Towards enhancing information dissemination in wireless networks

Rachit Agarwal

THÈSE DE DOCTORAT CONJOINT
TELECOM SUDPARIS et L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité: Informatique et Télécommunications

Ecole Doctoral: Informatique, Télécommunications et Electronique de Paris

Présentée Par
Rachit Agarwal

Pour obtenir le grade de
DOCTEUR DE TELECOM SUDPARIS

Vers une amélioration de la diffusion des informations dans les réseaux sans-fils

Soutenue le: 02/09/2013 **devant le jury composé de:**

Directeur de thèse:	Prof. Monique Becker	Telecom SudParis
Encadrant:	Dr. Vincent Gauthier	Telecom SudParis
Examinateur:	Prof. Marcelo Dias De Amorim	UPMC-Lip6
Examinateur:	Dr. Philippe Jacquet	Alcatel Lucent
Examinateur:	Prof. Nazim Agoulmine	Université d'Evry Val-D'Essonne
Examinateur:	Dr. Frederica Darema	Air Force Office of Scientific Research, USA
Rapporteur:	Prof. André Luc Beylot	ENSEEIHT
Rapporteur:	Prof. Xavier Lagrange	Telecom Bretagne

Thése numéro: 2013TELE0020

JOINT DOCTORATE OF
TELECOM SUDPARIS AND UNIVERSITY OF PIERRE & MARIE CURIE

Speciality: Computer Science and Telecommunications

Graduate School: Computer Science, Telecommunications and Electronics (Paris)

Presented by
Rachit Agarwal

For the Degree of
DOCTORATE OF TELECOM SUDPARIS

Towards Enhancing Information Dissemination in Wireless Networks

Thesis Defence Date: 02/09/2013

Jury:

Thesis Director:	Prof. Monique Becker	Telecom SudParis
Supervisor:	Dr. Vincent Gauthier	Telecom SudParis
Examinator:	Prof. Marcelo Dias De Amorim	UPMC-Lip6
Examinator:	Dr. Philippe Jacquet	Alcatel Lucent
Examinator:	Prof. Nazim Agoulmine	Université d'Evry Val-D'Essonne
Examinator:	Dr. Frederica Darema	Air Force Office of Scientific Research, USA
Reviewer:	Prof. André Luc Beylot	ENSEEIHT
Reviewer:	Prof. Xavier Lagrange	Telecom Bretagne

Thesis number: 2013TELE0020

to my father,
my sister and
in the loving memory
of my mother

Table of Contents

List of Tables

List of Figures

List of Abbreviation

Abbreviation	Expansion
ACK	Acknowledgement
AE	Antenna Element
APL	Average Path Length
BC	Betweenness Centrality
CC	Clustering Coefficient
$CCDF$	Complementary Cumulative Density Function
CDR	Call Detail Record
CLC	Closeness Centrality
$CTMC$	Continuous Time Markov Chain
DNA	Dynamic Network Analysis
DTN	Delay Tolerant Network
$D2D$	Device to Device
$D4D$	Data for Development
$\mathbf{E}$	Expected value
EBC	Ego betweenness centrality
eNB	Evolved Node B
FBC	Flow Betweenness Centrality
GCC	Giant Connected Component
$GSCC$	Giant Strongly Connected Component
GIN	Giant In Component
$IEEE$	Institute of Electrical and Electronics Engineers
LTE	Long Term Evolution
$MANET$	Mobile *Ad hoc* Network
$MATLAB$	Matrix Laboratory
$MIMO$	Multiple Input Multiple Output
MIS	Maximal Independent Set
$MNDB$	Maximum Node Degree Beamforming

Abbreviation	**Expansion**
PSN	Pocket Switched Network
PWS	Public Warning System
QoS	Quality of Service
RBC	Routing Betweenness Centrality
RDB	Random Direction Beamforming
$SET3$	$D4D$ DataSet 3
SIS	Susceptible-Infected-Susceptible spreading model
SIR	Susceptible-Infected-Recovered spreading model
SNA	Social Network Analysis
SP	Shortest Path
$SPLoc$	SubPrefecture Location Dataset
$SPLocComp$	SubPrefecture Region Boundary Dataset
TBC	Temporal Betweenness centrality
$TCLC$	Temporal Closeness Centrality
TD	Temporal Degree
$TNDB$	Two-hop Node Degree Beamforming
TSP	Temporal Shortest Path
ULA	Uniform Linear Antenna Array
UCA	Uniform Circular Antenna Array
WBA	Wireless Broadcast Advantage
WFB	Wireless Flow Betweenness

Abstract

In public warning message systems, information dissemination across the network is a critical aspect that has to be addressed. Dissemination of warning messages should be such that it reaches as many nodes in the network in a short time. In communication networks those based on device to device interactions, dissemination of the information has lately picked up lot of interest and the need for self organization of the network has been brought up. Self organization leads to local behaviors and interactions that have global effects and helps in addressing scaling issues. The use of self organized features allows autonomous behavior with low memory usage. Some examples of self organization phenomenon that are observed in nature are Lateral Inhibition and Flocking. In order to provide self organized features to communication networks, insights from such naturally occurring phenomenon is used.

Achieving small world properties is an attractive way to enhance information dissemination across the network. In small world model rewiring of links in the network is performed by altering the length and the direction of the existing links. In an autonomous wireless environment such organization can be achieved using self organized phenomenon like Lateral inhibition and Flocking and beamforming (a concept in communication).

Towards this, we first use Lateral Inhibition, analogy to Flocking behavior and beamforming to show how dissemination of information can be enhanced. Lateral Inhibition is used to create virtual regions in the network. Then using the analogy of Flocking rules, beam properties of the nodes in the regions are set. We then prove that small world properties are achieved using average path length metric. However, the proposed algorithm is applicable to static networks and Flocking and Lateral Inhibition concepts, if used in a mobile scenario, will be highly complex in terms of computation and memory.

In a mobile scenario such as human mobility aided networks, the network structure changes frequently. In such conditions dissemination of information is highly impacted as new connections are made and old ones are broken. We thus use stability concept in mobile networks with beamforming to show how information dissemination process can be enhanced. In the algorithm, we first predict the stability of a node in the mobile network using locally available information and then uses it to identify beamforming nodes. In the algorithm, the low stability nodes are allowed to beamform towards the nodes with high stability. The difference between high and low stability nodes is based

on threshold value. The algorithm is developed such that it does not require any global knowledge about the network and works using only local information. The results are validated using how quickly more number of nodes receive the information and different state of the art algorithms. We also show the effect of various parameters such as number of sources, number of packets, mobility parameters and antenna parameters etc. on the information dissemination process in the network.

In realistic scenarios however, the dynamicity in the network is not only related to mobility. Dynamic conditions also arise due to change in density of nodes at a given time. To address effect of such scenario on the dissemination of information related to public safety in a metapopulation, we use the concepts of epidemic model, beamforming and the countrywide mobility pattern extracted from the $D4D$ dataset. Here, we also propose the addition of three latent states to the existing epidemic model (SIR model). We study the transient states towards the evolution of the number of devices having the information and the difference in the number of devices having the information when compared with different cases to evaluate the results. Through the results we show that enhancements in the dissemination process can be achieved in the addressed scenario.

Acknowledgement

First, I would like to thank Prof. Monique Becker for her constant support and guidance during my research without which this thesis would have not been possible. I would like to also thank her for her patient updates and significant feedbacks to my numerous drafts and her hospitality to my numerous visits to her place in south and east. I would also like to thank her for her errands to University of Pierre and Marie Curie during my initial few month of the research. Post lunch interactions with her certainly helped me build on my research both conceptually and methodically. It was indeed my great fortune to carry out my PhD studies under her esteemed guidance.

I would then like to thank Dr. Vincent Gauthier, my supervisor, especially for his extensive help and friendly nature. His intrinsic detailing and thought process resulted in profound improvements in my research. His constant updates on the latest research helped me broaden my state of the art in my field of research. I would also like to thank him for his efforts towards making it possible for me to attend summer school at NECSI, MIT. Being a beginner in French, I would like to thank him for being occasional translator to me and his kindness for driving me to the University of Pierre and Marie Curie and helping me out with the contract signing.

I am also grateful to Prof. André-Luc Beylot, Prof. Xavier Lagrange, Prof. Marcelo Dias De Amorim, Dr. Philippe Jacquet, Dr. Frederica Darema and Prof. Nazim Agoulmine for being members of my thesis evaluation committee and providing their valuable comments on my thesis work.

I would like to thank Prof. Chai Kiat Yeo and Prof. Bu Sung Lee to providing me an opportunity to work with them during my initial months of my research. I thank Dr. Abhik Banerjee with whom I worked closely during my first year of research and for helping me with research methodologies.

At Telecom SudParis, I would like to specially thank Valerie Mateus and Françoise Abad for helping me with my numerous missions especially to US and with the administrative works. Nevertheless, I would like to thank other people at Telecom SudParis especially in “Réseaux et Services de Télécommunications” department for helping me during my thesis. These people include, Prof. Michel Marot, Prof Hossam Afifi, Dr. Charbel Nicolas and Dr. Mohammed Haykel Zayani. However, I am also thankful to other members of “Réseaux et Services de Télécommunications” department with whom I had many technical interactions.

I want to express special thanks to Dr. Ashish Gupta, Dr. Manoj Panda, Dr. Arshad Ali and

Dr. Faiz Hamid for helping me both technically and socially during the course of my thesis.

Further, I wish to thank the following people Prof. Emeritus Pierre Becker for his hospitality and being my tour guide during my frequent visits to Prof. Monique Beckers' home, Prof. Prem Kumar Kalra for getting me introduced to Prof. Monique Becker. I thank Prof. Emeritus Timothy Goldsmith of Yale, Prof. Elizabeth Wilmer of Oberlin College, Prof. Dan Braha of MIT, Prof. Yaneer Bar Yam of NECSI, Dr. Kristina Talbert-Slagle of Yale for providing me an opportunity to work with them during my NECSI days.

Of course, I am grateful to my father and sister for their patience and love, without them this work would never have come into existence.

Evry, France — Rachit Agarwal

DATE 2^{nd} September 2013

CHAPTER 1

Introduction

Due to the recent advancements in the technology and human needs, deployment of wireless networks in recent times has picked up a lot of attention. Many properties that affect the performance of wireless networks have been identified [3]. Further, many studies have been focused on how to address the scaling issues, constant supervision and technological development. Vast topological perils in the deployment of wireless networks have made it very hard for humans to constantly monitor the needs of the wireless systems in-terms of energy and technological advancements [4]. Not only the above mentioned limitations require great consideration but many other factors such as networking constraints, environment interactions, resource utilization, development of different types of heterogeneous network architectures also require great attention [4]. The evolution of wireless *ad hoc* network makes it tough for existing algorithms to answer many intrinsic details associated with it such as, increased contention and collision. This leads to a great need for the algorithms that should perform well in "self*" environment[1]. The growth of networks has made the researchers to look into complex, distributed and adaptive systems that could answer most of these questions and scale with the network. Thus, in this dissertation, we build the algorithms such that self organized features are provided to the network in order to achieve faster dissemination.

For a system to be considered adaptive, it should follow four paradigms [5]. The paradigms state that the entities in the system should operate in an autonomous manner using locally available information, implicitly coordinate, have short lived memory and should adapt to the changes in the environment. To achieve global performance characteristics using only local information in a distributed scenario is a hard task. Consider Nature, it very efficiently solves most of the complex problems using very simple rules and by following these four paradigms. The application of the

[1]Self* refers to self organized, self similar, self configure, self manage, self adapt, self diagnose, self protect and self repair

Natural phenomenon in computing dates back to the times of Sir Allan Turing when he modeled the patterns on the animal skin using simple rules. His research widely opened a field for researchers to investigate the use of naturally occurring phenomenon in their respective research areas.

1.1 Challenges

In Wireless Networks, despite the use of the local information in an autonomous environment and above mentioned deployment issues, efficient transmission of information from one node to another is still the most important issue. In a realistic environment, in wireless networks, transmission of information is affected by many factors like non uniform deployment, dynamic topology, lack of centralized infrastructure and limited resources. Moreover, in different types of wireless networks, several different associated properties affect the dissemination process. For example, in mobile wireless networks especially Pocket Switched Networks and Device to Device communication network, the characteristics of human behavior and human mobility patterns affect the dissemination of information.

Consider a public warning message that has to be disseminated across the population. In static wireless network, a network that does not change in a given time duration, the non-uniform deployment of nodes could lead to several disconnected network components. Information flow between two components will not be possible until a connection is established between the two components. This limits the dissemination process to a certain fraction of the population. In a scenario dealing with public warning message, a fraction of population only having the information is not sufficient. Full population should be able to receive the message.

Consider a static network where nodes do not move and lack mobility characteristics. Now consider that, at some instant, perturbation has occurred in the network causing changes in the network topology. Some factors that cause the network topology to change are: a bridge node within a component being removed from the component due to energy limitations causing the component to break into several smaller components and occurrence of a natural disaster, for example, landslides and earthquakes. In such a case, a strongly connected network component could break into several small components. Here we call the new formed network as a static network because the network will not change unless the above factors again occur. In such disaster hit scenarios, quick dissemination

of public warning message plays an important role. The quicker relief reaches, the better it is. However, due to changes in the network topology where the network broke into several components, the dissemination process is highly impacted as stated in the previous paragraph.

When the nodes in wireless network are mobile (Mobile *ad hoc* Network, Pocket Switched Networks, Delay Tolerant Networks, Device to device communication network), the topology of the underlying network changes frequently causing changes in the dissemination process. The mobility patterns of such networks is closely related to mobility patterns of human as devices are carried by humans. In low density scenarios, mobility pattern may leave the network disconnected and hamper the dissemination process. However, studies argue that information dissemination is enhanced through mobility. In a metapopulation (where the population is spatially divided into communities and the interact at some level), mobility has been proved to enhance the information dissemination process. However, in a scenario where a public warning message has to be disseminated quickly in the entire network, a challenging task is to formulate an efficient and effective algorithm that achieves the goal as soon as possible.

In the growing needs of frequent changes in the technology, the network is highly dynamic. In real scenarios dynamic population causes network wide changes thereby affecting the dissemination process. When the network is mobile, the information dissemination is also affected by the region characteristics like the density and the distribution of the population in the region.

The above challenges motivate us to study how the information dissemination process can be enhanced in the network where nodes lack global knowledge about the network and are constrained by local one hop information. If the network is static and disconnected, how properties of a node can be used to disseminate the information from one component to another and if the network is mobile how dissemination process can be further enhanced. Towards this, in the next section we provide more details on the motivation of the research and our contributions.

1.2 Motivation and Contributions

In an ever-changing environment of wireless networks, it is important for the researchers to design and analyze algorithms using local information that can minimize overheads and guarantee performance for wireless network operations. In a wireless network where nodes have limited energy and

limited memory the use of local information helps in lesser consumption of energy and memory. The interaction of nodes with their environment also plays a very important role in defining the node properties for efficient data transfer and information dissemination. Watts-Strogatz model [2] is one well established model where by rewiring few links in the network, the average path length of the network is reduced while the network structure is maintained. The reduction in the average path length of the network directly relates to the faster dissemination of the information in the network as the overall diameter of the network is reduced. In wireless networks, however, rewiring can be achieved through $MIMO$ technology by the adaptation of the transmission range and reception range. The adaptation of the transmission and reception range can be done using beamforming. Beamforming is a technique where radiation from antenna elements that are embedded in wireless nodes interfere constructively and destructively. The interference gives rise to a long directional beam. Adaptation of transmission range and reception range being one clear direction, a key aspect of this work is based on how nodes can self organize their beams to facilitate the information dissemination process. On the whole, in this research, we are focused on how performance enhancements can be achieved in wireless networks in terms of faster information dissemination in the network in an optimal way with nodes incurring minimal overheads.

Towards this, our objective in Chapter 3 is to propose how local information and natural phenomenon like Flocking and Lateral Inhibition can be used effectively to gain performance in terms of faster information dissemination with low message overheads in a static network. In this Chapter, we focus on the non-uniform deployment of the static nodes working autonomously in a network comprising of several components. The autonomous behavior leads us to propose a decentralized approach where the nodes in the network communicate directly with each other as in device to device communication. In order to achieve connection between the components for the transmission of the information we use beamforming. However, the use of beamforming in the wireless *ad hoc* networks introduces asymmetries and can reduce the neighborhood size of the node. If all nodes beamform, the network will be highly asymmetrical and will lose its structure. These factors lead us to limit the number of nodes that beamform. The determination of the beamforming nodes and the beam properties thus is a key part of Chapter 3. The determination of the beamforming nodes and the nodes towards which the beams are to be directed is achieved through the analogy of Flocking and

Lateral Inhibition. Analogous to the three Flocking rules, Alignment, Separation and Cohesion, in the algorithm design, we relate Alignment to the decision of the node to whether to beamform or not to beamform. We relate separation to no two neighbors beamforming in the same direction while we relate cohesion to setting the direction of the beam in the direction of the most central node. Further, as the size of network increases, memory usage also increases. In order to limit the memory usage, a component in the network can be further logically subdivided (not changing the component structure). This division can be achieved through Lateral Inhibition. The use of Lateral Inhibition creates logical regions in the component. This also help us in limiting the number of nodes that beamform. The beams are directed toward the centroid of a region. The centroid of the region is determined using averaging technique. Through the results obtained it is noted that the algorithm not only enhances network performance in terms of low average path length and maintaining the clustering coefficient but also increases the connectivity.

However, in mobile wireless network, the above algorithm technique incurs lot of overhead due to constant changes in the network topology. Motivated by the properties of Pocket Switched Networks, in Chapter 4, we propose how the concept of stability and beamforming can be used in order to achieve faster information dissemination. In Pocket Switched Network, the properties of the network are closely dependent on the human mobility patterns, the network topology changes at every instant and device to device communication occurs. Recently, it is found that the jump length distribution of human mobility fits truncated power law distribution [6].

Mobility in the system can cause the network to break into several components thereby hampering the information dissemination process. However, mobility can also enhance the dissemination process as it can enforce new connections within the network [7]. Using the properties of human mobility, in Chapter 4 we propose how, at each instant, stability of a node can be computed using local information and how the stability measure can be used towards enhancing information dissemination. In this Chapter we define the stability measure for a node as the fraction of connections kept intact by the node. We use the stability measure to determine whether the nodes should beamform or not and towards which nodes the beams should be directed.

However, in real scenarios, some researchers argue that truncated power law nature of human mobility is an estimate due to the sampling of the data [8] and depends on the region where the

samples are collected. In [6] the mobility data used are the samples collected from United States. While in much recent work [9] the samples are collected from Portugal. Csáji *et al.* in [9] found that the jump length distribution of human mobility has a cut off distance equal to 106 ± 10km and the power law exponent equal to 1.37 ± 0.06. Differences in the mobility pattern for a region, lead us to acquire $D4D$ (Data for Development) dataset [10] created by Orange Labs. The dataset featured Call Detail Records (CDR) and mobility patterns for the users of Orange network service in the country of Ivory Coast.

In a spatial distribution of large population where community structure is evident and interactions are based on the movements and proximity interactions, dissemination in such metapopulation has been well studied [11, 12] using epidemic models. As the traditional dissemination algorithms, in the two Chapters 3 and 4 of this dissertation we also considered fixed density of nodes over time. However, in real mobile device to device wireless communication scenario, variations in the density of the nodes has been overlooked. Thus, in Chapter 5, we propose how dynamic density affects the dissemination process and how beamforming and mobility can be used to enhance the dissemination process. Dynamic density can be achieved using sleep and wake of the mobile devices. In this Chapter we propose an update to a classical dissemination model (epidemic model), SIR model, with three additional states, E_S, E_I and E_R, in order to incorporate dynamic node density. We associate each state in the SIR model to an additional state. We calls these additional states as the latent states (sleeping states). The nodes in the latent state do not participate in dissemination process. However, once a node becomes active (awake state), it again participates in the dissemination process. The transitions between the three active states (S, I, R) and the latent states and vice versa occurs with a defined rate.

1.3 Dissertation Organization

This dissertation is organized into two parts. The first part is dedicated to the algorithms developed with related work while the second part contains a set of appendices dealing with the analysis of the data obtained for the $D4D$ challenge and similar work performed by us using flows in the network.

Part one is further divided such that it first provides an overview of the concepts used, like small world, beamforming, centrality, Flocking, Lateral Inhibition, human mobility, stability measure and

epidemic models to name a few, in the dissertation with state of the art pertaining to the concepts in Chapter 2. The three main contributions then form the three Chapters of the dissertation, Chapter 3, Chapter 4 and Chapter 5 respectively. Chapter 3 outlines the algorithm for the static network, Chapter 4 proposes to use stability measure and beamforming while Chapter 5 incorporates dynamic density scenario for achieving enhanced information dissemination. All Chapters contain a discussion on state of the art pertaining to the proposed idea, algorithm, simulation results and conclusion. The Chapters are then followed by a comprehensive conclusion and future perspectives.

Part two consists of a parallel work performed by us that relates to how small world properties in wireless networks can be achieved using flows in the network, the analysis of one of the dataset received from Orange Labs towards the *D4D* challenge, publications during the course of the thesis and an extensive bibliography.

CHAPTER 2

State of the Art and Useful Concepts

In this Chapter, we discuss the concepts that are used for achieving the goals of this dissertation and discuss the state of the art related to those concepts. We first present the concepts like Self Organization (Section 2.1), Small World (Section 2.2), Centrality measures (Section 2.2.1), Beamforming (Section 2.3), Lateral Inhibition (Section 2.4), Flocking (Section 2.5) and Non-uniform Node Deployment (Section 2.6) to provide an understanding for Chapter 3. We then provide an overview of Human Mobility (Section2.7), Stability Measures (Section 2.8) and Information Dissemination (Section 2.9) which form the basis for Chapters 4 and 5.

2.1 Self Organization

In the 1940's, Asby [13], in his work on cybernetics provided the notion and definition of self organization. According to him, self organization is the expression of global behavior in terms of cooperation among individual entities through simple means without any controlling agent.

In general, cooperation leads to different emergent behaviors such as grouping and synchronization in the system and achieves scalability due to the absence of a central controlling agent. This evolution of behavior also leads to changes and diversities in the system and vice versa. Evolution has been very well explored in the context of biology. Over the years, the focus of the researchers has remained towards the study of evolution over time and scale. An example of evolution over scale, also called *emergence* [14], is the change that occurs in the system due to the interaction of multiple subsystems. This means that the collective effort leads to the evolution of the system that has completely different properties from its subsystems. Some examples of collective effort over scale observed in biology are (i) different organs in a human body interact with each other to give an

individual that has different characteristics than the organs, (ii) fireflies synchronizing in unison (iii) Flocking and (iv) Fractals, highly scalable self organized systems, such as leaf of a fern, cauliflower and blood vessels. On the other hand, an example of evolution over time, also called as Self Organization [14], is the change that occurs *within* a system. Properties of self organized systems vary from system to system. Most of the systems show properties like positive feedback, negative feedback, leadership, inhibition, synchronization, coordination, cooperation and scalability. Learning is another example of evolution over time. Learning occurs when a system observes its surroundings and performs actions based on what can be beneficial to both surrounding and itself. During learning, a system can thus change its behavior depending on the environment. Nevertheless, evolution is not only restricted to biological systems, but also observed in many non-biological systems. For example, variations of population dynamics over food chain [15], self organized criticality [16, 17] and human behavior.

The history of self organization can be dated back to the evolution of the universe where the chaotic interaction of the particles gave rise to new structure and phenomena. Through time, this interaction has led the system to evolve and adapt to its surroundings. One of the first examples studied by mankind was the Belousov-Zhabotinsky reaction. In this reaction, the chaotic interactions among the chemicals lead to a pattern formation. During the same period Sir Allan Turing also published his paper on morphogenesis which showed that very simple interactions can lead to very complex patterns. These two pioneering works widely opened the field of self organization and various properties like *no central control*, *emerging structure*, *complexity*, *loose coupling*, *dependence on the environment* and *high scalability* were derived. These properties led researchers to formulate four basic paradigms for a system to be called self organized [5]. These paradigms are: the elements of the system should use locally available information, implicitly coordinate, have short lived memory and should adapt to the changes in the environment.

Many current technologies are based on the mechanisms applied by the naturally occurring phenomenon. Self organization in a system is applied for various purposes such as configuration, management, adaptability, diagnosis, protection, healing and repair and optimization. However, self organized systems have several limitations, for example, low predictability and slow convergence to the optimal state.

The interactions among the elements of the system lead to the formation of a network. Some researchers study the evolution of the network based on the growth and decay in the network [18, 19, 20, 21, 22, 23, 24]. In these studies, the growth and the decay patterns are based on the preferential and probabilistic attachment and detachment of nodes in the network. Some researchers study this effect on different types of underlying networks. Different studies have given many different names to the networks that evolve. Some of the studies refer the evolving network as a Dynamic Network [25, 26, 27]. Other names given to the evolving networks [28, 29, 30, 31] are Time Varying Graphs [32, 33], Temporal Networks [34, 35] and Adaptive Networks [36, 37].

Evolution also exists in mobile scenario where there are constant changes in the interactions in the network due to mobility. In the context of communication networks, Mobile *Ad hoc* NETworks ($MANET$) share the properties of the dynamic networks. In $MANET$s, a wireless device is considered as a node in the network with motion capabilities and spatial characteristics. Self organization techniques are used in communication networks mainly to provide better routing mechanisms, resource allocation, network security, synchronization, neighborhood discovery and fault tolerance [4].

2.2 Social Networking and Small World Networks

One of the key aspects of self organization observed in human community is that of social interactions or social networking. Social groups exist where social ties link individuals with common values and belief.

Social linking between individuals is called as social networking. In 1967, Stanley Milgram [38] conducted an experiment and analyzed the Average Path Length (APL) from one individual to another while working on his thesis on small world experiment. In his experiment, individuals were randomly selected and were asked to forward a letter to a target contact. If an individual knew the target contact the letter was directly forwarded after appending the individual's name. However, in the case when an individual did not know the target contact, the individuals were only allowed to forward the letter to one of their connections and append their name. This experiment of Milgram gave rise to the concept of Six degrees of Separation and widely opened the social networking research area. However, many shortcomings of the experiment have been reported since then, for example, (i) some letters did not reach the destination during specified time interval and (ii) the scale of the

experiment was small.

Consider a graph $G(V, E)$ where V and E represent the set of nodes and edges, respectively, and $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Suppose SP_{ij} is the shortest path between two node $i \in V$ and $j \in V$ such that $|SP_{ii}| = 0$. Then APL of the graph is defined as

$$APL = \frac{1}{|V|(|V|-1)} \sum_{i,j \in V} |SP_{ij}|. \tag{2.2.1}$$

Note that as there are $|V|(|V|-1)$ node pairs in the networks, the summation of all the shortest paths is normalized by $|V|(|V|-1)$.

Nevertheless, the analysis of social networks gave rise to the concept of Social Network Analysis (SNA). SNA is used in many research fields for example, epidemiology for knowing the spread of the diseases, mass surveillance, sociology and communication networks. SNA is extensively used to study network properties like centrality measures. SNA consist of many metrics like socio-centric betweenness centrality, closeness centrality, degree centrality, etc. that help in studying network features.

Other than above mentioned centrality measures and APL, metrics like Clustering Coefficient (CC) is also defined to characterize a network. CC is a measure of how well nodes are clustered and is defined in two ways, globally and locally. Global CC is defined as

$$CC_{global} = \frac{\text{number of cycles of length 3 in } G}{\text{number of paths of length 2 in } G}. \tag{2.2.2}$$

Consider a node $i \in V$ with N_i neighbors such that $0 \leq |N_i| \leq |V|$ and $N_i \subset V$. Note that the distance between nodes in N_i and i is one hop. Let e_{jk} be an edge between $j \in V$ and $k \in V$ such that $j, k \in N_i$. Let E_{N_i} be a set of all e_{jk}, $j, k \in N_i$ in the neighborhood of i. The local CC, is defined as

$$CC_i = \begin{cases} \frac{|E_{N_i}|}{|N_i|(|N_i|-1)} & \text{if directed graph.} \\ \frac{2|E_{N_i}|}{|N_i|(|N_i|-1)} & \text{if undirected graph.} \end{cases} \tag{2.2.3}$$

Consider the the subgraph restricted to the nodes in N_i and call it G_{N_i}. The eq. 2.2.3 refers to the fraction of edges in G_{N_i} that actually exist to the maximum number of edges that could have existed in G_{N_i} [2]. Using CC_i, the average CC in the graph is defined as

$$CC_{average} = \frac{1}{|V|} \sum_{i \in V} CC_i. \tag{2.2.4}$$

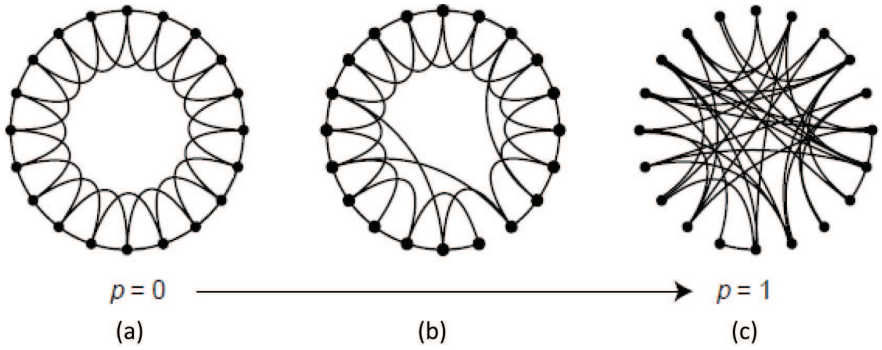


Figure 2.1: Source: [2], (a) regular graph with $p = 0$, (b) small world graph with $0 < p < 1$ where some of the edges from (a) were rewired, (c) random graph with $p = 1$ where complete rewiring of edges in graph (a) were performed.

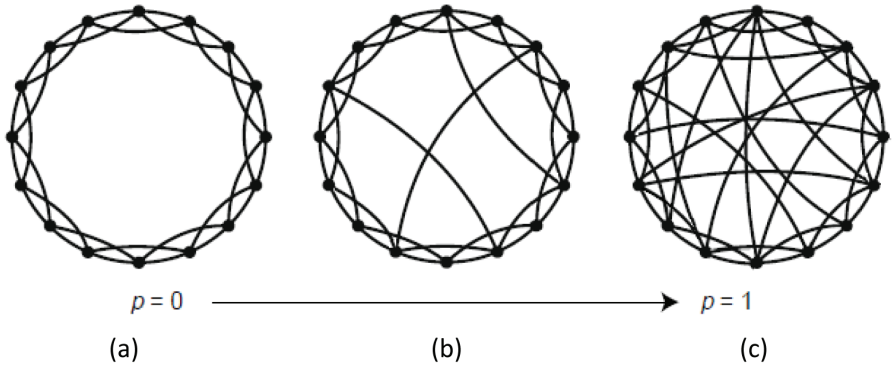


Figure 2.2: (a) regular graph with $p = 0$, (b) small world graph with $0 < p < 1$ where edges were added to few nodes in (a), (c) random graph with $p = 1$ where edges were added to all the nodes in (a).

Inspired by the experiment of Stanley Milgram [38] about the six degrees of separation, Watts and Strogatz [2] proposed a model for the concept of a small world and showed how dissemination process in a network can be enhanced. Watts and Strogatz [2] showed that by rewiring edges of the network with probability p, the APL in a network is reduced drastically while clustering coefficient is almost kept intact. Consider a regular undirected ring graph $G(V, E)$ with set of nodes V and average number of neighbors $\langle k \rangle$ such that $|V| \in \mathbb{N}$, $|V| > 2$ and $|V| \gg \langle k \rangle \gg 1$. For a regular graph $|E| = \frac{|V|\langle k \rangle}{2}$. Let E_i be the set of edges connecting node i to other nodes in V. Note that $|E_i| = \langle k \rangle$. In Watts and Strogatz model, a node i is uniformly chosen from V. Further, an edge $e_{ij} \in E_i$ is chosen such that the maximum distance between node i and node $j \in V$ is 1 (note that

there can be a path from node i to j such that distance between i and j using this path is greater that 1). The chosen edge is then reconnected to a randomly chosen node v with a probability p such that $v \in V - \{i, j\}$. If e_{iv} is already in E_i then the edge is not made. This process continues for all the edges where the maximum distance between i and j is 1. The process is done for all the nodes in the graph. Once all the nodes are considered, the process is again continued with the edges that connect two hop neighbors. The chosen edge is again rewired with probability p to a random vertex satisfying the above condition. This process is continued until all the edges are considered. In summary, every edge is rewired with a probability p such that there are no duplicate edges. As the rewiring process is dependent on p, the value of p shows an impact on the rewiring process. $p = 0$ means that no rewiring is done while $p = 1$ means complete rewiring of the graph is done (cf. Fig. 2.1). A rewiring probability $p = 1$ leads to a random graph. Note that, as no edge is deleted or added, the number of edges in the resultant graph after rewiring is also $\frac{|V|\langle k\rangle}{2}$. Watts and Strogatz showed that if $0 < p < 1$, then the APL is drastically reduced while the CC is almost kept intact. They also showed that regular graphs are highly clustered where the APL grows linearly with the number of nodes, while for a random network, the growth in APL is logarithmic. Variation to Watts *et al.* rewiring model has also been proposed. For example in [39, 40, 41], instead of link rewiring, link addition is performed between the nodes with a probability p (cf. Fig. 2.2).

Many researchers [18, 21, 39, 42, 43, 44, 45, 46, 47, 48, 49], since showed interest in small world concepts and analyzed many associated characteristics. A brief survey of some of these models is provided in [40]. A wireless network, due to its spatial and omnidirectional transmission range characteristics, can be considered as a regular network, and applying rewiring to wireless networks will lead to more randomness. Assuming a spatial wireless network, Helmy [48] performed small world analysis on wireless networks and showed that rewiring of links had less effect on the structure of the network. Helmy [48] also showed that the APL is reduced at a greater rate when shortcuts are 25% to 40% in length of the network diameter and the rate of APL reduction is more when only 0.2% to 2% shortcut links are created. The reduction rate stabilizes when more than 2% shortcut are created.

On the other hand, Barabási *et al.* in [42, 49] argued that a large network that has small APL exhibits small world characteristics and can be called as scale free network. In scale free network,

the network tends to have few nodes that have high number of connections, called the "hub" nodes, and more nodes that have less number of connections. Information is spread quickly within the network through the hub nodes. Note that the number of connections a node has is also called the degree of a node. Moreover, the degree distribution of nodes in a scale free network follows a power law distribution. Research has proved that information flow through these hub nodes is both advantageous as well as disadvantageous for the network. Despite low APL, these networks have increased probability of communication failures due to overloading at the hub nodes.

2.2.1 Social Network Analysis and Centrality Measures

Decades of research on graph theory has led researchers to be inspired from social networks and derive many metrics that identify importance of a node in the graph. This field of research commonly came to be known as SNA. In SNA, the concept of centrality is brought up and developed to address topological characteristics of the nodes in the graph. The centralities, depending on the parameters they use, can be both local as well as global. Some examples of global centrality measures include socio-centric betweenness [50, 51] and closeness centralities [51]. The egocentric betweenness centrality [52, 53] is an example of local centrality measure. All the above centrality measures are well-defined for static networks. Recently, centrality measures in dynamic conditions have also been proposed [54] where the field of research is popularly known as Dynamic Network Analysis (DNA).

Socio-Centric Betweenness Centrality

Socio-centric betweenness centrality [50, 51] is the measure of number of shortest paths passing through the node thereby expressing the most important node in the network through which most communications take place. Determination of the shortest path requires the knowledge about the network. Socio-centric betweenness is a global measure as it is computed using shortest paths. Nodes with high degree and nodes that are acting as the bridge nodes tend to have high socio-centric betweenness. Consider a graph $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let σ_{uv} be the number of all shortest paths between two nodes $u \in V$ and $v \in V - \{u\}$. Let σ^i_{uv} be the number of shortest paths between u and v passing through $i \in V$ such

that $u \neq v \neq i$. Socio-centric betweenness of a node $i \in V$ is defined as

$$BC_i = \sum_{u \neq v \neq i \in V, \sigma_{uv} > 0} \frac{\sigma^i_{uv}}{\sigma_{uv}}. \tag{2.2.5}$$

Variations to Socio-centric betweenness centrality have also been proposed. These include Egocentric Betweenness Centrality, Flow Betweenness Centrality (FBC) [55] that measures the importance of a node based on traffic flow in the network, Routing Betweenness Centrality (RBC) [56] that considers paths resulting from routing strategies and Wireless Flow Betweenness Centrality [57, 1], explained further in appendix A.

Egocentric Betweenness Centrality

Aiming to compute the Betweenness centrality using local properties, [52, 53] proposed Egocentric Betweenness Centrality measure (EBC). In a graph $G(V, E)$ with the set of nodes V and the set of edges E with no self edges, $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let N_i be the set of one-hop neighbors of $i \in V$, $0 \leq |N_i| \leq |V|$ and $N_i \subset V$. Let A_i be the adjacency matrix formed by all nodes in N_i. Each element, $(A_i)_{jk}$, in A_i represents the state (either 0 or 1) of the edge between two nodes j and k in N_i. The maximum distance between two nodes in N_i is equal to two hops as the nodes in N_i can be connected via i. The number of two hop paths between a pair of nodes in N_i is given by the corresponding element of A_i^2. Let $x_{jk} = [1 - (A_i)_{jk}](A_i^2)_{jk}$. When the path length between j and k is more than one hop, x_{jk} will provide number of paths between j and k that are of length 2 hops. However, if there is a one-hop path between j and k, then $x_{jk} = 0$. Everett *et al.* in [52] defined the Egocentric Betweenness of the node i as

$$EBC_i = \frac{1}{2} \sum_{j,k \in N_i, j \neq k, x_{jk} \neq 0} \frac{1}{x_{jk}}. \tag{2.2.6}$$

Note that because A_i is symmetric, EBC_i is normalized by 2. Also note that when nodes in N_i form a clique, $EBC_i = 0$. Next, consider the case when the nodes in N_i form a ring. If the ring contains 4 nodes then $EBC_i = 1$ because there will be two two-hop paths between any node and its diagonally opposite node. However, if $|N_i| > 4$, then $EBC_i = |N_i|$. As another special case, if nodes in N_i form a star topology, $EBC_i = \frac{(|N_i|-1)(|N_i|-2)}{2}$. This holds because each node, other than the central node, has one two-hop path to each of the other $|N_i| - 2$ nodes. As there are $|N_i| - 1$ non-central nodes, we have $EBC_i = \frac{(|N_i|-1)(|N_i|-2)}{2}$.

Marsden in [58] performed an empirical study to find the relation between the two types of Betweenness, Socio-Centric and Egocentric. He found that Egocentric Betweenness is strongly correlated with Socio-Centric Betweenness and can be used when global network information is missing.

Closeness Centrality

Closeness Centrality (CLC) [51] is used to measure how central a node is in the network. Closeness centrality is also a way to measure how quickly dissemination of information, in terms of hops, will take place through the node. Closeness Centrality is the inverse of the sum of the shortest path lengths between a node and all other nodes in the graph. Consider a graph, $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let SP_{ij} be the shortest path between node $i \in V$ and $j \in V - \{i\}$ such that $|SP_{ii}| = 0$, the CLC of the node i is defined as

$$CLC_i = \frac{1}{\sum_{j \in V - \{i\}} |SP_{ij}|}. \tag{2.2.7}$$

A node with the highest Closeness Centrality is also the centroid node in the graph. As all the centrality measures convey different information, it is not necessary that a node that has a high value for one centrality measure also has high values for the other centrality measures. Apart from Socio-centric Betweenness and Closeness Centrality, many other types of centrality measures, such as Eigen Value Centrality [59], Katz Centrality [60], Spectral Centrality and centrality measure in weighted networks [61] also exist but we refrain from defining them as we are only interested in closeness centrality and betweenness centrality. A comprehensive survey of some of the centrality measures is provided in [62].

Dynamic Centrality Measures

Recently, dynamic version of the centrality measures have been proposed using temporal graphs [25, 30, 32, 35, 63]. These measures mainly use temporal distance in order to formulate the centrality measures.

Definition 2.2.1. A temporal (dynamic) graph is a graph that changes with time. The characteristics of temporal graph are studied using a group of discrete snapshots of the graph each taken after

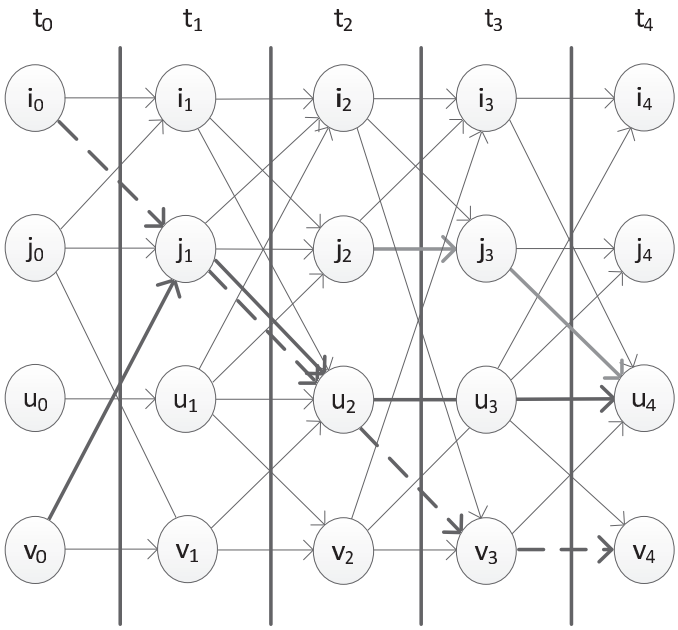


Figure 2.3: The connectivity and temporal distance in the time varying graph.

