorithm to the incoming messages. Messages to be forwarded are sent to the *Outgoing Messages Radio Port*.

5.3.3 System Hardware

In the actual implementation of the WSN system, we used Crossbow TelosB TPR2420CA motes as sensor nodes [Croa]. These motes have wireless communication capabilities, being equipped with a IEEE 802.15.4 compliant radio module, operating in the 2.4 to 2.4835 GHz ISM band, and supporting a data rate of 250 kbps [Crob]. Moreover, they are equipped with a sensor suite of humidity, light (visible and infrared), and temperature sensors. The motes run the TinyOS operating system [LMP+05, Tin12], and the software was written in nesC programming language [GLvB+03].

Moreover, we designed and implemented visual signaling devices intended to inform evacuees about the exit direction to follow, according to the order conveyed by the associated mote. Fig. 5.7 shows a TelosB mote connected to a visual signaling device.

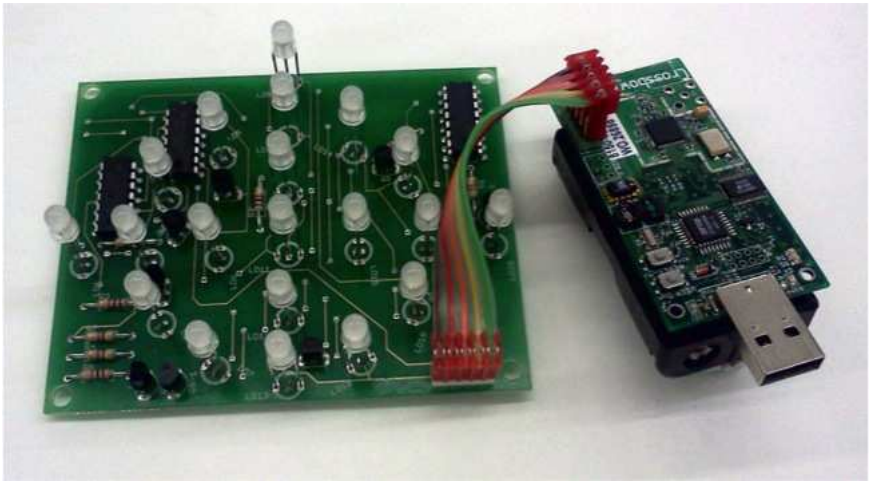


Figure 5.7: TelosB mote connected to a visual signaling device

5.3.4 Evaluation of the Sensor Data Dissemination

In this section we present a preliminary evaluation of sensor data dissemination using the WSN prototype for building evacuation support. The algorithms used in the dissemination of SGHS messages were Probabilistic Flooding (PF), with all nodes having the same forwarding probability ω, and Pure Flooding (which is a special case of PF in which $\omega = 1$).

The WSN is composed of 8 motes deployed over a part of a building[2]. The radio connectivity graph (determined by a set of measurements) has a mean node degree of 4.375 and a diameter of 2. Each mote was configured to periodically generate and transmit an SGHS message, such that the k-th SGHS message of a given node is transmitted at a time instant $t = k \cdot \tau \pm \Delta t$ with $\Delta t \in (-1/2\tau, 1/2\tau)$. The introduction of a jitter Δt in the period has as goal the mitigation of the correlation between transmission times of different source nodes. We conducted experiments for $\tau = 1$ s and $\tau = 5$ s.

We analyzed the following performance metrics :

- Delivery ratio (DR): ratio between number of messages successively received at a node and the number of sent messages;
- Relative frequency of global message outreaches (RFGO): relative frequency of global message outreaches — messages that were successively received by all nodes;
- Number of transmissions per generated SGHS message (Tx/msg);
- Number of redundant message receptions per source message per node (Rx/msg/node).

[2]The small number of nodes used in this experiment is due to restrictions on the available hardware.

In Fig. 5.8(b) and Fig 5.8(a) we see that the use of Pure Flooding ($\omega = 1$) does not lead to achieving a DR or a RFGO of 1. This is most likely due to the hidden terminal problem which is not addressed by the MAC layer reliability functions for broadcast transmissions. The carrier sensing mechanism of a transmitter is only able to infer an empty medium at the transmitter node. Two neighbors of a given node u, which are not in radio range of each other may initiate concurrent transmissions causing a collision at node u, since they are not able to ear the transmissions of each other. This effect is more severe under high network traffic loads ($\tau = 1$ s).

In Fig. 5.9(a) and Fig. 5.9(b) we can see that a reduction of the SGHS message period from $\tau = 5$ s to $\tau = 1$ s results in a non-negligible reduction of the number of replicas of a message that each node receives. At a first glance, this seems to be a desirable phenomenon (i.e. each node needs to receive only a copy of each source message). However, it may be a symptom of reception collisions at the MAC layer that may cause some nodes to fail the reception of some source messages.

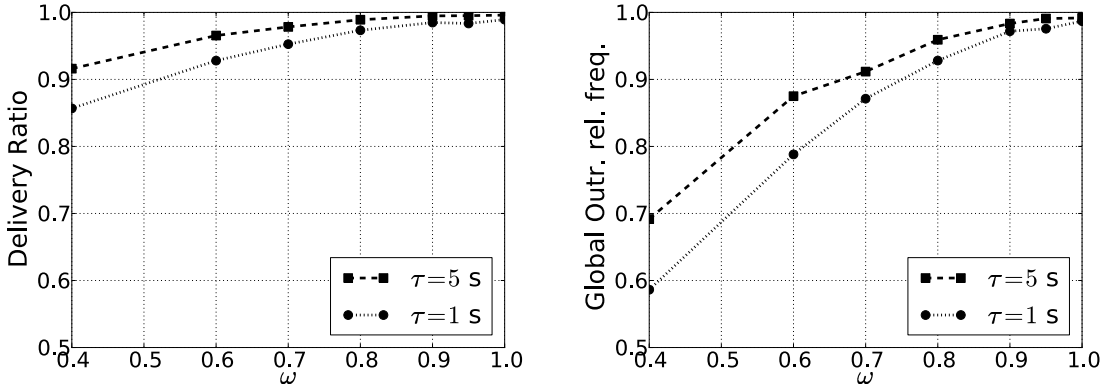
Since having a RFGO of 1 is a necessary condition for all nodes to derive the same emergency navigation graph, not surprisingly, Pure Flooding over an unreliable MAC does not give the necessary delivery guarantees. The same conclusions hold for PF.