Δt time step [63]. Temporal graph properties are characterized by changing graph properties over time. These properties mainly include APL, CC, number of components, hierarchical structure and centrality values.

Recently, [32, 35, 63] defined Temporal Shortest Path (TSP), a variation of shortest path for temporal graphs, as the path that takes the minimum number of time steps to send the information to the destination. The minimum number of time steps to any node is called the Temporal Distance. Consider a graph snapshot, $G_t(V, E_t)$, with set of nodes $V = \{i, j, u, v\}$ and set of edges E_t at time t such that $|E_t| \in \mathbb{Z}^+$. Note that V is fixed over all time instants. Let i_t, j_t, u_t and v_t be the time dependent representations of the nodes in V (cf. Fig. 2.3). To compute the temporal distance, the Aggregated Graph is generated as follows. The nodes i_t, j_t, u_t and v_t are vertically arranged at each time step t (cf. Fig. 2.3). If there is an edge between two nodes a and b in $G_t(V, E_t)$, then there are two directed edges in the aggregated graph, namely, an edge from a_t to $b_{t+\Delta t}$ and an edge from b_t to $a_{t+\Delta t}$. This step is performed because, if a node has a packet to transmit at time t, the neighbor nodes in $G_t(V, E_t)$ will only receive the packet at the next time instant $(t + \Delta t)$. It is also assumed that the nodes keep the packet with them and each transmission is completed in a single time step.

Fig. 2.3 shows a representation of aggregated graph generated from snapshot graphs for the period $t = 0$ to $t = 3$. Let G' be the resultant aggregated graph (Cf. Fig. 2.3). Suppose $\Delta t = 1$. Let node i generates a packet at time $t = 0$ that has to be sent to node v. Then the TSP between i and v is $(i_0 \to j_1 \to u_2 \to v_3)$, i.e., it takes 3 time steps to send the packet to the destination. Similarly, if node v generates a packet at time $t = 0$ to send to node u, then one of the TSPs that can be followed is $(v_0 \to j_1 \to u_2)$, i.e., it takes 2 time steps to reach to the destination. Further, if node j generates a packet at time $t = 2$ and wants to send it to v then it will not be able to do so because there is no path from j_2 to v.

Let i_* be an instance of the node $i \in V$ at time $t = *$. Let $G'(V', E)$ be the aggregated graph of n snapshots as explained above for time in $[t, t']$ such that $n \in \mathbb{N}$, $t' \in \mathbb{N}$, $t \in \mathbb{Z}^+$ and $t' > t$. Let $[t, t']$ be a discrete time interval with $\Delta t = 1$. In such a case $t' = t + n$. For time in $[t, t']$, Kim *et al.* in [63] defined TSP between two nodes i and j as

$$TSP_{ij} = \underset{t<x\leq t'}{\arg\min}\{(i_t \to \ldots \to k_{x-1} \to j_x)\}. \tag{2.2.8}$$

Note that due to the temporal nature, the graph G' is a directed graph. Also note that $TSP_{ij} \neq TSP_{ji}$ and $|TSP_{ij}| \in \mathbb{N}$ for $j \in V - \{i\}$. Further, if there is no path between i and j in the time interval $[t, t']$ then $|TSP_{ij}|$ is considered as ∞. This is because TSP is defined based on the time steps and it is assumed that in time $\to \infty$ the nodes i and j will be connected.

Let Y be the set of subintervals of the time interval $[t, t']$ such that the end time of each subinterval is t' but the starting time is strictly less than t', i.e.,

$$Y = \{[k, t'] \mid t \leq k < t'\}.$$

Let TSP_{ij}^{y} be the TSP between i and j in the time interval $y \in Y$. Temporal Closeness Centrality ($TCLC$) of a node i in the time interval $[t, t']$ is defined as

$$TCLC_i = \sum_{j \in V - \{i\}} \left(\sum_{y \in Y} \frac{1}{|TSP_{ij}^{y}|} \right). \tag{2.2.9}$$

The eq. 2.2.9, states to sum all the temporal paths between i and j which have occurred over time period $[t, t']$. Further, as $0 < |TSP_{ij}| < \infty$, when $|TSP_{ij}| \to \infty$, it is assumed that $\frac{1}{\infty} = 0$ and thus $\frac{1}{|TSP_{ij}|} = 0$.

Consider Y as discussed before and σ_{uv}^{i} as the number of all temporal shortest paths between two nodes in $u, v \in V$ passing through $i \in V$ such that $u \neq v \neq i$. Let σ_{uv} be the number of temporal shortest paths between two nodes $u, v \in V$ such that $u \neq v$ and $\sigma_{uv}^{i} \leq \sigma_{uv}$. Temporal Betweenness Centrality (TBC) of a node i is defined as

$$TBC_i = \sum_{u \neq v \neq i \in V, \sigma_{uv} > 0} \left(\sum_{y \in Y} \frac{\sigma_{uv}^{i}}{\sigma_{uv}} \right). \tag{2.2.10}$$

Kim *et al.* in [63] also defined Temporal Degree (TD) of a node i over a time period based on the degree of the nodes at each instant of time. Consider N_i^t as the neighborhood of the node i at time t such that $0 \leq |N_i^t| \leq |V|$ and $N_i^t \subset V$. The degree of node i at time t is given by $|N_i^t|$. The temporal degree of a node i, TD_i, is the average degree of the node over time period $[t, t')$ and is defined as

$$TD_i = \frac{\sum_{t}^{t'-\Delta t} N_i^t}{t' - t - \Delta t}. \tag{2.2.11}$$

In [54] Kim *et al.* proposed how centrality measures can be predicted using aggregated graphs over time rather than using TSP. In the algorithm, the prediction changes every instant as a new snapshot of the graph is aggregated to the previous graphs.

2.3 Antenna Model and Beamforming

[64, 65] provides an extensive study of antenna models. Consider an antenna with elevation θ and azimuth ϕ such that $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$. Let $u(\theta, \phi)$ be the radiation power per unit solid angle. For a lossless antenna, the gain, $g(\theta, \phi)$ is defined as

$$g(\theta, \phi) = \frac{u(\theta, \phi)}{\frac{1}{4\pi} \int_0^{2\pi} \left(\int_0^{\pi} u(\theta, \phi) sin\theta \, d\theta \right) \, d\phi}. \tag{2.3.1}$$

In wireless networks, long links are created using beamforming antenna. A beamforming antenna is an antenna array that consists of M antenna elements (AE)s. Consider an antenna array with M AEs such that $M \in \mathbb{N}$ where all AEs are isotropic radiators with the same phase shift, γ, between them. Individually, each AE has an omnidirectional radiation pattern. When one AE is used there is no superposition of the radiation. This leads to an omnidirectional radiation pattern. However, when more than one AEs are used, there is a constructive and destructive superposition of the

radiation due to the phase shift between the AEs and directional radiation pattern is observed. This directional pattern consists of a main lobe and secondary lobes. The main lobe is obtained in the direction of antenna boresight, B_b. The maximum gain is also achieved at the antenna boresight ($\theta = B_b$) and the radiation pattern is symmetric about the boresight B_b.

Using such setting, two basic antenna models, namely, Uniform Linear Array antenna model (ULA) and Uniform Circular Array antenna model (UCA), are defined. In the ULA model, the AEs are linearly arranged along the z axis. However, in the UCA model the AEs are circularly and symmetrically arranged on the yz plane. In both the models the distance between two AEs is Δ. Let c denote the velocity of light in vacuum, $c = 3 \times 10^8$m/s. Let f be the carrier frequency and $\lambda = c/f$. For simplicity, we define ULA model and will use it hereafter. For ULA model, radiation intensity is proportional to $\left(\frac{\sin(M\psi)}{M*\sin(\psi)}\right)^2$ where $\psi = \frac{1}{2}(\frac{2\pi\Delta}{\lambda}\cos\theta + \gamma)$. The radiation intensity is however independent of ϕ because of the rotational symmetry of the setup about the z axis [65]. We thus denote $g(\theta, \phi)$ as $g(\theta)$. For a ULA model, the boresight direction is defined as $B_b = \pm\arccos\left(-\frac{\gamma\lambda}{2\pi\Delta}\right)$. In order to achieve maximum gain in the direction of B_b, γ is set to $\gamma = -\frac{2\pi}{\Delta}\cos B_b$. That is, given B_b the phase shift between AEs is calculated. This simplifies ψ to $\psi = \pi\Delta(\cos\theta - \cos B_b)/\lambda$. Assuming $\Delta = \lambda/2$ as a common choice, $g(\theta)$ is defined as

$$g(\theta) = \frac{1}{M}\left(\frac{\sin(M\psi)}{\sin\psi}\right)^2. \tag{2.3.2}$$

It is observed that the maximum gain achieved for the ULA antenna model only depends on the number of AEs used while it is independent of the boresight direction (cf. Fig. 2.4). Note that the gain can also be expressed in terms of the maximum gain achieved. In such case, the normalized gain is given by

$$\hat{g}(\theta) = \frac{g(\theta)}{\max\{g(\theta)\}}.$$

Let the transmitter and receiver gains be g_t and g_r, respectively. Let p_t be the signal transmission power and p_r be the power with which the signal is received. Given a propagation environment, characterized by the path loss exponent α, in order to have a link setup between two nodes and for a transmitted signal to be correctly received by a node located at a distance s, p_r is defined as

$$p_r = \frac{p_t g_t g_r}{s^\alpha} > p_{r0} \tag{2.3.3}$$

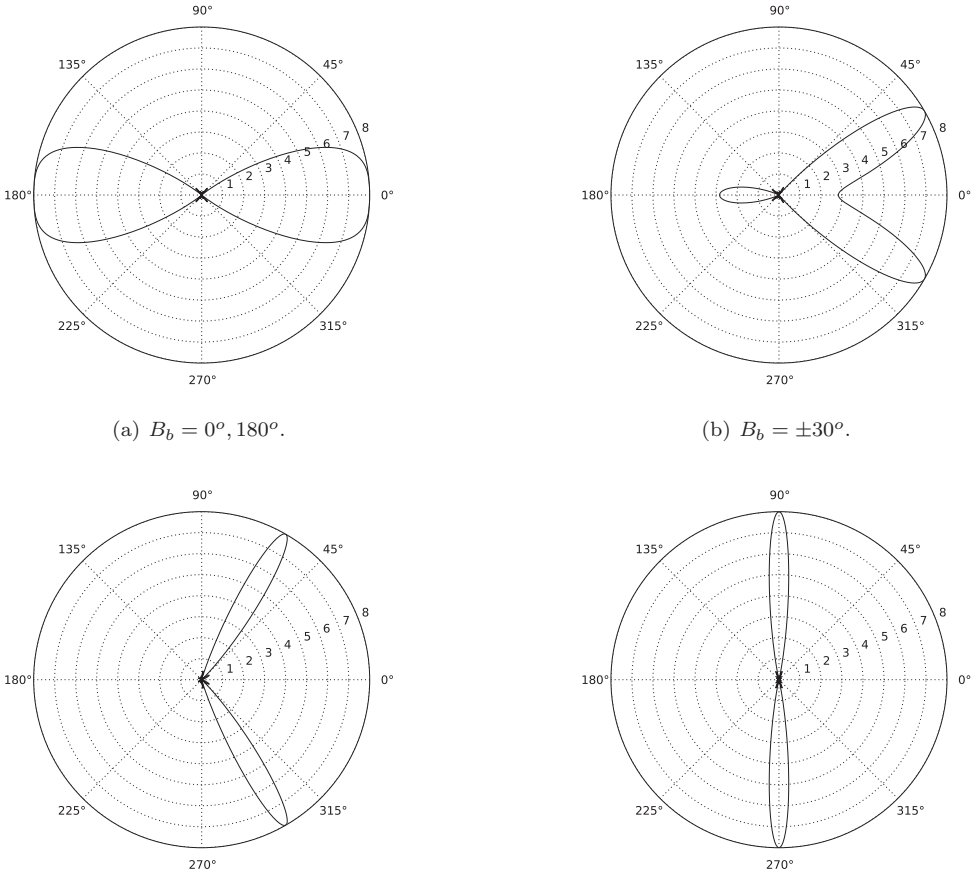
where p_{r0} is a threshold power called as receiver sensitivity. We assume that all nodes have the same p_t and p_{r0}.

In communication networks, Multiple Input Multiple Output ($MIMO$) technology as described in IEEE 802.11n represents a way to achieve beamforming using multiple antennas. However, in $MIMO$, beamforming can be achieved at both receiver and transmitter. In the dissertation, we only focus on achieving beamforming at the transmitter side. Thus we assume $g_r = 1$ as nodes always receive in omnidirectional mode. Further, we assume network setup in two dimensions, specifically the yz plane, which implies $\phi = \pm\pi/2$. Moreover, as in the dissertation our objective is to demonstrate the improvement that is to be achieved in the network performance when beamforming is used over that of omnidirectional antenna, for simplicity, we chose free space propagation model with path loss exponent $\alpha = 2$. In propagation environment other than free space there will be obstacles along certain directions and the transmission range cannot be increased in those directions even using directional antenna. Thus our basic premise of the tradeoff between range and angular coverage that exists between directional and omnidirectional antenna will no longer be valid. Hence, we assume free space propagation environment. For more realistic environments, the algorithms described in the dissertation can be studied using $2 < \alpha \leq 4$.

Nevertheless, our objective is to demonstrate the improvements that is achieved in the network performance when beamforming is used over that of omnidirectional antenna. Hence, it is necessary to present the relation between the beam length in the directional case and that of the transmission range in the omnidirectional case. Consider an omnidirectional transmission range, $r \in \mathbb{R}^+$, which is achieved by the use of one AE. The maximum beam length, B_l, in the B_b direction when M AEs are used is equal to $r \times \sqrt[\alpha]{M}$ for the ULA model. Note that when $\alpha = 2$, $B_l = r \times \sqrt{M}$.

In wireless *ad hoc* networks, beamforming using ULA and UCA model has been well studied. Beamforming techniques using these antenna models include Random Direction Beamforming (RDB) [65, 66]. Bettstetter *et al.* [65] studied the use of RDB with the path probability (cf. Definition 2.3.1 below) to improve the connectivity in the wireless networks. They showed that path probability between two nodes is increased using RDB.

Definition 2.3.1. Consider a graph, $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let there be ν components in the network such that $1 \leq \nu \leq |V|$. Let V_i be



(a) $B_b = 0^o, 180^o$.

(b) $B_b = \pm 30^o$.

(c) $B_b = \pm 60^o$.

(d) $B_b = \pm 90^o$.

Figure 2.4: Gain pattern obtained for different B_b and $M = 8$ when using ULA model.

set of nodes in the i^{th} component of the network such that $0 < |V_i| \leq |V|$. The Path probability is defined as

$$\text{path probability} = \frac{\sum_{i}^{\nu} |V_i|(|V_i| - 1)}{|V|(|V| - 1)}. \tag{2.3.4}$$

Note that if the network has a single component, i.e., if $\nu = 1$, then path probability is equal to 1. On the other hand, if all nodes are isolated, i.e., if $\nu = |V|$, then path probability is equal to 0.

Bettstetter *et al.* also showed that the increase in the path probability is not due to increase in the number of connections a node has but relates to the fact that beamforming connects distant nodes. These long links form bridges between unconnected components thereby increasing the component

Parameter\ Reference	**Vilzmann *et al.* [66]**	**Widmer *et al.* [67]**	**Kiese *et al.* [68]**	**Yu *et al.* [69]**	**Li *et al.* [70]**	**Durrani *et al.* [71, 72]**
Transmission mode	Directional	Directional	Directional	Directional	Directional Omni	Directional
Reception mode	Directional	Directional	Directional	Omni	Directional Omni	Directional
Mobility	No	Yes	No	No	No	No
Beam width	Depends on beam direction	Constant	Constant	Optional	Constant, switched beam antenna	Depends
Beam direction	Random	Optional	Optional	Optional	Random	Random
Antenna model	UCA	UCA	UCA as keyhole	Sector	Keyhole	UCA
Node distribution	Uniform	Uniform and Non Uniform	Non Uniform	Not specified	Uniform	Uniform

Table 2.1: Comparison between various research articles dealing with beamforming and connectivity.

size. Vilzmann *et al.* [67] derived low complexity techniques for beamforming and proposed Maximum Node Degree Beamforming ($MNDB$). In $MNDB$ the nodes direct their beams in the direction where they find maximum number of nodes. Vilzmann *et al.* found that $MNDB$ leads to less number of inter-cluster connections and more intra-cluster connections. To overcome this drawback, they proposed Two-hop Node Degree Beamforming ($TNDB$). In $TNDB$ the nodes direct their beams towards the direction in which maximum number of one-hop neighborhood and two-hop neighborhood is found. Vilzmann *et al.* showed that $TNDB$ outperforms both RDB and $MNDB$. Other works on beamforming include [67, 68, 69, 70]. Considering all nodes beamforming, [65, 66, 67, 68, 69, 70] addressed connectivity very well but do not discuss the impact on the APL and the CC. In [71, 72], authors study connectivity metrics with few nodes beamforming and show that using only few AEs connectivity is increased. However, again [71, 72] lacks impact on CC. Table 2.1 illustrates a comparison between some of these studies.

2.4 Lateral Inhibition

Lateral Inhibition is defined as the ability of a node to restrict its neighbors from performing a certain task. Various examples of Lateral Inhibition are found in biology. An example of Lateral Inhibition observed in biological process is where cells on the tissues of animals skin, depending on

the neighbor properties, decide either to grow a hair or not to grow a hair. Lateral Inhibition ensures that hairs are grown equidistant from each other and helps in producing regular patterns throughout the surface. When a cell grows a hair it inhibits its neighbors from growing a hair, resulting into equally spaced uninhibited cells [73]. Lateral Inhibition thus creates clusters where the cluster heads are the uninhibited nodes distributed uniformly over the region. The Lateral Inhibition process is also regarded as the identification of the maximal independent set (MIS).

Consider a graph $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. A maximal independent set, S, consists of nodes in V such that no two nodes in S are neighbors. Addition of any node from $V - S$ to S will break the property of independence. Note that the above property gives rise to another property whereby all nodes in $V - S$ are connected to a node in S. MIS is often used for clustering of nodes in the network.

In a decentralized environment, probabilistic approaches are used to select MIS. Here, the probability of getting selected to join nodes in S depends on the number of neighbors that are not connected to nodes in S. Moreover, the MIS algorithms have high message complexity and assume the knowledge of the network [74]. These features make it tough for the applications of MIS in the wireless *ad hoc* network.

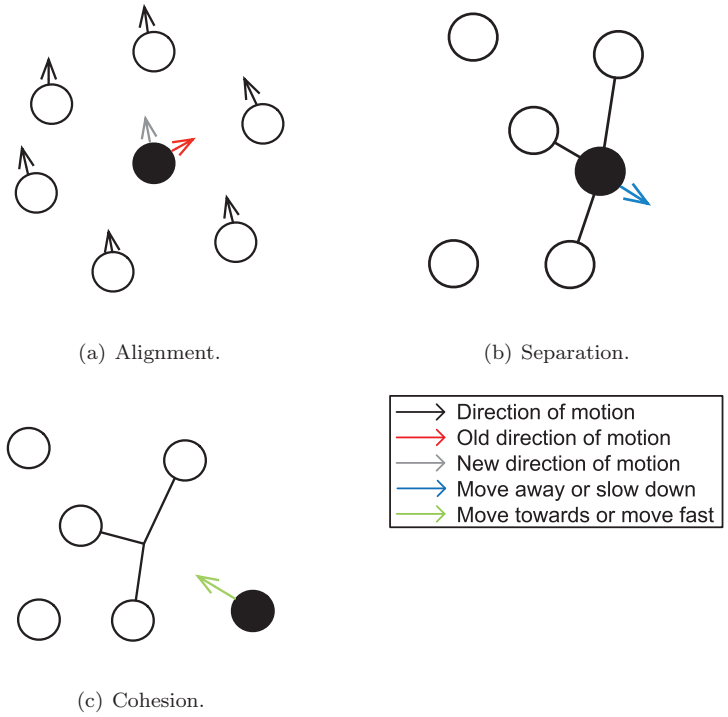
Nagpal *et al.* [75, 76] described a simple algorithm through which Lateral Inhibition is achieved with less message complexity and without the knowledge of the network. Consider $G(V, E)$ as defined above. A node $i \in V$ is said to be connected to another node $j \in V$ if the Euclidean distance between i and j is less than or equal to r where r is the transmission range of node i such that $r \in \mathbb{R}^+$. Suppose that all nodes have same transmission range r and all nodes are autonomous. At the start of the algorithm each node is assigned two different values. First, a same random value R_1 such that $R_1 \in \mathbb{R}^+$ and $R_1 > 0$ is assigned to all the nodes. Second, each node is assigned a random value R_2 such that $R_2 = [0, R_1)$. R_1 allows the algorithm to be run for R_1 time steps after which all nodes are either inhibited or uninhibited. The algorithm assumes discrete time steps. At each time step all nodes decrement R_1 by 1 and check if $R_2 = 0$ or not. If $R_2 \neq 0$, the node checks if an interrupt is received from the neighbors. If an interrupt is not received the node decrements R_2 by 1. However, if the interrupt is received the node exits the algorithm and becomes inhibited. However, if for a node R_2 becomes 0 before getting the interrupt, the node broadcasts the interrupt

signal to all its neighbors and becomes uninhibited. If an interrupt is received from two or more neighbors the inhibited node remains inhibited. Note that the inhibited node does not send any signal to other nodes in the network. The whole process runs for R_1 time steps. After running the algorithm, the nodes are either inhibited or uninhibited. The set of uninhibited nodes form the set S in the MIS. As an uninhibited node inhibits all its neighbor, the distance between two uninhibited nodes mainly lies between r to $2r$ hops where r is the transmission range of the nodes. However, the algorithm can be extended to h hops [76] where $h \in \mathbb{N}$. This will generate uninhibited node where the distance between the two uninhibited nodes mainly lies in $[hr, 2hr]$.

In another study, Afek *et al.* [74], developed an algorithm in which the selection of a node to join S depended on a probability. This probability is inversely related to the number of neighbors that are not connected to nodes in S. The aim of Afek *et al.* was to develop a MIS algorithm that was decentralized, had low communication complexity and was probabilistic. A low communication complexity is achieved using one bit message transfer between the neighbors. The message bit acts as an inhibition signal to the neighbors. Afek *et al.* developed Lateral Inhibition (MIS) algorithm using local information and two exchange mechanisms. The algorithm is executed for $\log D$ "phases" where D is the maximum node degree where each phase is again executed for $c \log n$ steps where n is the number of nodes and c is a constant. After each "phase" the probability of broadcasting the message is increased for those nodes that are not yet inhibited. After each phase, for a node, some of its neighbor nodes became inhibited. In each step two exchanges are done. First, all nodes broadcast one bit message to their neighbors with a probability and, second, all nodes check if a message from their neighbors is received or not. If no message is received, the node sends a one bit message telling its neighbors to inhibit and exist the algorithm. Thus, if the message is received, the node inhibits itself and exits the algorithm. The algorithm by Afek *et al.* differs from the algorithm developed by Nagpal *et al* in the following way. in the former case the nodes that are not yet inhibited join the set S in each discrete time slot with some probability whereas in the latter case a random value is sampled only once at the beginning and then is decremented deterministically. The runtime complexity of the algorithm is of the order $O(\log D \log n)$. The algorithm however, has a low message complexity as it is designed to send one bit messages over single hop.

Further, as Lateral Inhibition create regions around an uninhibited node, it is used to restrict

message complexity in the network. Moreover, in Section 3.1.3 we define our Lateral Inhibition algorithm based on Nagpal *et al* where the uninhibited nodes are separated by more than h hops in the network.



(a) Alignment.

(b) Separation.

(c) Cohesion.

Figure 2.5: Depiction of three Flocking rules.

2.5 Flocking

In nature, grouping is observed in many social living beings like birds, cattle, fishes and humans. A group of birds assembled together is called a flock. In nature it is observed that the flocks remain connected and move together. While moving, the birds exhibit a behavior which is called flocking. Flocking is thus a collective behavior of the birds that move together. Flocking was first modeled by Reynolds in order to simulate the collective behavior of the birds in motion [77]. In the model, each bird was considered as a *boid*. We call them as the nodes. In flocking, each node is considered to be autonomous and has three behavioral rules (a) Alignment (b) Separation and (c) Cohesion.

Definition 2.5.1. The Alignment rule states that if a node i is moving in different direction than that of the average direction of its neighbors, the node has to change its direction of motion and orient towards the average direction of motion of its neighbors to match that of its neighbors (cf. Fig. 2.5(a)). The Alignment rule thus consists of a decision whether to change the direction or not. As a consequence of this decision, the action for a node is to determine the new direction of motion and to move in that direction. Note that due to the definition of the alignment rule, it is independent of the magnitude of the velocity and the location of the node. Let N_i be the set of neighbors of node i. Let d_j be the direction of motion of a neighbor $j \in N_i$. The new direction of motion of i is defined as $d_i = \frac{\sum_{j \in N_i} d_j}{|N_i|}$.

Definition 2.5.2. The Separation rule states that a node should maintain a certain minimum distance with its neighbors (cf. Fig. 2.5(b)). Consider a node i and its neighbor j. Let D_{ij} be the Euclidean distance between two nodes such that $D_{ij} \in \mathbb{R}^+$. Let D_{min} be the minimum required separation distance such that $D_{min} \in \mathbb{R}^+$. Separation rule states that $D_{ij} \geq D_{min}$. If $D_{ij} < D_{min}$, then the node i should move away from node j either by slowing down or by moving in a different direction. Thus, Separation rule consists of a decision by a node whether to move apart. As a consequence of this decision, the action of the node is to either slow down or change the direction of motion. The separation rule ensure that there is no overcrowding at a given location and nodes do not collide. If D_{min} is small, the nodes will be overcrowded at a given location and there will be high number of collisions. However, if D_{min} is large the nodes will be too far apart. Among all the three Flocking rules, the Separation rule is of the highest priority [78].

The Separation rule is thus beneficial towards autonomous sensor deployment in the region where the minimum distance between two nodes should satisfy $D_{min} > 0$. Kadrovach *et al.* [79] used the Separation rule for efficient placement of nodes to maximize the coverage area.

Definition 2.5.3. The Cohesion rule states that if a node is far from its neighborhood it should move towards the centroid of the neighborhood to remain connected to all of its neighbors (cf. Fig. 2.5(c)). The cohesion rule thus consists of a decision whether to move towards the centroid of the neighborhood or not. As a consequence of this decision, the node should increase its velocity and should move in the direction of centroid. Let N_i be the set of neighbors of node i. Let (p_x^j, p_y^j, p_z^j) be the position of a neighbor $j \in N_i$. The new position i should move to is defined as

$(p_x^i = \frac{\sum_{j \in N_i} p_x^j}{|N_i|}, p_y^i = \frac{\sum_{j \in N_i} p_y^j}{|N_i|}, p_z^i = \frac{\sum_{j \in N_i} p_z^j}{|N_i|})$. Cohesion relates to the tendency of a node to remain as close to its neighbors as possible and do not stray. It also implies the group's strength to stay connected.

Note that the rules are applied at each instant. Further, recently, researches have revealed that flocking can be used to solve various problems in wireless networks. For example, Antoniou *et al.* [80] used flocking to provide efficient congestion control mechanism by computing the congestion at the neighbor nodes. In our case, we use analogy with Flocking rules in order to determine which nodes should beamform and in which direction the beams should be directed.

2.6 Non-Uniform Distribution of Nodes

Much of the research in wireless networks considers uniform node deployment. However, in reality, non-uniformity exists. Bettstetter *et al.* [81] proposed the use of thinning process to generate a non-uniform node deployment.

Consider a spatial graph $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let the nodes in V be uniformly distributed in a region of area A where $A \in \mathbb{R}^+$ and $A > 0$ (cf. Fig. 2.6). Thus the density of the nodes is $\rho = \frac{|V|}{A}$. An edge between $i \in V$ and $j \in V$ is created if j is within the Euclidean distance, $r \in \mathbb{R}^+$ of the node i. The algorithm prunes the nodes in the deployment using two parameters, r_b and N_{min} such that $r_b \in \mathbb{R}^+$ and $N_{min} \in \mathbb{Z}^+$.

In the algorithm, a new graph $G^{'}(V, E^{'})$ is generated such that every pair of nodes in V that are within a distance r_b are connected by an edge. Let N_i be the neighborhood of node $i \in V$ in the new graph $G^{'}$ such that $0 \leq |N_i| \leq |V|$ and $N_i \subset V$. A check on the number of neighbors is then performed on all the nodes. For all node $i \in V$ the algorithm checks if $|N_i| < N_{min}$. If the condition is true, then the node is marked as to be removed. Finally, when all the nodes have been checked for the condition, the nodes are then removed from the original graph G to obtain the final graph $G^*(V^*, E^*)$ (cf. Fig. 2.7). Note that the algorithm is only run once and the deployment achieved after the thinning process is used for non-uniform deployment. Also note that if $N_{min} = 0$, then the pruned graph will have no nodes left. However, if $N_{min} < \min\{|N_i|\}$ and $r_b = r$ then all nodes will remain in the pruned graph. The expected number of nodes remaining in the graph is

given by $\mathbf{E}(|V^*|) = |V|\left(1 - \frac{\Gamma(N_{min}, \rho r_b^2 \pi)}{(N_{min}-1)!}\right)$. Note that the probability of a node being removed is $p = \frac{\Gamma(N_{min}, \rho r_b^2 \pi)}{(N_{min}-1)!}$. However, this method faces with border effect, where nodes near the border are removed as the number of neighbors are less. This limitation of the algorithm is removed by using wrap around mechanism. However, we do not consider such case and consider thinning process with boarder effect to get the non uniform deployment.

Schilcher *et al.* [82], following the work of Bettstetter *et al.*, formulate the degree of non-uniformity of this pruned network. Consider a uniform distribution of a set of nodes V in an area $\mathcal{A}$. A grid of size $a \times a$ where $a \in \mathbb{N}$ is placed on the area. Let $\mathcal{A}_i$ be the area of the i^{th} square in the grid such that $\mathcal{A} = \bigcup_{i=1}^{a^2} \mathcal{A}_i$. Each subarea $\mathcal{A}_i$ has a set V_i of nodes associated with it. The expected number of nodes for all A_is is equal to $\frac{|V|}{a^2}$. A deviation from this quantity is measured as the non-uniformity in the deployment. Let V^* denote the set of nodes that are left after thinning. Each subarea $\mathcal{A}_i$ will now have V_i^* nodes associated with it. Non-uniformity is defined as

$$h_a = \frac{1}{2|V|} \sum_{i=1}^{a^2} absolute\left(|V_i^*| - \frac{|V|}{a^2}\right),$$

where $absolute(x)$ denotes the absolute value of x. Note that maximum non-uniformity is achieved when all nodes are located in one square in the grid. The normalization constant $\frac{1}{2|V|}$ is chosen such that $\lim_{a\to\infty} h_a = 1$.

Further, many other non-uniform deployment strategies also exist [83, 84, 85, 86, 87]. But in this dissertation, we use the thinning process proposed by Bettstetter *et al.* in [81] because of its simplicity.

2.7 Human Mobility

Recent analysis on the human mobility dataset reveals many different properties for human mobility. These properties include spatial properties like jump length and radius of gyration.

Definition 2.7.1. A jump length x such that $x \in \mathbb{R}^+$ is the distance moved by a person in a given time.

Definition 2.7.2. The radius of gyration, r_g, for a person is defined as the root mean square distance of the jump length from the mean jump length. Consider X to be the set of all jump

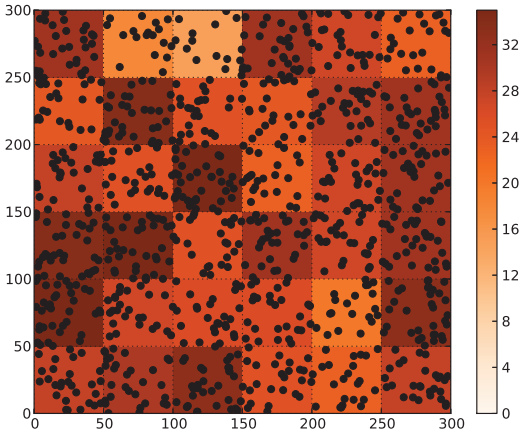


Figure 2.6: Uniform node deployment and the density of nodes per 50*50m^2 region. The density is shown by the color-bar.

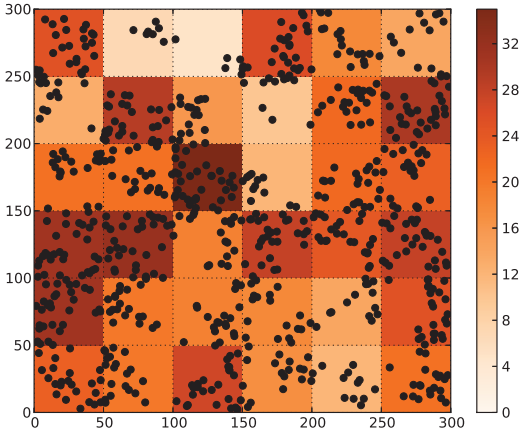


Figure 2.7: Deployment and the node density per 50*50m^2 region after applying the Thinning process on the deployment as shown by Fig. 2.6 with $r_b = 15$ and $N_{min} = 6$. The density is shown by the color-bar.

lengths for a person. The mean jump length is defined by $\overline{x} = \frac{1}{|X|}\sum_{i \in X} i$. Thus r_g is defined as

$$r_g = \sqrt{\frac{1}{|X|}\sum_{i \in X}(i - \overline{x})^2}.$$

Note that r_g will be different for different person.

Consider x as the jump length, x_{min} as the minimum jump length such that $x_{min} \in \mathbb{R}^+$ and $x_{min} > 0$, α as the power law exponent such that $\alpha \in \mathbb{R}^+$, β as a cutoff value such that $\beta \in \mathbb{R}^+$ and $\beta > 0$, $\lambda = \frac{1}{\beta}$ and $\Gamma(a, b)$ as the incomplete gamma function defined as $\Gamma(a, b) = \int_b^\infty t^{a-1}e^{-t}dt$. Using statistical analysis of mobility traces, Gonzalez *et al.* [6] observed that the distribution of x follows the truncated power distribution which is defined as [88]

$$p(x) = \begin{cases} \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda x_{min})} x^{-\alpha} e^{-\lambda x} & \text{if } x \geq x_{min}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.7.1}$$

Truncated power law distribution means that for small values of x the distribution follows a power law, for values of x sufficiently larger than β it follows an exponential distribution and this change in behaviour or cutoff occurs approximately at β. The exponential cutoff observed in the distribution of the jump length is mainly due to boundary effects and limited sampling of the mobility data [8]. In particular, [6] noted that human movement is limited within the radius of gyration r_g for 67% of the time. Human mobility has been closely related to the small world behavior. The few long jumps are mapped to long links while short jumps to regular links. This property of human mobility thus helps in the dissemination process. However, some researchers have argued that long jumps could cause disconnections and hence mobility could restrict the dissemination process. This mainly happens in low density scenarios.

For the human mobility dataset obtained for the region of the United States, the truncated power law distribution parameters for the jump length are $\alpha = 1.75 \pm 0.15$, $x_{min} = 1.5$km and a variable β [6]. The dataset used to obtain these parameters contains location coordinates of users recorded after each 2hrs. The dataset had low spatial resolution. In the dataset, the location coordinates are not the exact location coordinates of the person but they are the locations of the mobile towers they are associated with. If a user is in the region served by the mobile tower, the coordinates of the mobile tower is recorded in the dataset. The average distance between the mobile towers is found

to be 3km. It is revealed that the truncated power law distribution [6] is followed by r_g as well. The truncated power law distribution parameters for r_g are $\alpha = 1.65 \pm 0.15$, $x_{min} = 5.8$km and $\beta = 350$km. Note that the distribution of r_g is computed using r_g's for all the people in the dataset.

Recently, it is also noted that other than spatial properties many other properties are associated to the human mobility. These include temporal properties like the pause time [89].

Definition 2.7.3. The Pause Time is defined as the time a person stays at a given location.

For a similar dataset as used by [6], Song *et al.* in [89] revealed that the distribution of pause time also follows the truncated power law distribution. However, the difference in the datasets is the time interval (1hr) between the records. The truncated power law distribution parameters for the pause time are $\alpha = 0.8 \pm 0.1$, $x_{min} = 1$hr and $\beta = 17$hr. $x_{min} = 1$hr because the location of the person is recorded after each 1hr.

Apart from spatio-temporal properties, human mobility has also been associated to temporal behavioral characteristics in form of periodic movement, aperiodic movement and sporadic movement. Periodic movement corresponds to the case when a node traverses through the same sequence of locations periodically over time. In aperiodic movement the sequence of locations traversed before returning to a place may not be fixed whereas sparodic movement corresponds to the case when the sequence of locations traversed is never repeated. Further, human mobility also show spatial behavioral characteristics like centric, orbital, and random movement [90].

In another recent research [9], it is shown that the jump lengths for the country of Portugal follows a lognormal distribution thereby proving that the the distribution of the jump length depends on the region and boundaries associated to the region. Also note that human mobility pattern can be used to generate non-uniformity in the network. In an interaction network or a social network, application of mobility leads to links being intermittent thereby causing network structure to change over time. As explained before, characteristics of such networks can be studied using temporal snapshots of the network.

Mobility models that use truncated power law nature of spatio-temporal properties of human mobility include [91, 92, 93]. Consider a discrete time mobility model. Consider a person. We call the person as a node. Consider a sufficiently large square area with area A such that $A \in \mathbb{R}^+$. Within the bounds of the area, the mobility of the node is defined by the tuple (x,θ,t_p) where x is

the distance moved or the jump length such that $0 < x \leq \sqrt{A}$, θ is the direction of the motion such that $\theta \in [0, 2\pi]$ and t_p is the pause time such that $t_p \geq 0$. At the beginning of each step a node computes the values for the tuple. x and t_p are calculated from the truncated power law distribution while θ is uniformly chosen from $[0, 2\pi]$. Let each jump be completed in 1 time step. Note that for extracting a sample, x, from a truncated power law distribution, the sample is first extracted from an exponential distribution and then the sample is accepted with a probability $p = (\frac{x}{x_{min}})^{-\alpha}$ [88]. If $t_p > 0$, it means that the node stays at the current location for t_p time steps before it moves to some other location. During this period, the jump length x and the direction θ are not sampled [93]. $t_p = 0$ means that the node will change its position in the next time step.

For our mobility model we assume that $t_p = 0$. Further, we assume that $x_{min} = 1$ and $\beta < \sqrt{A}$. Note that when a node is at the boundary, choosing θ can cause the node to move out of the boundary of the given area. If such a situation happens θ is again computed until the jump is within the bounds of the area. Also note that if the jump length itself is large enough that it cannot be within the area bounds then the jump length is again sampled from the truncated power law distribution.

2.8 Stability Measures

Stability is complementary to the change occurring in the system. It is the measure that determines how much stable a system is in a dynamic environment. Stability can be calculated by determining the percentage change in the system. Stability is defined at different scales, for example, in a temporal graph, stability is defined for the graph, for the nodes and for the links in the graph. A temporal graph is characterized by a group of discrete snapshots of a graph each taken after Δt time step [63]. The stability measures are thus calculated using different snapshots.

For the stability of the graph, in [94, 95] the authors use the distance between the Laplacian spectrum defined by the eigen values of the Laplacian matrix of the snapshots. In other studies, [36, 37, 96], for each snapshot, the authors ranked the nodes based on their degrees to compute the stability of the graph. The top $x\%$ of the nodes are then selected from each snapshot and common nodes occurring in the selection are found. The number of common nodes provided the stability of the graph.

In [97,98] the authors use the adjacency matrix of the snapshots to define the stability of a link. In [97] the authors proposed that a link is stable if its state did not change in the considered time frame while [98] argued that the links that maintained their status of 0 should not be considered. Further, link stability can also be inferred from the entropy of the link [99]. Other ways of measuring stability includes the use of similarity measures. A high similarity means that two systems are similar and change that has occurred in the system over time is small. A survey of similarity measures is provided in [100].

We now provide a brief overview of some stability measures.

2.8.1 Similarity

Similarity is a well defined concept in data mining context. Representing a system as a vector, many similarity measures have been defined to determine the similarity between two vectors. These methods include Euclidian Distance, Cosine Similarity, Jaccard Coefficient, Pearson correlation, etc [100]. A high similarity means that two systems are similar and change that has occurred in the system over time is small. Consider a graph, $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let N_i^t be the neighborhood of node $i \in V$ at time t such that $0 \leq |N_i^t| \leq |V|$ and $N_i^t \subset V$. Cosine similarity of a node i in two instances of the graph, G^t and $G^{t^{'}}$ at time t and $t^{'}$, respectively, is defined as $\frac{|N_i^t \bigcap N_i^{t^{'}}|}{\sqrt{|N_i^t| \times |N_i^{t^{'}}|}}$ where $N_i^{t^{'}}$ is the neighborhood set for node i at time $t^{'}$.