Therefore information dissemination algorithms and MAC layer mechanisms encompassing reliability mechanisms are mandatory for rescue/evacuation support networked systems. Some possibilities for further study are:

- Reliability-focused development of new dissemination algorithms;
- MAC layer channel reservation for broadcast transmissions;
- Introduction of acknowledgment mechanisms for broadcasted messages;
- Use of implicit acknowledgment of messages by overhearing its retransmission by a neighbor;
- Retransmit important messages more than once (e.g. SGHS message related to an S-edge whose hazard state has changed).

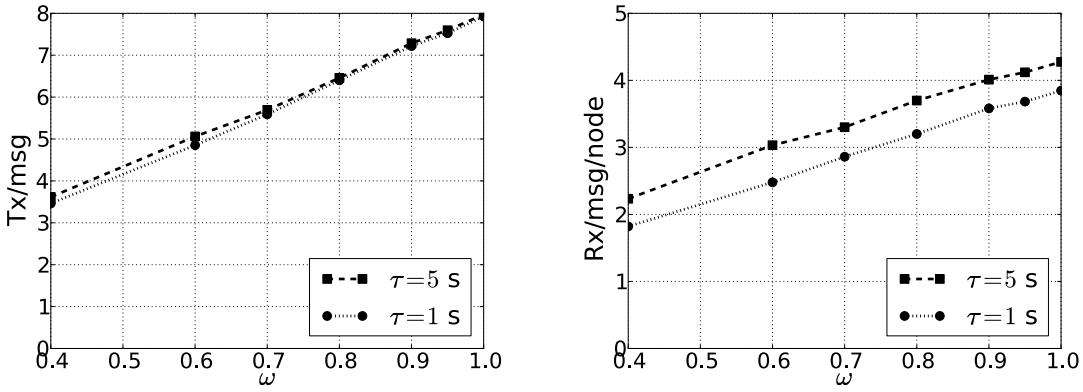
5.4 – Concluding Remarks

We applied the insights gained from the study of information dissemination algorithms to the design of a sensor-actuator networked system for emergency response in indoor scenarios.



(a) Delivery Ratio (b) Relative frequency of Global Outreaches

Figure 5.8: Experiment with real deployment



(a) Number of transmissions per message (b) Number of receptions per message and per node

Figure 5.9: Experiment with real deployment

We designed a sensor-actuator networked system for emergency response in indoor scenarios. The system guides people to the exits of a building via the shortest safe paths. These paths are computed independently by each node whenever a new measurement collected by a sensor is flooded throughout the network.

We proposed new path security metrics and new algorithms for the computation of the shortest safest exit paths from any point in a building. We implemented the proposed solutions in a wireless sensor network prototype using Pure Flooding for the dissemination of hazard state information collected by the sensor nodes. Pure Flooding over an unreliable MAC layer does not give the necessary delivery guarantees for this type of applications. Therefore, we plan to investigate new information dissemination algorithms and MAC layer solutions having reliable message delivery as the main requirement.

"It is not the answer that enlightens, but the question."

Eugène Ionesco

6

Main Contributions and Future Work

Motivated by the relevance of information dissemination algorithms in several networking scenarios, we characterized the efficiency of some of their main representatives in terms of transmission cost and reachability. We addressed both replication and network coded based approaches, devoting our main focus to the class of probabilistic algorithms.

Networks were modelled as random graphs generated by stochastic processes, facilitating the analyses of the interplay between the network topology and the process of disseminating information.

With the insights gained from the analysis, we applied information dissemination algorithms into specific networking and application scenarios. In particular, we designed a sensor-actuator networked system for emergency response in indoor scenarios. It computes the shortest safest paths to exits using the measurements collected by the sensors and disseminated throughout the network. The system was successfully tested in a prototype.

Probabilistic Flooding in Stochastic Networks

We analyzed how to set a system-wide forwarding probability ω of probabilistic flooding, such that all network nodes ultimately receive a message with high probability. For this purpose, we proposed a graph sampling method, which can be applied in arbitrary networks. This method yields an induced subgraph, whose node set is obtained by sampling the total node set uniformly at random with probability ω. We proved that the events "all nodes receive a flooded message" and "the induced subgraph is connected and its nodes dominate the network graph" have the same probability, and thus, the analysis of global outreach in probabilistic flooding can be performed by analyzing the properties of the induced subgraph.

In networks modeled as Erdős Rényi graphs, we derived the exact expression for the probability of global outreach. In random geometric graphs — as often used in modeling wireless ad-hoc networks — the local correlation among edges results in stochastic dependencies, but, in our model, these dependencies become asymptotically negligible with increasing node density. We derived an asymptotic expression for the global outreach probability, which is also a good approximation for high node density.

Moreover, we analyzed the impact of border effects in random geometric graphs and proposed a heuristic to overcome these effects. Finally, we studied probabilistic flooding in unreliable networks; erroneous links can simply be incorporated into both graph models, while the basic analysis and proofs remained in principle unchanged.

Network Coded Information Dissemination

Aiming at understanding how distinct forwarding paradigms influence the dissemination of information over communication networks, we selected one representative of each paradigm for our study: the network coding algorithm of [FWLB06] and replication based flooding MPR algorithm of [AQL02]. We evaluated (a) the number of transmissions per source message and (b) the incurred delay, and (c) the delivery ratio, under three relevant classes of random graph models.

Somewhat unintuitively, the analytical part of our work shows that over ERGs and RGGs, the number of transmissions required to flood a message with the NC flooding algorithm under consideration is asymptotically independent of the number of nodes. This observation becomes less surprising in retrospect, if we consider that in these classes of graphs the average node degree increases linearly with the number of nodes. Therefore, a higher number of nodes corresponds to a higher number of neighbors that can be reached by a single broadcast transmission. Since random linear network coding mixes multiple messages in a single transmission, it is very effective at exploiting the benefits of increased node density. With multipoint relays, however, the number of transmissions per message is not independent of

the number of nodes.

In contrast, for SWNs, the analysis shows that the number of transmissions per message of the network coding algorithm scales linearly with the number of nodes. The reason for these distinct results can be understood by the fact that in SWNs with rewiring the average node degree remains fixed irrespective of the number of nodes.

Naturally, the number of transmissions depends on other features of the network topology, as evidenced by Corollaries 1, 2, and 3, and by our simulation results. Consequently, the question as to which scheme should be preferred requires a nuanced answer.

In ERGs, NC flooding outperforms MPR flooding in terms of number of transmissions per source message; the extent of this gain is however deeply influenced by the diameter of the network. Reducing the diameter decreases both the number of transmissions and the delay gains. A unit diameter implies no gain at all.

In contrast, in general RGGs (non-toroidal distance metric) the considered NC flooding algorithm does not bring any benefits in terms of number of transmissions per message, when compared to MPR flooding. This appears to be in contradiction with the observation in [FWLB06]. However, it is worth noting that [FWLB06] focuses on RGGs on a torus and compares NC with probabilistic flooding. Our results thus indicate that the existence of border effects in general RGG topologies has a negative effect on the performance of the considered NC flooding technique.

In SWNs, the analytical expression for the number of transmissions per message of network coding shows no dependency on the rewiring probability of the SWN model. This result is corroborated by the simulation results. In fact, the simulations highlight the stability of the NC performance metrics (delivery ratio, number of transmissions per message and delay) within all the rewiring range of the SWN model (i.e. with distinct clustering coefficients and average geodesic distance values). NC flooding demonstrates to be relatively immune to changes in local connectivity parameters (i.e. presence or absence of strong local connectivity), suggesting that the strengths of the studied network coding algorithm in SWN topologies stems from its network-wide coding/decoding operation paradigm. In turn, the MPR flooding algorithm does not produce significant overhead reduction in terms of number of transmissions per source message, in poorly clustered SWN topologies. Its *Achilles' heel* resides on the scoped and limiting view of the topological properties centered in the neighborhood of each node.

Applications in Dynamic Sensor Networks

We designed a sensor-actuator networked system for emergency response in indoor scenarios. The system guides people to the exits of a building via the safest shortest paths. These

paths are computed by sensor nodes whenever a new measurement collected by a sensor is flooded throughout the network. We characterized the building evacuation problem with help of graph models. We proposed appropriate security metrics and algorithms that use flooded hazard information to compute the shortest safest paths to leave a building. Finally, we successfully implemented a prototype of the sensor-actuator networked system for emergency response. Since sensor data dissemination using Pure Flooding does not give the necessary delivery guarantees for this type of applications, we plan to investigate new information dissemination algorithms and MAC layer solutions that meet high reliability requirements.

Future Work

We now present an overview of possible lines of research based on the work presented in this thesis.

Further Analysis of Probabilistic Flooding

The analysis of Probabilistic Flooding, performed in this thesis, addressed a reference algorithm with a network-wide forwarding probability common to all nodes. A natural extension to this study is to analyze variants of Probabilistic flooding with non-constant forwarding probabilities (e.g. function of local topological properties such as node density or node degree and/or function of graph distance metrics).

Another line of research is to relax the reachability metric of the forwarding process. We considered scenarios that require a forwarding probability that ensures a given target for the global outreach of a flooded message. Less demanding applications may only require that each node independently gets the message with at least some target probability. This is an analytical open problem that deserves to be addressed.

A natural extension to the analysis of the reachability of Probabilistic flooding is to consider other network models. Some obvious candidates are Small-World Networks, Scale-Free Networks, and graphs with a specified degree distribution (i.e. the Configuration model).

Network Coded Information Dissemination

In Chapter 4 our study of Network Coded Information Dissemination addressed a class of algorithms that does not guarantee that the reception by a node of an encoded message that is linear independent of the ones received so far, will necessarily yield the decoding of (at least) a source message. As future work, we plan to analyze and propose new network coded dissemination algorithms in which every reception of a linear independent coded message

will result in the decoding of a source message with high probability. Moreover, we plan to propose probabilistic network coded dissemination algorithms that are aware of topological distance metrics, and that are adaptive and self-regulating depending on the dynamics of the dissemination process.

Adaptive Probabilistic Dissemination Algorithms

In this thesis we addressed mainly the interplay between Probabilistic Dissemination algorithms and network topology. As future work we plan to develop information dissemination algorithms that are topology aware and self-adaptive, adjusting the forwarding probability based on the dynamics of the dissemination process and based on target performance goals.

A

Proofs for Chapter 3

A.1 – Proof of Lemma 1

This proof is based on the Chen-Stein method and follows a similar approach as in [Pen97] and [FM08]. We first give definitions that are used in this and the following proofs. Consider the $\sqrt{A} \times \sqrt{A}$ square S_A used in the RGG definition. We partition the square S_A in m^2 disjoint sub-squares S_i of side $\sqrt{A}/m$ centered at $a_i \in S_A, i = 1..., m^2$. A *neighborhood of dependence* $\mathcal{N}_i$ for each $i \leq m^2$ is $\mathcal{N}_i \triangleq \{j : d(a_i, a_j) \leq 3\, r\}$, where r is the transmission range. We define D_i as disks of radius r centered at $a_i, i = 1, ..., m^2$. Finally, we define $D(r, x)$ as the area of the union of two disks of radius r with centers at toroidal distance x apart.

From the Chen-Stein method we have

$$d_{\mathrm{TV}}\left(W^*, \mathrm{Po}\left(\mathrm{E}(W^*)\right)\right) \leq 2\left(b_1 + b_2\right). \tag{A.1}$$

Now we show that (a) W^* is a sum of Bernoulli random variables, (b) $\mathrm{E}(W^*) = e^{-\alpha^*}$, (c) $\lim_{\lambda\to\infty} b_1 = 0$, and (d) $\lim_{\lambda\to\infty} b_2 = 0$. For this purpose we partition the square S_A as

described in 2.4.

For $i = 1, \ldots, m^2$, define X_i^* to be the indicator of the event that there is a single point of Π_s^* in a sub-square $S_i \subset S_A$ and no points of Π_s^* in the region of all sub-squares intersecting $D_i \setminus S_i$. We have

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_i^*)}{\frac{A\,\lambda^*}{m^2}\, e^{-\lambda^*\pi\, r^2}} = 1. \tag{A.2}$$

If the two disks of radius r centered at a_i and a_j cover each other's centers, i.e. $d(a_i, a_j) \leq r$, we get

$$\mathrm{E}(X_i^* X_j^*) = 0, \tag{A.3}$$

and if $d(a_i, a_j) > r$, we have

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_i^* X_j^*)}{\left(\frac{A\,\lambda^*}{m^2}\right)^2 e^{-\lambda^* D(r,\, d(a_i,a_j))}} = 1, \tag{A.4}$$

Therefore, the total number of isolated nodes of G^* is $W^* = \lim_{m\to\infty} \sum_{i=1}^{m^2} X_i^*$, and

$$\mathrm{E}(W^*) = \lim_{m\to\infty} \sum_{i=1}^{m^2} \mathrm{E}\left(X_i^*\right) = A\,\lambda^*\, e^{-\lambda^*\pi\, r^2} = e^{-\alpha^*}, \tag{A.5}$$

where $\alpha^* = \lambda^*\pi\, r^2 - \ln(A\,\lambda^*)$.