2.8.2 Spectral Distance

Changes in the graphical structure is also computed using the difference in their Laplacian spectrum [94,95]. Laplacian spectrum is defined through the eigen values of the Laplacian matrix.

Definition 2.8.1. Let A be the adjacency matrix of the graph $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$, $|E| \in \mathbb{Z}^+$ and dimension of $A = |V| \times |V|$. Let $(A)_{ij}$ denote the element in the i^{th} row and j^{th} column such that $(A)_{ij} \in \{0, 1\}$ and $(A)_{ii} = 0$ for all i. Let D be the degree matrix such that $size(D) = |V|\text{x}|V|$, $(D)_{ii} = \sum_{j \in V} (A)_{ij}$ and $\forall j \neq i, (D)_{ij} = 0$. The Laplacian matrix L of a graph is defined as $L = D - A$.

The distance between the eigen values of the Laplacian matrices corresponding to time t and t' is defined as the spectral distance between the two graphs corresponding to time t and t'. Let λ_i be the i^{th} eigen value of the Laplacian matrix of the graph at time t while μ_i be the i^{th} eigen value of the Laplacian matrix of the graph at time $t^{'}$. The spectral distance is defined as

$$sd = \begin{cases} \sqrt{\frac{\sum_{i=1}^{|V|}(\lambda_i-\mu_i)^2}{\sum_{i=1}^{|V|}(\lambda_i)^2}} & \text{if } \sum_{i=1}^{|V|}(\lambda_i)^2 \geq \sum_{i=1}^{|V|}(\mu_i)^2, \\ \sqrt{\frac{\sum_{i=1}^{|V|}(\lambda_i-\mu_i)^2}{\sum_{i=1}^{|V|}(\mu_i)^2}} & \text{otherwise.} \end{cases} \tag{2.8.1}$$

For a system where no change has occurred, the eigen values of the two Laplacian matrices will be identical. This will result in $sd = 0$.

2.8.3 Rank Overlap

Stability can also be measured by computing rank overlap as suggested in [36,37]. For email network, Braha *et al.* [36, 37] suggested to compute the degree of all nodes in the network at a given time instant. Once the degrees are computed, the nodes are ranked based on their degree. As the email networks are large, top $x = 1000$ nodes are selected from the ranking. This process is done for different time instants and respective lists are made. Commonalities between two selections are then found in the form of same nodes appearing in the two lists. Let G^t_{Top} and $G^{t^{'}}_{Top}$ be the lists of all top x nodes at time t and $t^{'}$, respectively. Braha *et al.* in [96] defined the Rank Overlap as

$$S = \frac{G^t_{Top} \bigcap G^{t^{'}}_{Top}}{x}. \tag{2.8.2}$$

2.8.4 Adjacency Matrix Based Measures

Stability of a link can be measured using changes in the adjacency matrix [97,98]. Consider a graph, $G(V, E)$ with set of nodes V and set of edges E such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let A^t be the adjacency matrix of the graph at time t and $A^{t^{'}}$ be the adjacency matrix of the graph at time $t^{'}$. Let $(A^t)_{ij}$ denote the element in the i^{th} row and j^{th} column such that $(A^t)_{ij} \in \{0, 1\}$. Hanneke *et al.* proposed that a link is stable if it maintains its state in both t and $t^{'}$. However, a link is not

stable if it changed its state from 1 to 0 or vice versa. In terms of adjacency matrix, it is represented as, $(A^t)_{ij}(A^{t'})_{ij} + (1-(A^t)_{ij})(1-(A^{t'})_{ij})$.

In another related study, Tang *et al.* [98] argued that the link with status 0 in both the instances should not be considered for the stability as the link did not exist in any of the considered time instances. That is, $(1-(A^t)_{ij})(1-(A^{t'})_{ij})$ term should not be considered.

2.8.5 Entropy

Zayani *et al.* [99] showed that stability of a link can be computed using entropy measure. Consider a graph $G^t(V, E^t)$ with set of nodes V and set of edges E^t at time instance t such that $|V| \in \mathbb{N}$ and $|E| \in \mathbb{Z}^+$. Let E^t_{ij} be the state of an edge between two nodes $i \in V$ and $j \in V$ at time t such that $E^t_{ij} \in \{0, 1\}$. Let Λ be the set of all the states of the edge between i and j over time T. Thus $\Lambda = \{E^1_{ij}, \cdots, E^T_{ij}\}$. [99] calls Λ as a string of length T consisting of either 0 or 1. Note that $|\Lambda| = T$. Zayani *et al.* first extracted unique substrings from Λ. Let Λ_x denote the x^{th} element in Λ. For extraction of the set of unique substrings, W, the algorithm they applied is given in Algorithm 1.

Algorithm 1 W finding

```
Let W ← ∅, Y ← {Λ_1}
for x from 1 to T do
  if Y ∈ W then
    if Λ_{x+1} ≠ ∅ then
      Y ← {Y, Λ_{x+1}}
    end if
  else
    W ← W ⋃ Y
    Y ← ∅
    if {Λ_{x+1}} ≠ ∅ then
      Y ← {Λ_{x+1}}
    end if
  end if
end for
```

For example, consider $\Lambda = \{0,1,0,0,0,1,1,0,1,1,0,0,0,0,0,1,0,1,0,0\}$. Using Algorithm 1, $W = \{\{0\},\{1\},\{0,0\},\{0,1\},\{1,0\},\{1,1\},\{0,0,0\},\{0,0,1\},\{0,1,0\}\}$. This will result in $|W| = 9$ where W consists of unique substrings extracted from Λ. Zayani *et al.* [99] then defined entropy

Ent_{ij} of a link between nodes i and j as

$$Ent_{ij} = \frac{|W| * \ln(|W|)}{T}. \tag{2.8.3}$$

If the value of Ent_{ij} is low the link is said to be stable however if not then it is said to unstable. Note that a link, E_{ij} will have highest entropy if $\Lambda = \{0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, \cdots\}$. A link, E_{ij} will have low entropy if $\Lambda = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \cdots\}$. However, this method requires huge amount of memory as there is a need to store the history of the link. If the stability of each link in G is to be computed this method will require memory greater than $O(|V|^2 T)$ as there can be $\frac{|V|(|V|-1)}{2}$ links in the graph. A similar entropy method is used in [101] where location preferences are used instead of links to predict the human mobility.

2.8.6 Neighborhood Based

Let $G^t(V, E^t)$ be a graph with set of nodes V and set of edges E^t at time t such that $|V| \in \mathbb{N}$ and $|E^t| \in \mathbb{Z}^+$. Given two snapshots of this graph at t and t' a way to compute the stability of a node, $i \in V$, is through finding the intersection between the neighborhood nodes at time t and $t^{'}$. Let N_i^t be the neighborhood of $i \in V$ at time t and $N_i^{t^{'}}$ be the neighborhood of $i \in V$ at time t' such that $N_i^t \subset V$ and $N_i^{t^{'}} \subset V$. The stability of node is defined as

$$S_i = \begin{cases} \frac{|N_i^t \bigcap N_i^{t^{'}}|}{|N_i^t|} & \text{if } N_i^t \neq \emptyset, \\ \text{undefined} & \text{if } N_i^t = \emptyset. \end{cases} \tag{2.8.4}$$

The numerator of the case when $N_i^t \neq \emptyset$ in the eq. 2.8.4 refers to common links that existed both at t and $t^{'}$. Let E_{ij}^t be the state of an edge from $i \in V$ to $j \in V - \{i\}$ at time t and $E_{ij}^{t^{'}}$ be the state of an edge from $i \in V$ to $j \in V - \{i\}$ at time t'. In terms of edges the numerator is expressed as $\sum_{\forall j \in N_i^t \bigcup N_i^{t^{'}}} E_{ij}^t \wedge E_{ij}^{t^{'}}$ where $\wedge$ denotes the logical AND operator.

Note that for $N_i^t = \emptyset$, S_i is not defined. If $N_i^{t^{'}} = \emptyset$ and $N_i^t \neq \emptyset$ then stability is 0.

A similar method for computing node stability to that of eq. 2.8.4 is proposed in [102, 103]. However, the computation of the stability measure using the eq. 2.8.4, requires the knowledge of two snapshots of graphs in time, t and $t^{'}$. This means, the node is first required to find its neighborhood at time $t^{'}$ and only then the computation of the stability can be performed. To reduce

the time required for the algorithm to compute stability, it is beneficial if a node is able to predict the links before. This forms the basis of our proposed stability estimation method in Chapter 4.

2.9 Information Dissemination

Spreading models across population have always been in focus since its introduction in 1766 and various models have been proposed. Models related to spreading in the network have been well studied. Most of these models are primarily inspired from epidemic spreading within the population that is divided into communities where the interaction between the communities is based on mobility [11, 104]. Epidemic spreading models are often studied in terms of reach in the population where reach is defined as the number of people under the influence of the epidemic. A node in the population is considered to be in either of the following states, *S*usceptible, *I*nfected, *R*ecovered and *E*xposed. A node is said to be in state S if it is not infected by the disease but is likely to gain the disease in future. A node is said to be in state I when the node has the disease. State R corresponds to the recovered state of the node, meaning that a node was infected before but has now gained immunity, while the state E means that a node is exposed to the disease and is highly likely to gain infection. The transition from one state to another occurs with a defined rate.

One of the main studies focusing on the spreading in the network is that of Watts and Strogatz's [2] where they showed that, in static network, rewiring few links can reduce the APL of the network. The smaller the APL, the smaller the number of hops between two nodes. This relates to the fact that less time is taken to spread the information in the network in terms of hops. We define the spreading of information as the information dissemination process in the network.

Information dissemination is not limited to static networks. It is also studied in dynamic networks. In dynamic scenarios, dynamicity affects the dissemination of the information in the network [7]. It is thus critical, while modeling dissemination, to know what application is targeted. For example, the spread of a warning signal in the community should be very fast. However, the dissemination of viruses should be very slow. Dissemination is also affected by many other factors like, strength of the ties [105], network structure [106], activity pattern [107], content type [108], node characteristics [109], altruism [110], etc.

In Communication networks such as Delay Tolerant Networks, information dissemination has

been closely related to epidemics across the population where epidemic models like Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered (SIR) have been considered [111, 112, 113, 114]. Information dissemination, however, has also been studied using rumor spreading or the gossip based algorithms [115, 116].

We now describe our methods for enhancing information dissemination in the network. In Chapter 3, we use the concepts like small world, beamforming, Lateral Inhibition, analogy to flocking rules and non-uniform distribution primarily to build our algorithm. In Chapter 4, however, we use beamforming, human mobility model and stability concept. In Chapter 5, we use data from mobile Call Detail Records (CDR) to extract human mobility patterns. We also use beamforming to build our model. Note that all Chapters use beamforming and information dissemination concept.

CHAPTER 3

Achieving Small World Properties in Wireless Network using Bio Inspired Techniques

Decades of academic and industrial research in wireless networks [117] has led to a tremendous growth of wireless networks requiring researchers to address manageability and scalability issues. Due to these issues, most of the research work has been oriented towards autonomous wireless networks. The autonomous behavior of the wireless nodes made decentralized computing and cost efficient topology deployment possible [118]. It was also proved that self-organization of the network can lead to better performance.

An attractive model to achieve better network performance is the Small World network. Small world networks are characterized by reduced APL and maintained CC. Drawing inspiration from the experimental work of Stanley Milgram [38], Watts *et al.* [2] proposed a model that could achieve small world properties. In the model, Watts *et al.* proposed, small world properties could be reached by randomly rewiring a few existing links within the network. Watts *et al.* showed that the dynamics of these small world networks lie between that of a regular network and a random network [2, 119]. To prove the findings, however, Watts *et al.* used a regular wired network and called the rewired links as shortcuts. Many complex real world networks such as internet, biological networks, food web and social networks also demonstrate small world properties [41, 42, 49]. Section 2.2 provides more details on small world networks.

However, Watts' model cannot be applied directly to wireless *ad hoc* networks because of the spatial nature of such networks. In wireless *ad hoc* networks, addition of a shortcut between any two nodes should depend on the distance between two nodes. Helmy in [48] first studied the effect

of adding few distance-limited links in the network. He showed that, upon introduction of distance-limited links, wireless *ad hoc* networks show small world properties. He concluded that, when the shortcut lengths are $\frac{1}{4}$th of the network diameter, there is a maximum reduction in the APL. Thus, proving that realization of small world properties in a wireless *ad hoc* network depends crucially on the length of the shortcuts created. Another important factor in the realization of small world properties is the choice of nodes among which shortcuts are to be created. One method to obtain these nodes is that of *preferential attachment* [42, 120], typically observed in real world networks, wherein links are created to nodes with high structural importance. It was shown that, analogous to real world networks, using *preferential attachment* for creation of distance-limited links in a spatial network resulted in reduced network diameter [121, 122]. This is accompanied by high clustering coefficient and a shift in the node degree distribution towards power law. These results motivate us to say that, creation of links to nodes having high structural importance in the network can result in the desired small world characteristics.

The creation of a wireless *ad hoc* network with the small world properties also depends on the manner in which distance-limited shortcuts are added. Such shortcuts can be added through different techniques like: (i) creating a directional beam using the same power as when the node is operating in the omnidirectional mode; (ii) increasing the omnidirectional transmission range of the node; (iii) introducing few long wired links [123]; (iv) introducing special nodes with higher omnidirectional transmission range deterministically in the network [124]; (v) using another antenna for beamforming in addition to the omnidirectional antenna.

Talking about the self-organization characteristics, [5], of the nodes, only techniques (i) and (ii) mentioned above qualify. However, even though other techniques help in achieving desired network characteristics, they lack self-organization capabilities. In addition, the second technique suffers from the problem of early death of the node due to increased energy consumption. Thus leaving only the first technique. Achieving reorganization or rewiring in a wireless *ad hoc* network through the first technique is hard due to the spatial nature of the wireless *ad hoc* network. Finding the beam direction, the beam length and determining the new neighborhood are the primary issues associated with rewiring in a wireless *ad hoc* network.

Most of the literature in this direction has been limited to the introduction of an external infras-

tructure and considers global knowledge of the network. Acquiring such global information incurs additional overheads, thereby impacting the network performance. A related study, [1], proves that the use of distance-limited long links in wireless *ad hoc* network to achieve small world properties is beneficial, (cf. Fig. 3.1). Let $APL(p)$ and $CC(p)$ be the APL and the CC of the network when $p\%$ of the nodes create long-range links. $p = 0$ means no node is beamforming. Fig. 3.1 shows a reduction in the APL while almost no change in the CC for the case when a realistic antenna model (ULA antenna model) is used. On the other hand, for the theoretical model (sector antenna model[1]), the reduction in the APL is relatively less while the reduction in CC is considerably large. The number of nodes that beamform is shown with a help of probability value in the log scale. The results also show that the reduction in the APL increases with an increase in the number of beamforming nodes, [2, 119]. In fig. 3.1, we normalize $APL(p)$ with $APL(0)$ and $CC(p)$ with $CC(0)$ to account for the variation in the APL and the CC. In [1] it is also shown that fraction of asymmetric paths increases with the increase in the probability of rewiring for the ULA and the sector model (cf. Fig. 3.2).

Definition 3.0.1. Consider two nodes i and j, an asymmetric path is a path when i is able to communicate with j while j is not able to communicate with i.

It is also noted that as the diameter of the network, D, increases, when beamforming is used using ULA and sector model, the APL grows as $O(\log D)$ (cf. Fig. 3.3).

Table 2.1 illustrats the comparisons between various studies performed with the objective of increasing connectivity using beamforming. We now formulate table 3.1 to illustrate comparisons and bring out the differences between various approaches in literature that focus on achieving small world characteristics in wireless networks and the proposed algorithm in this Chapter. There are various additional approaches other than those mentioned in the table, but they are not considered since they deal with external infrastructure of at least two radios.

In this Chapter, we investigate how increase in connectivity, reduction in the APL and almost retention of the CC in a non-uniformly distributed wireless *ad hoc* network using only locally available information and beamforming can be achieved. Through the algorithm, we proposes to achieve the goals by creating long-range directional beams between nodes that have low and high structural

[1]Sector model approximates realistic antenna models

Parameter\ Reference	**Proposed Algorithm**	**Abhik *et al.* [1]**	**Guidoni *et al.* [124]**	**Helmy *et al.* [48]**	**Sharma *et al.* [123]**	**Verma *et al.* [125]**
Shortcut Creation	Rewiring	Rewiring	Addition	Addition	Addition	Addition
Node distribution	Non Uniform	Uniform	Uniform	Uniform	Uniform	Uniform
External infrastructure	No	No	High range Sensor	-	Wired	Two radios for each node
Global knowledge	No	No	Yes	Yes	Yes	Yes
Density of nodes	Low	High	High	High	-	Low
Shortcut Edge	Directed	Directed	Undirected	Undirected	Undirected	Directed
Shortcut direction	Towards centroid of other region	Longest Traffic Flow path	Random, towards sink	Random	Random	Random
Shortcut length	Function of antenna elements	Function of node density	Constant	Limited	Constant	Constant
Shortcut width	Depends on Shortcut Length	Depends on Shortcut Length	Constant	-	-	Constant
Probability of Shortcut creation	$(0, 1]$ based on model parameters	Function of β	$\in (0, 1]$	$\in (0, 1]$	function of network size	$\in (0, 1]$
Performance metric	APL, CC, Connectivity	APL, Connectivity	APL, CC	APL, CC	APL, Energy	APL, CC, Degree

Table 3.1: Comparison between various researches related to achieving Small World characteristics in wireless networks. More details on Abhik *et al.* algorithm [1] are provided in appendix A

.

importance. The decentralized computing and self-organizing requirements of such an approach motivates us to draw inspirations from nature and use natural phenomenon like Lateral Inhibition [73, 75, 76, 74] and Flocking [77]. In conjunction with these natural techniques, the concept of centrality also provides valuable insights in building a solution to the current problem.

We use Lateral Inhibition to create small logical regions within a static network. The use of Lateral Inhibition not only reduces the message complexity but also enables us to apply the Flocking rule analogy for beamforming successfully. We use the analogy of Flocking rules to identify the nodes that beamform and the beam properties. According to the rules, explained in Section 2.5 of Chapter 2, it is important to identify stray nodes, align the nodes and move them towards the centroid of

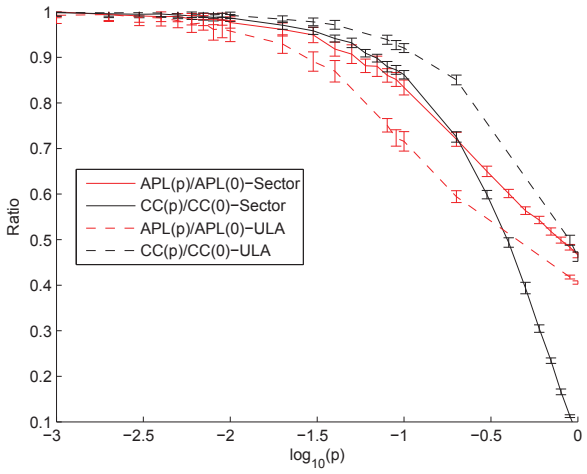


Figure 3.1: Effect of beamforming on the APL and the CC when the nodes are using different antenna models.

their neighborhood. Analogous to this, after region formation in a non-uniformly distributed wireless *ad hoc* network, we use Flocking rules to identify the beamforming nodes and direct the beams of these beamforming nodes towards the centroid of the region. The centroid node in the region has a high structural importance. Beamforming towards the centroid node of the region contributes towards reducing the APL because the centroid node of the region is the most connected node and has the highest Closeness Centrality measure. Thus, beamforming towards the centroid node is the *preferential attachment* behavior of the beamforming node, thereby making centroid finding a prerequisite to Flocking.

Thus, we design the algorithm such that it first identifies regions using Lateral Inhibition, then identifies the centroid nodes of the regions and then uses the analogy of Flocking rules to identify the nodes that will beamform along with their beam properties.

The rest of the Chapter is organized as follows. Section 3.1 presents the assumptions used for the proposed algorithm along with the algorithm specifications. Section 3.2 presents the formal definitions. Sections 3.3 and 3.4 discuss the simulation setup and the results respectively while Section 3.5 provides concluding remarks. Further, table 3.2 lists the notations used in this Chapter.

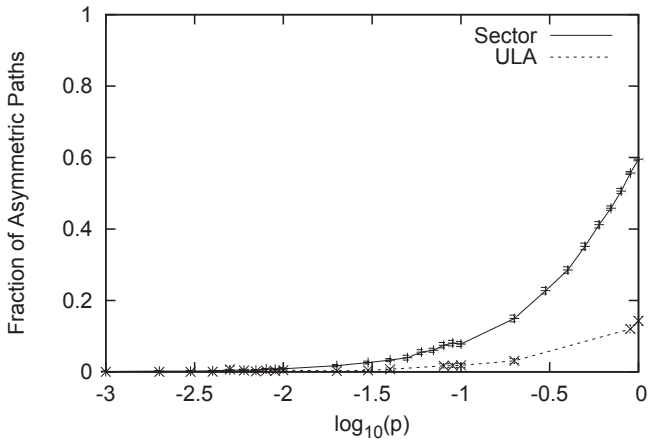


Figure 3.2: Asymmetric paths as a function of varying probability of rewiring.

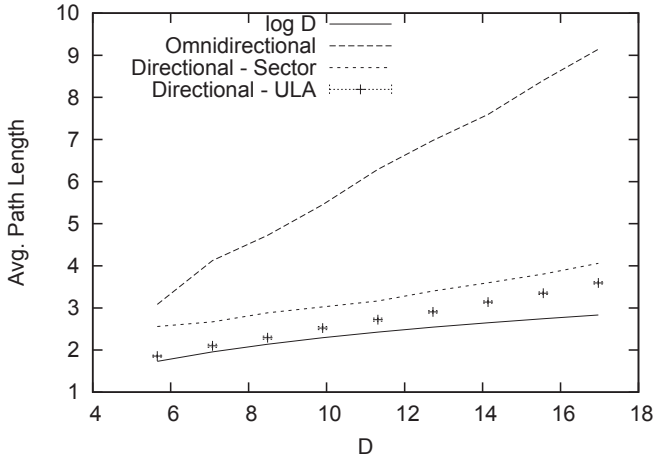


Figure 3.3: Growth of APL with increase in the network diameter.

Notation	Meaning
A	simulation area
B_b	boresight direction
B_w	beam width
B_l	beam length
C	set of all centroid nodes
c_i	centroid node of the region $G_i \vert c_i \in C$
D	diameter of the network
D_{jk}	distance between two nodes
D_{min}	minimum separation distance
$deg(i)$	degree of node i
EBC_i	Egocentric Betweenness of i w.r.t. its cluster
$G(V, E)$	network with set of vertices V and set of edges E
g	hop limited desired region size
g_{max}	maximum g
$G_i(V_i, E_i)$	region $G_i \vert G_i \subset G$ with set of vertices V_i and set of edges E_i
H	set of all region heads
h_i	head node of the region $G_i \vert h_i \in H$
$hopcount_{ji}$	hopcount between node j and i
ID_i	identification number of node i
M	max antenna elements available with v
m	number of antenna elements used by v to beamform $\vert m \in [2, M]$
N_i	set of neighbor of $i \vert i \in V$
$N_{j,i}$	set of neighbor of j in the region $G_i \vert j \in V_i$
N_{min}	minimum number of neighbors used for creating a non-uniform distribution
N_r	number of regions formed
P	set of all peripheral nodes
P_i	set of peripheral nodes in the region $G_i \vert P_i \in P$
p_t	transmission power
p	probability of rewireing
RC_i	set of centroid nodes reachable from i with their *hopcount* that are within g_{max} hops from i when i is not beamforming
RC_i^*	set of centroid nodes reachable from i with their *hopcount* when i is beamforming
r	transmission radius

Continued on Next Page...

Table 3.2 – Continued	
Notation	**Meaning**
r_b	Bettstetter transmission radius
S	standard node set
T	time
(x_j, y_j)	coordinates of j in the region G_i
(x_j^*, y_j^*)	updated coordinates of v in the region G_i
ε	error margin for identifying centroid node
α	path loss exponent
ρ	average node density
$\varrho_{\wp_i}$	peripheral neighbor of $\wp_i \vert \varrho_{\wp_i} \in P_i$
$\wp_i$	peripheral node $\vert \wp_i \in P_i$

Table 3.2: Notations and their meaning.

3.1 Assumptions and Algorithm

3.1.1 Assumptions

To address issues mentioned above, we focus towards the deployment of homogenous and autonomous wireless *ad hoc* nodes with no central entity controlling the nodes. This type of deployment enables us to easily apply self-organizing features, achieve global consensus with very limited local information, make any eligible node the group leader, make the system highly fault tolerant, ease the topological maintenance, lower the deployment cost and extend to incorporate the mobility of the nodes in the future. Let each nodes have an omnidirectional transmission range r such that $r \in \mathbb{R}^+$. We assume a non-uniform distribution of the nodes generated using thinning process defined by Bettstetter *et al.* [81] (cf. Section 2.6) with parameters N_{min} and r_b such that $N_{min} \in \mathbb{Z}^+$ and $r_b \in \mathbb{R}^+$ in order to mimic realistic scenario. We assume the deployment of the nodes on a finite size 2-D plane of an area A.

As part of the network setup, we assume that each node has an antenna array consisting of M isotropic elements such that $M \in \mathbb{N}$. Individually each AE has an omnidirectional radiation pattern.

The use of single antenna element, AE, results into omnidirectional beam while use of more than one AE results into a long-range directional beam. A node, however, decides to use more than one AE using simple local rules mentioned later in this section. A node uses beamforming only to transmit data but uses omnidirectional beams for reception. Consider the total transmission power available to nodes as p_t such that $p_t \in \mathbb{R}^+$. When a node decides to beamform using $m \in [2, M]$ AEs out of M available AEs, each AE transmit with power $\frac{p_t}{m}$ [65]. Note that when one AE is used, the AE transmits with full power p_t and an omnidirectional beam is achieved. We use the theoretical antenna model, Sector model [69], and assume transmission of data to be synchronous. As our objective is to demonstrate the improvements that is achieved in the network performance when beamforming is used over that of omnidirectional antenna, we assume a free space propagation model with path loss exponent $\alpha = 2$.

Each node operates using local information, i.e., information available with one hop neighbors and itself. Determining single hop neighborhood to build the local information is thus essential for the correct operation of the algorithm. As our aim is to study the effect on the APL and CC, we assume that the nodes have information about their neighborhood.

It is also essential to address the self-organizing paradigms, [5], to claim the self-organizing behavior of the network. The paradigms state: designing local rules to achieve global properties, implicit coordination, minimizing the use of historic information about the state of the network and designing an algorithm that changes with environment parameters. All nodes operate using only locally available information to determine whether they should beamform or not, decide beam properties and form regions. The nodes implicitly coordinate with their neighbors to determine the node with the highest *hopcount* from the centroid of the region. For a given region, the nodes also coordinate implicitly to determine the centroid node of that region.

In the following sections, we describe the system model and the algorithm.

3.1.2 System Model

Given a network, $G(V, E)$, where V is the set of vertices such that $|V| \in \mathbb{N}$ and E is the set of edges such that $|E| \in \mathbb{Z}^+$, we visualize G as a network consisting of N_r logical regions, $\{G_1, G_2, \ldots, G_{N_r}\}$,

i.e., $G = \bigcup_{i=1}^{N_r} G_i$ such that $1 \leq N_r \leq |V|$. Each region, G_i, consists of the set of nodes, V_i such that $V_i \subset V$ where $V = \bigcup_{i=1}^{N_r} V_i$, and set of edges, E_i such that $E_i \subset E$ where $E = \bigcup_{i=1}^{N_r} E_i$. Let all nodes in V have (x, y) coordinates and a unique ID associated to them.

The set of vertices, V, is classified into three sets. These are termed as the peripheral node set (P), the centroid node set (C) and the standard node set (S). Each set is provided with separate role. P contains the nodes that beamform. C contains the nodes towards which the nodes in the Peripheral node set beamform. The nodes in P and C are called as the peripheral nodes and the centroid nodes respectively. The nodes that are not in P or in C are said to be in set S. Note that $V = P + C + S$.

Let $hopcount_{ij}$ be the distance in hops between nodes i and j, the closeness centrality of a node, $i \in V$, in a graph G is equal to $\frac{1}{\sum_{j \in V} hopcount_{ij}}$. The node having maximum Closeness Centrality is the centroid of the graph and has a high structural importance. For the vertex sets defined above, nodes in the set P have lowest value of Closeness Centrality, i.e., $\arg\max_{i \in V} \left\{ \sum_{j \in V - \{i\}} hopcount_{ij} \right\}$. However, the nodes in the set C have highest value of closeness centrality, i.e., $\arg\min_{i \in V} \left\{ \sum_{j \in V - \{i\}} hopcount_{ij} \right\}$. A node in P beamforms towards a node in C in order to minimize the distance to other nodes and reduce APL.

We model directional beam using Sector model. A sector model is a simplified antenna model well suited for analysis purposes. In sector model, a beam is characterized by an area of a sector bounded between two radii of length B_l each such that $B_l \in \mathbb{R}^+$ and an arc connecting the two radii. Note that the two radii have one common starting point which is also the location of the node. The length of the arc is obtained by the angle between the two radii. We call the angle as the width of the beam B_w. Let the total transmission power available to the node be p_t. As discussed above, when a node decides to beamform using m AEs out of M available AEs, each AE transmits with power $\frac{p_t}{m}$ [65]. Note that when only one AE is used, the AE transmits with full power p_t. We assume that in sector model, $B_l = r\sqrt{m}$. Further, as the overall p_t is same, we assume the area under the omnidirectional beam and the sector are equal. This results into a sector where the B_w

is defined as

$$B_w = \frac{2\pi r^2}{B_l^2}. \tag{3.1.1}$$

3.1.3 Algorithm

The algorithm presented here is divided into two parts:

A) Use of Lateral Inhibition technique and averaging scheme for the identification of regions and the centroid nodes of the regions, so that there are less message overheads and nodes can beamform towards the centroid node to achieve reduced APL.

B) Use of Flocking rules to identify the nodes that beamform, to determine beam properties that realize small world properties and improve connectivity.

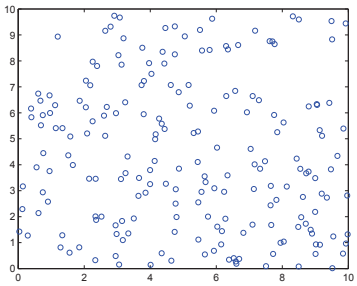
We describe these parts in detail in the next subsections.

Region formation

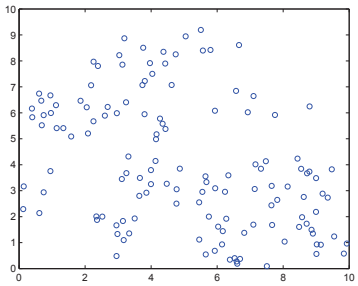
The Closeness Centrality [50, 51] identifies the structural importance of the node in the network. The node with the highest Closeness Centrality value is the most central node in the network. Through this node, the spread of the information to other nodes is fast. To determine the Closeness Centrality of the node, the node requires the knowledge of other nodes in the region as suggested by the definition of Closeness Centrality, Section 2.2.1. This makes the Closeness Centrality a global measure. Storing information about all the nodes in the network can consume a lot of node's memory. When there is a lack of global information, gathering such information can also be time consuming and the message complexity could be high. To overcome these problems, we create small logical regions. The creation of regions not only reduces the message complexity of the network but also reduces the effect on the APL due to the failure of a node, thereby making the network more manageable, efficient and tolerant to failures [126]. Some algorithms designed in this direction are centralized where the Base Station chooses the region heads based on the energy and the position of the nodes. Other techniques use either the transmission power of the node or the degree of the node [127]. On the contrary to centralized approaches, some algorithms are either distributed, [128], or probabilistic [129].

Let us consider a node $i \in V$. As nodes lack global information, computing global centrality measure and using it will not be possible. Thus we use degree of a node as to build our algorithm. For Lateral Inhibition, a node broadcasts and stores a message containing the following information: the identity of the head node to which i is associated (h_i), its *hopcount* from h_i and the degree of h_i ($deg(h_i) = |N_{h_i}|$, where N_{h_i} is the set of neighborhood nodes of h_i such that $|N_{h_i}| \geq 0$). Let H be the set of all head nodes. Initially, all the nodes consider themselves as heads, i.e. $H = V$, and store their own information, i.e., $h_i = i$, $hopcount = 0$ and $deg(h_i) = |N_i|$. Each node, $i \in V$, then broadcasts this information to its neighbors, N_i. Similarly, i receives information from each of its neighbors and subsequently updates the information stored in it. A node updates its stored values, if the stored degree, $deg(h_i)$, is less than that of the received value and $hopcount + 1$ is less than g, where g is the hop limited desired region size such that $g \in \mathbb{Z}^+$. If the stored and the received $deg(h_i)$ are same, the node decides to update the stored information based on lower *hopcount* value. If the stored and the received *hopcount* is also the same, then the node updates the stored information using the lower h_i. The node i then broadcasts the information it has after incrementing the *hopcount* by 1. Subsequently, i removes itself from H, i.e., $H = H - \{i\}$, and inhibits itself from acting as the regional head if the information was updated. Due to g and autonomous behavior, there is another check that has to be performed. It is possible that a node has an information about a head node that is no longer the head. In order to check for this condition, a node when inhibiting itself broadcasts a message saying that it is not an uninhibited node anymore. The nodes associated to this node thus reset their information and the process is again performed. The process continues until sufficiently large time T after which all nodes are either uninhibited or inhibited. We assume that each node knows this value of time. Each uninhibited node will thus have inhibited nodes associated to it forming a virtual region G_i. All nodes in G_i will be within g hops from the uninhibited node and will have same head node information.

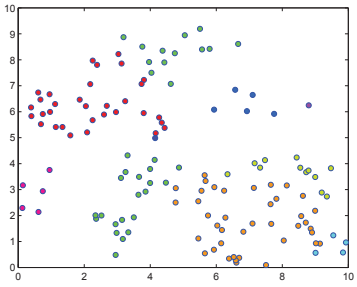
We call the nodes having same head node to belong to one region, G_i. Let V_i be the set of nodes in G_i. The nodes at different *hopcount* from the head node, h_i, of the region G_i virtually forms levels of different hops around h_i, (cf. Fig. 3.4(d)). Note that a node with no neighborhood will be tagged as the head as it will uninhibited, (cf. Fig. 3.4(c)). However, the described technique does not guarantee that the head nodes identified above have a high Closeness Centrality value and are



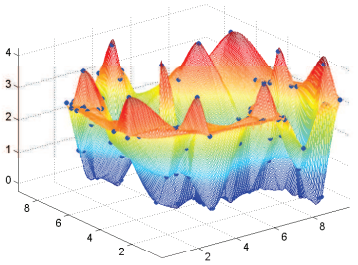
(a) Uniform Node Distribution.



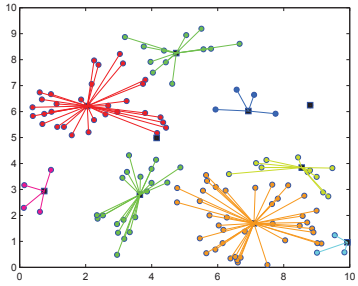
(b) Distribution after applying Thinning process with $r_b = 1$ and $N_{min} = 5$.



(c) Identified regions in the deployment shown by the fig. 3.4(b).



(d) The levels of the nodes created using the *hopcount* for the regions created in the fig. 3.4(b). The peaks show the centroid nodes while the valley shows nodes with the max. distance from its associated centroid nodes.



(e) Association of the nodes to the centroid nodes. The centroid of the region is marked with a black square.

Figure 3.4: Region Formation and Centroid identification using $g = 5$.

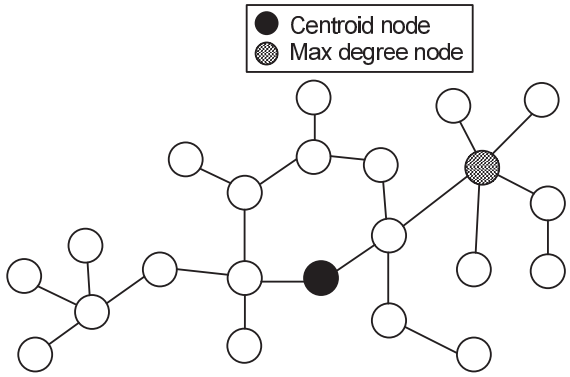


Figure 3.5: The max degree nodes are not at the center of the region. The Closeness Centrality of these nodes is less.

the most central nodes, (cf. Fig. 3.5).

To address this issue, the steps for the centroid node identification in a given region, G_i, created using Lateral Inhibition are next described. As the centroid node is the most central node in the region, the coordinates of the centroid node can be approximated through averaging. Due to the lack of global information averaging is studied through reaching consensus over time.

Centroid Finding

Let the coordinates of the node $j \in V_i$ be (x_j, y_j). Let the node have 2 copies of these coordinates. One which is fixed and one which can be modified. Let the copy that can be modified be called as (x_j^*, y_j^*). The nodes then communicate (x_j^*, y_j^*) to their neighbors, $N_{j,i} \in V_i$. At the same time the node j receives such coordinate information from its neighbors in V_i. Each node $j \in V_i$ then checks if the received coordinates are same as the stored one, (x_j^*, y_j^*). If it is not same then using this information (the received coordinates of the neighbors and (x_j^*, y_j^*)), each node $j \in V_i$ computes an average of the coordinates and updates (x_j^*, y_j^*). This updated information is then broadcasted to their neighbors in G_i. When the process is run multiple times each node will have stored coordinates approximately same as those it receives from its neighbors. This process is continued until all nodes in G_i have same average coordinates. This new average coordinate, (x_j^*, y_j^*), approximates the coordinates of the centroid in G_i. Note that the time taken to reach consensus can be large. However, in a G_i with finite number of nodes, the consensus will be reached in finite time.

This technique reveals the location of the centroid node but not the identity of the node that is to be termed as centroid. In order to identify the centroid node of the region, nodes use their fixed copy of the coordinates and (x_j^*, y_j^*) coordinates. Each node j checks if $(x_j, y_j) = (x_j^*, y_j^*) \pm \varepsilon$, where ε is the error margin such that $\varepsilon \in (0, r]$, and declares itself as the centroid. This process might result into multiple nodes declaring themselves as the centroid as two or more nodes can lie within the ε range of (x_j^*, y_j^*). To avoid this, a node also considers their degree and egocentric betweenness. The nodes within ε range of (x_j^*, y_j^*) share this information among themselves. Subsequently, the node having maximum sum of degree and egocentric betweenness declares itself as the centroid, c_i of the region G_i. Note that the above identification of centroid node is not valid for some topologies. For example, consider G_i with 5 nodes having coordinates (r, r), $(2r, 2r)$, $(3r, 3r)$, $(4r, 2r)$ and $(5r, 1r)$ respectively. The coordinates of centroid node will be $(3r, \frac{9r}{5})$. The above technique will result into no centroid node if $\varepsilon = r$. This problem will also occur in ring G_i. Thus, to avoid such scenarios where no centroid node is identified we assume that the head node is the centroid node of the region.

After the identification of the centroid nodes, the centroid nodes broadcast their information in the network. All nodes then update their stored head information to their respective c_i and the *hopcount* to their *hopcount* from c_i.

Let the information about a centroid node be disseminated to nodes g_{max} hop apart. We make $g_{max} > g$. This dissemination of the centroid node information enables the nodes to build a set RC_i which contains set of all centroid nodes within g_{max} hops of the node i for future use. Algorithm 2 represents the algorithmic description of the region formation and algorithm 3 represents the centroid identification process. Fig. 3.4(e) shows the centroid nodes for the regions identified in Fig. 3.4(c).

Beamforming

According to the results of [48], it requires only a small fraction of nodes with long link capabilities to achieve small world properties. In a self organizing environment where all nodes possess beamforming capabilities, it is essential to identify nodes that create long-range beams along with the direction and the width of the beam. Flocking provides valuable insights in determining the answers to these questions. The nodes use insights from the **Alignment** rule of Flocking to identify themselves if

they belong to the set P or not.

In our case the Alignment rule is the decision making of whether to beamform or not to beamform. However, we do not determine the new direction of the beam using alignment rule. The Alignment rule thus refers to the identification of the set of peripheral nodes, P_i in the region G_i. A node uses the *hopcount* of the neighborhood nodes to decide whether or not it is a peripheral node, $\wp_i$, in the region G_i. If all nodes in the neighborhood, $N_{j,i}$, of the node j in the region G_i have *hopcount* less than or equal to the node's *hopcount* to the c_i, then the node declares itself as a peripheral node. The above statement means, for a given region G_i with centroid c_i, $\wp_j \in P_i \iff hopcount_{\wp_j c_i} \geq hopcount_{N_{\wp_j c_i}}$. A single unconnected node will always become a peripheral node because it does not have any neighborhood. Note that two peripheral nodes can be neighbors of each other due to the equality in the condition.

The peripheral nodes randomly choose the number of AEs, $m \in [2, M]$, to beamform. As described above for the sector model, $B_l = r \times \sqrt{m}$ and $B_w = \frac{2\pi}{m}$. Note that to cover all the directions, minimum number of sectors that are needed are m. The dependency of B_l and B_w on m affects the connectivity of the network. Fig. 3.6(a) shows the variation in B_l and B_w when $m > 1$. When B_l is small, i.e., when few AEs are used, the probability of connecting to the nodes in the original neighborhood is high as the beam is wider, (cf. Fig. 3.6(b)). However, when B_l is long, i.e., when multiple AEs are used, the probability of connecting to the nodes in the original neighborhood is low as the beam is narrow, (cf. Fig. 3.6(c)).

As the number of sectors increases with an increase in the number of AEs, there is an increase in the time taken to decide the best sector. Checking all the sectors formed for all $m \in [2, M]$ requires a test of $\frac{(M)(M+1)}{2} - 1$ sectors. The complexity of such a test is $O(M^2)$. This results into more energy consumption at the node. To reduce this energy consumption and the complexity to $O(M)$, the node selects randomly the number of AEs, $m \in [2, M]$, and only tests the corresponding set of m sectors.

Non-uniformity reduces the size of the giant component in the wireless *ad hoc* network. In order to increase connectivity, it is thus important for the nodes to find different network components and connect them using beamforming. **Separation** rule of Flocking provides insight towards this problem.

Algorithm 2 Region formation

Let $U \leftarrow uninhibited$;
Let $I \leftarrow inhibited$;
Let $ID \leftarrow identity\ of\ node$;
\\ Region formation;
for all $i \in V$ **do**
  set $i_{Status} \leftarrow U$, $i_{Head} \leftarrow i$, $deg_i \leftarrow |N_i|$, $i_{hopcount} \leftarrow 0$, $i\_coordinates \leftarrow (x_i, y_i)$, $B_i = 0$
  Initially broadcast$(ID_i, hopcount = 0, deg_i)$
end for
for all $i \in V$ **do**
  repeat
    $recv \leftarrow$ receive$(ID, hopcount + 1, degree, B)$
    if $deg_i < degree$ & $hopcount < g$ **then**
      $i_{Status} \leftarrow I$, $i_{Head} \leftarrow ID$, $i_{hopcount} \leftarrow hopcount + 1$, $deg_i \leftarrow degree$, $B_i \leftarrow 1$
      broadcast$(recv, B_i)$
    else
      if $deg_i == degree$ & $i_{hopcount} > hopcount$ & $hopcount < g$ **then**
        $i_{Status} \leftarrow I$, $i_{Head} \leftarrow ID$, $i_{hopcount} \leftarrow hopcount + 1$, $deg_i \leftarrow degree$, $B_i \leftarrow 1$
        broadcast$(recv, B_i)$
      else
        if $deg_i == degree$ & $i_{hopcount} == hopcount$ & $hopcount < g$ & $i_{Head} > ID$ **then**
          $i_{Status} \leftarrow I$, $i_{Head} \leftarrow ID$, $i_{hopcount} \leftarrow hopcount + 1$, $deg_i \leftarrow degreeB_i \leftarrow 1$
          broadcast$(recv, B_i)$
        end if
      end if
    end if
    if $B_{Head} == 1$ **then**
      Reset
    end if
  until Time
end for

In the proposed algorithm, similar analogy to address the connectivity issue is applied. In order to increase connectivity, nodes create beam in different directions from their peripheral neighbors. In terms of Flocking separation rule, it is the decision to beamform in another direction. Consider $\varrho_{\wp i} \in P_i$ as a peripheral neighbor of $\wp_i$ then for all $\varrho_{\wp i}$'s, $\wp_i(B_b) \neq \varrho_{\wp i}(B_b)$ must hold where B_b is the boresight direction. To make this decision, if $\varrho_{\wp_i}$ of a $\wp_i$ decides to create the beam in certain direction, $\varrho_{\wp_i}$ informs $\wp_i$ about the chosen direction before it actually creates the beam. $\wp_i$ then tries to create the beam in another direction. $\wp_i$ gives preference to connect to the nodes in other region rather than those of its own. This increases the possibility of connecting to an isolated region. Fig. 3.7 shows two nodes w and x which were initially neighbors of each other, create beams in different direction in order to increase connectivity.

Algorithm 3 Centroid finding

```
Let i be a region formed using region formation algorithm;
for all j ∈ V_i do
  (x*_j, y*_j) ← (x_j, y_j)
  repeat
    Let flag ← 0
    for all k ∈ N_{j,i} do
      recv_k ← receive((x_k, y_k))
    end for
    for all k ∈ N_{j,i} do
      if recv_k!=(x*_j, y*_j) then
        set flag ← 1
      end if
    end for
    if flag = 1 then
      for all k ∈ N_{j,i} do
        (x*_j, y*_j) ← (x*_j, y*_j) + recv_k
      end for
      (x*_j, y*_j) ← (x*_j, y*_j)/|N_{j,i}|
      broadcast((x*_j, y*_j))
    end if
  until flag = 1
end for
Let X ← {}
for all j ∈ V_i ∈ V|(x_j, y_j) − ε < j(x*, y*) < (x_j, y_j) + ε do
  X ← X + {j}
  compute sum_j ← sum(|N_j|, EBC_j)
end for
if X == {} then
  C ← C + {h_i}
else
  for all j ∈ X do
    c_i ← j|j = max{sum_j}
    C ← C + {j}
  end for
end if
for all k ∈ V do
  formulate RC_k
end for
```

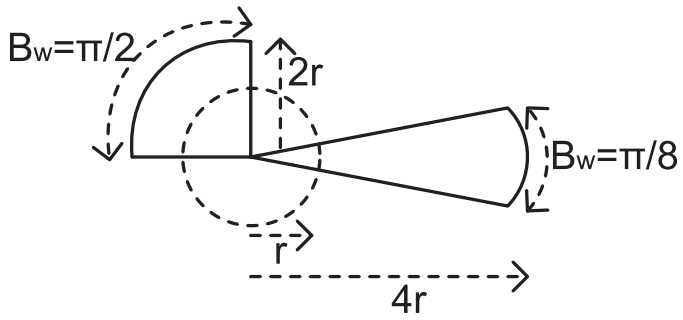
Nevertheless, the choice of best direction of the beam and the knowledge of whether a $\wp_i$ has a node within its 1 hop is still not provided. These problems are addressed next in this section. To the above-mentioned problem, the analogy of **Cohesion** rule of Flocking to determine the best direction of the beam is used.

We apply similar analogy to cohesion rule in the proposed algorithm. We state that the cohesion

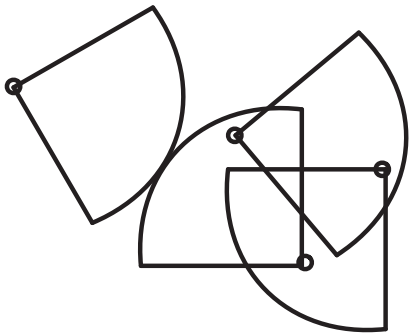
rule is the decision to bind a peripheral node with a centroid node. From the previous Chapter, it can be noted that the centroid node has the highest Closeness Centrality value in a given region. As the APL is the sum all shortest distances, any reduction in this summation will lead to a reduced APL. In order to have maximum reduction in the path length, the node should connect to the farthest centroid. Directing the peripheral node's beam towards the centroid node will thus help reduce the average distance of the peripheral node to other nodes of the region in which the centroid node lies.

If the centroid node chosen by the peripheral node and the peripheral node itself were not initially connected, connecting them will help in increasing the connectivity, (cf. Fig. 3.7 where two peripheral nodes w and x which were initially neighbors of each other, create beams in different directions. In order to have increased connectivity, the node w creates a beam towards the region containing the centroid node y, while the node x creates a beam towards the region containing the centroid node z. Here $g = 4$.). On the other hand, if the centroid node chosen by the peripheral node and the peripheral node were connected and were some finite hops apart, beamforming will lead to the reduction in the APL.

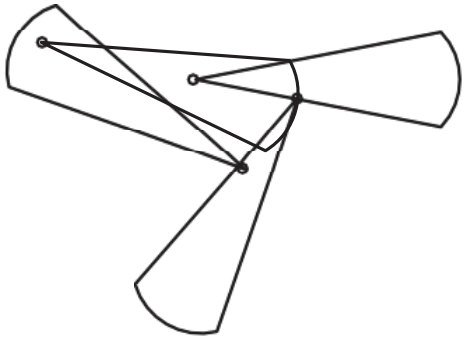
To account for choosing the correct centroid to connect, the peripheral node, $\wp_i$, builds $RC^*_{\wp_i}$, a set of all centroid nodes reachable when it is beamforming. To determine $RC^*_{\wp_i}$, the peripheral nodes sweep through all the sectors (m) created with the chosen number of AEs except the sectors in which $\varrho_{\wp_i}$'s have created the beam. If $RC^*_{\wp_i} - RC_{\wp_i} \neq \emptyset$ and $|RC^*_{\wp_i} - RC_{\wp_i}| > 1$, i.e., $\wp_i$ identified two or more potential centroid nodes, assuming the *hopcount* to these centroid nodes as ∞ the decision to connect to one of them is randomly made. However, if $RC^*_{\wp_i} - RC_{\wp_i} = \emptyset$, i.e., no new centroid is found, the $\wp_i$ decides to connect to farthest centroid node in $RC_{\wp_i}$. However, this decision also depends on the *hopcount* between c_i and $\wp_i$. Creating the beam toward the centroid that is less than two hops away will only reduce the initial neighborhood but not the APL. In this case $\wp_i$ drops the decision of being the peripheral node and remains omnidirectional. Fig. 3.8(a) and fig. 3.8(b) depicts the same. In fig. 3.8(a), node x is 5 hops away from y while it is 3 hops away from z and 2 hops away from the centroid of the region in which x lies. Thus, in order to have a reduced path length, node x decides to create beam towards y. On the contrary, in the case when the node x does not have the previously stored information about the centroid nodes y and z, the node considers



(a) The difference in the beam properties when a $\wp_i$ uses different number of AEs.



(b) Connectivity when 4 AEs are used.



(c) Connectivity when 16 AEs are used.

Figure 3.6: Relationship between beam properties and connectivity.

hopcount to these centroid nodes as ∞ and randomly chooses one of them to connect to, (cf. Fig. 3.8(b)).

Whenever a peripheral node creates a beam towards a centroid node that is more than 1 hop away, asymmetric link may arise. This is due to the fact that the B_l of peripheral node is $r \times \sqrt{m}$ while B_l of a centroid node is r, in other words, $\frac{B_l \; of \; Centoid}{B_l \; of \; peripheral} = \frac{1}{\sqrt{m}}$. Due to this difference, peripheral nodes will not know if they got connected to the centroid of other region or not. We propose to solve this issue by performing following steps, when a centroid node receives information about the node trying to connect to it, it just for one time instant, to acknowledge the reception, creates the beam back to the node. This can be easily done after determining angle of incidence of

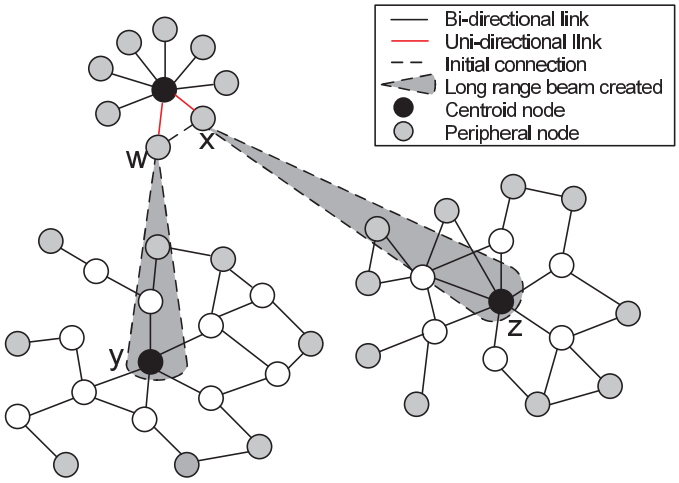


Figure 3.7: Nodes beamforming in different directions.

the beam. This works well for both connected and unconnected components. Algorithm 4 represents a brief algorithmic description of beamforming using Flocking rule analogy. Fig. 3.9 shows the new network created as the output of the algorithm on the network shown in fig. 3.4(b). Fig. 3.9 uses $g = 5$ to create regions. In fig. 3.9 nodes marked in green beamform. Their directional beams are also shown in green. The nodes marked with black triangle do not beamform and the nodes marked with red square are the centroid nodes.

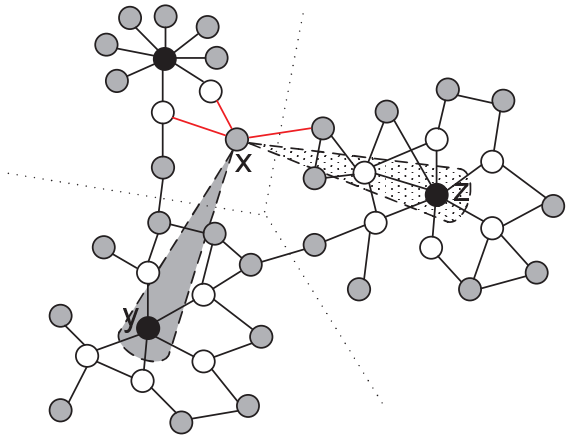
3.2 Definitions and Lemmas

Definition 3.2.1. Assume a centroid c_i of the region G_i, and a node j in V_i which has the highest Closeness Centrality, then

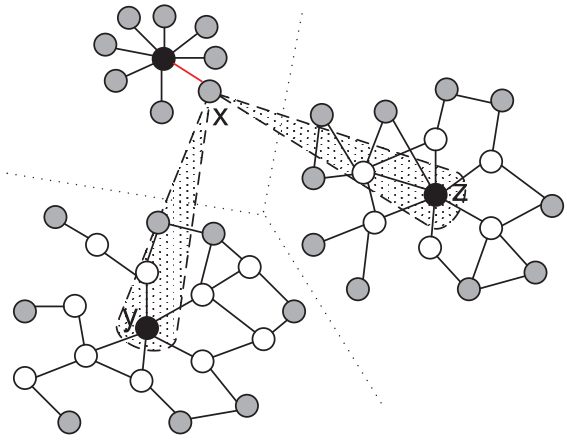
$$\begin{aligned} j &= \arg\max_{\forall k \in V_i} \{CLC_k\} \\ CLC_{c_i} &\simeq CLC_j \end{aligned} \tag{3.2.1}$$

Definition 3.2.2. The node j with neighborhood $N_{j,i}$ of the region G_i with centroid c_i is a peripheral node $\iff hopcount_{jc_i} \geq hopcount_{N_{j,i}c_i}$.

Lemma 3.2.1. Let ρ be the initial node density in a given area A, r_b be the radius chosen for the thinning process, N_{min} be the minimum number of neighbors required and $\Gamma(N_{min}, \rho r_b^2 \pi)$ be



(a) One component with three regions when $g = 3$. Here, the node x can create the beam towards y or z, but because the *hopcount* to y is more than z, node x creates the beam towards y.



(b) Three unconnected components with three regions when $g = 4$. Here, the node x can create the beam either towards y or z, but because the *hopcount* to y and z is same as ∞, x randomly decides between y and z to connect to.

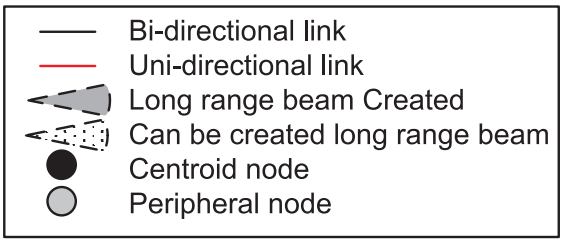


Figure 3.8: Beamforming priority.

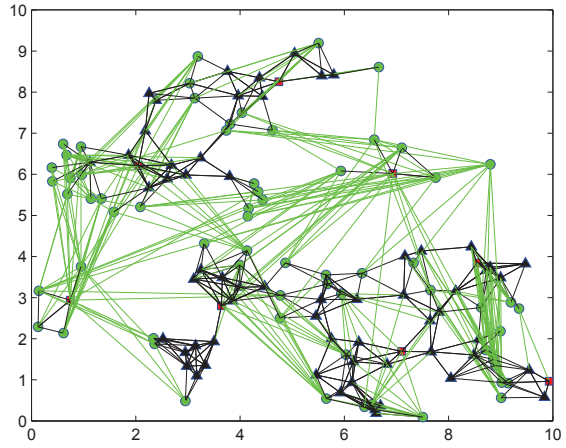


Figure 3.9: Beamforming nodes network depiction.

the incomplete gamma function. The expected number of nodes remaining, n, after applying the thinning processes, [81], on a uniformly distributed network is defined by

$$\mathbf{E}(n) = \rho A \left(1 - \frac{\Gamma(N_{min}, \rho r_b^2 \pi)}{(N_{min} - 1)!}\right). \tag{3.2.2}$$

Lemma 3.2.2. The separation between any two head nodes is between $(g, 2g + 1]$ where g is the *hopcount* used to create the region, [76].

Proof. Consider a head node with all nodes within g hops around itself in its region. A node which is more than g hops away will lie in another region. If in the neighboring region, a head node does not have any region of its own around it, then the distance between the two head nodes in hops will be $g+1$. On the other hand, if the neighboring region also has a region of size g around it, then the distance between two head nodes in hops will be $2g + 1$. □

Lemma 3.2.3. The number of regions is equal to number of centroid nodes and each region has exactly one centroid node.

Proof. The algorithm computes the centroid of the region based on average of coordinates, Degree and Egocentric Betweenness of the node for each region. According to the algorithm, the nodes are termed as centroid if the node falls within ε range of the centroid coordinate estimation algorithm and have maximum sum of Degree and Egocentric Betweenness. If still there are multiple nodes

Algorithm 4 Beamforming using Flocking Analogy

\\ Alignment;
for all $j \in V_i \in V$ **do**
  if $hopcount_{jc_i} > hopcount_{N_{j,i}c_i}$ **then**
    $P_i \leftarrow P_i + \{j\}$
    $P \leftarrow P + \{j\}$
  end if
end for
\\ Separation;
for all $\wp_i \in P_i \in P$ **do**
  set m
  for all m Sectors $|\varrho_{\wp_i}(B_b) \notin$ Sectors **do**
    $RC^*_{\wp_i} \leftarrow RC^*_{\wp_i} + \{reachable\ centroid\ nodes\}$
  end for
end for
\\ Cohesion;
for all $\wp_i \in P_i \in P$ **do**
  if $RC^*_{\wp_i} - RC_{\wp_i} \neq \emptyset$ **then**
    for all $c \in RC^*_{\wp_i} - RC_{\wp_i}$ **do**
      $h \leftarrow h + hopcount_{\wp_i c}$
    end for
  else
    if $RC_{\wp_i} \neq \emptyset$ **then**
      for all $c \in RC_{\wp_i}$ **do**
        $h \leftarrow h + hopcount_{\wp_i c}$
      end for
    else
      $P_i \leftarrow P_i - \{\wp_i\}$
    end if
  end if
  $beamtonode \leftarrow \max\{h\}$
  $\theta \leftarrow$ Sector containing $beamtonode$
end for

that are termed as centroid nodes, the nodes randomly decide for being the centroid and thus only one node is chosen as centroid. The value of ε is thus an important factor in the estimation of the centroid node. Also, smaller ε will tend to provide better estimation of the centroid nodes. As there is only one centroid node per region, the number of centroid nodes is equal to the number of regions. However, if no centroid is identified through the algorithm then the head node is termed as centroid node. □

Lemma 3.2.4. If a node is not a centroid node, it is connected to a centroid node.

Proof. The algorithm identifies regions and their centroid nodes. An identified region is always

connected, i.e., all the nodes in the identified region are connected to each other and there is one and only one centroid node in a region, lemma 3.2.3. Thus for a given region, all nodes that are not centroid are connected to the centroid node. □

Lemma 3.2.5. An unconnected node is both the centroid node as well as the peripheral node.

Proof. A single unconnected node does not have any neighborhood. It thus remains uninhibited at the end of the region formation phase and becomes the head. As it is lacking any neighborhood, the node does not have any other node in the region around itself and is the only node in the region. In this region, the average coordinates perfectly match the average coordinates of the node. Thus it requires no further computation for correctly identifying the centroid node.

This node is also the peripheral node as the condition of definition 3.2.2 holds true because of the unavailability of the neighborhood. □

Lemma 3.2.6. For a node distribution and fully connected network with average node density ρ and total number of nodes $|V|$, then $|C|$ is bounded by $\frac{|V|}{\rho g^2 r^2 \pi}$ and $\frac{|V|}{\rho g^2 r^2 \sqrt{3}}$.

Proof. From lemma 3.2.2, the hop distance between two heads is bounded by $(g, 2g+1]$.

Lower Bound: When the heads are separated by $2g+1$ hops, the number of regions formed are less. The number of nodes in one region is $\rho g^2 r^2 \pi$. Thus, the total number of nodes in all the N regions is $|N| \rho g^2 r^2 \pi$. As the total number of nodes are $|V|$, $\therefore |N| = \frac{|V|}{\rho g^2 r^2 \pi}$. From lemma 3.2.3, $|C| = |N|$, $\therefore |C| = \frac{|V|}{\rho g^2 r^2 \pi}$

Upper Bound: When all the heads are separated by $g+1$ hops, the number of regions formed are more. A head in such a case is connected to only 6 other heads. This can be visualized as a hexagon with vertex-vertex distance equal to $g+1$ and a node at the center of hexagon. Each of the vertex nodes are shared between 3 other hexagons. Thus, the total number of heads that are exclusive for the hexagon are $\frac{6}{3}+1 = 3$. In other words, there are 3 heads in an area of $\frac{6g^2 r^2 \sqrt{3}}{2}$. Thus, for the area=$\frac{|V|}{\rho}$, $|C|$=$\frac{|V|}{\rho g^2 r^2 \sqrt{3}}$. □

Lemma 3.2.7. Consider a network with j components ($j > 1$), average density of the nodes as ρ_k and number of nodes as $|V_k^j|$ for $k \in j$, $|C|$ is bounded by $\sum_{k=1}^{j} \frac{|V_k^j|}{\rho_k g^2 r^2 \pi}$ and $|V|$, where $|V| = \sum_{k=1}^{j} |V_k^j|$.

Proof. From lemma 3.2.2, the hop distance between two heads is bounded by $(g, 2g+1]$.

Lower Bound: Consider k^{th} component of the network. When the heads are separated by $2g+1$ hops, the number of regions formed is less. The number of nodes in one region is $\rho_k g^2 r^2 \pi$. Thus, the total number of nodes in all the regions in the component is $N_k \rho_k g^2 r^2 \pi$, where N_k are the number of region in k^{th} component. But as the total number of nodes were assumed to be $|V_k^j|$, $\therefore N_k = \frac{|V_k^j|}{\rho_k g^2 r^2 \pi}$. Thus for all the components, the number of regions formed is $|N| = \sum_{k=1}^{j} N_i = \sum_{k=1}^{j} \frac{|V_k^j|}{\rho_k g^2 r^2 \pi}$. From lemma 3.2.3, $|C| = |N|$, $\therefore |C| = \sum_{k=1}^{j} \frac{|V_k^j|}{\rho_k g^2 r^2 \pi}$

Upper Bound: Upper bound to the number of regions arises when all nodes in the network are disconnected. Thus, all nodes in such a case will be uninhibited thereby becoming region heads. Thus $|C| = |V|$. □

Lemma 3.2.8. For a node distribution and fully connected network, and using lemma 3.2.6, the number of peripheral nodes in the network is bounded by $\frac{|V|(2g-1)}{g^2}$ and $\frac{|V|(2g-1)\pi}{g^2\sqrt{3}}$.

Proof. Peripheral nodes are the nodes lying in the outer most part of the region. Thus, the number of nodes g hops away from head node of a region $= \rho g^2 r^2 \pi - \rho(g-1)^2 r^2 \pi = \rho(2g-1) r^2 \pi$

Now using lemma 3.2.6, the number of peripheral nodes for all regions thus varies between $\frac{|V|(2g-1)}{g^2}$ and $\frac{|V|(2g-1)\pi}{g^2\sqrt{3}}$. □

Lemma 3.2.9. For a node distribution and network with j components ($j > 1$), and using lemma 3.2.7 and lemma 3.2.5, the number of peripheral nodes in the network is bounded by $\sum_{k=1}^{j} \frac{|V_k^j|(2g-1)}{g^2}$ and $|V|$.

Proof. Peripheral nodes are the nodes lying in the outer most part of the region. Thus, the number of nodes g hops away from head node of a region in k^{th} component $= \rho_k g^2 r^2 \pi - \rho_k (g-1)^2 r^2 \pi = \rho_k (2g-1) r^2 \pi$.

Now using lemma 3.2.7 and lemma 3.2.5, the number of peripheral nodes for all regions thus varies between $\sum_{k=1}^{j} \frac{|V_k^j|(2g-1)}{g^2}$ and $|V|$. □

3.3 Simulation Setup

We consider a simulation area of $A = 500\text{x}500$ to simulate the algorithm. r_b and N_{min} are set to 30 and 5 respectively to achieve the non-uniform distribution of node throughout the simulation area. The non-uniform node distribution enables us to represent the real world scenarios. Initially, each node operates in omnidirectional mode using $m = 1$ AE with the omnidirectional radius as $r = 30$. The maximum number of AEs that the nodes are equipped is set to $M = 6$. The separation between two AEs computed using $WiFi$ frequency, $f = 2.4Ghz$.

Using such setting we first simulate the variation of percentage number of nodes in the giant connected component (GCC) for non uniform deployment for various average node densities, ρ, in the region. The simulation reveals that the percentage number of nodes in $GCC \in (0.1, 0.9)$ for the range of $\rho \in [1\text{x}10^{-3}, 2.5\text{x}10^{-3}]$ (cf. Fig. 3.10). This is an ideal range for us to test our algorithm as we are interested in a disconnected network. Further, as non uniform deployment strategy is applied, nodes can become disconnected. This leads to a network that is less connected than the uniformly deployed network. However, when the density increases the size of the GCC also increases.

Further, through simulations, we explore the effect on connectivity, APL and CC by varying the node densities and the region size. We use MATLAB to simulate the algorithm with a confidence interval of 95%. We average all the results over 50 topologies and set $\varepsilon = 0.66r$.

3.4 Results

First, we prove the correctness of the centroid finding in the region. For this, we compute the relation between the nodes that have maximum Socio-Centric Betweenness and the centroid nodes in the region. If the centroid node has the highest Socio-Centric Betweenness in the region, then the algorithm found the centroid node correctly, (cf. Fig. 3.11). This depends on the value of the region size. Larger regions decrease the Socio-Centric Betweenness rank of the centroid node in the region. As the region size increases, more nodes are now associated to a region thereby increasing the possibility of occurrences of the bridge nodes (bridge nodes have high Socio-Centric Betweenness value). Thus, we also calculate the distance in hops between the centroid node and the maximum Socio-Centric Betweenness node. According to the results, (cf. Fig. 3.11), the percentage of centroid

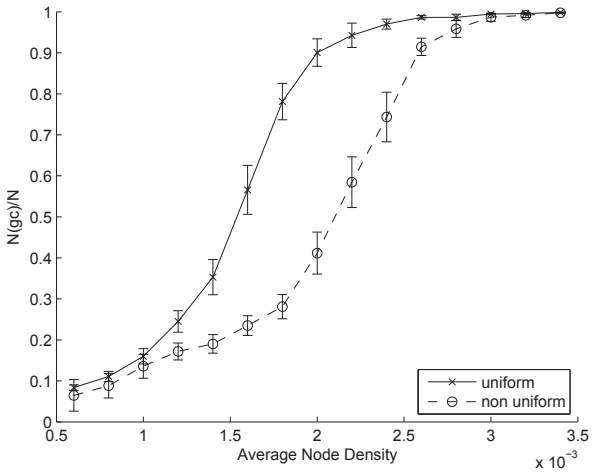


Figure 3.10: % number of nodes in *GCC* for nodes distributed uniformly and non-uniformly.

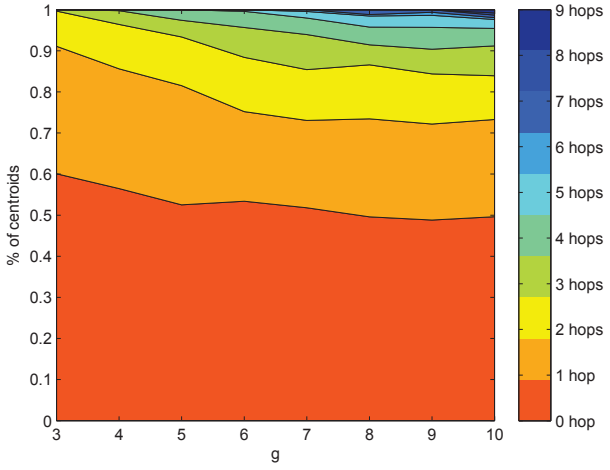


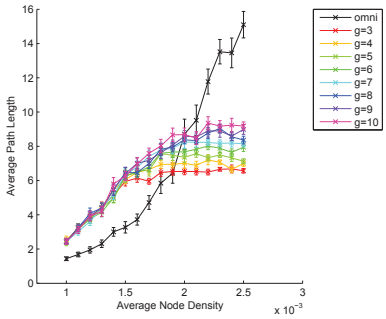
Figure 3.11: Relationship between centroid nodes and the nodes having maximum Socio-Centric Betweenness.

nodes that also have high Socio-Centric Betweenness is between 50% and 60% for $g \in [3, 10]$. Further, all the centroid nodes in the network are within $hopcount < g$. Fig. 3.11, however, also shows that for any $g \in [3, 10]$, more than 95% of the time the centroid node is within 4 hops from the maximum Socio-Centric Betweenness value node.

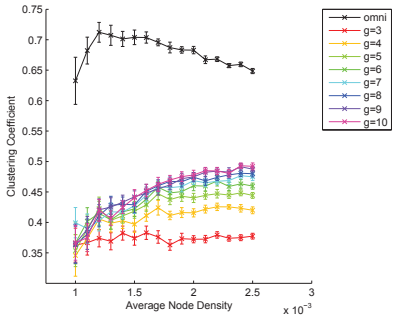
We use $g \in [3, 10]$ and Sector model to obtain the results, (cf. Fig. 3.12). Fig. 3.12(a) shows the effect of beamforming on the APL. Note that the APL is calculated for the largest component in the network. The APL obtained in omnidirectional case is initially less than that obtained for the directional cases because the density of the nodes in the component is low. When the algorithm induces directional beams, due to the inclusion of the nodes in other network components, there is an increase in the APL. The APL for the directional case is less than that of the omnidirectional case when $\rho > 2 * 10^{-3}$ due to the fact that the nodes connect to the centroid node of other regions in the different component as well as in the same component. The region size affects the APL. The lower the value of the region size, higher is the number of nodes that beamform, (cf. Fig. 3.12(e)), leading to more shortcuts and in turn more reduction in the APL. For $\rho = 2.5 * 10^{-3}$ and $g = 10$, there is a reduction of almost 40% in the APL while there is a reduction of almost 55% for $g = 3$, (cf. Fig. 3.12(f)). However, for $\rho = 1 * 10^{-3}$ and $g \in [3, 10]$ when most nodes are unconnected, there is an increase of 70% in APL due to the above-mentioned facts.

The introduction of the long-range beams also causes the CC to change, (cf. Fig. 3.12(b)). For very low-density networks, the CC for the directional case is less because beamforming leads to loss in the initial neighborhood. However, for higher density networks, the CC does not vary as much as the APL (cf. Fig. 3.12(f)). For $\rho = 2.5 * 10^{-3}$, there is a reduction of 25% and 38% for $g = 10$ and $g = 3$ respectively. However, for $\rho = 1 * 10^{-3}$ and any $g \in [3, 10]$, the reduction in CC is almost 40%.

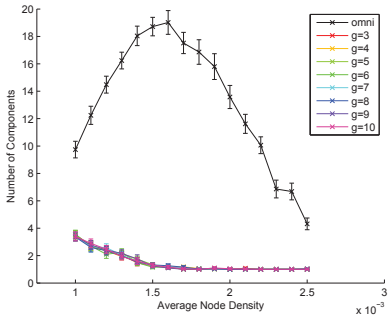
The number of components in the network can define connectivity. In a very low-density omnidirectional network, the number of disconnected components is higher, (cf. Fig. 3.12(c)). The number of disconnected components increases to a certain maximum and then decreases as the density increases. This is because, for a high density, all nodes can find at least one neighborhood node within their reach. In addition, as the number of components decreases, the connectivity increases. For the directional case however, as nodes beamform towards different components with the ob-



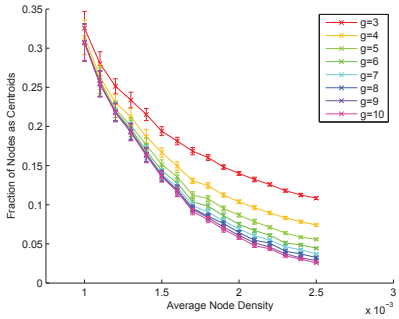
(a) Average path length in hops.



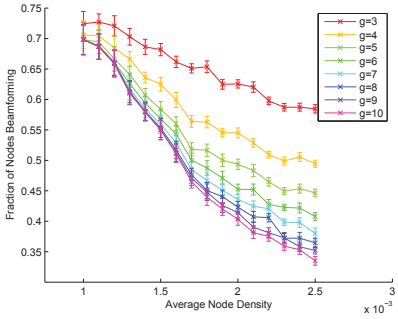
(b) Clustering Coefficient.



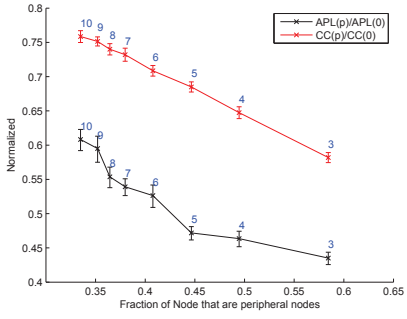
(c) Number of components in the network.



(d) Fraction of nodes designated as centroid nodes.



(e) Fraction of nodes designated as Peripheral nodes.



(f) Normalized directional APL and CC for $\rho = 2.5 * 10^{-3}$ showing the effects of the region size.

Figure 3.12: Results obtained for $g \in [3, 10]$, when Sector model and non-uniform node distribution are used.

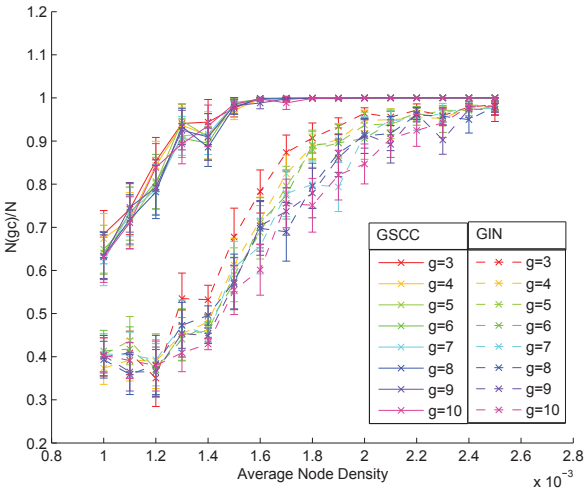


Figure 3.13: Variation in the size of the $GSCC$ and the GIN for different density of nodes and $g \in [3, 10]$.

jective of increasing connectivity, the number of disconnected components is less than that of the omnidirectional case.

The size of the giant component can also explain the connectivity of the network. For the directed graphs however, [130] defined the giant component using the Giant Strongly Connected Component ($GSCC$)[3] and the Giant In-Component (GIN)[4]. Thus, we calculate the size of $GSCC$ and GIN. We show the difference between the size of the giant component for omnidirectional network, $GSCC$ and GIN. As stated in [130] that $GSCC \subset GIN$, we also observe that GIN is a bigger set and contains more nodes than $GSCC$. The percentage of number of nodes in GIN reaches 1 very early, (cf. Fig. 3.13). Comparing the size of the $GSCC$ of directional network with the giant component of the omnidirectional network, (cf. Fig. 3.10), we find that the size of $GSCC$ varies between $[0.84, 0.94]$ for $\rho = 2 * 10^{-3}$ for different values of the region size while the size of giant component for the omnidirectional network is 0.41. Thus, an increase of almost 2.1 times. Fig. 3.14 shows an increase of almost 2.2 times when size of the $GSCC$ and the GIN for $g = 6$ with the giant component of the omnidirectional network are compared.

The number of centroid nodes ($|C|$) depends on the value of the region size, (cf. Fig. 3.12(d)).

[3] $GSCC$ in a directed graph is a component where there is a path between every pair of vertices.
[4] GIN is the set of nodes in the component which can connect to $GSCC$.

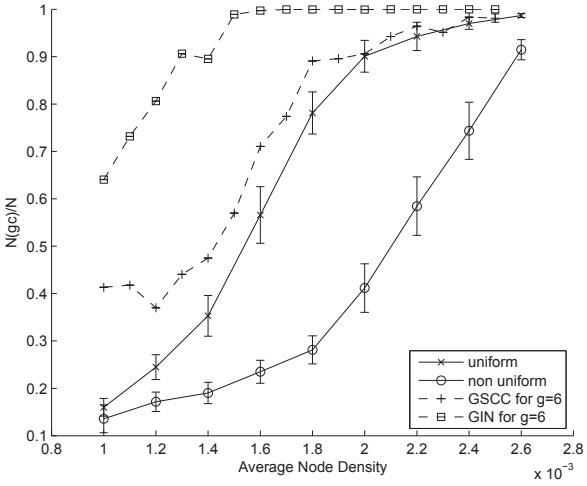
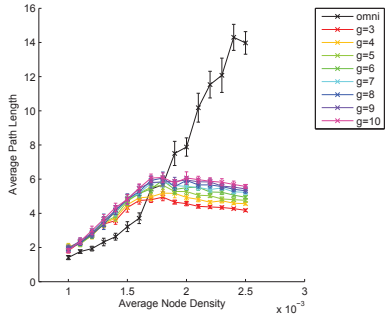


Figure 3.14: Comparison of size of the $GSCC$ and the GIN for directed network with that of omnidirectional network for $g = 6$.

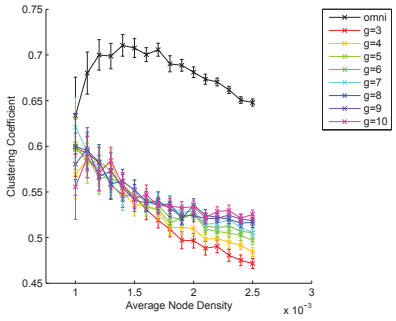
For a low-density network, the value of the region size does not matter, while as the density increases the value of the region size affects the number of regions formed. As the region size increases, more nodes are inhibited leading to less number of regions. The difference between the number of regions formed for $g = 3$ and $g = 10$ is of 40 for $\rho = 2.5 * 10^{-3}$ while the difference for $\rho = 1 * 10^{-3}$ is very less.

The value of the region size used also affects the number of peripheral nodes ($|P|$) identified, (cf. Fig. 3.12(e)). For a low value of region size, as there are more regions, more nodes are included in P because of the reduced neighborhood with respect to the region. However, when the value of the region size is large, $|P|$ is less because there are more nodes in the region and the nodes have relatively more neighbors to check before making the decision of beamforming. $|P|$ greatly affects the number of unidirectional paths. However, it has an adverse effect on the CC. As the number of peripheral nodes increases, unidirectional paths between the nodes also increases leading to more loss in the CC. For $\rho = 1 * 10^{-3}$ and $g \in [3, 10]$, the difference between the number of peripheral nodes is almost negligible. For $\rho = 2.5 * 10^{-3}$, however, the number of peripheral nodes varies by more than 120 as the regions formed for lower region size are more.

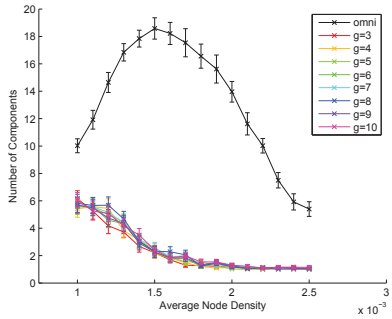
The algorithm affects the APL and the CC of the network when ULA model is used, (cf. Fig.



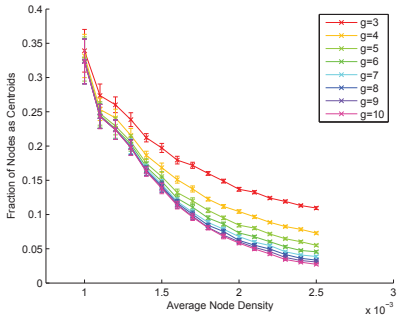
(a) Average path length in hops.



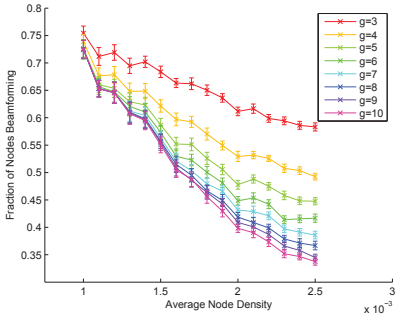
(b) Clustering Coefficient.



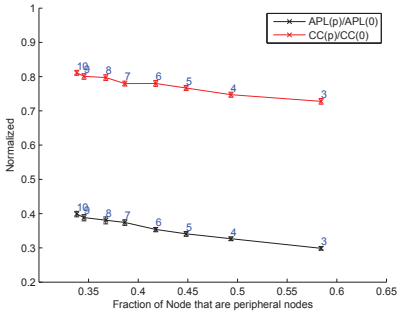
(c) Number of components in the network.