We now show that $\lim_{\lambda\to\infty} b_1^* = 0$ and $\lim_{\lambda\to\infty} b_2 = 0$. Combining (A.2) and (2.4) we get

$$\begin{aligned}
\lim_{m\to\infty} b_1 &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{N}_i} \mathrm{E}(X_i^*)\, \mathrm{E}(X_j^*) \\
&= \lim_{m\to\infty} \sum_{i=1}^{m^2} \frac{\pi\,(3r)^2}{\frac{A}{m^2}} \left(\frac{A\,\lambda^*}{m^2}\, e^{-\lambda^*\pi r^2}\right)^2 \\
&= \frac{\pi\,(3r)^2}{A}\, e^{-2\,\alpha^*} \xrightarrow[\lambda\to\infty]{} 0.
\end{aligned} \tag{A.6}$$

Defining an annular neighborhood $\mathcal{O}_i$ for each $i \leq m^2$ as

$$\mathcal{O}_i \triangleq \{j : r \leq d(a_i, a_j) \leq 3\, r\}, \tag{A.7}$$

and combining (2.5) with (A.3), (A.4), (A.7), we get

$$\begin{aligned}
\lim_{m\to\infty} b_2 &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{N}_i, j\neq i} \mathrm{E}(X_i^* X_j^*) \\
&= \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{O}_i, j\neq i} \left(\frac{A\,\lambda^*}{m^2}\right)^2 e^{-\lambda^* D(r,\, d(a_i,a_j))}.
\end{aligned} \tag{A.8}$$

Due to the spatial stationary property of the Poisson point process, we can re-write (A.8) as

$$\begin{aligned}\lim_{m\to\infty} b_2 &= \lim_{m\to\infty} m^2 \sum_{j\in\mathcal{O}_1, j\neq 1} \left(\frac{A\,\lambda^*}{m^2}\right)^2 e^{-\lambda^* D(r,\, d(a_1,a_j))} \\ &= A\,(\lambda^*)^2 \int_{r\leq|x|\leq 3r} e^{-\lambda^* D(r,\,|x|)} dx \\ &\leq A\,(\lambda^*)^2 \pi\,(3\,r)^2\, e^{-\lambda^* \frac{3}{2}\pi r^2} \xrightarrow[\lambda\to\infty]{} 0. \end{aligned} \tag{A.9}$$

Combining (A.1) with (A.6) and (A.9) yields

$$\lim_{\lambda\to\infty} d_{\mathrm{TV}}\left(W^*, \mathrm{Po}\left(e^{-\alpha^*}\right)\right) = 0\,. \tag{A.10}$$

□

A.2 – Proof of Lemma 2

Similarly, applying the Chen-Stein method we have

$$d_{\mathrm{TV}}\left(W^\diamond, \mathrm{Po}\left(\mathrm{E}(W^\diamond)\right)\right) \leq 2\,(b_1 + b_2)\,. \tag{A.11}$$

For $i = 1, \ldots, m^2$, define $X_i^\diamond$ to be the indicator of the event that there is a single point of $\Pi^\diamond$ in a sub-square $S_i \subset S_A$, and that there are no points of Π_s^* in the region of all sub-squares intersecting D_i. We have

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_i^\diamond)}{\frac{A\,\lambda^\diamond}{m^2}\, e^{-\lambda^*\pi r^2}} = 1, \tag{A.12}$$

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_i^\diamond X_j^\diamond)}{\left(\frac{A\,\lambda^\diamond}{m^2}\right)^2 e^{-\lambda^* D(r,\, d(a_i,a_j))}} = 1. \tag{A.13}$$

Thus, the total number of non-dominated nodes of G is $W^\diamond = \lim_{m\to\infty} \sum_{i=1}^{m^2} X_i^\diamond$, and

$$\mathrm{E}\,(W^\diamond) = \lim_{m\to\infty} \sum_{i=1}^{m^2} \mathrm{E}\,(X_i^\diamond) = A\,\lambda^\diamond\, e^{-\lambda^*\pi\, r^2} = e^{-\alpha^\diamond},$$

where $\alpha^\diamond = \lambda^*\pi\, r^2 - \ln\,(A\,\lambda^\diamond)$.

We now show that $\lim_{\lambda\to\infty} b_1 = 0$ and $\lim_{\lambda\to\infty} b_2 = 0$. Combining (A.12) with (2.4), we

get

$$\begin{aligned}\lim_{m\to\infty} b_1 &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{N}_i} \mathrm{E}(X_i^\diamond)\,\mathrm{E}(X_j^\diamond) \\ &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \frac{\pi\,(3r)^2}{\frac{A}{m^2}} \left(\frac{A\,\lambda^\diamond}{m^2}\, e^{-\lambda^*\pi r^2}\right)^2 \\ &= \frac{\pi\,(3r)^2}{A}\, e^{-2\,\alpha^\diamond} \xrightarrow[\lambda\to\infty]{} 0. \end{aligned} \tag{A.14}$$

Combining (2.5) with (A.13), and taking into account the spatial stationary property of the Poisson process yields

$$\begin{aligned}\lim_{m\to\infty} b_2 &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{N}_i, j\neq i} \left(\frac{A\,\lambda^\diamond}{m^2}\right)^2 e^{-\lambda^* D(r,\, d(a_i,a_j))} \\ &= A\,(\lambda^\diamond)^2 \int_{|x|\leq 3r} e^{-\lambda^* D(r,\,|x|)} dx \\ &\leq A\,(\lambda^\diamond)^2 \pi\,(3\,r)^2\, e^{-\lambda^*\pi r^2} \xrightarrow[\lambda\to\infty]{} 0. \end{aligned} \tag{A.15}$$