(d) Fraction of nodes designated as centroid nodes.



(e) Fraction of nodes designated as Peripheral nodes.



(f) Normalized directional APL and CC for $\rho = 2.5 * 10^{-3}$ showing the effects of the region size.

Figure 3.15: Results obtained for different $g \in [3, 10]$, using ULA model and non-uniform node distribution.

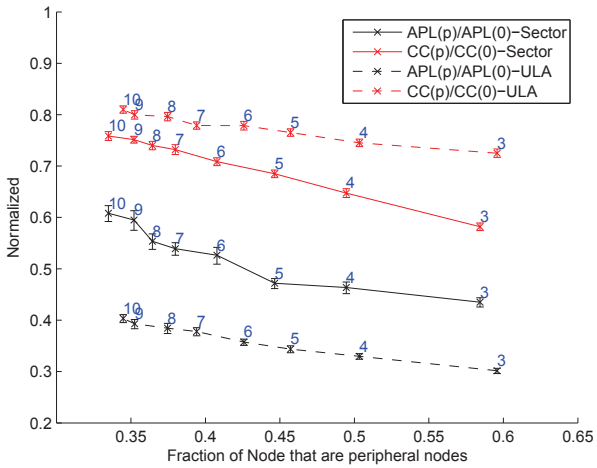


Figure 3.16: Normalized APL and the CC for $\rho = 2.5\text{x}10^{-3}$ showing the effects of the region size for both the Sector and the ULA model.

3.15). For high density, there is a reduction of almost 60% and 68% in the APL for $g = 10$ and for $g = 3$ respectively. On the other hand, the reduction in the CC is only between 19% to 22%. Due to variation in B_w for different B_b in ULA model and the presence of a secondary lobe (cf. Fig. 2.4), the values obtained for the APL, the CC and connectivity are different from that of the Sector model. However, antenna model does not affect $|P|$ and $|C|$. No dependency of the ULA model on $|P|$ and $|C|$ is rightly justified because these sets are build when the network is omnidirectional, (cf. Fig. 3.15(e), 3.15(d)). From fig. 3.16 it can be seen that, for higher density networks, the change in the APL for the ULA model is more than that of the Sector model while the CC changes at a much lower rate.

3.5 Conclusion

In this Chapter, we present an algorithm for achieving small world properties using beamforming and bio-inspired techniques in a wireless *ad hoc* network. The algorithm uses locally available information and does not require the global knowledge of the network. The algorithm operates in decentralized environment where the nodes do not require any external infrastructure to achieve the goals. We also show that isolated communities can be connected using beamforming with each other

to achieve more dissemination of information. In the algorithm, bio-Inspired techniques like Lateral Inhibition and Flocking are used to form virtual communities within the network for the reduced message complexity and to determine beam properties respectively. Through the results it is also shown that for both theoretical and realistic antenna models and relatively high-density networks, there is a reduction in the APL by almost 40% to 68% for $g \in [3, 10]$. On the other hand, reduction in the CC is between 19% to 38%. The results also show improvements in the connectivity. Through the use of the algorithm we show that the increase in the size of the $GSCC$ for the non-uniformly distributed directional network is around 10% for high density network while it is around 61% for relatively low density networks.

CHAPTER 4

Enhancing Information Dissemination in Dynamic Wireless Network using Stability and Beamforming

Recent studies have shown that mobility can both speed up as well as slow down the dissemination process [131, 132, 133, 7]. The speed up in the dissemination process results from the increased probability of devices meeting each other causing more exchanges of data and thereby enhancing the dissemination. However, mobility can also adversely affect the rate of dissemination if it causes devices to get isolated quickly. A scenario where isolation can happen is when the node density is low. In such scenarios, the likelihood of devices exchanging information is reduced due to few instances of meeting.

However, in a mobile communication network, such as Mobile *Ad hoc* NETwork ($MANET$), Pocket Switched Networks (PSN) and Delay Tolerant Networks (DTN), we are interested in studying how information could be quickly transmitted through these time varying networks. Example scenarios can involve dissemination of traffic updates, which may not only involve single source of information but can also consist of emergency alerts and safety information generated from multiple sources. Key aspects of mobile communication network are that the communication is based on the contacts between the energy and memory limited devices. Further, information about the network is typically limited to that of the immediate neighborhood. This motivates us to study how information could be quickly transmitted through these time varying networks where devices lack global knowledge of the network, have limited energy and memory.

Networks characterized by high dynamicity have been the studied using Time Varying Networks. A Time Varying Network is characterized as a network where the contacts between the devices are

intermittent and change with time. In such networks, connectivity of the network changes every Δt instant, due to the changes in the network structure. This results into time varying path length between two devices and hence affects the flow of information from one device to another [63]. The properties of time varying graphs are well studied using discrete time snapshots of the network [63,134]. Other than mobility, other factors studied in literature that affect the dissemination process in the network are, bursty data [135], strength of the connection [105], network structure [106], activity pattern [107], content type [108], node characteristics [109] and altruism [110]. Note that time varying are also termed as Evolving Network [132, 28], Evolving Graphs [29, 30, 31], Adaptive Network [36, 37], Dynamic Networks [27] and Temporal Networks [34, 35].

In [112, 113, 114] it is shown that when beamforming is combined with mobility, much faster information dissemination is achieved in terms of broadcasting time. The broadcasting time is defined as the time at which all devices have received the information. These studies also prove that (i) only a small fraction of devices are needed to beamform for enhancing the information dissemination process in the mobile network and (ii) better performance in terms of information dissemination is achieved when the beams are longer, narrower, use same power as that of omnidirectional beam and when the antenna rotates while the devices move. Another work related to beamforming in the mobile communication networks is by Li *et al.* [136]. The study by Li *et al.* [136] focuses on the effect of beamforming in a human centric vehicular network. Li *et al.* show that beamforming can enhance information dissemination process in a mobile network with similar observations as that of [112, 113, 114]. However, the only difference between [112, 113, 114] and [136] studies is in the selection of the devices that beamform. However, in these studies only single source information dissemination scenario was considered. Contrary to [112, 113, 114], where the direction of the beam and the choice of the beamforming devices was random, we believe that, if the selection of the devices that beamform is made based on a measure and the beams are directed in a direction where there are more devices that do not have the information, the dissemination process can be enhanced. That is, if *preferential attachment* features are incorporated in mobile networks information dissemination process can be enhanced.

Determining which device should beamform is surely the first question that has to be answered in this context. After finding the beamforming devices, the direction of the beams has to be determined.

In Chapter 3, we proposed the use of centrality measure with other concepts to solve the problem in a static network. In Chapter 3, the *preferential attachment* feature was incorporated as the preference to attach to the devices that has the highest closeness centrality. As discussed in Chapter 3, the computation of closeness centrality requires global network knowledge. In an autonomous environment where devices lack global knowledge the computation of centrality measure is time consuming. In mobile environment, where the position of the devices change frequently, it is hard to directly apply steps in Chapter 3 as it will be a slow process (the device can move by the time the highest closeness centrality device is found). Other methods for computing closeness centrality in dynamic conditions will be computationally expensive, complex, energy wasteful and will require historical information [63, 54]. Storing huge amount of historic information in a memory limited device could raise computational issues.

In the context of multi-packet dissemination and mobile networks, we are particularly interested in the measure of stability in networks. A key distinguishing feature of the approach taken here is that low stability is accompanied by diversity in content and can therefore be a catalyst for information dissemination. Stability is the complement of the change in the system. One way of defining Stability is using the fraction of neighborhood devices kept intact. Let S be a measure that quantifies stability of a device and let it be given by the eq. 2.8.4 as defined in Section 2.8.6. Let us assume discrete time steps where devices are able to jump from one location to another. This definition results into $S \in [0, 1] + \{undefined\}$. Note that $S = 1$ implies (i) the device is still connected to all of its neighbors at time t and no new device has become its neighbor and (ii) the device is still connected to all of its neighbors at time t and new neighbors have joined. In the first case the device will be limited to the information flowing within its neighbors at time t as no new device has joined the neighborhood set.

If $S = 0$, it implies (i) the device has lost all of its neighbors at time t completely and no new neighbor has joined and (ii) the device has lost all of its neighbors at time t and new neighbors joined it. Here, in the second case it will mean that the device has met more devices over time and will have the information flowing among neighbors at time t, while being able to receive the information flowing among new neighbors at time $t + \Delta t$. In the case of different types of information flowing in the network, this gives more diversity to the device in terms of knowledge about different types of

information flowing in the network. However, $0 < S < 1$ implies (i) the device has maintained some of its neighbors at time t and no new neighbor has joined and (ii) the device has maintained some of its neighbors at time t and new neighbors have joined. Further, $S = undefined$ means that the device had no neighbors at time t and thus would have not received any information flowing in the network at time t.

However, the stability measure proposed in Section 2.8.6 requires two network snapshots. One initial and one at the time when the stability measure has to be computed. In a mobile scenario where beamforming is also applied, before the information can be transmitted using stability as defined in Section 2.8.6, it requires following steps: (i) identification of the new neighborhood at time $t + \Delta t$; (ii) the computation of stability as defined in Section 2.8.6; (iii) beamforming and (iv) finding new neighborhood under beamforming. Performing these steps will consume time, energy and will stall the dissemination until the above steps are performed. Thus, it is beneficial for a device if it can estimate its stability before performing step "(i)" (knowing its neighbors at $t + \Delta t$) using its neighborhood at t only and mobility characteristics.

In order to determine which devices should beamform, we use the stability measure and proposes which devices should beamform and in which direction. One straight forward choice for the beamforming devices are the devices with low stability. This choice relates to the fact that when different types of information are flowing in the network, low stability devices are likely to have large diversity of information. If such devices beamform they will be able to transmit more information to far away devices. Alternatively, one intuitive choice of devices that can serve the purpose of *preferential attachment* are the devices with high stability. High stability devices are likely to be limited to the information flows within limited set of neighbors. Connecting to them will make them aware of more information flowing in network and will help them gain all the information flowing in the network more quickly. Thus, in this Chapter, we use the above mentioned discussion to build our algorithm. We focus on how information dissemination can be further enhanced in a dynamic spatial network where devices lack global knowledge about the network and have limited memory. Towards this, we first show how devices can predict their stability and use it to identify if they should beamform and where the beams should be directed. More explanation is provided in Section 4.1.

Table 4.1 details the notations used in this Chapter hereafter. The rest of this Chapter is

Notation	Meaning
A	simulation area
a_j	jump of j such that $j \in N_i^t$
B_b	boresight direction
b_j	jump of i
C_1	circle of radius a_j centered at j^t
C_2	circle of radius r centered at $j^{t+\Delta t}$
D_1	disk of radius a_j centered at j^t
D_2	disk of radius r centered at $j^{t+\Delta t}$
E^t	set of edges at time t
G	initial network of set of nodes V and set of edges E
G^t	instance of G at t with set of nodes V and set of edges E^t
i	a device in V
i^t	node i at time t
l_j	distance between initial position of neighborhood node and the final position of i
M	maximum number of antenna elements available with $i \in V$
m	number of antenna elements used by $i \in V$ to beamform $\|m \in [2, M]$
N_i^t	set of neighborhood nodes of i at time t
r	transmission radius
S_i^*	stability of the node i
S_{max}	maximum stability threshold
S_{min}	minimum stability threshold
t	time
x_{min}	minimum jump length
θ_j	connectivity angle
Δt	change in time
α	power law exponent constant
β	exponential cutoff constant
ρ	density of devices in the region

Table 4.1: Notations and their meaning.

structured such that it first provides the algorithm outline in Section 4.1 and the results obtained in Section 4.2. We conclude the Chapter by presenting conclusion of the study in Section 4.3.

4.1 Algorithm

In this Section we provide a method to estimate the stability of the node in the dynamic environment and how this stability can be used toward enhancing the information dissemination process in a mobile network. We first list assumptions used to build the algorithm and then propose the algorithm.

4.1.1 Assumptions

Let, initially at time $t = 0$, $|V|$ autonomous devices be distributed in a square area A, where V is the set of autonomous devices such that $|V| \in \mathbb{N}$ and $A \in \mathbb{R}^2$. In graph terminology, these devices are the nodes in the graph. Here, we consider V to be fixed. Let, each device in V have a transmission radius r such that $r > \sqrt{\frac{A}{|V|\pi}}$. We assume that the probability of connection in all the directions is 1, i.e., we consider a disk model with radius r. Note that choosing such r ensures the existence of a giant connected component (GCC) where number of devices in GCC are greater than 1 [137, 138]. Consider a device $i \in V$, for all devices in $V - \{i\}$ that have Euclidian distances less than or equal to r with i are considered to be connected to i and in the neighborhood of i. Let the set of neighbors at a given time t be denoted by N_i^t. Note that $0 \leq |N_i^t| \leq |V|$ and $N_i^t \subset V$. The connectivity of the devices in V constitutes a network (graph) G^t with $|V|$ devices and $|E^t|$ edges, where E^t is the set of edges in the graph at time t such that $|E^t| \in \mathbb{Z}^+$.

Let all the devices in V possess only local knowledge (knowledge limited to one hop neighborhood) and lack global knowledge of the network like, number of devices in the network and the location of all the devices. Let each device be memory limited. Consider, at time t, a device i possess the following information: its own location coordinates and two threshold measures, maximum stability threshold, S_{max}, and minimum stability threshold, S_{min} such that $0 < S_{max} < 1$ and $0 < S_{min} < 1$. Further, we assume that each device is equipped with an antenna array consisting of maximum M antenna elements (AE) arranged in a linear manner as in the ULA antenna model where, individually, each AE has an omnidirectional radiation pattern. An omnidirectional beam, is achieved using single AE, $m = 1$, while a long range beam is achieved when $m \in [2, M]$. As we are interested in the dissemination process and the benefit achieved while beamforming, we set the path loss exponent to 2 and assume free space propagation. Further, in this Chapter, we assume directional transmission and consider omnidirectional reception.

Let x_{min} be the minimum jump length such that $x_{min} \in \mathbb{R}^+$ and $x_{min} > 0$. Let the direction of the jump be uniformly chosen from $[0, 2\pi]$. As discussed in Section 2.7, the jump length of human mobility follows truncated power law distribution. We use this fact to state that the distribution of the jump lengths is known. Thus, we assume that x_{min}, α and β used in the eq. 2.7.1 are known where $\alpha > 0$ and $\beta \geq x_{min}$. Following this we assume that the jump lengths are sampled from the

truncated power law distribution (cf. eq. 2.7.1). Let the devices be carried by humans. This will mean that a jump made by a device also mimic jumps made by humans. We assume that these jumps are made in discrete time, i.e., a jump is made by a device $i \in V$ from a given location in $\Delta t > 0$ time causing changes in the graphical structure. Let after the jump, at time $t + \Delta t$, the device i possess information about its location at time t, its location at time $t+\Delta t$, and the acquired location coordinates of the devices in N_i^t. After the devices have moved, each device estimates its stability for the time instant $t + \Delta t$ based on the neighborhood at t. Further, we assume that the network operates in synchronous mode.

4.1.2 Stability Estimation

As terms nodes and devices can be used interchangeably, in the current Section we use the term node to represent a device. We now propose a method to estimate the stability measure. Let i^t denote the coordinates of the location of node i at time t. Let $i^{t+\Delta t}$ denote the coordinates of location of node i at time $t + \Delta t$. Following the assumptions discussed above, let the distance traveled by a node $i \in V$ in time Δt be b_i such that $b_i \geq x_{min}$. Let N_i^t be the set of neighbors of node i at time t. Let the distance between the location of a neighbor j ($j \in N_i^t$) at time t, j^t, and that of node i at time $t + \Delta t$, $i^{t+\Delta t}$, be l_j such that $l_j \geq 0$. $l_j = 0$ relate to the fact that the node i has moved to the location j^t. Let the distance traveled by the neighbor j in Δt time be a_j such that $a_j \geq x_{min}$.

In order to estimate the stability, it is essential to know if node j is connected to node i at time $t+\Delta t$. Let C be the event that node i is connected to node j at $t+\Delta t$. To calculate the probability of event C, $P_{i,j}(C)$, one requires the knowledge of l_j, r and a_j. Note that node i does have the knowledge of l_j and r. However, node i does not know the exact displacement of node j; it has only distributional knowledge of a_j. The distribution of a_j follows human mobility pattern and is given by the truncated power law distribution (cf. eq. 2.7.1 in Section 2.7).

Suppose that node i knows the exact value of a_j. We calculate the probability of event C given a_j denoted by $P_{i,j}(C|a_j)$. Consider a circle C_1 of radius a_j centered at j^t. Note that the new position of node j at time t lies on C_1. Consider another circle C_2 of radius r centered at $i^{t+\Delta t}$. Let D_1 and D_2 denote the circular disk areas covered by C_1 and C_2 respectively. Note that D_2 is the area in which if a node j, $j \neq i$, lies, then it is connected to node i at time $t + \Delta t$, because D_2 is the

transmission area of node i. For the nodes i and j to be connected with a positive probability, node j must lie in D_2 at time $t + \Delta t$, i.e., D_2 must have nonempty intersection with C_1.

First, we would like to determine the conditions when we can conclude with certainty that the two nodes i and j are connected or disconnected, i.e., when $P_{i,j}(C|a_j) = 1$ or $P_{i,j}(C|a_j) = 0$. If D_2 completely encapsulates C_1, then $P_{i,j}(C|a_j) = 1$. If the intersection of D_2 and C_1 is empty, then $P_{i,j}(C|a_j) = 0$. In all other cases, $0 < P_{i,j}(C|a_j) < 1$.

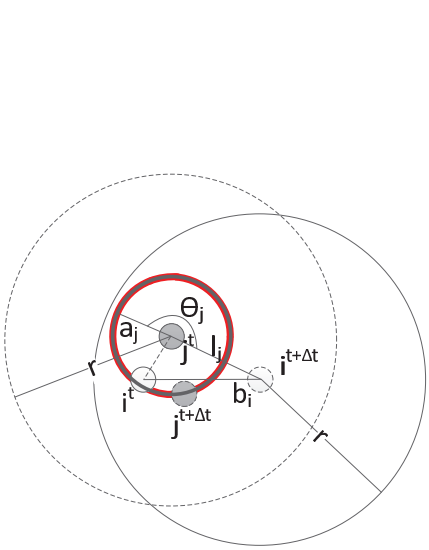
Fig. 4.1(a) depicts the scenario when D_2 encapsulates C_1. This corresponds to the situation when the transmission radius r is large enough so that all possible locations of node j fall within the transmission area of node i at time $t + \Delta t$. In this case, we have $P_{i,j}(C|a_j) = 1$.

However, if C_1 encapsulates D_2 then $P_{i,j}(C|a_j) = 0$ (cf. Fig. 4.1(c)). Further, if $D_1 \bigcap D_2$ is empty, then $P_{i,j}(C|a_j) = 0$ (cf. Fig; 4.1(e)). This means that C_1 and C_2 are far apart. In all other cases $0 < P_{i,j}(C|a_j) < 1$ (cf. Fig. 4.1(b) and Fig. 4.1(d)).

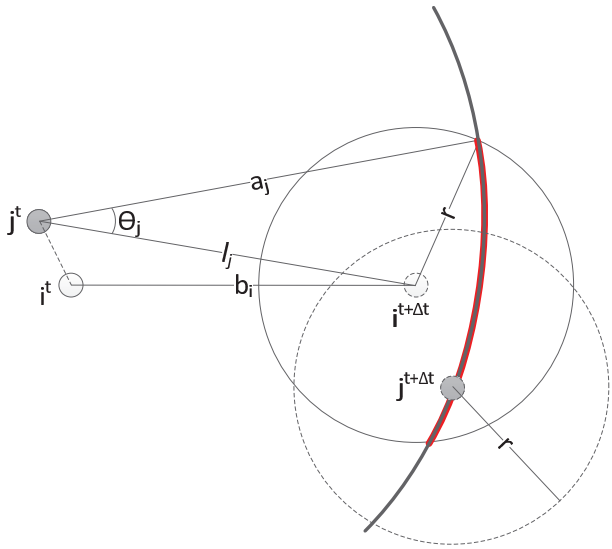
This motivates us to draw a check on the relation between a_j, l_j and r. Towards this we first compare if $a_j + l_j \leq r$. That is, if the total distance moved by the two nodes is smaller than transmission range r. If the condition holds, it means that D_2 encapsulates C_1 and thus $P_{i,j}(C|a_j) = 1$ (cf. Fig. 4.1(a)). If the condition is not true then we compare if $a_j \leq l_j + r$. This relates to the fact that the distance between the positions of i at time $t + \Delta t$ and j at time $t + \Delta t$ is less than r. If the condition is also true then there is a $P_{i,j}(C|a_j) > 0$ (cf. Fig. 4.1(b)) as j will lie within the transmission range of i. However, if $a_j > l_j + r$ then the $P_{i,j}(C|a_j) = 0$ (cf. Fig. 4.1(c)). This means that node j at time $t + \Delta t$ is beyond the transmission range of i at time $t + \Delta t$. Further, we check if $a_j + r \geq l_j$, if the condition holds then $P_{i,j}(C|a_j) > 0$ else the $P_{i,j}(C|a_j) = 0$ (cf. Fig. 4.1(d) and 4.1(e) respectively).

Let E_1 be the event such that $E_1 = \{a_j > l_j + r \ or \ a_j < l_j - r\}$ where $P_{i,j}(C|a_j) = 0$ and E_2 be the event such that $E_2 = \{a < r - l_j\}$ where $P_{i,j}(C|a_j) = 1$. For computing $P_{i,j}(C|a_j)$ when the complement of event $(E_1 \bigcup E_2)$ occurs, let θ_j denote the angle between the displacement vectors of node j and node i taking the location of node j at time t as the reference. Note that the displacement vector of node i (resp. node j) w.r.t. the location of node j at time t is given by $i^{t+\Delta t} - j^t$ ($j^{t+\Delta t} - j^t$). Using the law of cosine, θ_j is given by

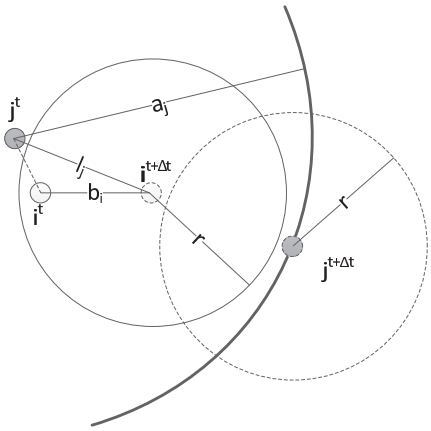
$$\theta_j = cos^{-1}\left(\frac{a_j^2 + l_j^2 - r^2}{2a_j l_j}\right). \tag{4.1.1}$$



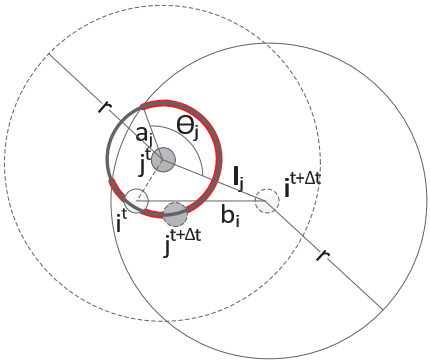
(a) Case: $a_j + l_j \leq r$.



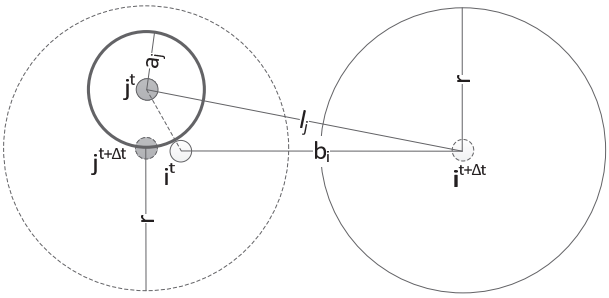
(b) Case: $a_j \leq l_j + r$.



(c) Case: $a_j > l_j + r$.



(d) Case: $a_j + r \geq l_j$.



(e) Case: $a_j + r < l_j$.

Figure 4.1: Probability of connection between the node i^t and $j \in N_i^t$ after Δt time is given by the size of the dark line with red border. The circle in bold line represents C_1, the circle in solid line represents C_2. The dotted circle is the transmission range for $j^{t+\Delta t}$.

Note that $0 \leq \theta_j \leq \pi$.

The above discussion can be summarized into the following eq.:

$$P_{i,j}(C|a_j) = \begin{cases} 0 & \text{if } E_1, \\ 1 & \text{if } E_2, \\ \frac{2\theta_j a_j}{2\pi a_j} = \frac{\theta_j}{\pi} & \text{otherwise.} \end{cases} \tag{4.1.2}$$

We now remove the conditional probability about the knowledge of a_j. In order to remove the condition about the knowledge of a_j, a distribution of a_j is needed. We remove the conditioning on a_j by using the fact that the jump lengths are distributed according to a truncated power law. Let $f_A(a_j)$ be the distribution of a_j given by eq. 2.7.1. Note that $\int_{x_{min}}^{\infty} f_A(a_j)da_j = 1$. We obtain the unconditional $P_{i,j}(C)$ as $P_{i,j}(C) = \int_{x_{min}}^{\infty} f_A(a_j)P_{i,j}(C|a_i)da_j$. This reduces to

$$P_{i,j}(C) = \begin{cases} \begin{cases} \int_{l_j-r}^{l_j+r} f_A(a_j)\frac{\theta_j}{\pi}da_j & \text{if } l_j - r \geq x_{min} \\ \int_{x_{min}}^{l_j+r} f_A(a_j)\frac{\theta_j}{\pi}da_j & \text{if } l_j - r < x_{min} \end{cases} & \text{if } l_j \geq r \\ \begin{cases} \int_{x_{min}}^{r-l_j} f_A(a_j)da_j + \int_{r-l_j}^{l_j+r} f_A(a_j)\frac{\theta_j}{\pi}da_j & \text{if } r - l_j \geq x_{min} \\ \int_{x_{min}}^{l_j+r} f_A(a_j)\frac{\theta_j}{\pi}da_j & \text{if } 0 \leq r - l_j < x_{min} \end{cases} & \text{if } l_j < r \end{cases} \tag{4.1.3}$$

Further, as the eq. 4.1.3 refers to the connection probability between i and its neighbor $j \in N_i^t$ in $t + \Delta t$, this probability directly relates to the probability of link occurrence between i and $j \in N_i^t$ at $t + \Delta t$. If $P_{i,j}(C) \to 1$, it means that the link is highly likely to exist at time $t + \Delta t$. However, if $P_{i,j}(C) \to 0$, it means that the link may not exist at time $t + \Delta t$.

The stability of node i is estimated by the ratio of the expected value of the number of links between node i and its neighbors in N_i^t that can exist at time $t + \Delta t$ over the number of neighbors

that existed at time t. Thus, we define stability for a node i as

$$S_i^* = \frac{\sum_{j \in N_i^t} P_{i,j}(C)}{|N_i^t|}. \tag{4.1.4}$$

Using the eq. 4.1.4, all the nodes in the network perform the stability estimation task independent of each other. This stability estimation does not require any global knowledge and can be achieved using local information available to the node.

4.1.3 Beamforming using S_i^*

As stated before, only few nodes with beamforming capabilities are needed to enhance the dissemination process [112]. In order to determine the few beamforming nodes, we use the above proposed stability measure. Once a node i measures its stability, it then decides whether to beamform or not to beamform. If $S_i^* < S_{min}$ then i decides to switch its omnidirectional beam to directional beam. In order to let few nodes beamform we let S_{min} to be a small. A $S_{min} = 1$ will allow all the nodes to beamform, will cause lot of asymmetric links and a loss in clustering coefficient. A $S_{min} = 0$ will allow no node to beamform. Hence, in order to have only few beamforming nodes S_{min} should be small. An interpretation for choosing nodes that have $S_i^* < S_{min}$ to beamform in terms of information dissemination is: such nodes are more likely to have information from much bigger set of neighbors (case (ii) of $S = 0$ and case (i) of $0 < S < 1$, as discussed at the beginning of the Chapter). However, note that due to the definition of S_i^* some nodes might have undefined stability because $|N_i^t| = 0$. We let such nodes to beamform also. The idea behind such a rule is, at time t a node i might be the source of information. As $|N_i^t| = 0$, node i will not be able to transmit the information until it gets connected to some other node at some time. Letting such node to beamform will ensure the information propagates from the node early.

After the decision of whether to beamform or not, it is necessary to determine which is the best direction to beamform and what are the beam properties (length and width). In the previous Chapter, we stated that there are some preferred nodes to which when beamforming is done network average path length can be reduced hence requiring less time to disseminate the information. In order to determine preferred nodes, we state that the nodes that have $S_i^* > S_{max}$ are the preferred nodes. As from the discussion in the beginning of the Chapter, when $S = 1$, (i) the node is still

Beamform From		Beamform To	
Stability	**Degree at t**	**Stability**	**Degree at t**
Low	Low/High	High	Low/High
Low	Low/High	Undefined	Zero
Undefined	Zero	High	Low/High
Undefined	Zero	Undefined	Zero

Table 4.2: Beamforming strategies.

connected to all of its neighbors at time t and no new node has become its neighbor and (ii) the node is still connected to all of its neighbors at time t and new neighbors have joined. Further, when $S_i^* > 0.5$ it implies more than 50% of its neighbors at time t are still neighbors. Beamforming to nodes that have $S_i^* > S_{max}$ will let nodes with $S_i^* > S_{max}$ to get information from a bigger set of nodes thereby making them aware of more information flowing in the network. As the beamforming will increase the in-degree of the preferred nodes, the preferred nodes will be able to disseminate more information. Also, if the S_i^* is undefined, it means that $|N_i^t| = 0$. Further, we let such nodes to also be the preferred nodes.

Definition 4.1.1. In a directed graph, in-degree of a node is the number of connections made to the node.

For a more clear understanding in Table 4.2 we present all possible combinations of stability measures for the nodes that should beamform and the nodes to which the beams should be directed. Here we add the degree of a node i at time t to help explain the beamforming strategy.

Note that in a dynamic scenario the edges in the network are intermittent. The choice of beamforming node will change every time as it depends on the stability measure. This will cause different set of nodes to beamform at a given instant.

In the algorithm, we assume that m *AE*s are randomly selected from the available M *AE*s by the beamforming nodes. In order to determine the direction of the main beam, B_b, the node sweeps 2π angles and determines the best direction as the direction in which maximum number of *beamforming to* nodes are found. While determining the best direction, we also condition that the *hopcount* between *beamforming to* nodes and beamforming nodes should be greater than 1. If *beamforming to* nodes are one hop away, beamforming will have no affect on enhancing the information dissemination process and will increase the network path length by reducing the neighborhood. If the node i is not able to find any *beamforming to* nodes more than one hop away, the node reverts back to the

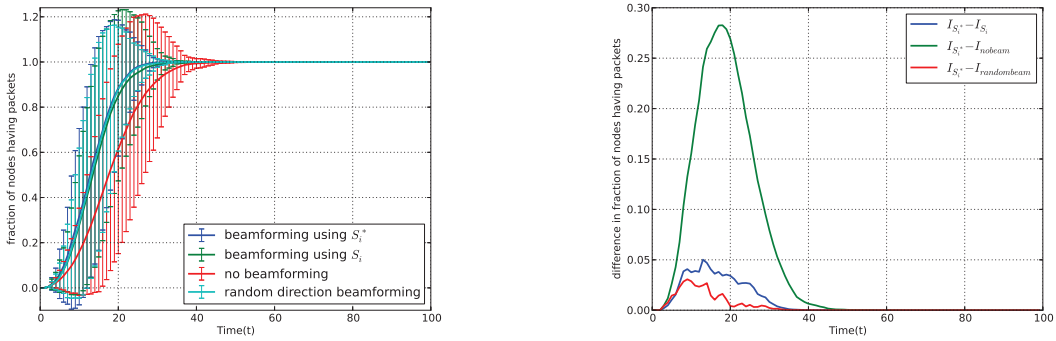
omnidirectional case.

Further, beamforming makes the network asymmetric. A node will be able to transmit the data but will not be able to receive the ACK for the transmission made. Also beamforming nodes will not be able to know if there are any *beamforming to* nodes in the region of directional beam. As we are just concerned about the dissemination of the information in a broadcast medium, the problem of asymmetry is of less importance. For the second problem, same as Chapter 3, we assume that the *beamforming to* nodes, just to let the beamforming node know about the connection setup, creates the beam in the direction of the beamforming node and then reverts back to omnidirectional scenario. Also note that as the network is directed, degree of a node can not be used to estimate the stability value. Towards this, we use in-degree of the node to compute stability.

4.2 Simulation Results

We use a simulation area of $A = 500$x500 to simulate the algorithm. We set the average density of nodes, ρ, to 1x10^{-3}. Initially, each node operates in omnidirectional mode using $m = 1$ AE with the omnidirectional radius as $r = 30$. We set the maximum number of AEs that the nodes are equipped with to $M = 6$. The separation between two AEs is computed using $WiFi$ frequency, $f = 2.4Ghz$. We set $\alpha = 1.6$, $\beta = 300$, $x_{min} = 1$, $S_{min} = 2$x10^{-3}, $S_{max} = 0.7$ and use a ULA antenna model to simulate the algorithm. Through the simulations, we explore the effect on time taken to disseminate the information in the network using various scenarios.

We validate the results and provide comparison between four different cases: (i) when a node i uses S_i^* with beamforming strategy as discussed in Section 4.1.3; (ii) when a node i uses stability measure S_i as discussed in Section 2.8.6 instead of S_i^* with beamforming strategy as discussed in Section 4.1.3; (iii) when a node i does not use beamforming; and (iv) when choice of beamforming nodes and the beam direction is random but the number of nodes that beamform is same as that found by the algorithm. We refer to these cases as Case A, Case B, Case C and Case D respectively. Note that mobility is applied in all the four cases. For all the above mentioned cases we then simulate four different scenarios. These include: (i) single source having single packet to transmit; (ii) single source sending update information; (iii) multiple sources with each having single packet to transmit; and (iv) multiple sources with multiple packets where sources can join over time. Using these cases



(a) fraction of nodes having packet for different cases.

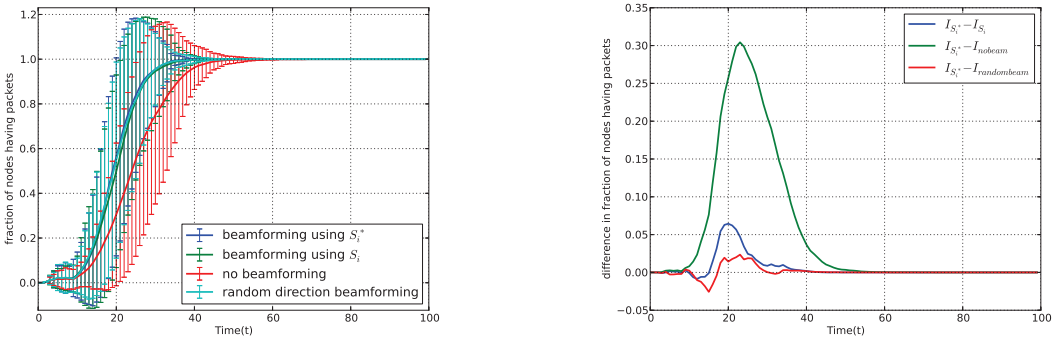
(b) difference in fraction of nodes having packet.

Figure 4.2: Scenario when only one node acts as a source and has a single packet to transmit.

and scenarios we study the fraction of nodes having information at a given time and compare the differences. We also show the effect of various parameters on the time taken to disseminate the information in the network. We use Python to simulate the algorithm. We average the results for these cases and scenarios over 50 topologies for a $T = 100$. We use a confidence interval of 95%.

We first show the results obtained for the scenario when there is only one source having only one packet to transmit. In this scenario, a node is first randomly selected to be the source of information. The selected node then generates a packet at $t = 0$ (start of the simulation). Fig. 4.2 shows an improvement achieved when the algorithm is used. The results show that in almost 30 time steps all nodes in the network have the packet when beamforming is done using Case A. We then plot the difference in the fraction of nodes having packets for different cases. Let $I_{S_i^*}$ be the fraction of nodes having packet at at a given time for Case A. Let I_y be the fraction of nodes having packet at at a given time for other cases. Let $y \in \{S_i, nobeam, randombeam\}$. Note that y denotes cases B, C and D respectively. Thus we plot, $I_{S_i^*} - I_y$ (cf. Fig. 4.2(b)). Fig. 4.2(b) shows the maximum improvement achieved over cases B, C and D is 0.05, 0.282 and 0.031, respectively. Note that as the scenario is single packet dissemination in the network, random direction beamforming performs equally well.

In the above scenario, the dissemination process starts as soon as the packet is generated. However, when an update is generated, the previous packet is now obsolete. All the nodes that received the first packet will now be considered as nodes not having the packet. The dissemination process



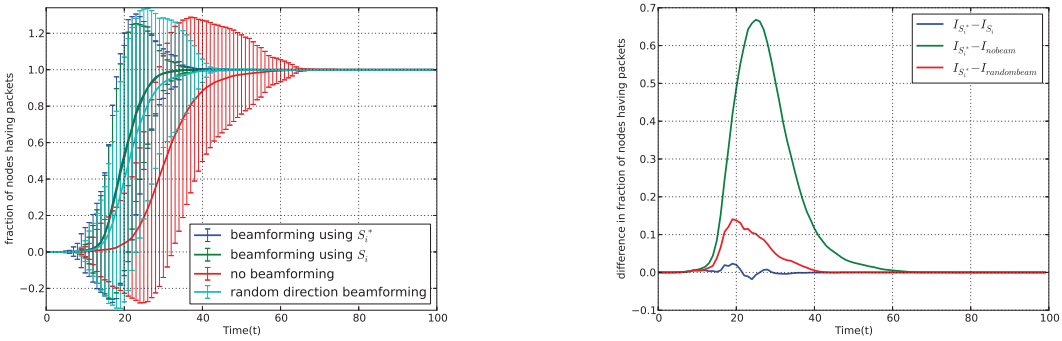
(a) fraction of nodes having packet for different cases.

(b) difference in fraction of nodes having packet.

Figure 4.3: Scenario when only one node acts as a source and has an update packet to transmit.

will thus have a new initial starting condition. If the source node generates an update after t_x time then the dissemination process effectively starts from t_x. Further, if an update is again generated after another t_x time, the effective starting point will now be $2t_x$. Fig. 4.3(a) is generated when there is only single source having an update packet to be transmitted in the network. Fig. 4.3(a) also shows an improvement achieved in this scenario. In this scenario, the initial packet is generated at $t = 0$ (start of the simulation) while the last update is generated at $t = 10$. Here time taken to completely disseminate the update has a shift by ≈ 10 time steps than the previous scenario, thereby confirming our explanation above. However, other observations for the fraction of nodes having the packet after $t = 10$ remains same to that of the previous scenario. From $t \in [0, 10]$, there is a hill because the update packet is randomly generated between $t \in [0, 10]$. When an update packet is not generated, the network assumes that there is no update packet in the network and the dissemination of the previously generated packet occurs. Similar to the previous scenario, we also plot $I_{S_i^*} - I_y$ (cf. Fig. 4.3(b)). Fig. 4.3(b) shows the maximum improvement achieved over cases B, C and D is 0.065, 0.305 and 0.024, respectively.

Further, we now assume that there are multiple sources each having single packet to disseminate in the network. This will increase the diversity of information in the network. We set the number of sources to be 40 for the next two scenarios dealing with multiple sources. Similar to the scenario of single source, all the sources generate the packet at $t = 0$. For this scenario, we define the fraction of nodes having the information as the set of nodes that have received all the packets from all the



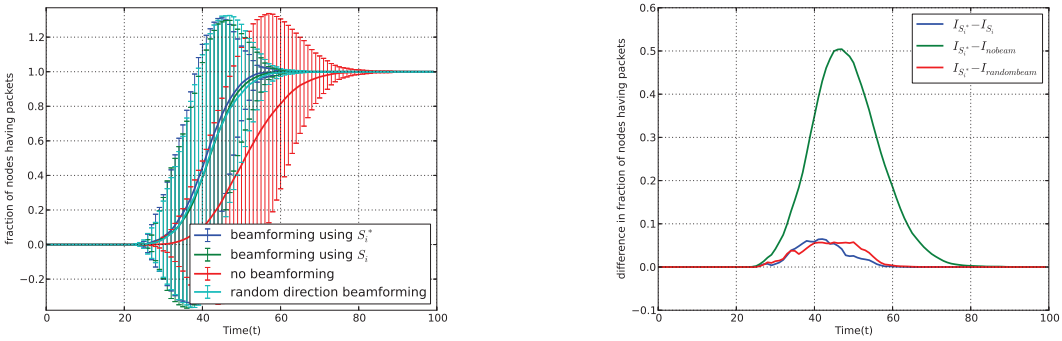
(a) fraction of nodes having packet for different cases.

(b) difference in fraction of nodes having packet.

Figure 4.4: Scenario when multiple nodes act as source and each has single packet to transmit.

sources in the network. Consider $n = 2$ sources each having one packet to transmit at time $t = 0$. The time taken for all nodes to receive information from all the sources is $\max\{T_1, T_2\}$, where T_1 and T_2 are the time taken for packets from source 1 and source 2. This is generalized for n sources where the total time taken for all the nodes to receive the information from all the sources is given by $\max\{T_1, T_2, \ldots, T_n\}$. The results show that in ≈ 38 time steps the packets from all the sources can be disseminated in the network when beamforming is used using S_i^*. However, only $\approx 83\%$ nodes have received all the packets for Case C (cf. Fig. 4.4(a)). As there are multiple sources, the fraction of nodes having all the packets has a shift in time. It takes ≈ 6 time steps for one node to have all the packets from all the sources for the beamforming case while ≈ 9 time steps when beamforming is not used. The time taken for all the nodes to receive the packets from all the sources, however, is ≈ 36 time steps for Case A. While only $\approx 77\%$ nodes received all the packets in 36 time steps for Case C. We now show the difference in the fraction of nodes having packets, $I_{S_i^*} - I_y$, for different cases for this scenario (cf. Fig. 4.4(b)) as a representation for Fig. 4.4(a). Fig. 4.4(b) shows the maximum improvement achieved over cases B, C and D is 0.0256, 0.665 and 0.141, respectively.

In the above scenario number of sources were fixed and no other source was allowed to join the dissemination process. We now relax this assumption and allow sources to join as time progresses. Similar to the previous scenarios, we define the fraction of nodes having the information as the set of nodes that have received all the packets from all the sources in the network. Fig. 4.5(a) is generated towards such scenario. We set a time limit until sources can join to $t = 20$ so that we can obtain



(a) fraction of nodes having packet for different cases.

(b) difference in fraction of nodes having packet.

Figure 4.5: Scenario when multiple nodes act as source and each has multiple packets to transmit. Sources can join as time ticks.

the results in desired time. Here each source i randomly generates n_i packets until $t = 25$ as in the previous scenario. The results show that in ≈ 60 time steps all packets can be disseminated in the network when beamforming is used with S_i^*. However, only $\approx 81\%$ nodes have received all the packets for Case C (cf. Fig 4.5(a)). When compared to the results obtained for multiple source single packet scenario, (cf. Fig. 4.4(a)), there is a shift of almost ≈ 20 time steps before at least one node can have packets from all the sources for Case A. This is because, sources are being added between $t \in [0, 20]$ and last packet is generated at $t = 25$. We now show the difference in the fraction of nodes having packets, $I_{S_i^*} - I_y$, for different cases for this scenario (cf. Fig. 4.5(b)) as a representation for Fig. 4.5(a). Fig. 4.5(b) shows the maximum improvement achieved over cases B, C and D is 0.068, 0.506 and 0.0597, respectively. Note that, when the population is large and spread in larger area (as in real country wide population), these differences on the actual population scale can be large. Further, results obtained for the multiple source scenario suggest that beamforming using S_i^* performs better than random direction beamforming.

In the above four scenarios, the fraction of nodes that beamform is shown in Fig. 4.6 for Case A. From the results it is noted that for different scenarios the fraction of nodes beamforming remain mostly in the range 0.043 ± 0.006 (cf. Fig. 4.6). This is because the selection of the beamforming nodes is independent of the number of sources in the network.

We next provide results for varying ρ, α, β, r, M, S_{min}, S_{max} and antenna type when there are

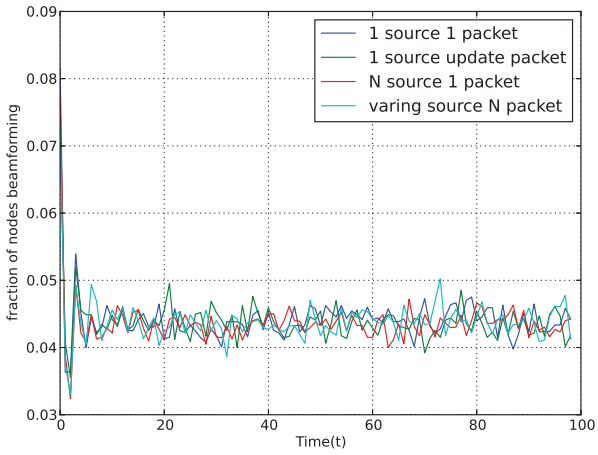


Figure 4.6: fraction of nodes beamforming for different scenarios for Case A.

multiple sources that have multiple packets to transmit and sources can join. Through this set of simulations we show the effect of the aforementioned parameters on the information dissemination process when the algorithm is used. We first provide the effect of varying ρ on the dissemination process. Note that all other parameters are kept constant. As ρ increases, the number of time steps taken to disseminate the packet reduces, (cf. Fig 4.7(a)). This is because (i) when ρ is high more nodes are in the giant connected component and the connectivity is high and (ii) the average number of connections for a node is higher. This enables a node to forward the information to more nodes.

Varying α, it is observed that when α is high the time taken to disseminate the information in the network also increases, (cf. Fig. 4.7(b)). This is because as α increases the number of long jumps decreases and there are more short jumps. When jumps are short devices are not able to meet other devices as frequently as when α is less. Varying β, variations in time taken for all the nodes to receive the information is shown in Fig. 4.7(c). For small β, i.e. when the jump length cutoff is small, the nodes are not able to move far. This affects the nodes meeting different nodes in the network. Thus more time is required to disseminate the information in the network. While when β is high, the nodes are able to move far and meet more nodes and disseminate the information in the network much faster. Note that in the above two sets of simulations all the other parameters are kept constant.

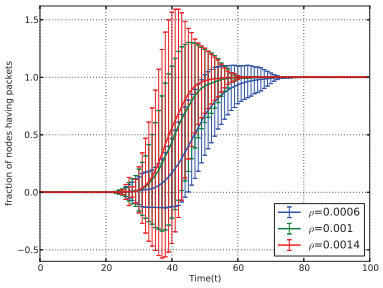
Parameter		Time taken
	value	**value**
ρ	$\uparrow$	$\downarrow$
α	$\uparrow$	$\uparrow$
β	$\uparrow$	$\downarrow$
r	$\uparrow$	$\downarrow$
m	$\uparrow$	$\downarrow$
S_{min}	$\uparrow$	$\downarrow$
S_{max}	$\uparrow$	$\uparrow$
Number of sources	$\uparrow$	$\uparrow$
Number of packets	$\uparrow$	$\uparrow$
Other Parameters		
Stability measure type		different
Type of antenna array		different

Table 4.3: Factors affecting the dissemination of packet. The table shows the effect on the time taken to disseminate the information in the network completely when the parameter value increases.

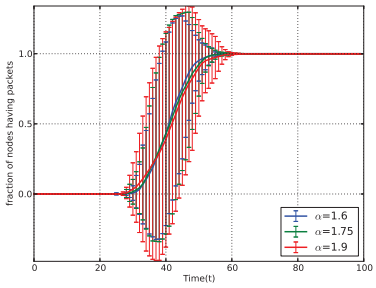
Next we show the results obtained when r and M vary. Results clearly state that when r is high, more nodes are connected and the dissemination of packets is faster, (cf. Fig. 4.7(d)). Similar results are obtained for varying M (cf. Fig. 4.7(e)). When M is high the beam length is large and thus more distant nodes can be connected.

Variations in S_{min} and S_{max} affect the number of nodes beamforming and to whom they beamform. Low value of S_{min} reduces the number of nodes beamforming while a high value increases that number. This affects the information dissemination in the network, (cf. Fig 4.7(f)). However, low value of S_{max} increases the number of nodes to which the nodes can beamform to, while a high value decreases that number. This affects the dissemination process (cf. Fig 4.7(g)).

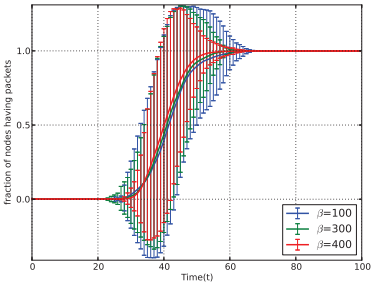
Next we show results obtained when different types of antennas are used. We use two different types of antennas, namely, ULA and a theoretical antenna model (Sector Model). ULA achieves better results than the Sector model because in ULA there is a secondary lobe with substantial beam length and width, (cf. Fig 4.7(h)). This increases the connectivity of the node as compared to when the Sector model is used. These results lead us to summarize the effect of each of the parameters, described above, on the information dissemination in Table 4.3.



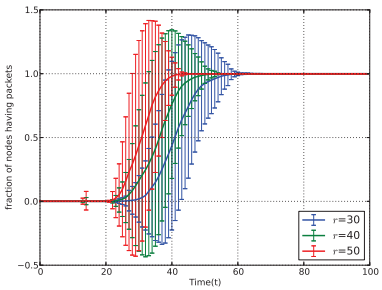
(a) for variable ρ.



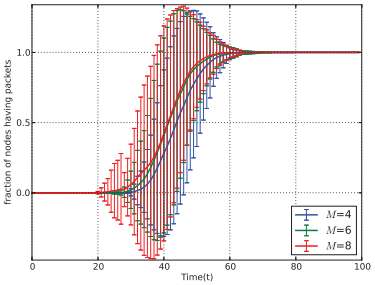
(b) for variable α.



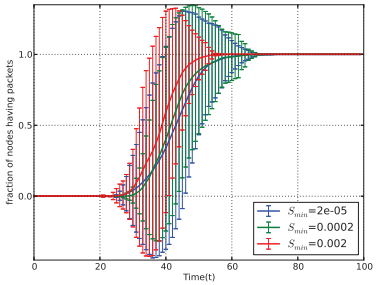
(c) for variable β.



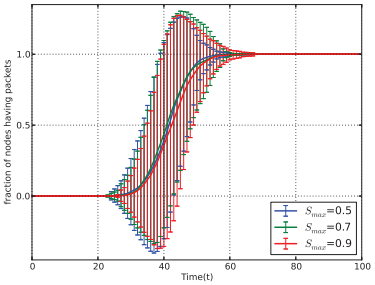
(d) for variable r.



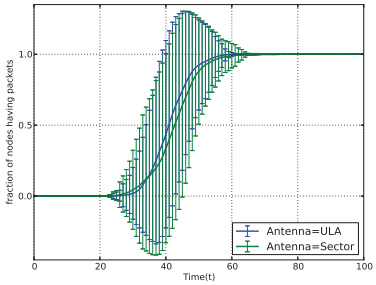
(e) for variable M.



(f) for variable S_{min}.



(g) for variable S_{max}.



(h) for different antenna models.

Figure 4.7: fraction of nodes having packets when multiple nodes act as source and have multiple packets to transmit.

4.3 Conclusion

In this Chapter, we propose an algorithm that can enhance information dissemination in terms of number of nodes having the information in less time in mobile scenarios. There are three main contributions in this Chapter. First, in this Chapter we propose a probabilistic approach to compute the stability of a node in a mobile environment. Second, based on the stability measures beamforming is performed and we show that large number of nodes can receive the information quickly in the mobile environment when compared to other related studies. Third, the algorithm operates in an environment where nodes do not require any global knowledge of the network, takes decisions autonomously and requires less memory to operate.

The results show an improvement in the information dissemination process when the proposed stability measure is used over three different cases including the case when there was no beamforming and when nodes were randomly chosen to beamform. Moreover, our proposed stability measure does not require the entire history of links. It only requires latest link status, i.e. links at time t to compute stability at time $t+\Delta t$. The algorithm works in fully decentralized manner with no entity controlling the decisions made by the nodes. The results show that information dissemination is affected by many parameters, such as number of sources, number of packets, mobility parameters, density, radius of communication, number of antennas, antenna type, stability parameters and stability type used.

CHAPTER 5

Enhancing Information Dissemination in Device to Device Metapopulation using Human Mobility Trace and Beamforming

In communication networks devices in each other's physical proximity can communicate. Information dissemination in such proximity based networks have been the focus of a lot of studies. Through this Chapter, we are interested in how safety information can be quickly disseminated in a countrywide network with the help of Device-to-Device ($D2D$) communication. $D2D$ communication is defined as a short range communication between devices in physical proximity without any involvement of the network infrastructure. $D2D$ has many advantages like, autonomous communication, improved performance and spectrum reuse, low energy consumption and reduced load on the infrastructure. Since, the interest lies in the dissemination process on a countrywide scale, mean field approximation is best suited to model the dissemination process in the contact network within the country. The information dissemination process can be closely related to the study of epidemic spreading across the population through Susceptible-Infected-Recovered (SIR) model. A comprehensive survey about the epidemic model is provided in [139,140]. In the context of message dissemination in a communication network, the S (Susceptible), I (Infected) and R (Recovered) states correspond to the state of a device 'not having the message', 'having the message and ready to transmit' and 'having the message but no longer willing to transmit the message to other devices to save energy', respectively. In this context, it has been shown that information dissemination is influenced by many factors like communities in the network, bursty data [135], strength of the connection [105], source of the infection, human mobility parameters (Chapter 4), location preferences [141], network structure [106], activity pattern [107], altruism [110], social measures [142], antenna parameters as discussed

in Chapter 4 and other device characteristics [109]. However, as noted in [7], the characteristics of the human mobility impact the dissemination process the most when compared to others.

For a countrywide scale mobility pattern, we use the data provided by the Data for Development ($D4D$) group and Orange Labs [10]. The analysis of the dataset reveals that in a large population, communities are formed. The density of the population per community changes over time due to the movement of individuals from one community to another. Such an organization of a large population into community structure where the interaction between the communities is based on the mobility, has already been studied through a synthetic model of population called metapopulation [11, 12]. It has also been discussed that communities formed in a large population affect the dissemination process. An individual, if restricted to a community, will not be able to move freely to other communities. This will affect the message dissemination process. For example, the movement of individuals from one country to another country is restricted by the border controlling officers. Fig. 5.1 shows the difference in the number of people having information when c communities are present and when there is only one large community. The results are normalized by the total number of people N. Initially, a large difference in the curves is noticed because more people meet other people when they are not restricted to certain community and thus the dissemination process is fast. Also, the rate of spreading information in each community is different. Some communities are more resistant to the dissemination process while some are not. We assume similar features in our model too. Recently, in [143], it is shown that not only the community structure affects the dissemination process but the density of the communities also has an important role in the dissemination process.

In communication networks, dynamic density can be achieved using two scenarios. First, addition and removal of devices and second, sleep and wake process. We are interested in investigating how the information dissemination process at the countrywide scale is impacted and how it can be enhanced in a variable density scenario due to sleep and wake type variation alone (not considering addition and removal of devices). To capture the variable density effect, we use the Latent state (E). To distinguish between devices that are Latent but Susceptible, Latent but Infected and Latent but Recovered, we subdivide E into three states: E_S, E_I and E_R. Devices in E_S, E_I and E_R states do not participate in the dissemination process. Only active devices in S, I and R states participate

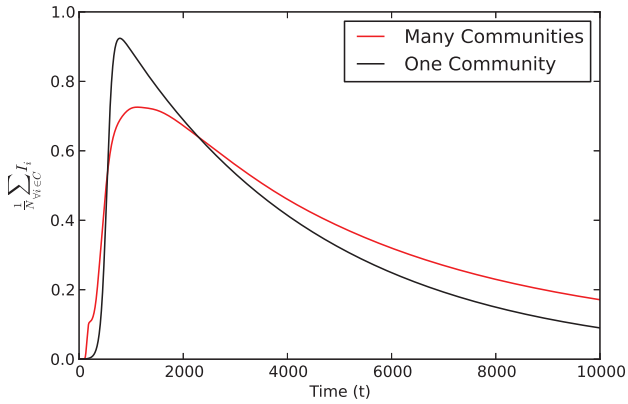


Figure 5.1: The evolution of the number of devices having information when the region is sub divided into many small communities and when the region has only one community. The results are obtained using the $D4D$ dataset and SIR type epidemic model.

in the dissemination process. Further, devices in E_S, E_I and E_R can later change their states and become active.

The datasets provided by the $D4D$ group includes the information about the user mobility and the association of a user to a community, termed as subprefecture, in the region of Ivory Coast. In order to study the model on a countrywide population, we use the population data provided in [144] to populate the subprefectures in the country of Ivory Coast. In Ivory Coast, a group of subprefectures form a department, collection of these departments form a region and collection of these regions form the country. In this Chapter, however, we are more interested at the subprefecture level as the dataset contains the mobility patterns at that level. More details on how mobility patterns are calculated using the dataset are provided in Section 5.2.2 and in Appendix B.

Based on the context and proximity information, a device discovers other devices in its physical proximity and establishes a link with them [145]. In some situations, a device acts as a relay in a multi-hop $D2D$ communication scenario. This increases the network coverage because the information can now be transmitted to the devices beyond the coverage zone and even in a poor cellular coverage zone. Based on the context, devices can also form groups, where a device can act as a leader or the member of a group. Using collaborative forwarding mechanisms, the devices within a group share data efficiently for optimized resource utilization and higher throughput [146].

A device in a network, however, can be equipped with an antenna array with multiple antenna elements embedded in some geometrical pattern, e.g., linear or circular. When a device uses multiple antenna elements, larger transmission range with certain angular coverage is achieved due to constructive and destructive interference in the radiation from the antenna elements. The technique whereby multiple antenna elements are used to achieve larger transmission range is known as beamforming. The effects of beamforming have already been studied on the information dissemination process for both static and mobile networks with positive results [112, 136] (Cf. Chapters 3 and 4) and considerable enhancement in the dissemination process is achieved with a few devices beamforming. We use the idea of a few devices beamforming, mobility and the proposed six state model, and show how it can help in enhancing the dissemination process.

In a human population, a user is associated with a home location. Every time a user moves away from its home location it has a rate to return to its home location from its new location. We thus update the mobility model to incorporate rate of return for a user to return to its home location. This return rate is also derived from the *D4D* dataset. More details on how the return rates are computed are mentioned in Section 5.2.2. Using two different mobility models (with and without home locations) we then show the effect of the mobility models on the dissemination process.

Thus, in this Chapter, we use insights from the dissemination in the metapopulation model, data provided by the *D4D* organizers, epidemic model and beamforming to develop the analytical model. A table of notations and their meanings as used in this Chapter, hereafter, is provided as Table 5.1. The remainder of this Chapter is organized as follows. In Section 5.1 we first provide an overview of the *D2D* scenario. In Section 5.2 we describe how the dataset provided by the *D4D* organizers is used to get useful information. A detailed description of the model is then provided in Section 5.3 which is followed by the results obtained in Section 5.4. The Chapter finally concludes with Section 5.5.

Notation	**Meaning**
$Area$	area of the country.
A_i	area of community $i \vert i \in C, \sum_{\forall i} A_i = Area$
$A_{b,q}$	Area under the beam for the beamforming device q
B_b	boresight direction
C	set of communities or subprefectures

Continued on Next Page...

Table 5.1 – Continued

Notation	Meaning
MD	device density matrix
E_{I_i}	number of devices in the community $i \vert i \in C$ that are latent, contain the information and will transmit it when becoming active
E_{S_i}	number of devices in the community $i \vert i \in C$ that are latent and do not contain the information
E_{R_i}	number of devices in the community $i \vert i \in C$ that are latent, contain the information and will not transmit it
$E_{I_{ij}}$	number of devices in the community $i \vert i \in C$ that are latent, contain the information, will transmit it when becoming active and have home community $j \vert j \in C$
$E_{S_{ij}}$	number of devices in the community $i \vert i \in C$ that are latent, do not contain the information and have home community $j \vert j \in C$
$E_{R_{ij}}$	number of devices in the community $i \vert i \in C$ that are latent, contain the information, will not transmit it and have home community $j \vert j \in C$
G_c	complete graph obtained by joining all subprefectures
G_v	graph obtained after delaunay triangulation of subprefectures position
g_{q,θ,B_b}	gain achieved using antenna array
H	home community
I_i	number of devices with information those who will transmit the information in the community $i \vert i \in C$
I_{ij}	number of devices with information those who will transmit the information in community $i \vert i \in C$ that have home community $j \vert j \in C$
M	maximum antenna elements available
MP_v	conditional probability matrix for movement on G_v
MP_c	conditional probability matrix for movement on G_c
MR_v	movement rate matrix on G_v
MR_c	movement rate matrix on G_c
MRR_v	return rate matrix on G_v
MRR_c	return rate matrix on G_c
m	number of antenna elements used by the device to beamform $\vert m \in [2, M]$
N_{ij}	number of devices that have home community as i but are in community $j \vert i, j \in C$
N_i^*	steady state population of community $i \vert i \in C$
N_i	initial population of community $i \vert i \in C$
N	total population considered

Continued on Next Page...

Table 5.1 – Continued

Notation	Meaning
P	sum of all P_u
P^d	weight matrix for calculating density
P_u	weight matrix for user u
$(P_u)_{ij}$	the number of times a user u traversed through edge between node i and j. We also call it as the weight of the edge
P_u^r	weight matrix for user u for calculating rates
$(P)_{ij}$	weight of an edge between i and j
R_i	number of devices with information those who will not transmit the information in community $i \vert i \in C$
R_{ij}	number of devices with information those who will not transmit the information in community $i \vert i \in C$ that have home community $j \vert j \in C$
r	transmission radius
S_i	number of devices without information in community $i \vert i \in C$
S_{ij}	number of devices without information in community $i \vert i \in C$ that have home community $j \vert j \in C$
SP_{ij}	shortest path between community i and $j \vert i, j \in C$
$Time$	duration of SET3
U	set of subprefecture entries related to the user u
u	a user
u_i	subprefecture traversed by u
X	set of beamforming devices
x	beamforming device $\vert x \in X$
(x, y)	location coordinates of the central location in the subprefecture region
α_{S_i}	transition rate from state E_{S_i} to S_i
α_{I_i}	transition rate from state E_{I_i} to I_i
α_{R_i}	transition rate from state E_{R_i} to R_i
β_i	transition rate from state S_i to I_i or in other words transmission rate in the community $i \vert i \in C$
γ_i	transition rate from state E_{I_i} to R_i
δ_i	transition rate from state I_i to R_i or in other words recovery rate in the community $i \vert i \in C$
ε	initial number of nodes in state I
ζ_{ji}	rate of return from community j to community $i \vert i \in C$

Continued on Next Page...

Table 5.1 – Continued	
Notation	**Meaning**
μ_{S_i}	transition rate from state S_i to E_{S_i}
μ_{I_i}	transition rate from state I_i to E_{I_i}
μ_{R_i}	transition rate from state R_i to E_{R_i}
ν_{ij}	conditional probability of moving from community i to $j \vert i, j \in C$
ρ_i	density of devices in community $i \vert i \in C$
σ_i	rate of moving out of community $i \vert i \in C$
θ	direction
$\langle k_i \rangle$	expected number of connections for a device in community $i \vert i \in C$
$\langle k_i^* \rangle$	expected number of connections for a device in community $i \vert i \in C$ when beamforming is applied

Table 5.1: Notations and their meaning.

5.1 Loosely Controlled *D2D* Support for Information Dissemination

We assume the following *D2D* scenario. Using a Public Warning System (*PWS*), a national emergency warning has to be broadcast by some *eNB* (evolved Node B, a radio interface in Long Term Evolution (*LTE*) network) and should reach the maximum number of people in the country. However, in such situations, the operator's network is massively overloaded by a huge traffic of users trying to use their mobile devices at the same time. Also, the warning message may not reach a part of the population due to network coverage issues in rural zones. Using *D2D* communication, devices communicate directly with other devices in the physical proximity and transmit the information. This mechanism not only allows the extension of the broadcasting area but also enhances the traffic on the network infrastructure and avoid network saturation and waste of resources.

A *D2D* communication has two phases: the Neighbor Discovery phase and the Communication phase for the data exchange. A neighborhood discovery can be based on either direct discovery or a network assisted discovery. In some studies related to *D2D* communication systems, these discovery models are also called distributed and centralized discovery, respectively [147, 148].

In the distributed approach, each device sends beacons to all devices in its surrounding to an-

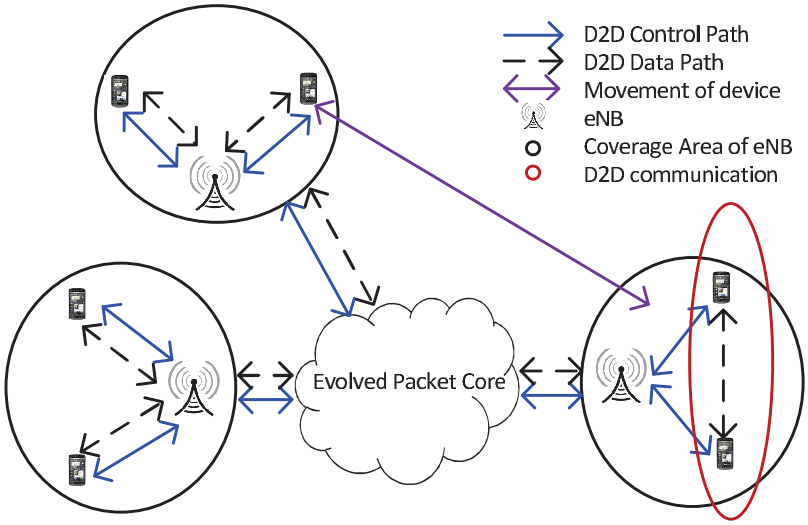


Figure 5.2: $D2D$ communication diagram.

nounce its presence. This approach is highly scalable as the operator has less control on the communication between the devices and does not control the QoS of the communication. This approach is also a solution to avoid communication disruption when some eNBs are down in a disaster situation like Earthquake or when an eNB fails. The users could still use their devices and be connected to the operator network using direct communication. However, in such $D2D$ loosely-controlled mode, operators could face several issues like performing lawful interception, ensuring the security of the data transfer when devices are exchanging the information directly and whether the communication and discovery should be allowed on the licensed or unlicensed spectrum [147, 149].

In the centralized approach, however, some entities in the operator core network could be responsible for the peer discovery phase and the communication phase. For instance, the eNB can be informed about the devices that are in the physical proximity and could provide the devices with the necessary authentication and resource configuration to setup a $D2D$ link. Nevertheless, this approach is less scalable than the distributed approach as the network could face load balancing issues in handling huge number of $D2D$ links.

Using a loosely controlled $D2D$ mechanism, some eNBs can allow direct communication between devices that are in physical proximity. Fig. 5.2 shows, some eNBs allowing the direct communication between devices that are in physical proximity while some do not. Fig. 5.2 also shows that a device could move from one eNB to another. However, the device could also move within the associated eNB. In this Chapter, we assume that all eNBs allow direct communication between devices that

are in physical proximity. Let a public warning message be generated in some area associated to an eNB. In this Chapter, we are interested in investigating how fast dissemination of the warning message can be achieved through a large population in a $D2D$ environment where the devices are also mobile and links are intermittent. By fast dissemination we mean how, after the warning message is generated, a large set of devices can quickly receive the warning message.

5.2 Data

In this Section we extract useful information from the data collected by Orange Labs for the region of Ivory Coast. The data is based on the calls made by using mobile devices in the region of Ivory Coast [10]. Ivory Coast is divided into a number of regions which are subdivided into departments where departments are again subdivided into subprefectures. A subprefecture contains antennas. The $D4D$ dataset provided to us as part of a research project has the following structure. It contains the central location of the subprefecture area and the position of antennas in longitude and latitude format along with four sub-datasets (SET1, SET2, SET3 and SET4) out of which we are interested only in SET3 which contains mobility information. SET3 contains information of 500,000 users over 5 months from 5th December 2011 until 28th April 2012. Each tuple in SET3 contains subprefectureID from where the call was made, `time` when the call was made and userID of the users who made the calls. Mobility of a user from this dataset can be estimated. Consider a user u. A query "SELECT subprefectureID from SET3 where userID=u" on the SET3 results into a list of subprefectureIDs that the user made calls from. The list may have different subprefectureIDs. Whenever a change in the subprefectureID is observed in the list, it is said that the user has moved to the new subprefecture and has called from there. Due to this reason SET3 is also called mobility dataset by the $D4D$ Organizers. Further, as the actual coordinates of a user are missing and only subprefecture coordinates are available, the mobility of the users can only be determined approximately.

Another dataset (SET2) contains similar information as SET3. SET2, however, is much smaller as compared to SET3 and contains relatively fine grain information than SET3. SET2 contains tuple with information antennaID from where the call was made, time when the call was made and userID of a person who made the call. However, SET2 contains information for only 50,000 users

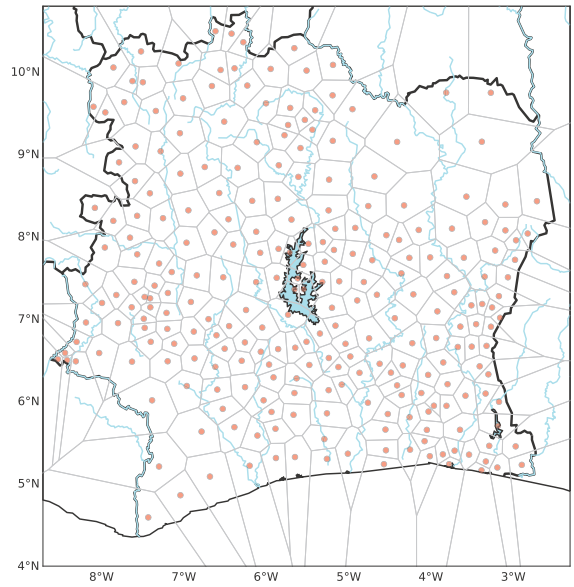
over 5 months from 5th December 2011 until 28th April 2012. Thus, we are interested in SET3 as it is bigger and is well complemented by the subprefecture location dataset (SPLoc).

5.2.1 Dataset Subprefecture Locations (SPLoc)

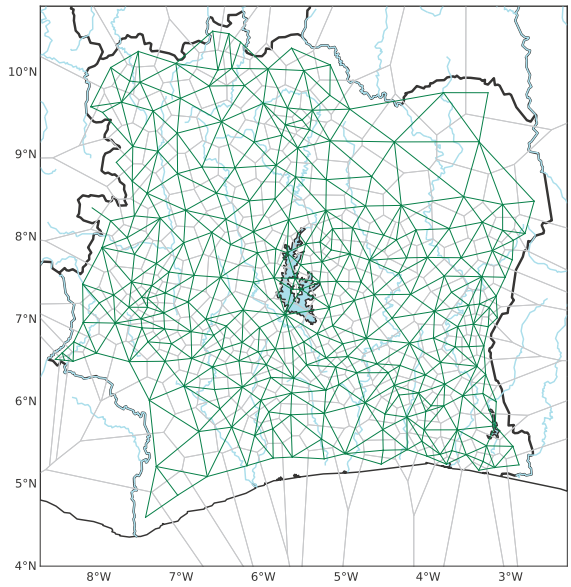
Dataset SPLoc contains the information about the central position in a subprefecture region. A tuple in SPLoc contains, subprefectureID, x coordinate of the central position in the subprefecture region and the y coordinate of the central position in the subprefecture region. These x, y coordinates are in longitude and latitude format. Consider x, y coordinates in SPLoc as the locations of the set of nodes V in a spatial graph where V is the set of subprefectureIDs in SPLoc. Using this information, we visualize a Voronoi tessellation for the subprefectures in the region of Ivory Coast (cf. Fig. 5.3(a)). A cell in Voronoi tessellation could represent a cell (coverage area of an eNB) in a $D2D$ communication network setup (cf. Fig. 5.2). Actual subprefecture boundaries also provide hypothetical coverage area and can be related to the coverage area of an eNB. Using this Voronoi tessellation we generate the dual of Voronoi tessellation, Delaunay triangulation. The Delaunay triangulation, in this context, represents a graphical structure that connects all neighboring subprefectures in the same country. We call the graphical structure G_v, (cf. Fig. 5.3(b)). This dataset is complemented with another dataset (SPLocComp) which contains the actual subprefecture boundaries. The SPLocComp dataset also contains information about which subprefecture lies in which department. This leads us to visualize the department-subprefecture relation based on the departments to which subprefectures are assigned, (cf. Fig. 5.3(c)).

5.2.2 Interpreting SET3

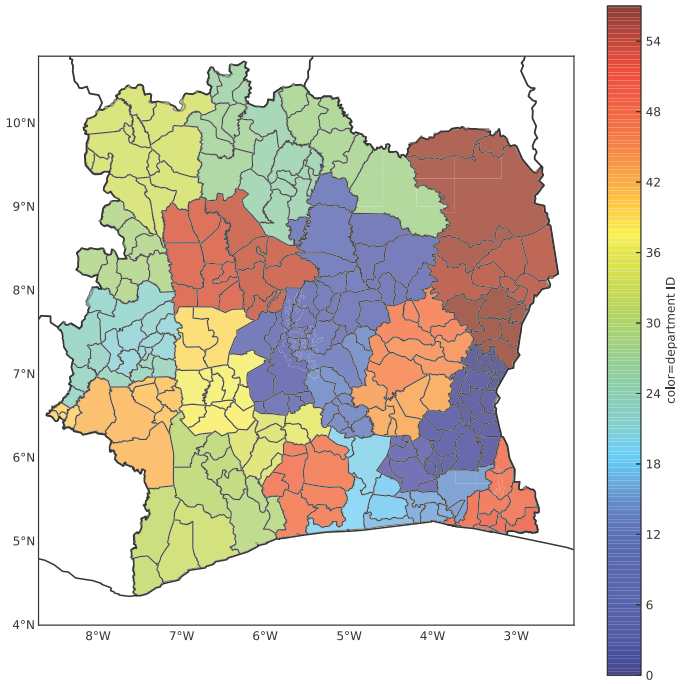
As stated before, mobility of a user can be interpreted and estimated using SET3. We map the mobility of the user on a complete graph (G_c) formed by connecting all the subprefectures to all other subprefectures and also map the mobility of the user on G_v. As human mobility often observes few long jumps to communities that are not near (neighbors), it is essential to map the mobility pattern of the users on G_c. However, the mapping of the mobility of the user on G_v provides the route which the user might have taken while it moved.



(a) Voronoi tessellation for the subprefectures in Ivory Coast with subprefecture locations.



(b) G_v.



(c) Association of subprefecture to departments in Ivory Coast.

Figure 5.3: Visualization of datasets SPLoc and SPLocComp.

Mapping of the user mobility either using G_c or G_v also gives valuable information like how much time a user stays at a location and how many times the user takes a certain path. To interpret the mobility using G_v and G_c we use shortest path between two subprefectures in the graph G_v and G_c. Consider a user u in SET3. Using the query "SELECT subprefectureID from SET3 where userID=u" a list U of subprefectureIDs is obtained. Let $U = (u_1, u_2, \ldots, u_n)$. The first entry, u_1 marks the starting point of the track followed by the user. Let us consider G_v first. To track a user on G_v, first the shortest path, $SP_{u_1u_2}$, between u_1 and u_2 in G_v is found. This process is done for all consecutive pair of entries in U. Each entry in the shortest path provides the information about the subprefecture that the user would have encountered while moving from u_1 to u_2. Let $SP_{u_1u_2} = (u_1, u_x, u_{x+1} \cdots, u_2)$. Note that the successive subprefectures in $SP_{u_1u_2}$ are connected by an edge in the graph G_v. If $u_1 = u_2$, it is inferred that the user has not moved.

Let C be the set of subprefectures. For user u, let P_u be a weight matrix of size $|C| \times |C|$ where $(P_u)_{ij}$ represents the weight of the edge between i and j. We initialize P_u as $(P_u)_{ij} = 0, \forall i, j \in C$. Consider the shortest path $SP_{u_1u_2}$ between two consecutive entries in U. $(P_u)_{ij}$ is incremented by 1 if the edge from i to j is encountered in the shortest path $SP_{u_1u_2}$. Note that i and j corresponds to a pair of successive entries in $SP_{u_1u_2}$. This process is then repeated for $SP_{u_2u_3}$ and so on until all elements in U have been considered. If $u_1 = u_2$, then we increment $(P_u)_{u_1u_1}$ by 1. Users are also tracked on G_c using the same method.

As an example, we track the user $u = 297412$ in the SET3 dataset on both G_v and G_c (cf. Fig. 5.4(a) and 5.4(b)). In Fig. 5.4 the blue edges mark the edges that user traversed. The thickness of the edges is equal to the number of times a user traversed through that edge. We also call it as the weight of the edge. The size of the node in the graph is equal to the weight of the node, $(P_u)_{ii}$. The yellow node marks the subprefecture k which satisfies

$$k = \arg\max_{\forall i \in C} \{(P_u)_{ii}\}.$$

The above process is performed for all the users in the SET3 dataset for the entire time period and the weights of the edges are determined for both G_v and G_c. Let $P = \sum_{\forall u} P_u$.

We now create the conditional Probability Matrix for movement, MP_v, and the Rate Matrix for movement, MR_v, from P that are obtained while tracking users on G_v. Further, we create the conditional Probability Matrix for movement, MP_c, and the Rate Matrix for movement, MR_c, from

P that are obtained while tracking users on G_v. As the method to obtain MP_v and MP_c are same, we only explain how MP_v is obtained. In P, let $P_{ii} = 0$ for all communities i as we are interested in the conditional probability of movement into and out of communities and not how long they stay there.

Definition 5.2.1. MP_v is a $|C| \times |C|$ matrix where the ij-th element represents the conditional probability of moving from subprefecture i to subprefecture j given that it has moved out of subprefecture i. We have, for all $i \neq j$,

$$(MP_v)_{ij} = \frac{(P)_{ij}}{\sum_{\forall k \in C}(P)_{ik}}. \tag{5.2.1}$$

We now describe the method to obtain MR_v and MR_c. As the method to obtain MR_v and MR_c are the same, we only explain how MR_v is obtained.

Definition 5.2.2. MR_v is a $1 \times |C|$ matrix where the i^{th} element represents the rate of moving out of subprefecture i. We have, for all $i \in C$,

$$(MR_v)_i = \frac{\sum_{\forall k \in C}(P)_{ik}}{Time} \tag{5.2.2}$$

where $Time = 150 \times 24 \times 60$ minutes, the total duration in minutes over which SET3 was generated.

The matrices MP_v and MP_c are shown in Fig. 5.5(a) and 5.5(b). Note that because the number of edges are less in G_v than G_c the MP_v is more sparse than MP_c.

We now generate the Return Rate matrices, MRR_v and MRR_c, i.e., the rate at which the user returns to its home subprefecture which is defined as follows.

Definition 5.2.3. Recall the list $U = (u_1, u_2, \ldots, u_n)$ defined earlier for user u. For all $j \in C$ and $1 \leq i \leq n$, let Z_{ij} be the indicator variable such that $Z_{ij} = 1$ if $u_i = j$ and $Z_{ij} = 0$ otherwise. Let $M_j = \sum_{i=1}^{n} Z_{ij}$. Note that subprefecture j appears exactly M_j times in the list U. The home subprefecture, H, of the user u is the subprefecture k that satisfies

$$k = \arg\max_{\forall j \in C} M_j.$$

We now describe the steps for generating MRR_v. Consider the home subprefecture H of a user u. Let P_u^r be a weight matrix of size $|C| \times |C|$ similar to the weight matrix P_u with the only difference

being that the increment now is done whenever a user jumps to H from any other location $j \neq H$, i.e., we increment $(P_u^r)_{jH}$ by 1. We then perform the same task for all the users and determine $P^r = \sum_{\forall u} P_u^r$.

Definition 5.2.4. MRR_v is a $|C| \times |C|$ matrix where its ij-th entry represents the rate of coming back to the home subprefecture j from another subprefecture i. For each $i, j \in C$, we have

$$(MRR_v)_{ij} = \frac{(P^r)_{ij}}{Time},$$

where $Time = 150 \times 24 \times 60$minutes.

Steps for generating MRR_c are similar. Note that H for a user will be same for G_v and G_c. As an example, here we only provide the return rate matrix for G_c (cf. Fig. 5.5(c)).

The information in SET3 can be used to infer the density of people in subprefecture over time. A tuple in SET3 provides which user is associated to which subprefecture at a certain time. Using this information we determine the density of each subprefecture at a certain time. From SET3 using query "SELECT DISTINCT time FROM SET3" a set of all distinct times in the SET3 are extracted. We then generate a matrix P^d with distinct times as row indices and subprefectureIDs as column indices. Note that the distinct times in SET3 have a granularity of 1 minute (which forms the natural bin size). We initialize P^d with a zero matrix. We then read tuples in SET3 one by one and add 1 to the $(P^d)_{time,subprefectureID}$ for each tuple.