Combining (A.11) with (A.14) and (A.15), we get

$$\lim_{\lambda\to\infty} d_{\mathrm{TV}}\left(W^\diamond, \mathrm{Po}\left(e^{-\alpha^\diamond}\right)\right) = 0\,. \tag{A.16}$$

□

A.3 – Proof of Lemma 3

The Chen-Stein method yields

$$d_{\mathrm{TV}}\left(W^{\diamond *}, \mathrm{Po}\left(\mathrm{E}(W^{\diamond *})\right)\right) \leq 2\,(b_1 + b_2)\,. \tag{A.17}$$

We partition the square S_A as described in 2.4. For $i = 1, \ldots, m^2$, define $X_i^{\diamond *}$ to be the indicator of the event that there is a single point of Π_s^* or $\Pi^\diamond$ in $S_i \subset S_A$, and that there are no points of Π_s^* in the region of all sub-squares intersecting $D_i \setminus S_i$. We have

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_i^{\diamond *})}{\frac{A\,\lambda+1}{m^2}\, e^{-\lambda^*\pi r^2}} = 1. \tag{A.18}$$

Let $\beta_{ij}^* \triangleq \lambda^* D(r,\, d(a_i, a_j))$. For $d(a_i, a_j) \leq r$, we get

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_i^{\diamond *} X_j^{\diamond *})}{\left(\frac{A\,\lambda^\diamond}{m^2}\right)^2 \left(1 - \frac{A\,\lambda^*}{m^2}\right)^2 e^{-\beta_{ij}^*}} = 1, \tag{A.19}$$

and if $d(a_i, a_j) > r$, we have

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_i^{\diamond *} X_j^{\diamond *})}{\left(\frac{A\,\lambda+1}{m^2}\right)^2 e^{-\beta_{ij}^*}} = 1. \tag{A.20}$$

The sum of all non-dominated nodes and all isolated forwarding nodes of G is $W^{\diamond *} = \lim_{m\to\infty} \sum_{i=1}^{m^2} X_i^{\diamond *}$ with

$$\mathrm{E}\left(W^{\diamond *}\right) = \lim_{m\to\infty} \sum_{i=1}^{m^2} \mathrm{E}\left(X_i^{\diamond *}\right) = (A\,\lambda + 1) e^{-\lambda^* \pi r^2} = e^{-\alpha}, \tag{A.21}$$

where $\alpha = \lambda^* \pi\, r^2 - \ln\left(A\,\lambda + 1\right)$.

We now show that $\lim_{\lambda\to\infty} b_1 = 0$ and $\lim_{\lambda\to\infty} b_2 = 0$. Combining (A.18) with (2.4) we get

$$\begin{aligned}
\lim_{m\to\infty} b_1 &= \sum_{i=1}^{m^2} \sum_{j\in\mathcal{N}_i} \mathrm{E}(X_i^{\diamond *})\, \mathrm{E}(X_j^{\diamond *}) \\
&= \lim_{m\to\infty} \sum_{i=1}^{m^2} \frac{\pi\,(3r)^2}{\frac{A}{m^2}} \left(\frac{A\,\lambda+1}{m^2}\, e^{-\lambda^*\pi r^2}\right)^2 \\
&= \frac{\pi\,(3r)^2}{A}\, e^{-2\,\alpha} \xrightarrow[\lambda\to\infty]{} 0.
\end{aligned} \tag{A.22}$$

For each $i \le m^2$ define neighborhoods $\mathcal{O}_i^{(1)}$ and $\mathcal{O}_i^{(2)}$ as

$$\mathcal{O}_i^{(1)} \triangleq \{j : d(a_i, a_j) \le r\}, \text{ and} \tag{A.23}$$

$$\mathcal{O}_i^{(2)} \triangleq \{j : r \le d(a_i, a_j) \le 3\,r\}. \tag{A.24}$$

Combining (2.5), (A.19), (A.20), (A.23), (A.24), and considering the stationarity of the Poisson processes, we get

$$\begin{aligned}
\lim_{m\to\infty} b_2 &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{O}_i^{(1)}, j\neq i} \left[\frac{A\,\lambda^\diamond}{m^2}\right]^2 \left(1 - \frac{A\,\lambda^*}{m^2}\right)^2 e^{-\beta_{ij}^*} \\
&+ \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{O}_i^{(2)}, j\neq i} \left(\frac{A\,\lambda+1}{m^2}\right)^2 e^{e^{-\beta_{ij}^*}} \\
&= A\,(\lambda^\diamond)^2 \int_{|x|\le r} e^{-\lambda^* D(r,|x|)} dx \\
&+ A\left(\lambda + \frac{1}{A}\right)^2 \int_{r\le|x|\le 3r} e^{-\lambda^* D(r,|x|)} dx \\
&\le A\,(\lambda^\diamond)^2 \pi\, r^2\, e^{-\lambda^*\pi r^2} \\
&+ A\left(\lambda + \frac{1}{A}\right)^2 \pi (3r)^2 e^{-\lambda^* \frac{3}{2}\pi r^2} \xrightarrow[\lambda\to\infty]{} 0.
\end{aligned} \tag{A.25}$$

Combining (A.17) with (A.22) and (A.25) leads to

$$\lim_{\lambda\to\infty} d_{\mathrm{TV}}\left(W^{\diamond *}, \mathrm{Po}\left(e^{-\alpha}\right)\right) = 0\,. \tag{A.26}$$

□

B

Symbols, Mathematical Notation, and Abbreviations

B.1 – List of Symbols and Mathematical Notation

$\lvert \cdot \rvert$	Cardinality of a set
$\lfloor x \rfloor$	Floor function: gives the largest integer that is smaller than or equal to x
$\lceil x \rceil$	Ceiling function: gives the smallest integer that is larger than or equal to x
$[a, b]$	Interval reaching from a to b, including both a and b
$[a, b)$	Interval reaching from a to b, including a and excluding b
A	Area
$A(\cdot)$	Area of a region
$\mathrm{Bin}(n, p)$	Binomial distribution with parameters n and p
$\mathrm{c}(\cdot)$	Cost of an edge or a path
C_v	Clustering coefficient of a node v

$C(p)$	Average clustering coefficient in a Small-World network with rewiring probability p
$d(\cdot)$	Degree of a node
$\mathrm{d}(e)$	Length of an edge e
$d(\cdot,\cdot)$	Euclidean distance between two points or two nodes
$d_E(\cdot,\cdot)$	Euclidean distance between two points or two nodes
$d_T(\cdot,\cdot)$	Toroidal distance between two points or two nodes
$d_{\mathrm{TV}}(X,Y)$	Total variation distance between distributions of two integer-valued random variables X, Y
D_i	Disk centered at point i
$D(r,x)$	Area of the union of two disks of radius r with centers at toroidal distance x apart
$\mathrm{E}(\cdot)$	Expected value
$\mathbb{F}_q$	Finite field of size q
G	Graph
G_B	B-graph (Building graph)
G_N	N-graph (Emergency navigation graph)
G_S	S-graph (Spatial graph)
G_R	R-graph (Radio connectivity graph)
$g()$	Antenna gain
H	Hazard state set
$\mathrm{h}(e)$	Hazard state of an edge e
E	Set of edges
k	Number of neigboors of a node in a regular lattice
L	Average geodesic distance in a graph
$L(p)$	Average path length in a Small-World network with rewiring probability p
$L_{\cdot,\cdot}$	Geodesic distance between a pair of nodes
n	Number of nodes
$N(u)$	Set of the 1-*hop* neighbors of u
$N^l(u)$	Set of the l-*hop* neighbors of u
$\mathcal{O}_i$	Annular neighborhood centered at point i
$O(f(x))$	Big O notation: there exists some constant c such that $O(f(x)) \leq c \cdot f(x)$, for large enough x
p	Edge probability
$\mathrm{P}(\cdot)$	Probability
$\mathrm{Po}(\lambda)$	Poisson distribution with parameter λ
r	Transmission radius Neighborhood radius
$\mathbb{R}$	Set of the real numbers

$\mathcal{R}$	Region
S_A	Square of area A
u	Node
v	Node
V	Set of nodes
V_{x}	Set of exit nodes
w	Node
$\mathbb{Z}^+$	Set of all positive integer numbers
γ	Scaling factor of Algorithm NC_{FWB}
λ	Node density
ω	Forwarding probability
Ψ	Probability of global outreach
Π	Poisson point process
Π^*	Thined poisson point process
$\Pi^\diamond$	Thined poisson point process
ζ	Message erasure probability

B.2 – Abbreviations

B-graph	Building graph
B-RGG	Binomial Random Geometric Graph
CHM	Continuous Hazard Metric
DHM	Discrete Hazard Metric
DR	Delivery ratio
ERG	Erdős Rényi Random Graph
GS	Graph sampling
MAC	Media Access Control
MDSP	Multiple-Destination Shortest Paths
MPR	Multipoint relays
N-graph	Emergency navigation graph

NC	Network Coding
NR	Normalized Rank
OLSR	Optimized Link State Routing
P-RGG	Poisson Random Geometric Graph
PF	Probabilistic Flooding
PP	Point Process
PPP	Poisson Point Process
R-graph	Radio connectivity graph
RFGO	Relative frequency of global message outreaches
RLNC	Random Linear Network Coding
RGG	Random Geometric Graph
RGGT	Random Geometric Graph on a Torus
S-graph	Spatial graph
SGHS	S-graph Hazard State
SSSP	Single-Source Shortest Paths
SDSP	Single-Destination Shortest Paths
SWN	Small-World Network
WSN	Wireless Sensor Network Network

B

Bibliography

[ACLY00] R. Ahlswede, N. Cai, S.Y.R. Li, and RW Yeung. Network information flow. *IEEE Trans. on Information Theory*, 46(4):1204–1216, July 2000. [pp. 4, 19, and 56]

[AGG89] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: The Chen-Stein method. *Ann. of Prob.*, 17(1):9–25, January 1989. [pp. 12 and 51]

[AGG90] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the Chen-Stein method. *Statistical Science*, 5(4):403–424, November 1990. [pp. 12 and 51]

[AQL02] L. Viennot A. Qayyum and A. Laouiti. Multipoint relaying for flooding broadcast messages in mobile wireless networks. *Proc. HICSS, Big Island, HI, USA*, January 2002. [pp. 4, 56, 57, 58, 76, and 104]

[AWF02] K. Alzoubi, P.-J. Wan, and O. Frieder. New distributed algorithm for connected dominating set in wireless ad hoc networks. *Proc. HICSS, Big Island, HI, USA*, January 2002. [p. 4]

[BBB+08] Z. Benenson, M. Bestehorn, E. Buchmann, F. Freiling, and M. Jawurek. Query dissemination with predictable reachability and energy usage in sensor networks. *Ad-hoc, Mobile and Wireless Networks*, 5198:279–292, 2008. [p. 4]

[Bet02] C. Bettstetter. On the minimum node degree and connectivity of a wireless multihop network. In *Proc. ACM MobiHoc*, Lausanne, Switzerland, June 2002. [pp. 68 and 71]

[Bet04] C. Bettstetter. On the connectivity of ad hoc networks. *The Computer Journal*, 47(4):432–447, July 2004. [p. 51]

[Bol98] B. Bollobás. *Modern Graph Theory*. Springer, 1998. [p. 15]

[BW00] A. Barrat and M. Weigt. On the properties of small-world network models. *The European Physical Journal B - Condensed Matter*, 13(3):547–560, January 2000. [p. 64]

[CBB08a] S. Crisóstomo, J. Barros, and C. Bettstetter. Flooding the Network: Multipoint Relays versus Network Coding. In *Proc. IEEE ICCSC 2008*, Shanghai, China, May 2008. [pp. 6 and 57]

[CBB08b] S. Crisóstomo, J. Barros, and C. Bettstetter. Network coding with shortcuts. In *Proc. IEEE Intern. Conf. on Communication Systems (ICCS)*, Guangzhou, China, November 2008. [pp. 6 and 57]

[Che75] L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. of Prob.*, 3(3):534–545, June 1975. [pp. 12 and 51]

[CJA+03] T. Clausen, P. Jacquet, C. Adjih, A. Laouiti, P. Minet, P. Muhlethaler, A. Qayyum, and L.Viennot. Optimized link state routing protocol (OLSR). RFC 3626, October 2003. Network Working Group. [pp. 4 and 56]

[Cre91] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, 1991. [p. 16]

[Croa] Crossbow Technology, Inc. Crossbow Technology Webpage. http://www.xbow.com. Accessed: July 30, 2012. [p. 97]