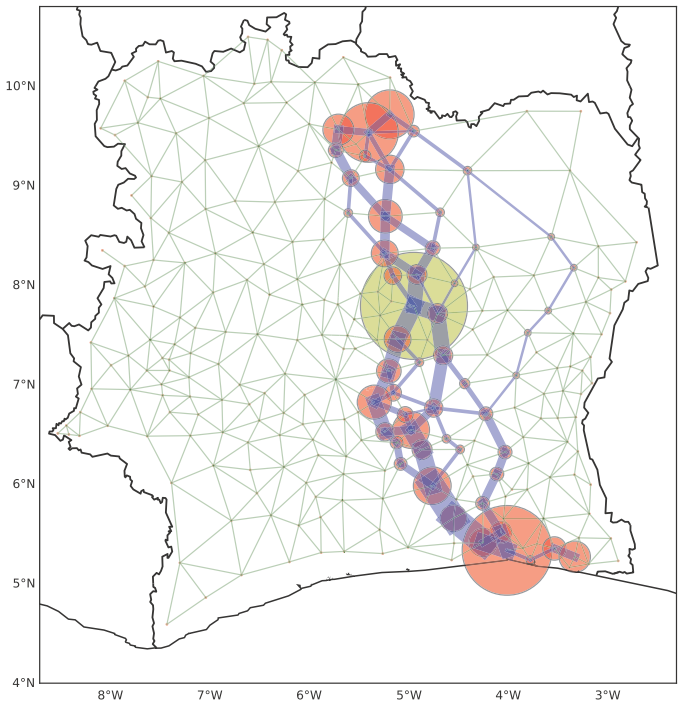
Definition 5.2.5. The density matrix, MD, is defined as

$$(MD)_{time,subprefectureID} = \frac{(P^d)_{time,subprefectureID}}{A_{subprefectureID}} \tag{5.2.3}$$

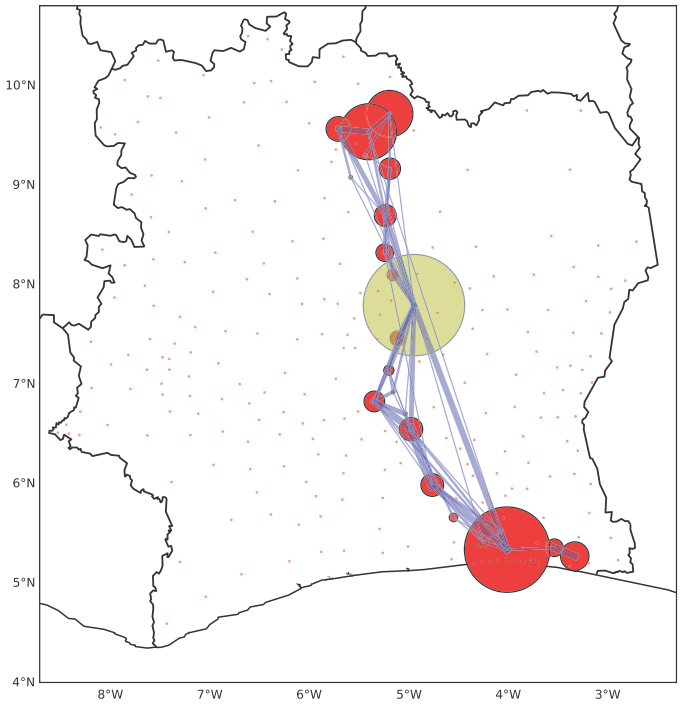
where $A_{subprefectureID}$ is the area of a subprefecture in Ivory Coast.

In Fig. 5.6 we provide a visual representation for the density of users on a log scale for the period from 16th December 2011 starting at 00:00:00 until 31st December 2011 ending at 23:59:00[1]. Fig. 5.6 shows that the number of calls is high during the day time while it is low during the night time. In a $D2D$ perspective this change in density causes the network structure to change over time which provides us the motivation to incorporate the changes in the density through the concept of Latent States explained in Section 5.3.

[1] An animated version of active number of users from 16th December 2011 starting at 00:00:00 until 31st December 2011 ending at 23:59:00 can be viewed at http://complex.luxbulb.org/content/d4d-challenge

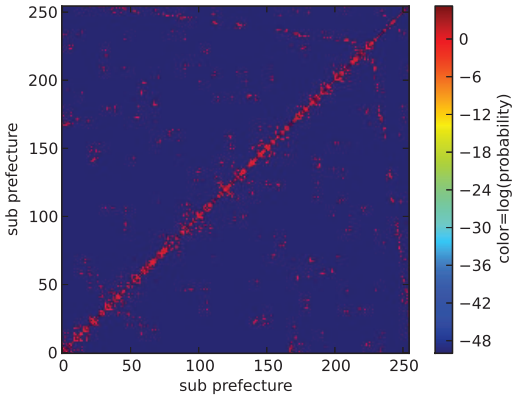


(a) Estimate of subprefectures reached by user 297412 using G_v.

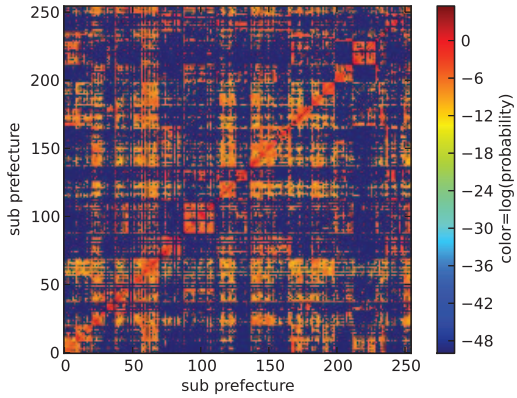


(b) Subprefectures reached by user 297412 using G_c.

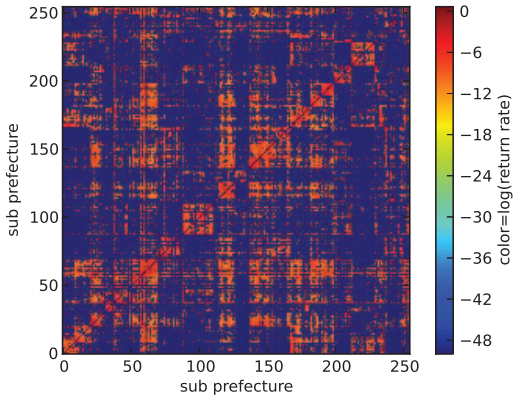
Figure 5.4: SET3 user mobility on the graph.



(a) MP_v.



(b) MP_c.



(c) MRR_c.

Figure 5.5: SET3 movement matrices.

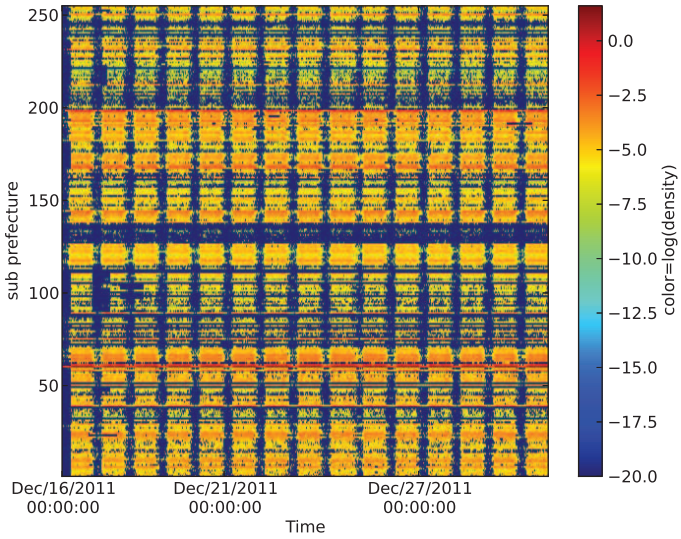


Figure 5.6: Density of users from 16th December 2011 to 31st December 2011.

We will now describe the dissemination model using MP_c and MR_c in the next Section.

5.3 Model

We represent our model in three layers. The bottom layer is the population layer where population of devices is spatially divided into communities. Above the population layer is the mobility layer where devices are allowed to move from more community to another with the rates extracted from the SET3. The dissemination model is then applied on the top of the mobility model. Thus, each device is associated to a community and a epidemic state (S, I or R).

Consider a set N of devices, $|N| \in \mathbb{N}$, that are non-uniformly distributed in a region of area $Area$, $Area \in \mathbb{R}^+$ and $Area > 0$. The non-uniformity leads to a community structure in the region. We assume that the set of these communities, C, is the set of subprefectures in the Ivory Coast. Let N_i be the number of devices that are in community i such that $\sum_{\forall i \in C} N_i = N$. Let A_i be the area of the community i such that $\sum_{\forall i \in C} A_i = Area$. Let density $\rho_i = \frac{N_i}{A_i}$.

5.3.1 Mobility Model

In a $D2D$ context a subprefecture can be envisioned as an eNB where the eNBs have allowed the $D2D$ communication. At a given time, a device is associated to one and only one eNB, i. From the dataset, $|C| = 255$. Further, as argued by Watts *et al.*, community structure is evident in the population in a realistic scenario [12]. The movement of a device or the jump of a device from a community i to another community j occurs with a rate $\sigma_i \nu_{ij}$ where σ_i denotes the total rate out of community i and ν_{ij} denotes the conditional probability of going to community j given that there is a transition out of community i. It was also shown that the jump also depends on the nature of the community [141] and the activity pattern [150]. This rate plays an important role in determining which new community has to be joined by the device. As each device is carried by humans this allows us to state that each device closely follows the mobility pattern of humans. Towards this, we use $(MR_c)_i$ calculated in Section 5.2.2 to determine if the device has moved out of community i or not. The device then has a conditional probability to go to the desired community. We use $(MP_c)_{ij}$ as the conditional probability to go from community i to community j. Note that the devices can move within the bounds of a community, but it does not affect the parameters that describe inter-community movement. However, the contact rate between devices moving inside the same community is affected by this intra-community movement. We shall see that the contact rates affect the rate of message dissemination in the same community.

Definition 5.3.1. We assume that the random time spent in community i, called the *dwell time* in community i, is exponentially distributed with mean $\frac{1}{\sigma_i}$. The rate of going out of community i is defined by $\sigma_i = (MR_c)_i$. Note that when $\sigma_i \to \infty$ it means that the device does not stay in i and the movement out of the community occurs instantaneously. However, when $\sigma_i \to 0$, it means that a device stays permanently in community i.

Definition 5.3.2. The conditional probability to move to community j given that there is a transition out of community i is defined by $\nu_{ij} = (MP_c)_{ij}$ where $\sum_{\forall j \in C} \nu_{ij} = 1$ and $\nu_{ii} = 0$.

Fig. 5.7 shows a mobility model with only three communities where movement from one community i to another community j is represented with an edge defined by $\sigma_i \nu_{ij}$. All other notations for movements are similarly defined. Note that each community i in the above model consists of devices

(shown by small circles inside big circles). Moreover, this can be generalized to $|C|$ communities. This type of mobility model is a simple depiction of the inter-community movements. However, the mobility model can be enriched by including the information of home location for each node. A model introducing the return rates in this simple mobility model is described in the Section 5.3.5.

We assume that the behavior of one device is Markovian. For N devices, the superposition of these Markov processes is Markovian. It leads to a Continuous Time Markov Chain ($CTMC$).

Due to mobility, the number of the devices in a community i change. This change in the number of devices in a community is defined as

$$\frac{dN_i}{dt} = \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} N_j - \sum_{\forall j \in C, j \neq i} \sigma_i \nu_{ij} N_i. \tag{5.3.1}$$

In eq. 5.3.1, the term $\sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} N_j$ represents the number of devices coming to the community i. However, the term $\sum_{\forall j \in C, j \neq i} \sigma_i \nu_{ij} N_i$ represents the number of devices going out of community i. Using the definition 5.3.2, the eq. 5.3.1 can be represented as

$$\frac{dN_i}{dt} = \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} N_j - \sigma_i N_i. \tag{5.3.2}$$

In order to visualize the dissemination process, we first identify the steady state of the mobility model. Note that there are two models, mobility model and dissemination model. For the mobility model, we look for the steady state limit solution. Let N_i^* be the steady state population in community i. Note that N_i^* is a limit and will not change over time. Then, we use the differential equations in order to derive the transient solution for the dissemination process where we study the change in the number of devices in S, I and R state. In the differential equations, we use the parameters that are derived from a mean field approximation and do not change over time. However, note that the number of devices in state S, I and R will change over time.

Further, note that, the steady state does not mean that the devices are not moving out of a community or the devices are not coming in to the same community. It only states that the expected number of devices in a community does not change over time, i.e. expected number of devices coming in to a community is equal to the expected number of devices moving out of the same community.

At the steady state, $\frac{dN_i}{dt} = 0$ for all $i \in C$. The steady state population in the community i, N_i^*,

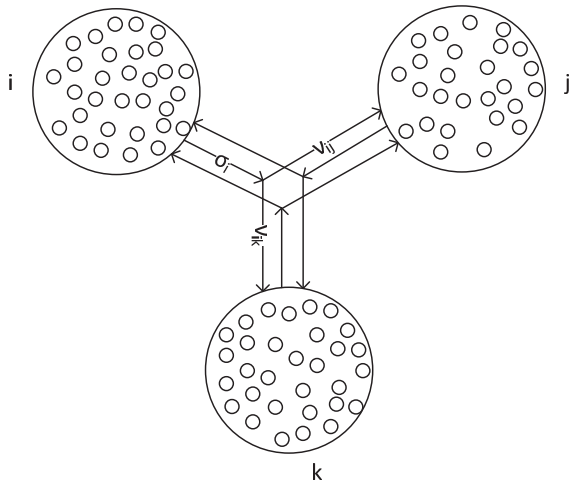


Figure 5.7: Mobility model showing the movements from one community to another one.

is thus defined as

$$N_i^* = \frac{\sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} N_j^*}{\sigma_i}. \tag{5.3.3}$$

At the steady state, $\rho_i^* = \frac{N_i^*}{A_i}$ where ρ_i^* is the density of devices at steady state.

5.3.2 Dissemination Model

We now describe the dissemination model. Let each device in a community be in either of the three states S, I and R. Note that these states are different from the communities to which the devices are associated to. State S means that the device does not have the information. State I means that the device has the information and is transmitting it. State R means that the device has the information and is not transmitting it. Let S_i, I_i and R_i be the number of devices in state S, I and R, respectively, in community i at time t. As we apply the dissemination model on top of the mobility model where we approximated the number of devices by the steady state limit, it should be noted that $N_i^* = \mathbf{E}[S_i] + \mathbf{E}[I_i] + \mathbf{E}[R_i]$. Under the mean field assumption, we can safely assume that N_i^* will remain constant over the simulation duration.

Let each device have an omnidirectional transmission range, r such that $r \in \mathbb{R}^+$. Due to the transmission range each device has a neighborhood in the community i to which they can communicate. We assume the interactions among the devices to occur under mean field approximation where

the expected neighborhood size is given by $\langle k_i \rangle$.

Definition 5.3.3. The expected neighborhood size for a device in a community i is defined as $\langle k_i \rangle = \rho_i^* \pi r^2$.

In real scenarios, communities affect the global dissemination of the information. In a system with a community structure, the information transfer is bounded within the community unless a device in the community moves to another community. This makes us to constrain the devices in state I to be able to only transmit the information to the devices in the associated community (in associated eNB) that are in state S and are within their transmission range or the physical proximity, i.e., a device in state I will be able to transfer information to $\langle k_i \rangle$ device in its neighborhood. Note that some devices in $\langle k_i \rangle$ might already have the information. In such case the device would not be able to transfer information to that device.

Definition 5.3.4. The transition rate with which a message is received successfully is given by β such that $\beta \in \mathbb{R}^+$. It can also be understood as the contact rate between two devices.

We assume that the devices in different communities have different β represented as β_i. This is because of many factors like, the density of the population and willingness of other devices to accept the transmitted information. This idea is analogous to the epidemic spreading in different communities which happen at different rates. Thus, devices in each community have different contact rate.

Due to the energy constraints, a device can only transmit information for a certain period of time after which the device does not transfer the information. This can be understood as the rate with which devices change their state from I to R.

Definition 5.3.5. $\forall x$ associated with community i, the rate with which devices in community i change their state from I to R is given by δ_i such that $\delta_i \in \mathbb{R}^+$.

The β_i and δ_i associated with a device change when the device moves from one community to another. This can be understood as when a device moves from one community to another its willingness to transmit the information or not to transmit the information may increase or decrease.

In a community i, let initially at time $t = 0$, $S_i = N_i^*$, $I_i = 0$ and $R_i = 0$. In order to study the dissemination process, let one community i at time $t = 0$ be the source of information,

i.e., $S_i = N_i^* - \varepsilon$, $I_i = \varepsilon$ and $R_i = 0$ where $0 < \varepsilon < N$. Under mean field approximation, $S_i I_i / N_i^*$ provides the fraction of interactions between devices in state S and devices in state I. As each device has a neighborhood $\langle k_i \rangle$ and the contact rate between devices is β_i, the number of devices that will change their state from S to I is given by $\beta_i \frac{S_i I_i \langle k_i \rangle}{N_i^*}$. Let us consider the change in the number of devices in state S. Due to mobility, some devices in state S will move from other communities to community i while some devices in state S will move from community i to other communities. The number of devices moving from other communities to i is given by $\sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} S_j$ while those moving from i to other communities is given by $\sigma_i S_i$. Thus the rate of change in the number of devices in state S in community i in dt time is given by $\frac{dS_i}{dt} = -\beta_i \frac{S_i I_i \langle k_i \rangle}{N_i^*} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} S_j - \sigma_i S_i$. Using such interpretation and above explanation the rate of change in the number of devices in state I and state R can be formulated. We define the rate of change in number of devices in state S, I and R in community i as

$$\left.\begin{aligned} \frac{dS_i}{dt} &= -\beta_i \frac{S_i I_i \langle k_i \rangle}{N_i^*} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} S_j - \sigma_i S_i \\ \frac{dI_i}{dt} &= \beta_i \frac{S_i I_i \langle k_i \rangle}{N_i^*} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} I_j - \sigma_i I_i - \delta_i I_i \\ \frac{dR_i}{dt} &= \delta_i I_i + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} R_j - \sigma_i R_i \end{aligned}\right\} \tag{5.3.4}$$

In eq. 5.3.4, the fourth term in $\frac{dI_i}{dt}$ represents the number of devices changing their state to R. Note that the set of equations in eq. 5.3.4 is for one community. As there are $|C|$ communities the actual number of equations in eq. 5.3.4 will be $3 \times |C|$. Further, note that due to steady state of mobility, $\sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} S_j + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} I_j + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} R_j - \sigma_i S_i - \sigma_i I_i - \sigma_i R_i = 0$.

As we are interested in knowing how many devices are in state I over time, we do not determine the steady state of dissemination. However, we study the transient states.

5.3.3 Adding Latent States

Irrespective of whether a device is in state S, I or R, the capability of the device to transmit or receive also depends on whether the device is switched on or off. It is possible that a device has the information but has been switched off by its user. This hampers the transmission of the information from the device to other devices. Following the same argument, if a device does not have the

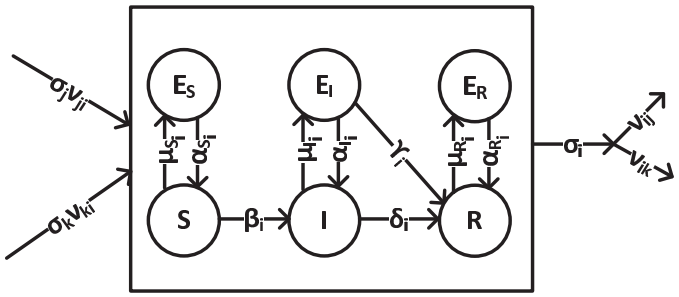


Figure 5.8: State diagram with states S, I and R and their latent states E_S, E_I and E_R respectively with transition rates between states. The square box represents the community i.

information and it is switched off it will not be able to receive the information from other devices. We call such state of a device as latent state and term them as E_S, E_I and E_R to represent latent states pertaining to each active state, S, I and R respectively. When a device is switched on, this will mark the transition in the state of the device from either E_S, E_I or E_R to S, I or R respectively.

Definition 5.3.6. The transition rate to change the state from S to E_S in a community i is defined as μ_{S_i} such that $\mu_{S_i} \in \mathbb{R}^+$. The transition rate to change the state from I to E_I in a community i is defined as μ_{I_i} such that $\mu_{I_i} \in \mathbb{R}^+$. The transition rate to change the state from R to E_R in a community i is defined as μ_{R_i} such that $\mu_{R_i} \in \mathbb{R}^+$. The transition rate to change the state from E_S to S in a community i is defined as α_{S_i} such that $\alpha_{S_i} \in \mathbb{R}^+$. The transition rate to change the state from E_I to I in a community i is defined as α_{I_i} such that $\alpha_{I_i} \in \mathbb{R}^+$. The transition rate to change the state from E_R to R in a community i is defined as α_{R_i} such that $\alpha_{R_i} \in \mathbb{R}^+$.

Further, a device in state E_I can wake up and decide not to transmit information.

Definition 5.3.7. The transition rate to change the state from E_I to R in a community i is defined as γ_i such that $\gamma_i \in \mathbb{R}^+$.

Note that due to addition of three new states, $N_i^* = \mathbf{E}[S_i]+\mathbf{E}[I_i]+\mathbf{E}[R_i]+\mathbf{E}[E_{S_i}]+\mathbf{E}[E_{I_i}]+\mathbf{E}[E_{R_i}]$. Similar to initial condition as in Section 5.3.2, we assume at time $t = 0$ $S_i = N_i^* - \varepsilon$, $I_i(t = 0) = \varepsilon$ and $R_i(t = 0) = E_{S_i}(t = 0) = E_{I_i}(t = 0) = E_{R_i}(t = 0) = 0$, where $0 < \varepsilon < N$ and t is the starting time.

Thus, the state diagram for this case is given by Fig. 5.8. This addition of latent states lead us

to modify the equations mentioned in the eq. 5.3.4 for a community i to

$$\left.\begin{aligned}
\frac{dS_i}{dt} &= -\beta_i \frac{S_i I_i \langle k_i \rangle}{N_i^*} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} S_j - \sigma_i S_i - \mu_{S_i} S_i + \alpha_{S_i} E_{S_i} \\
\frac{dI_i}{dt} &= \beta_i \frac{S_i I_i \langle k_i \rangle}{N_i^*} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} I_j - \sigma_i I_i - \delta_i I_i - \mu_{I_i} I_i + \alpha_{I_i} E_{I_i} \\
\frac{dR_i}{dt} &= \delta_i I_i + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} R_j - \sigma_i R_i - \mu_{R_i} R_i + \alpha_{R_i} E_{R_i} + \gamma_i E_{I_i} \\
\frac{dE_{S_i}}{dt} &= \mu_{S_i} S_i - \alpha_{S_i} E_{S_i} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} E_{S_j} - \sigma_i E_{S_i} \\
\frac{dE_{I_i}}{dt} &= \mu_{I_i} I_i - \alpha_{I_i} E_{I_i} - \gamma_i E_{I_i} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} E_{I_j} - \sigma_i E_{I_i} \\
\frac{dE_{R_i}}{dt} &= \mu_{R_i} R_i - \alpha_{R_i} E_{R_i} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} E_{R_j} - \sigma_i E_{R_i}
\end{aligned}\right\} \tag{5.3.5}$$

Note that the set of equations in eq. 5.3.5 is for one community. As there are $|C|$ communities the actual number of equations in eq. 5.3.5 will be $6 \times |C|$. Further, it can be noted that there will be a growth in the population of devices in I if $\beta_i \frac{S_i I_i \langle k_i \rangle}{N_i^*} - \delta_i I_i - \mu_{I_i} I_i + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} I_j - \sigma_i I_i + \alpha_{I_i} E_{I_i} > 0$.

Due to the addition of latent states the information dissemination is highly impacted. Let the case explained in the Section 5.3.2 be case A and the case explained in this Section be case B. We will maintain this nomenclature throughout this Chapter. The difference in number of devices in state I for case A and case B at a given time can be formulated. Let $I_{i,t,A}$ and $I_{i,t,B}$ be the number of devices in state I community i at time t for case A and B respectively. The difference in the number of devices in state I between two cases A and B can be computed as $\Delta I_i(t) = I_{i,t,A} - I_{i,t,B}$. Using the initial condition, it is known that $I_{i,0,A} = I_{i,0,B}$. The difference in the number of devices in state I for all the communities is defined as $\Delta I(t) = \sum_{i \in C} \Delta I_i(t)$. As stated in Section 5.3.2 here also we study the transient states of the dissemination.

5.3.4 Adding Beamforming

As we are interested in enhancing information dissemination, one way of enhancing the information dissemination in a community based model is by the introduction of mobility. Another way of enhancing dissemination process studied in literature is through beamforming. Effects of beamforming have already been studied on the information dissemination process in the static population in

Chapter 3 as well as on the mobile population Chapter 4 and in [112, 136] with positive results.

In this Subsection, our objective is to apply beamforming to improve information dissemination. We assume that each device is equipped with an antenna array with M antenna elements (AEs, each capable of achieving omnidirectional transmission range r) such that $M \in \mathbb{N}$, where M can be different for different devices. Initially, all devices use one AE for omnidirectional transmission. Here, r can be different for different devices. As we are interested in studying the benefits on information dissemination while using directional antenna, we assume path loss exponent to be equal to 2 and assume free space propagation. Further, only the devices in state I are allowed to beamform. In [112, 136] and in Chapter 4 it has been shown that considerable enhancement in the dissemination process can be achieved using very few long range links. Using this result, we state that only $x\%$ of the devices in the state I are randomly chosen to beamform. The selected devices randomly choose $m \in [2, M]$ AEs from M available AEs and beamform in a random direction. In the model, we use Uniform Linear Array Antenna model (ULA) [64].

Adding beamforming to the model only changes $\langle k_i \rangle$, as $x\%$ devices beamform. In terms of $D2D$ communication scenario, $\langle k_i \rangle$ relates to the number of $D2D$ links for the device. Let $\langle k_i^* \rangle$ be the new expected neighborhood size in a community i after beamforming is added. $\langle k_i^* \rangle$ is modeled using the expected number of devices under the beam in the community i. Let X be the set of $x\%$ devices in community i that beamform. Let the antenna gain be $g(q, \theta, B_b)$ in the direction θ such that q in X with the antenna boresight direction as B_b. The total area covered under the beam of q in X is given by $A_q = \frac{r^2}{2} \int_0^{2\pi} g(q, \theta, B_b) d\theta$. Note that the gain in at the reception side is considered to be 1. For a device q in X, its $\langle k_{i,q} \rangle = \rho_i A_q$. Thus, for all the devices in X, $\langle k_{X,i} \rangle = \frac{\sum_{q \in X} \langle k_{i,q} \rangle}{|X|}$. This implies $\langle k_{X,i} \rangle = \frac{\rho_i \sum_{q \in X} A_q}{|X|}$. Note that the assumption here is that the area covered by the beamforming device lies within the area of the community the device is associated to.

Definition 5.3.8. When $x\%$ of the devices in the community i have expected number of connections $\langle k_{X,i} \rangle$ while others have $\langle k_i \rangle$, the overall expected number of connections in a community i is defined as $\langle k_i^* \rangle = \frac{\langle k_i \rangle (N_i^* - |X|) + \langle k_{X,i} \rangle (|X|)}{N_i^*}$.

As only $\langle k_i^* \rangle$ has changed, the rate eq. 5.3.5 are now defined using $\langle k_i^* \rangle$ instead of $\langle k_i \rangle$. The new

rate equation for the change in number of devices in state S, I, R, E_S, E_I and E_R is defined as

$$\left.\begin{aligned}
\frac{dS_i}{dt} &= -\beta_i \frac{S_i I_i \langle k_i^* \rangle}{N_i^*} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} S_j - \sigma_i S_i - \mu_{S_i} S_i + \alpha_{S_i} E_{S_i} \\
\frac{dI_i}{dt} &= \beta_i \frac{S_i I_i \langle k_i^* \rangle}{N_i^*} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} I_j - \sigma_i I_i - \delta_i I_i - \mu_{I_i} I_i + \alpha_{I_i} E_{I_i} \\
\frac{dR_i}{dt} &= \delta_i I_i + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} R_j - \sigma_i R_i - \mu_{R_i} R_i + \alpha_{R_i} E_{R_i} + \gamma_i E_{I_i} \\
\frac{dE_{S_i}}{dt} &= \mu_{S_i} S_i - \alpha_{S_i} E_{S_i} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} E_{S_j} - \sigma_i E_{S_i} \\
\frac{dE_{I_i}}{dt} &= \mu_{I_i} I_i - \alpha_{I_i} E_{I_i} - \gamma_i E_{I_i} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} E_{I_j} - \sigma_i E_{I_i} \\
\frac{dE_{R_i}}{dt} &= \mu_{R_i} R_i - \alpha_{R_i} E_{R_i} + \sum_{\forall j \in C, j \neq i} \sigma_j \nu_{ji} E_{R_j} - \sigma_i E_{R_i}
\end{aligned}\right\} \tag{5.3.6}$$

To see a positive effect of beamforming a necessary condition is $\langle k_i \rangle < \langle k_i^* \rangle$ which reduces to $\langle k_i \rangle < \langle k_{X,i} \rangle$. This means that $\langle k_{X,i} \rangle$ for a beamforming device should be greater than $\langle k_i \rangle$ for the same device when it was not beamforming. Let the current beamforming scenario be case B'. Let $I_{i,t,B'}$ and $I_{i,t,B}$ be the number of devices in state I in the community i at time t for case B' and B respectively. The difference in the number of devices in state I at time t for B' and B for all the communities on the population scale is defined as

$$\overline{\Delta I}(t) = \sum_{\forall i \in C} I_{i,t,B'} - I_{i,t,B}. \tag{5.3.7}$$

When the initial condition is same as defined before. Then $I_{i,0,B'} = I_{i,0,B}$.

5.3.5 Mobility Model with Return Rate

In this Subsection, we enrich the mobility model presented in Section 5.3.1 by considering rate of return to home community for each device [104,151]. Let N_{ii} be the number of devices having home community as i that are in community i and N_{ij} be the number of devices having home community as i that are in community j. As defined in definition 5.3.1 and 5.3.2, let σ_i be the rate of moving out of community i and ν_{ij} as the conditional probability of going to community j given the dive has moved out from community i, respectively.

Definition 5.3.9. For a device, the rate to return to its home community i from community j is defined as $\zeta_{ji} = (MRR_c)_{ji}$. Note that when $\zeta_{ji} \to \infty$, it means that the device returns to its home

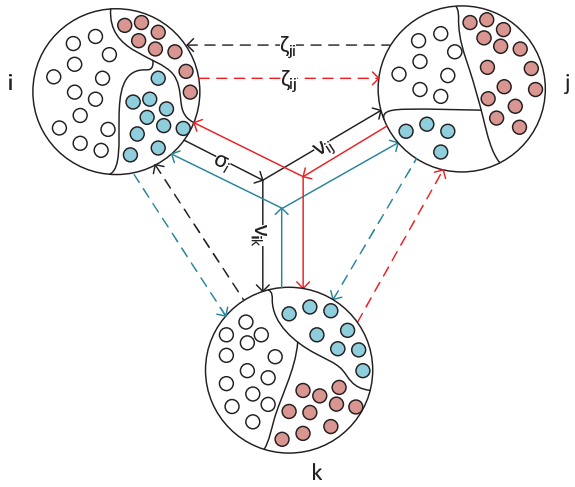


Figure 5.9: Associating return rate to the mobility model. The dotted line shows the return rate from another community and is marked with $\zeta_{c'c}$ and dotted line, rate of moving out of a community is given by σ_c, while rate of going out to another community is marked with $\sigma_c\nu_{cc'}$ and solid lines. Each community has associated number of devices. For different communities these devices are marked in different color (white for i, red for j, blue for k). The color of the lines also represent which community the devices are associate to.

community i instantaneously after visiting community j. However, when $\zeta_{ji} \to 0$, it means that a device never moves back to its home community.

Note that $\sum_{\forall i}\sum_{\forall j} N_{ij} = N$. As represented in Fig. 5.9 the travel pattern leads to [104]

$$\left.\begin{aligned} \frac{dN_{ii}}{dt} &= \sum_{\forall j \in C, j \neq i} \zeta_{ji} N_{ij} - \sigma_i N_{ii} \\ \frac{dN_{ij}}{dt} &= \sigma_i \nu_{ij} N_{ii} - \zeta_{ji} N_{ij} \end{aligned}\right\} \tag{5.3.8}$$

The first term on the right hand side of the first equation in eq. 5.3.8, that is, $\sum_{\forall j \in C, j \neq i} \zeta_{ji} N_{ij}$, relates to the number of devices returning to their home community i from other communities while the second term of the same equation, that is, $\sigma_i N_{ii}$, refers to devices that are moving out of community i that have home community i. The second equation in eq. 5.3.8 can also be similarly explained. Note that similar equations apply to all other communities. Also note that the total number of such equations is $|C| \times |C|$ and $N_i = N_{ii} + \sum_{j \in C, j \neq i} N_{ij} = \sum_{j \in C} N_{ij}$ and $\sum_{j \in C} N_i = N$. As described in Section 5.3.1, we first determine the steady state of mobility and then study the transient states in the dissemination process using the steady state population. Using eq. 5.3.8 the

number of devices in the steady state is given by [104]

$$\left.\begin{aligned} N_{ii}^{*} &= N_{i}^{*}\left(\frac{1}{1+\sigma_i \sum_{\forall k\in C, j\neq i} \frac{\nu_{ik}}{\zeta_{ki}}}\right) \\ N_{ij}^{*} &= N_{i}^{*}\left(\frac{\sigma_i \nu_{ij}}{\zeta_{ji}\left(1+\sigma_i \sum_{\forall k\in C, j\neq i} \frac{\nu_{ik}}{\zeta_{ki}}\right)}\right) \end{aligned}\right\} \tag{5.3.9}$$

Dissemination model

The addition of return rates cause the system to change whereby new set of equations for dissemination process are to be written. Similar to the explanation of N_{ii} and N_{ij}, let S_{ii}, I_{ii}, R_{ii}, $E_{S_{ii}}$, $E_{I_{ii}}$ and $E_{R_{ii}}$ be the number of devices in state S, I, R, E_S, E_I and E_R, respectively, that have the home community as community i and are in the community i while S_{ij}, I_{ij}, R_{ij}, $E_{S_{ij}}$, $E_{I_{ij}}$ and $E_{R_{ij}}$ be the number of devices in state S, I, R, E_S, E_I and E_R, respectively, that have the home community as community i and are in the community j. Thus, $N_i^* = \sum_{\forall j\in C} N_{ij}^* = \sum_{\forall j\in C} \left[S_{ij} + I_{ij} + R_{ij} + E_{S_{ij}} + E_{I_{ij}} + E_{R_{ij}}\right]$. The new rate equations for the change in number of devices in state S, I, R, E_S, E_I and E_R are given as

$$
\left.\begin{aligned}
\frac{dS_{ii}}{dt} &= -\beta_i \frac{\langle k_i^* \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} S_{ij} - \sigma_i S_{ii} - \mu_{S_i} S_{ii} + \alpha_{S_i} E_{S_{ii}} \\
\frac{dS_{ij}}{dt} &= -\beta_j \frac{\langle k_j^* \rangle}{N_j^*} \sum_{\forall q \in C} S_{ij} I_{qj} + \sigma_i \nu_{ij} S_{ii} - \zeta_{ji} S_{ij} - \mu_{S_j} S_{ij} + \alpha_{S_j} E_{S_{ij}}
\end{aligned}\right\}
$$

$$
\left.\begin{aligned}
\frac{dI_{ii}}{dt} &= \beta_i \frac{\langle k_i^* \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} I_{ij} - \sigma_i I_{ii} - \delta_i I_{ii} - \mu_{I_i} I_{ii} + \alpha_{I_i} E_{I_{ii}} \\
\frac{dI_{ij}}{dt} &= \beta_j \frac{\langle k_j^* \rangle}{N_j^*} \sum_{\forall q \in C} S_{ij} I_{qj} + \sigma_i \nu_{ij} I_{ii} - \zeta_{ji} I_{ij} - \delta_j I_{ij} - \mu_{I_j} I_{ij} + \alpha_{I_j} E_{I_{ij}}
\end{aligned}\right\}
$$

$$
\left.\begin{aligned}
\frac{dR_{ii}}{dt} &= \delta_i I_{ii} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} R_{ij} - \sigma_i R_{ii} - \mu_{R_i} R_{ii} + \alpha_{R_i} E_{R_{ii}} + \gamma_i E_{I_{ii}} \\
\frac{dR_{ij}}{dt} &= \delta_j I_{ij} + \sigma_i \nu_{ij} R_{ii} - \zeta_{ji} R_{ij} - \mu_{R_j} R_{ij} + \alpha_{R_j} E_{R_{ij}} + \gamma_j E_{I_{ij}}
\end{aligned}\right\}
$$

$$
\left.\begin{aligned}
\frac{dE_{S_{ii}}}{dt} &= \mu_{S_i} S_{ii} - \alpha_{S_i} E_{S_{ii}} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} E_{S_{ij}} - \sigma_i E_{S_{ii}} \\
\frac{dE_{S_{ij}}}{dt} &= \mu_{S_j} S_{ij} - \alpha_{S_j} E_{S_{ij}} + \sigma_i \nu_{ij} E_{S_{ii}} - \zeta_{ji} E_{S_{ij}}
\end{aligned}\right\}
$$

$$
\left.\begin{aligned}
\frac{dE_{I_{ii}}}{dt} &= \mu_{I_i} I_{ii} - \alpha_{I_i} E_{I_{ii}} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} E_{I_{ij}} - \sigma_i E_{I_{ii}} - \gamma_i E_{I_{ii}} \\
\frac{dE_{I_{ij}}}{dt} &= \mu_{I_j} I_{ij} - \alpha_{I_j} E_{I_{ij}} + \sigma_i \nu_{ij} E_{I_{ii}} - \zeta_{ji} E_{I_{ij}} - \gamma_j E_{I_{ij}}
\end{aligned}\right\}
$$

$$
\left.\begin{aligned}
\frac{dE_{R_{ii}}}{dt} &= \mu_{R_i} R_{ii} - \alpha_{R_i} E_{R_{ii}} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} E_{R_{ij}} - \sigma_i E_{R_{ii}} \\
\frac{dE_{R_{ij}}}{dt} &= \mu_{R_j} R_{ij} - \alpha_{R_j} E_{R_{ij}} + \sigma_i \nu_{ij} E_{R_{ii}} - \zeta_{ji} E_{R_{ij}}
\end{aligned}\right\}
\tag{5.3.10}
$$

Thus, the total number of equations in the system being $6 \times |C| \times |C|$. From the eq. 5.3.10 the total number of devices in state I for a community i that have home community as community i is defined by $\sum_{\forall j \in C} I_{ij}$. Thus, the total number of devices in state I in all the communities is given by $\sum_{\forall i \in C} \sum_{\forall j \in C} I_{ij}$. Similarly the number of devices for other states can be computed.

5.4 Simulation Results

The simulations are performed using the Python environment. Initially, each device operates in omnidirectional mode using $m = 1$ AE. The maximum number of antenna elements that any device could be equipped with is set to $M = 5$ with transmission range $r = 0.5km$. We assume free space propagation with path loss exponent equal to 2. The separation between two AEs is computed using $WiFi$ frequency, $f = 2.4Ghz$. We consider, $N = 15686986$ in an area of $Area \approx 323096km^2$ (Ivory

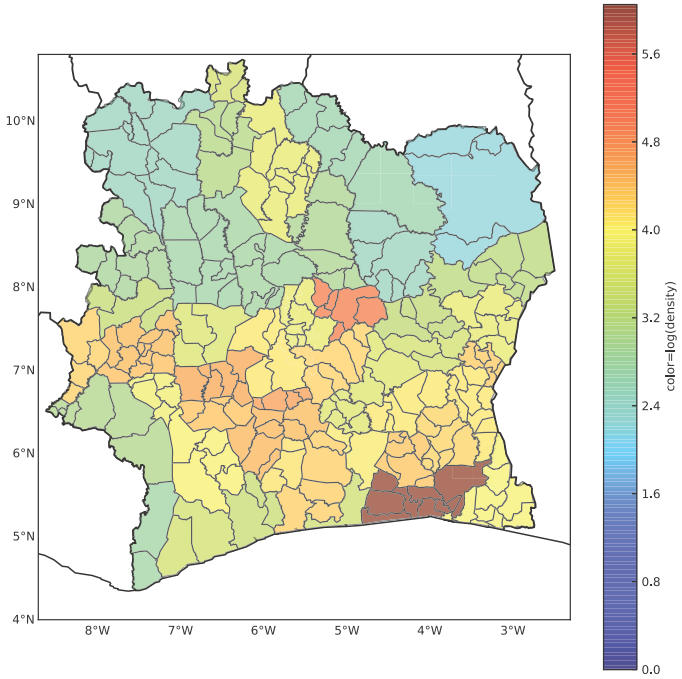


Figure 5.10: Population density in each subprefecture on a log scale.

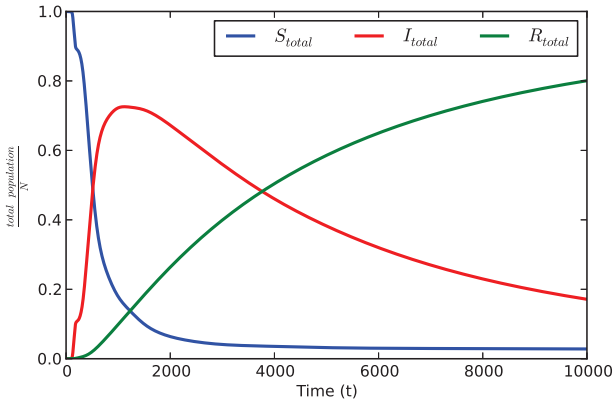
Coast region). These values are the census data for the year 1998 for Ivory Coast and are collected from the source [144]. The source [144] contains population count for each subprefecture area for the year 1998. Using this information and the actual subprefecture area from SPLocComp, we get the population density in each subprefecture. Let N_i and A_i be the population in the subprefecture i and the area of the subprefecture i, respectively. The population density, ρ_i, for a subprefecture i is equal to $\frac{N_i}{A_i}$. The population density in subprefectures in log scale is shown in Fig. 5.10. Further, we consider each subprefecture as a community where each person in the population is equipped with a device.