[Crob] Crossbow Technology, Inc. TelosB mote platform. www.willow.co.uk/TelosB_Datasheet.pdf. Accessed: July 30, 2012. [p. 97]

[CSBB09] S. Crisóstomo, U. Schilcher, C. Bettstetter, and J. Barros. Analysis of probabilistic flooding: How do we choose the right coin? In *Proc. IEEE Intern. Conf. Commun., Dresden, Germany*, Dresden, Germany, June 2009. [pp. 6 and 27]

[CSBB12] S. Crisóstomo, U. Schilcher, C. Bettstetter, and J. Barros. Probabilistic flooding in stochastic networks: Analysis of global information outreach. *Computer Networks*, 56(1):142 – 156, 2012. [pp. 6 and 27]

[CSRL01] T. Cormen, C. Stein, R. Rivest, and C. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001. [pp. 91 and 92]

[CT06] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006. [p. 47]

[CWJ03] P. Chou, Y. Wu, and K. Jain. Practical network coding. In *Proc. Allerton Conf. Communication, Control and Computing*, October 2003. [pp. 19, 57, and 66]

[DEH+05] S. Deb, M. Effros, T. Ho, D. Karger, R. Koetter, D. Lun, M. Médard, and N. Ratnakar. Network coding for wireless applications: A brief tutorial. In *Proc. of the International Workshop on Wireless Ad-hoc Networks (IWWAN)*, London, UK, May 2005. [p. 21]

[DGWR07] A. G. Dimakis, P. Brighten Godfrey, M. J. Wainwright, and Kannan Ramchandran. Network coding for distributed storage systems. In *Proc. of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007)*, Anchorage, Alaska, USA, May 2007. [p. 20]

[DYT05] S. Dixit, E. Yanmaz, and O. K. Tonguz. On the design of self-organized cellular wireless networks. *IEEE Communications Magazine*, 43(7):86–93, 2005. [p. 17]

[EFSC+06] A. El Fawal, K. Salamatian, D. Cavin, Y. Sasson, and J. Y. Le Boudec. A framework for network coding in challenged wireless network. In *MobiSys 2006*, Uppsala, Sweden, June 2006. [p. 21]

[FB08] A. Faragó and S. Basagni. The effect of multi-radio nodes on network connectivity - A graph theoretic analysis. *Proc. IEEE PIMRC, Cannes, France*, September 2008. [p. 16]

[FKG71] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys*, 22:89–103, 1971. [pp. 11, 36, and 51]

[FLBW06] C. Fragouli, J.-Y. Le Boudec, and J. Widmer. Network coding: an instant primer. *SIGCOMM Comput. Commun. Rev.*, January 2006. [p. 56]

[FM08] M. Franceschetti and R. Meester. *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2008. [pp. 51 and 109]

[FWLB06] C. Fragouli, J. Widmer, and J. Y. Le Boudec. A network coding approach to energy efficient broadcasting: From theory to practice. In *Proc. INFOCOM*, Barcelona, Spain, April 2006. [pp. 5, 21, 56, 58, 59, 63, 66, 68, 76, 104, and 105]

[GHH06] R. Gowaikar, B. Hochwald, and B. Hassibi. Communication over a wireless network with random connections. *IEEE Trans. Inform. Theory*, 52:2857–2871, 2006. [p. 16]

[Gil59] E. N. Gilbert. Random graphs. *Ann. of Math. Stat.*, 30(4):1141–1144, December 1959. [p. 30]

[GLvB+03] D. Gay, P. Levis, R. von Behren, M. Welsh, E. Brewer, and D. Culler. The nesC language: A holistic approach to networked embedded systems. *SIGPLAN Not.*, 38(5):1–11, May 2003. [p. 97]

[Gol05] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005. [pp. 13 and 14]

[GR05] C. Gkantsidis and P. Rodriguez. Network coding for large scale content distribution. In *Proc. of IEEE Infocom*, Miami, USA, 2005. [p. 20]

[Gri99] G. Grimmett. *Percolation*. Springer-Verlag, 1999. [p. 11]

[GS11] R. Gaeta and M. Sereno. Generalized probabilistic flooding in unstructured peer-to-peer networks. *IEEE Trans. Parallel Distrib. Syst.*, 99, 2011. [p. 4]

[Hel03] A. Helmy. Small worlds in wireless networks. *IEEE Communications Letters*, 7(10):490–492, October 2003. [p. 17]

[HHL06] Z. Haas, J. Halpern, and L. Li. Gossip-based ad hoc routing. *IEEE/ACM Trans. Netw.*, 14(3):479–491, 2006. [pp. 4, 27, and 51]

[HLY04] K. Hui, J. Lui, and D. Yau. Small world overlay P2P networks. In *Proc. Int. Workshop Quality of Service*, Montreal, Canada, June 2004. [p. 17]

[HMK+06] T. Ho, M. Médard, R. Koetter, D.R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, 2006. [pp. 19, 20, 57, and 58]

[JLMV01] P. Jacquet, A. Laouiti, P. Minet, and L. Viennot. Performance analysis of OLSR multipoint relay flooding in two ad hoc wireless network models. Technical Report 4260, INRIA, 2001. [pp. 57 and 58]

[Kin93] J. Kingman. *Poisson Processes*. Oxford University Press, 1993. [p. 11]

[Kle00] J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proc. Thirty-second annual ACM symposium on Theory of computing*, New York, NY, USA, 2000. [p. 18]

[KM03] R. Koetter and M. Médard. An algebraic approach to network coding. *IEEE/ACM Transactions on Networking*, 11:782–795, 2003. [p. 19]

[KWB01] B. Krishnamachari, S. B. Wicker, and R. Bejar. Phase transition phenomena in wireless ad hoc networks. *Proc. IEEE GLOBECOM, San Antonio, TX, USA*, November 2001. [pp. 4, 27, and 51]

[LMHK04] D. Lun, M. Medard, T. Ho, and R. Koetter. Network coding with a cost criterion. In *Proc. Intern. Symp. on Information Theory and its Applications*, Parma, Italy, October 2004. [pp. 5 and 56]

[LMKE05a] D. Lun, M. Médard, R. Koetter, and M. Effros. Further results on coding for reliable communication over packet networks. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1848–1852, Adelaide, Australia, September 2005. [p. 19]

[LMKE05b] D. Lun, M. Médard, R. Koetter, and M. Effros. On Coding for Reliable Communication over Packet Networks. *Proc. of the 42nd Annual Allerton Conference on Communication, Control, and Computing*, September 2005. [pp. 19 and 20]

[LMP+05] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler. TinyOS: An Operating System for Sensor Networks Ambient Intelligence. In *Ambient Intelligence*, chapter 7, pages 115–148. Springer Berlin Heidelberg, 2005. [p. 97]

[LRK+05] D. Lun, N. Ratnakar, R. Koetter, M. Medard, E. Ahmed, and H. Lee. Achieving minimum-cost multicast: a decentralized approach based on network coding. In *Proc. IEEE Infocom*, Miami, FL, USA, March 2005. [pp. 5 and 56]

[LRM+06] D. Lun, N. Ratnakar, M. Médard, R. Koetter, D. Karger, T. Ho, E. Ahmed, and F. Zhao. Minimum-cost multicast over coded packet networks. *IEEE/ACM Trans. Netw.*, 14:2608–2623, 2006. [pp. 5 and 56]

[LW02] W. Lou and J. Wu. On reducing broadcast redundancy in ad hoc wireless networks. *IEEE Trans. Mobile Comput.*, 1(2):111–123, 2002. [p. 4]

[MAA08] D. Miorandi, E. Altman, and G. Alfano. The impact of channel randomness on coverage and connectivity of ad hoc and sensor networks. *IEEE Trans. Wireless Commun.*, 7:1062–1072, 2008. [p. 16]

[Mad08] U. Madhow. *Fundamentals of Digital Communication*. Cambridge University Press, 2008. [p. 13]

[Man69] N. Mantel. Functional averages of a variable. *The American Statistician*, 23(1):21–22, February 1969. [p. 64]

[MNW04] G. S. Manku, M. Naor, and U. Wieder. Know thy neighbor's neighbor: the power of lookahead in randomized P2P networks. In *In Proc. of the 36th ACM Symp. on Theory of Computing (STOC)*, pages 54–63, 2004. [p. 17]

[New03] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167, 2003. [pp. 17 and 71]

[NTCS99] S.-Y. Ni, Y.-C. Tseng, Y.-S. C., and J.-P. Sheu. The broadcast storm problem in a mobile ad hoc network. *Proc. ACM/IEEE MobiCom, Seattle, WA, USA*, August 1999. [p. 27]

[NW99] M. Newman and D. Watts. Scaling and percolation in the small-world network model. *Physical Review E*, 60(6):7332–7342, December 1999. [p. 18]

[OKS10] K. Oikonomou, D. Kogias, and I. Stavrakakis. Probabilistic flooding for efficient information dissemination in random graph topologies. *Computer Networks*, 54(10):1615 – 1629, 2010. [p. 4]

[Pen97] M. Penrose. The longest edge of the random minimal spanning tree. *Ann. Appl. Prob.*, 47(4):432–447, July 1997. [pp. 37, 51, and 109]

[Pen03] M. Penrose. *Random Geometric Graphs*. Oxford Univ. Press, July 2003. [pp. 12, 16, and 35]

[Rem04] G. Rempala. Asymptotic factorial powers expansions for Binomial and Negative Binomial reciprocals. *American Mathematical Society*, 132(1):261–272, 2004. [p. 62]

[RKV04] A. Reznik, S. Kulkarni, and S. Verdú. A small world approach to heterogeneous networks. *Communication in Information and Systems*, 3(4):325–348, 2004. [p. 17]

[SB07] A. Stauffer and C. Barbosa. Probabilistic heuristics for disseminating information in networks. *IEEE/ACM Trans. Netw.*, 15(2):425–435, 2007. [p. 4]

[Sch05] M. Schwartz. *Mobile Wireless Communications*. Cambridge University Press, 2005. [pp. 13 and 15]

[SCS03] Y. Sasson, D. Cavin, and A. Schiper. Probabilistic broadcast for flooding in wireless mobile ad hoc networks. *Proc. IEEE Wireless Comm. Netw. Conf., New Orleans, LA, USA*, March 2003. [pp. 4, 27, and 51]

[SKM85] D. Stoyan, W. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Wiley & Sons, 1985. [p. 10]

[SRS07] A. Sangwan, V. Ramaiyan, and R. Shorey. Reliable multihop broadcast protocols for inter-vehicular communication in a fading channel. *Proc. Intern. Conf. Commun. Sys. Softw. Middlew.*, January 2007. [p. 4]

[Tin12] TinyOS team. TinyOS Webpage. http://www.tinyos.net/, 2012. Accessed: July 30, 2012. [p. 97]

[TMA09] X. Ta, G. Mao, and B. Anderson. On the phase transition width of k-connectivity in wireless multihop networks. *IEEE Trans. Mobile Comput.*, 8(7):936–949, July 2009. [p. 16]

[Vie98] L. Viennot. Complexity results on election of multipoint relays in wireless networks. Technical Report RR-3584, INRIA, 1998. [p. 57]

[vJPHE02] M. Čagalj, J.-P. Hubaux and C. Enz. Minimum-energy broadcast in all-wireless networks: NP-completeness and distribution issues. *Proc. ACM MobiCom, Atlanta, GA, USA*, September 2002. [p. 4]

[Wat99] D. Watts. Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105(2):493–527, September 1999. [p. 71]

[WLB05] J. Widmer and J. Le Boudec. Network coding for efficient communication in extreme networks. In *Proc. of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, WDTN '05, pages 284–291, 2005. [pp. 20 and 21]

[WS98] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684), June 1998. [pp. 17 and 18]

[YOKM+06] M. Yassein, M. Ould-Khaoua, L. Mackenzie, S. Papanastasiou, and A. Jamal. Improving route discovery in on-demand routing protocols using local topology information in MANETs. *Proc. ACM PM2HW2N, Torremolinos, Spain*, October 2006. [pp. 4 and 51]

[YOKP05] M. Yassein, M. Ould-Khaoua, and S. Papanastasiou. Performance evaluation of flooding in manets in the presence of multi-broadcast traffic. *Proc. IEEE Intern. Conf. Parallel Distr. Sys., Fukuoka, Japan*, July 2005. [pp. 4 and 51]

[YWLF06] C. Yi, P. Wan, X. Li, and O. Frieder. Asymptotic distribution of the number of isolated nodes in wireless ad hoc networks with Bernoulli nodes. *IEEE Trans. Commun.*, 54(3):510–517, March 2006. [p. 51]