Unless otherwise stated in all the results we use G_c and associated movement rate matrix MR_c and probability matrix MP_c. Initially, zt time $t = 0$ we set $S_i = N_i^*$, $I_i = R_i = E_{S_i} = E_{I_i} = E_{R_i} = 0$ for all communities. In order to start the dissemination, we set $S_0 = N_i^* - \varepsilon$ and $I_0 = \varepsilon$ where $\varepsilon \to 0$. In $D2D$ communication scenario, this means that a public warning message is generated somewhere in the area covered by eNB having ID equal to 0. Through the evolution of fraction of devices in any of the three states, in this Chapter, we show the effect of the model on the dissemination of the

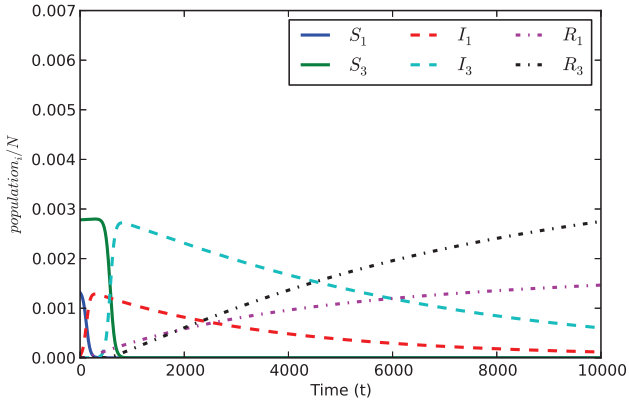
public warning message across a large population. Further, the population size in each community is normalized by N.

As preliminary results, we first provide the results for case A using the data provided (cf. Fig. 5.11(a)). Fig. 5.11(a) shows the evolution of number of devices normalized by N in the three active (S, I, R) states. Let $S_{total} = \sum_{i \in C} S_i/N$, $I_{total} = \sum_{i \in C} I_i/N$ and $R_{total} = \sum_{i \in C} R_i/N$. Initially, when there are many devices without information, the devices with information are able to send the information to them. Thus, dI/dt is high and more devices change their state from S to I. However, as time increases, the fraction of devices in state S decreases as most of devices now have the information. This reduces dI/dt. From Fig. 5.11(a) it can be seen that initially dI/dt is very high. However, it becomes equal to 0 for some time and then again increases. This is due to the mobility pattern of the users and the community structure. Consider some communities where there are devices in state I and the devices are not moving out of the community. Also consider all devices in other communities to be in state S. When all the devices in the communities with devices in state I are in state I, there are no devices in state S in those communities and thus $dI/dt = 0$. However, when the devices in state I move to another communities where there were no devices in state I, they are able to meet devices in state S. This again increases dI/dt. Fig. 5.11(b) shows the evolution of S_i, I_i and R_i for community numbered one and three. Due to the difference in the number of devices in steady state of the mobility, the curves for two considered communities differ.

As we add latent states to case A, we next show the evolution of number of devices in S, I and R states normalized by N when latent states are added, case B. We then bring out the differences in the fraction of devices in S, I and R. Fig. 5.12(a) shows the evolution of the fraction of devices in states S, I and R for the case A and case B. We call the devices in states S, I and R for the case when latent states are added as active devices. Let $S_{total,active} = \sum_{i \in C} S_i/N$, $I_{total,active} = \sum_{i \in C} I_i/N$ and $R_{total,active} = \sum_{i \in C} R_i/N$. The maximum difference between the fraction of devices in state I for the two cases at any time is found to be ≈ 0.23186, i.e., $\max\{I_{total} - I_{total,active}\} \approx 0.23186$. However, this is not a correct measure as there is a population difference in the three active states because of the addition of the latent states. We thus normalize the population in the active states by $S_{total,active} + I_{total,active} + R_{total,active}$. When the results are normalized, the $\max\{I_{total} - I_{total,active}\}$ at any time reduces to ≈ 0.043694 (cf. Fig. 5.12(b)). We then plot $S_{total} - S_{total,active}$, $I_{total} -$



(a) Evolution of S_{total}, I_{total} and R_{total} in Ivory Coast using simple SIR model.



(b) Evolution of fraction of devices in S, I and R states in community one and three in Ivory Coast using simple SIR model.

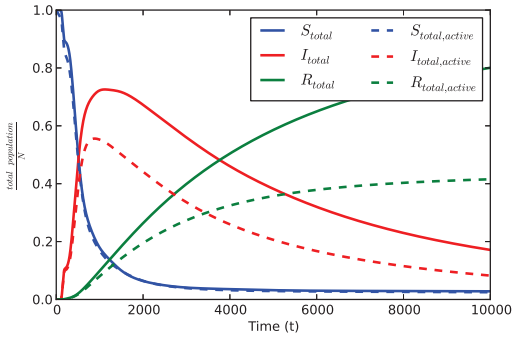
Figure 5.11: Evolution of fraction of devices in S, I and R states in Ivory Coast.

$I_{total,active}$ and $R_{total} - R_{total,active}$ on the actual population scale (cf. Fig. 5.12(c)) as another representation for Fig. 5.12(b). It can be seen that even a small difference of ≈ 0.04369, on the actual population scale can be of high significance. This difference is ≈ 685440 on the actual population scale.

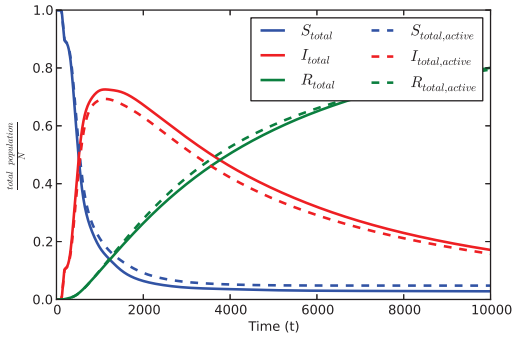
Using case B' we next show how beamforming can effect the dissemination process. We first present the comparison of the number of devices in states S, I or R when normalized by the active population for case B and case B'. We set the percentage of devices that are beamforming in a community as $x = 6\%$ (cf. Fig. 5.13(a)). Let $x = 0\%$ be the case B. Fig. 5.13(a) shows that the $\max\{I_{total,x=6\%} - I_{total,x=0\%}\} \approx 0.1792$. This difference in the actual population scale is ≈ 281240 (cf. Fig. 5.13(b)). The difference in the curves of I for the two cases clearly suggests that beamforming[2] can provide better results in terms of more devices receiving information quickly when $x = 6\%$. We then investigate the effect of x on the dissemination process. As suggested in [2], the number of long links decreases the average path length and hence more number of devices are able to receive information in less time, i.e., faster dissemination is achieved, similar results for this case is also obtained. We plot $I_{total,x=y\%} - I_{total,x=0\%}$, where $y \in 2, 6, 10$ (cf. Fig. 5.13(c)). Note that when $x = 2\%$ there are less number of devices that beamform and hence the curve takes off slowly when compared to when $x = 10\%$.

This proves that the information dissemination is enhanced in terms of number of devices receiving information over time. All the above results are generated considering the case when only one community, one eNB, initially has the source of information. However, in realistic scenarios, different public warning messages could originate in many different eNBs. Let two public warning message be generated in 2 communities (1 and 2) that are far apart. Here, we show the evolution of the number of devices in the active states normalized by N. Here we consider devices in state I as the devices that have received at least one public warning message. We then plot the difference in the fraction of devices in different active states on the actual population scale (cf. Fig. 5.14). The results show, the fraction of devices in state I for the case where sources are in two different communities, differ between $[-13365, 236460]$ in the non normalized case ($[-0.00085, 0.015073]\%$ in the normalized case) from the case when the sources are only in a single community. The results

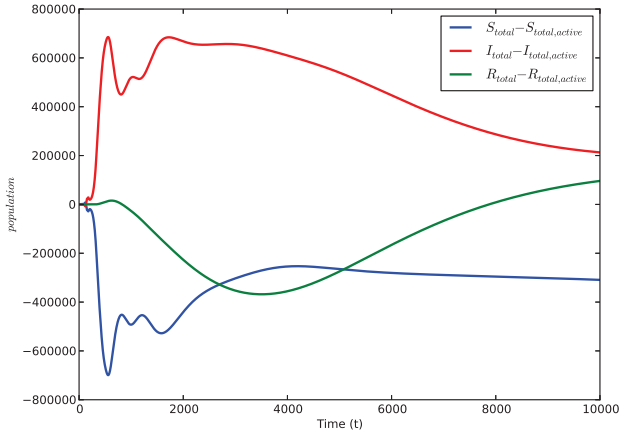
[2]A movie showing the dissemination process in Ivory Coast using the model can be viewed from http://complex.luxbulb.org/content/d4d-challenge



(a) Evolution of fraction of devices in S, I and R states for case A and case B.

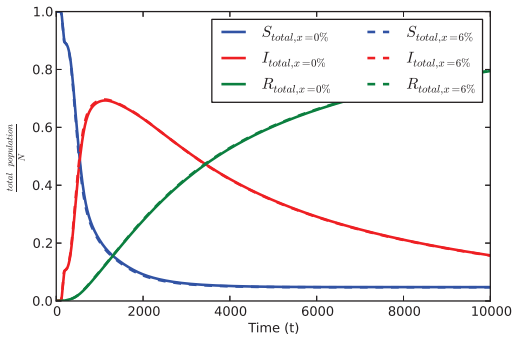


(b) Evolution of fraction of devices in S, I and R states for case A and case B when normalized by the fraction of active devices.

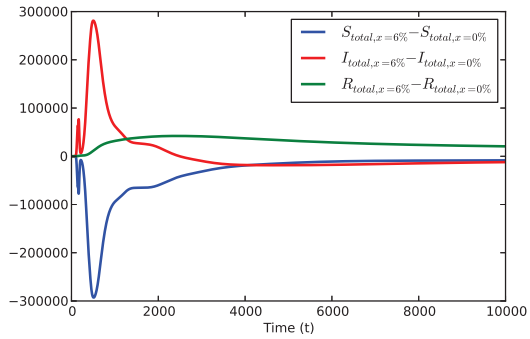


(c) Difference in number of devices in active states normalized over active population for case A and case B.

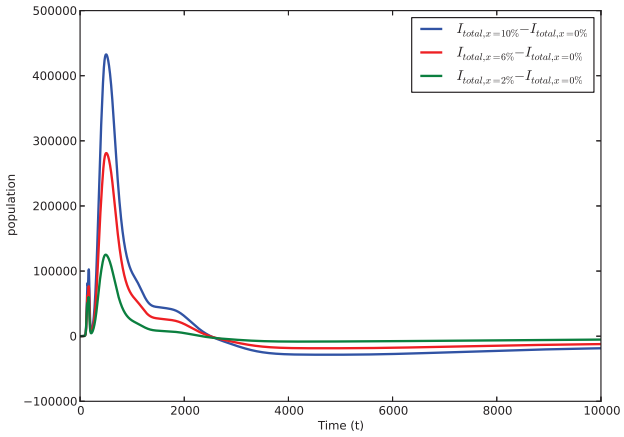
Figure 5.12: Evolution of fraction of devices in active states in Ivory Coast.



(a) Evolution of fraction of devices in the active states in Ivory Coast for case B' and case B. The results are normalized over active population.



(b) Evolution of difference in the number of devices in the active states in Ivory Coast for case B' and case B. The results are normalized over active population.



(c) $I_{total,x=y\%} - I_{total,x=0\%}$ when $y = 2\%$, $y = 6\%$ and $y = 10\%$.

Figure 5.13: Effects of beamforming.

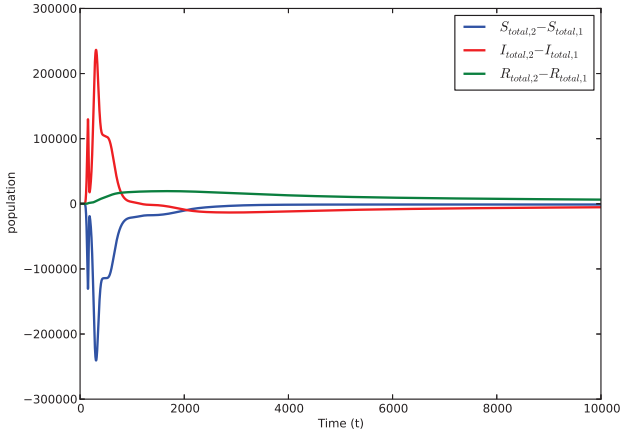


Figure 5.14: Evolution of difference in the number of devices in the active state. The difference is calculated between the case when sources are in two different communities and when the sources are in only one community. The number 2 and 1 in the legend suggest the how many communities generate the public warning message.

suggest that more devices have at least one message with them.

Further, all the above results are obtained using MP_c generated through G_c. Let for the two cases (when MP_c is used and when MP_v is used), the total number of devices in the three active states normalized by total number of devices in active state be termed as $S_{total,y}$, $I_{total,y}$ and $R_{total,y}$ where $y \in MP_c, MP_v$. When MP_v is used, different results in terms of number of devices in state I are obtained[3] (cf. Fig. 5.15). This is because now the movement of the device is only restricted to neighboring community and there are no long jumps as in the case when MP_c was used. However, it is observed that some time communities far apart also start disseminating information. Consider two communities i and j more than one hop away. Assume that community i have devices in state I initially. Over time, these devices from community i move to other neighboring communities. Now suppose that a device in state E_I has moved and remained in state E_I until it reached community j. On reaching community j the device switches to state I and start disseminating the information. The same phenomenon is observed in the obtained results[3]. As discussed before constraining the mobility affects the dissemination process, the same phenomenon is observed in the case when MP_v

[3]A movie showing the dissemination process in Ivory Coast using the model can be viewed from http://complex.luxbulb.org/content/d4d-challenge

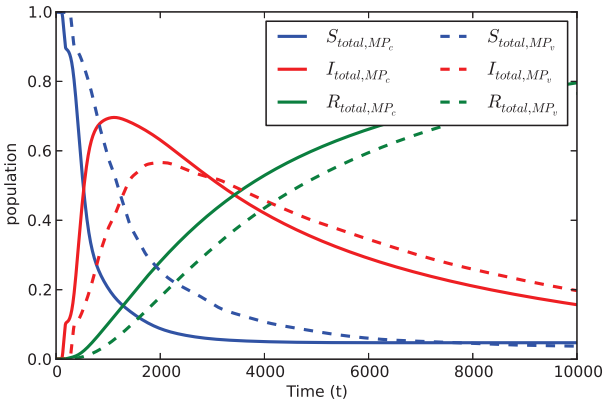


Figure 5.15: Effect of constraining mobility on information dissemination. Both the results are obtained under similar setting and when $x = 6\%$ but one uses MP_c while other uses MP_v.

is used, (cf. Fig. 5.15). It can be noted that there is a high difference in the devices in state I for the two cases.

We now show results obtained using the mobility model when return rates are associated. Towards this, we first show the difference in the fraction of active devices in states S, I and R when different mobility models are used with the addition of latent states and beamforming. Let the mobility model when return rates are used be called as *mobility*2 and mobility model when return rates are not used be called as *mobility*1. Here we show the difference in the fraction of active devices for the two cases. As return rates are included, a device in state I will return back to its community where the dissemination process has already been started and the fraction of devices not having the information will be low. This lowers dI_i/dt. When compared to when *mobility*1 is used, the $I_{mobility1,total} - I_{mobility2,total}$ is large. Fig. 5.16 shows the same. We then show the change in fraction of devices in the three active states when beamforming is used and when beamforming is not used on the *mobility*2. The percentage number of devices beamforming in a community are set to $x = 6\%$. Note that when there is no beamforming $x = 0\%$. As discussed previously that beamforming enhances information dissemination, same phenomenon was observed in the case of *mobility*2 also (cf. Fig. 5.17).

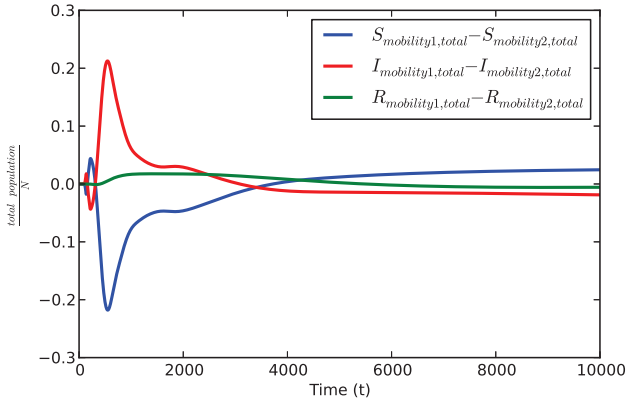


Figure 5.16: Effect of mobility model on information dissemination. Both the results are obtained under similar setting and when $x = 6\%$.

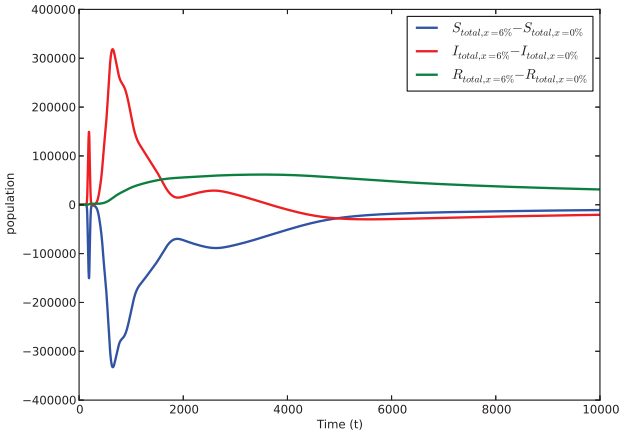


Figure 5.17: Evolution of differences in number of devices in the three active states. The difference is calculated between the case when $x = 6\%$ and $x = 0\%$ for the mobility model with return rate.

5.5 Conclusion

In this Chapter, we present a model that could help in enhancing the information dissemination in terms of the fraction of devices having the information across the population in a $D2D$ based communication network. We combine the concepts of epidemic model, beamforming and metapopulation to build our model. We use the data provided by the $D4D$ organizers to determine the movement rates for the movement of the devices from one community to another. To realize the information dissemination process we use SIR type epidemic model with addition of latent states E_S, E_I and E_R. We then add beamforming on the above proposed variation of the SIR model. Through the results, we show that by allowing small fraction of devices to beamform, dissemination process can be enhanced in terms of devices having information in the environment where the density of active devices is constantly changing. Thus, this Chapter presents three main contributions, first the design of the two models: a human mobility model for which we study the steady state limit and the transient behavior of an information dissemination model under the mean field assumption. Secondly, we introduced the latent states to account for variable density and finally, we showed how beamforming and mobility in a network with community structure could be applied to achieve large scale dissemination in a dynamic $D2D$ based communication network.

CHAPTER 6

Conclusion

The evolution of wireless networks and the need for providing self organizational capabilities to wireless devices for achieving better performance has lead us to propose algorithms in Chapters 3, 4 and 5. Better performance in wireless networks is achieved via reduction in average path length, increase in connectivity and how quickly information is disseminated in the network. The algorithms are designed such that they operate in a decentralized manner, use only locally available information, concepts from self organization such as Lateral Inhibition and Flocking, human mobility, stability, beamforming and dynamic density.

In Chapter 3 we investigate how small world properties can be efficiently achieved in static wireless networks in an autonomous environment by using concepts from self organization such as Lateral Inhibition and Flocking analogy. We also use the concept of beamforming to achieve the goals. The algorithm works using locally available information and does not require the knowledge of the whole network. Lateral Inhibition is used to form virtual communities within the network for reducing message complexity while the Flocking analogy is used to determine beam properties. Results show that, high performance can be achieved not only in terms of reduced APL but also in terms of connectivity. Reduced APL relates to smaller number of hops from one node to another. Hence, the transmission of information from a source node to the destination node requires less time.

However, due to limitations of the algorithm developed in Chapter 3, it is not applicable to mobile networks. Mobility creates diversities and clusters, and causes changes in the network structure. When mobility is associated, interactions between devices are intermittent. However, there is some percentage of certainty in the interactions. With respect to all the interactions a device had, we call the percentage certainty as stability. Using this definition of stability, we define an algorithm

in Chapter 4 that achieves better results in terms of dissemination. In Chapter 4 we propose a probabilistic approach to compute the stability of a node in a mobile environment. The algorithm uses beamforming to achieve the goal of faster dissemination of information in the network and operates using only locally available information. Beamforming is performed by devices in the network that have low stability and in the direction of high stability devices. In this Chapter, faster dissemination relates to how quickly large number of devices are able to receive the information flowing in the network. In this Chapter, it is also discussed that the proposed algorithm shows better results over other related studies. Further, the effect of the variation of the related parameters on the information dissemination is also clearly presented.

The algorithm in Chapter 4, however, does not consider the scenario when the density in mobile network is changing with time. Thus in order to incorporate dynamic density scenario, in Chapter 5 we define three latent states where the devices in the latent states do not participate towards disseminating information. We consider SIR type dissemination model and relate each state in SIR model with a latent state. The set of devices that are in S, I or R states are only considered in relation to the dissemination process. The transitions between active states and latent states happen with a certain rate. In order to achieve quicker dissemination, mobility is combined with beamforming on top of the dissemination process described by the updated SIR model. The results show that the process of dissemination in dynamic density scenarios can also be enhanced in terms of more number of devices being able to receive information in less time.

In all the three contributions, the number of beamforming devices are limited. In Chapter 3, Lateral Inhibition and Flocking alignment rule is used to identify the beamforming devices while in Chapter 4 low stability devices were allowed to beamform. In Chapter 5 the number of beamforming devices was set to a low value. Limiting the number of beamforming devices allows the small world phenomenon to occur as discussed in [2].

6.1 Future Directions

The algorithms proposed in this thesis provide a way to enhance information dissemination in an autonomous environment. The scenarios in which our algorithm is useful include spreading of public warning messages across the population efficiently. Enhancing information dissemination process

helps in obtaining quicker relief in a disaster hit scenario. Another, scenario where the proposed algorithms can be useful include, dissemination of traffic updates over a mobile network. The $D2D$ communication use-case currently shows the necessity of efficient dissemination of public warning message system. We think that our algorithms can bridge this gap.

In an autonomous environment, the decision of devices to beamform using local interactions where global favorable results are achieved can be studied using game theory. Thus, as a possible extension of the present work, game theoretic approach can be applied and studied to achieve the goals. In a game theoretic approach, the decision to beamform is a game that is played by the device. Whether to cooperate with other devices and whether changing the beam property can help in achieving global results, are some of the questions that can be interesting to research on. Towards this, one possible extension will be the application of evolutionary game theory to achieve the goals.

In Chapter 5, we have not used changes in the population size, i.e., addition of new devices and removal of old devices. This, in a communication environment can be explained through many ways. For example, due to hardware failures, change in technology and growing needs for the devices, to name a few. Addition of such concept would add more realism to our model. In order to model such processes, there is a need for the knowledge about the rates at which the devices are added or removed from the network. Capturing such information in CDRs could help in enhancing the model and will introduce more realism towards the study of the dissemination process. Also to model more realistic scenarios, we suggest the use of probability of connection in the given direction for the calculation of $\langle k \rangle$ [138].

The mobility characteristics used in this dissertation are obtained from the analysis of the sampled human mobility data. In Chapter 4, the mobility characteristics used are based on the mobility patterns observed for the data collected in the region of USA. However, in Chapter 5 we use the mobility characteristics that are derived from samples collected over the region of Ivory Coast. In both Chapters, due to the use of CDR, mobility can only be approximated. The sampling of the mobility data should be conclusive and extensive to exactly map human mobility pattern. The above requirements would lead to large number of samples extracted at a very high rate and would require efficient data storage mechanisms.

Applications studying human mobility are diverse and range from studying traffic flows, cascad-

ing behaviors, road/rail network planning, understanding urban developments (growth of spatial networks like roads and railways [152]), epidemic control and outbreaks, evolution of networks etc. The study of such applications is only possible if the data collected on human mobility is conclusive and has both timing and behavioral aspect. The study of such applications is hampered if the sampled data is a CDR. In CDR a tuple is only created when a call is made and not when the location is changed. Thus, mobility can only be approximated using CDR. If the above mentioned applications are considered, CDR needs to be replaced with more active entries of the mobility information. If such active datasets on human mobility are made available, better understanding of human mobility can be achieved and the above mentioned applications can be made more robust.

Considering individuals with autonomous behavior, the datasets collected thus far have failed to record any behavioral aspect attached to human mobility. Researchers cite security issues towards having such kind of information in the dataset. However, if such information were to be made available, mobility models could be made more subtle. Using such models will surely enhance the algorithms proposed in Chapter 4 and 5. The lack of such dataset and efficient storage of dataset enables us to propose another future direction whereby mobility datasets could be made more comprehensive and manageable.

APPENDIX A

A Self-Organization Framework for Wireless *Ad hoc* Networks as Small World

This Appendix provides detailed description of algorithm proposed in [1]. In this Appendix, we propose a distributed algorithm for small world creation that achieves path length reduction while maintaining connectivity. In this Appendix, we first define a new centrality measure that estimates the structural importance of nodes based on traffic flows in the network, which is used to identify the set of nodes that beamform. We show, using simulations, that this leads to significant reduction in path length while maintaining connectivity.

Central to the design is a new measure of centrality defined in this Appendix that allows distributed estimation of the structural importance of nodes in the network. We define Wireless Flow Betweenness (WFB) which gives an accurate estimate of the Flow Betweenness Centrality (FBC) [55]. The proposed measure enhances the earlier measure defined in [57] by identifying key redundancies. The key aspect of WFB is that it can be computed in a completely distributed manner by exploiting the wireless broadcast advantage (WBA) [153] for information regarding traffic flow, thereby incurring negligible overheads.

The rest of the Appendix is organized as follows. Section A.1 introduces WFB after the introduction of table of notations used. A centralized scheme for choosing beamforming nodes is shown to give promising results for small world creation. Subsequently, a distributed algorithm for small world creation using beamforming is proposed and evaluated using simulations in Section A.2. Finally, the Appendix concludes in Section A.3.

Notation	Meaning
$a_f(v)$	additional flows for node v
B_w	beam width
B_w^*	chosen beam width
D	diameter of the network
d_i	absolute difference in ranks of the node i
f	fraction of nodes generating packets
$g(v)$	number of packets forwarded by node v
$L(p)$	Average path length of the network corresponding to p
m	number of antenna elements in a directional antenna
$\mathcal{N}(v)$	Set of neighbors of node v
$o(v)$	number of packets overheard by node v
p	fraction of edges rewired or nodes beamforming
p_{nf}	probability that at least one node is located in the first sectoral region
p_{lf}	probability that at least one node is located in the last sectoral region
r	omnidirectional transmission range
V	the set of all nodes in the network
$w(v)$	betweenness of node v
β	similarity factor
ρ	Spearman's correlation coefficient

Table A.1: Notations and their meaning.

A.1 A Distributed Definition of Centrality for Wireless *Ad hoc* Networks

For the problem of distributed self-organization of the network, there is a need for nodes to accurately estimate their importance with respect to paths in the network. This constraint makes it unsuitable to use any of the conventional measures as they consider a centralized view of the network. In addition to the distributed computation of the centrality itself, nodes need to decide on their beamforming behavior based on these values. As nodes with high betweenness are ideal for shortcut creation, nodes need to estimate the rank of their betweenness in the network. In order to satisfy these requirements, the algorithm design in this Appendix is centered on a new measure of centrality that allows nodes to decide on shortcut creation depending on the structure of the network.

A.1.1 Wireless Flow Betweenness

In wireless *ad hoc* networks, the computation costs are particularly costly because of the transmission overheads involved. In order to ensure that network performance is not affected, thus, the computation of a betweenness measure should ideally incur zero overheads. Using this as the motivation, We propose a measure of betweenness computed by nodes by exploiting the wireless broadcast advantage (WBA), thereby minimizing the overheads incurred.

The proposed measure WFB, makes use of information extracted from traffic flows being routed through the network to estimate the structural importance of nodes in the network. This allows the computed measure of centrality to adapt to the traffic flow patterns in the network. Further, traffic flow information can be obtained locally by nodes, thereby minimizing the transmission overheads. Finally, such a mechanism can proceed in parallel with regular network operations such as routing and packet delivery, therefore increasing its robustness. After starting off with a simplistic expression of locally computable centrality we identify the redundancies involved in such a measure and describes how they are addressed.

The broadcast nature of the wireless medium results in implicit sharing of information as nodes can overhear all transmissions in their one-hop neighborhood. Based on this neighborhood information, a node estimates its importance in terms of how often it transmits its information, either as a source or as a forwarding node. Consider a network $G(V, E)$ with set of nodes V and set of edges E. Let $g(u)$ denotes the number of packets forwarded by a node $u \in V$ for distinct source-destination pairs while $\mathcal{N}(v)$ denotes the set of neighbors of $v \in V$. A straightforward measure of betweenness of a node v, can therefore be computed as the ratio of the number of times it acts as a forwarding node to the total number of unique traffic flows in its neighborhood. This can be expressed as

$$w(v) = \frac{g(v)}{\sum\limits_{u \in \{\mathcal{N}(v) \cup v\}} g(u)}. \tag{A.1.1}$$

The denominator in the eq. A.1.1 gives the total number of packets forwarded in the neighborhood of v. The eq. A.1.1, though straightforward in design, only gives an egocentric measure of importance of the node v. This is because a count of the transmissions in the neighborhood does not give any indication of the likelihood of any flow in the network passing through v. Identifying the exact number of traffic flows in the entire network, including those that do not pass through

the neighborhood of v, would be non-trivial and costly. Instead, we propose to estimate the betweenness of a node by propagating network information as part of traffic flows which uses recursive computation at nodes to obtain an accurate estimate.

The betweenness of a node v based on the set of traffic flows in the network, was defined in [55] as the FBC which refers to the share of the maximum flow between all node pairs in the network that passes through v. However, FBC requires nodes to maintain global network state. Many other proposed measures of centrality for wireless *ad hoc* networks [154] requires knowledge of multi-hop neighborhood of a node, which is again costly in terms of transmission overheads. In order to minimize the state information, the approach centers around estimation of the number of flows by each node using neighborhood values of centrality.

For any node u, the number of transmissions it is aware of is obtained as $\frac{g(u)}{w(u)}$. Thus, a node v can gain an estimate of the number of transmissions in the network using the above estimation for each of its neighbors $u \in \mathcal{N}(v)$. Nodes piggyback their self-computed values of centrality whenever they transmit a packet, either as a forwarding node or as source. However, this involves multiple redundancies that can result in inaccurate estimation.

The first redundancy arises due to the term $g(u)$ that gives the number of packets transmitted by a node u, which lies in the neighborhood of v. As v can overhear all of u transmissions, it counts all of these as part of the number of traffic flows $g(v)$ it counts in its neighborhood and uses to compute $w(v)$. Thus, $g(v) \geq g(u) \forall u \in \mathcal{N}(v)$. As $g(u)$ is counted again as part of the expression $\frac{g(u)}{w(u)}$, this results in overestimation. Thus, node v needs to ensure that the term $g(u)$ is not counted multiple times so as to accurately estimate the number of traffic flows which u is aware of but v is not. Therefore, the estimated number of additional flows that u is aware of is given as

$$a_f(u) = \frac{g(u)}{w(u)} - g(u). \tag{A.1.2}$$

The betweenness of v can be computed using the a_f values of all its neighbors as

$$w(v) = \frac{g(v)}{o(v) + \sum\limits_{u \in \mathcal{N}(v)} a_f(u)}.$$

where $o(v)$ denotes the number of traffic flows overheard by v, including those transmitted by v itself. A node updates its betweenness if it either transmits a packet or overhears one from its neighbors.

However, the above summation over all neighbors of v introduces a second redundancy which needs to be taken care of. Just as v is aware of all packets transmitted by a neighbor u, multiple neighbors of v also affect each other's betweenness values. Neighbors of v which are also neighbors of each other estimate the term a_f based on the values of w observed for each other. Consider two nodes u and u' which are neighbors of v as well as neighbors of each other. As part of their centrality calculations, u and u' estimate $a_f(u')$ and $a_f(u)$ respectively. Both these values include the number of transmissions made by v. In the absence of a one-hop neighborhood map, neither node is aware of v being a common neighbor and therefore, $a_f(v)$ is implicitly counted more than once, resulting in overestimation of the number of flows at all the three nodes. This effect is aggravated over multiple iterations and for networks with high node density.

In order to handle the second redundancy, the set of neighbors whose betweenness values are used for computation is limited. Nodes that infrequently act as forwarding nodes spend the majority of time overhearing transmission from other nodes. Computing the a_f measure for such a flow is likely to result in a lot of redundancy as the majority of estimated additional flows are likely to be redundant. Thus, a the definition of WFB is revised to only consider the neighbor which acts as a forwarding node most often. The possibility of redundancy is reduced with such a choice as the node chosen is unique within a neighborhood. Further, since a majority of traffic in a given network region flows through this node, it gives a better estimate of the number of flows. Based on this discussion, the WFB value of a node v is obtained as

$$w(v) = \frac{g(v)}{o(v) + (\frac{g(u)}{w(u)} - g(u))} \qquad \text{(A.1.3)}$$
$$\text{where } u = \arg\max_{u' \in \mathcal{N}(v)} \{g(u')\}.$$

A node computes its WFB value whenever it acts as a forwarding node or overhears a transmission in its neighborhood.

By considering a single node in the neighborhood, the chances of overestimation are reduced. However, it also raises the possibility of some flows not being counted. Looking at the denominator of the expression in (A.1.3), it is observed, as before, that the term $o(v)$ accounts for all transmissions in the neighborhood of v, and therefore includes $g(u)$. Thus, the term $(\frac{g(u)}{w(u)} - g(u))$ gives the number of additional flows in the network estimated recursively. As with node v, the value of $w(u)$ takes

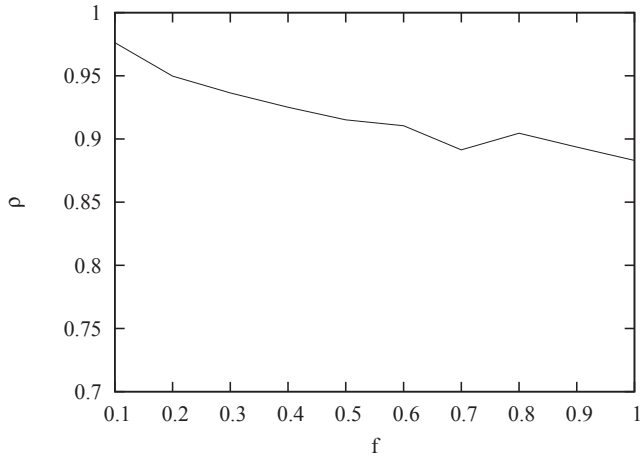


Figure A.1: Correlation between WFB and FBC for varying traffic load.

into consideration the number of flows overheard by u as well as the node with maximum forwarding count in $\mathcal{N}(u)$. The consideration of the neighbor with maximum forwarding count implies that a majority of flows are always accounted for in a neighborhood, thereby minimizing the probability of flows not being counted. It can be noted that this estimation over successive hops can stop when a node itself has the maximum forwarding count in its neighborhood, leading to an estimation error. On further consideration, however, it is observed that this is unlikely to occur unless the node is the global maxima, i.e. has the maximum centrality value in the entire network. Thus, the possibility of inaccuracies in estimation is minimized. This is verified in the next Section based on the rank correlation of WFB values in the network with that of centrally computed FBC.

A.1.2 Correlation with FBC

To verify the validity of the proposed WFB measure,we compare the values obtained with the FBC. Since the relative importance of a node with respect to either betweenness measure is given by its rank in the network, rank correlation of the two measures is obtained. The nodes are ranked in the network separately according to their WFB and FBC values and Spearman's correlation coefficient is found between two as

$$\rho = 1 - \frac{6 \sum_{i \in V} d_i^2}{n(n^2 - 1)} \tag{A.1.4}$$

where V is the set of all nodes in the network and d_i is the absolute difference in rank between the two rankings for the i^{th} node. A value $\rho = 1$ implies perfect correlation between the two rankings.

We vary the amount of traffic in the network and obtain the value of ρ for the WFB and FBC values obtained thereafter. A fraction f of all nodes in the network generate a packet to randomly chosen node as the destination. The results are shown in Fig. A.1. A very high correlation is seen for lower values of f but it drops slightly for higher values. Since the computed WFB measure does not make use of explicit transmissions to propagate information across the network, as f increases to 1, a greater percentage of information does not propagate over multiple hops. This results in lower values of ρ as f increases. However, even for high values of f, $\rho > 0.88$ implies high correlation between WFB and FBC.

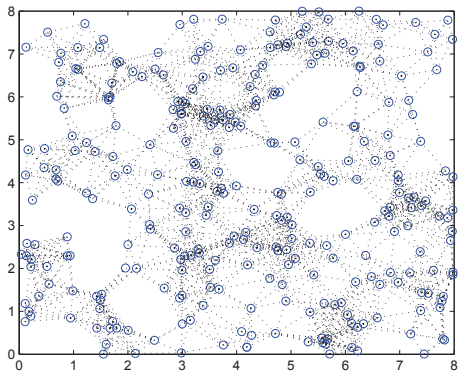
A.1.3 Overhead and Buffer Costs

While the above results show that there is a close correlation between the WFB values computed at individual nodes and the corresponding FBC values, it is necessary to understand the additional costs required for such computation.

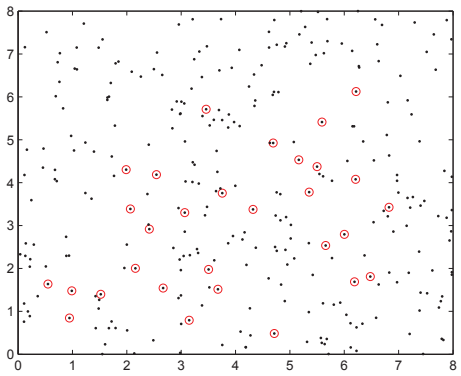
The transmission overhead costs are minimal since only WFB values are piggy-backed on to packets by the forwarding nodes, involving one additional field. Thus, additional transmission costs are of constant order. The design, however, requires nodes to store the WFB values of their neighbors along with the corresponding forwarding count. The buffer requirements, thus, scale with increase in the node density. Given a neighborhood size of n nodes, a node needs to store three fields for each neighbor, namely the node identity along with the WFB value and the forwarding count, resulting in a buffer size of the order of $O(3n)$. However, since a lot of existing *ad hoc* network mechanisms rely on the presence of neighborhood knowledge, which is of $O(n)$, the additional costs involved for computation of WFB are unlikely to impose a significant burden.

A.1.4 Using WFB for Small World Creation with Beamforming

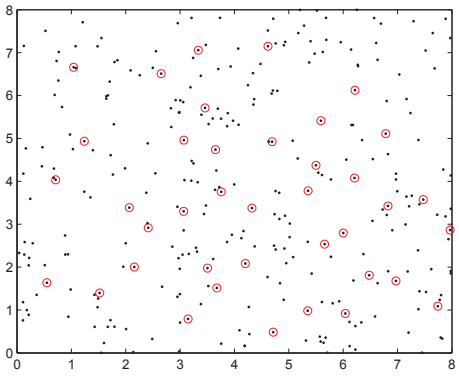
We now explore the use of WFB for small world creation using directional beamforming. Motivated by the close correlation between WFB and FBC, we evaluate the performance benefits of using



(a) Initial network setup with omnidirectional beams

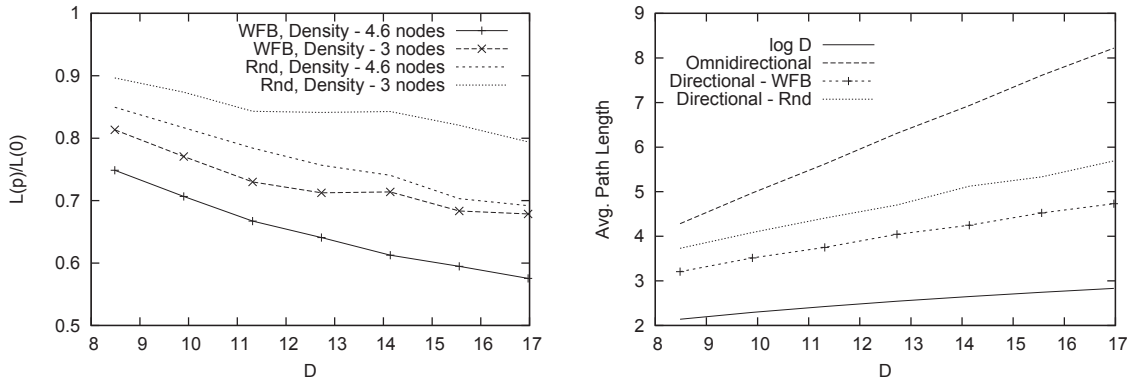


(b) Centralized choice of nodes ranked by WFB values



(c) Distributed choice of beamforming nodes

Figure A.2: Selection of network using centralized and distributed choices of beamforming nodes using WFB.



(a) Path length reduction

(b) Growth of APL

Figure A.3: Path length reduction by beamforming at nodes with high values of WFB.

nodes with high values of WFB as beamforming nodes. The issue of distributed identification of beamforming nodes is discussed later. Instead, it is considered that somehow the top ranking nodes with respect to WFB values are identified and create beams. Based on the insights obtained in this Section, we formulate an algorithm for distributed beamforming in the next Section.

We consider that the top 10% (i.e. $p = 0.1$) of all nodes in the network create directional beams using sector model. Let r be the omnidirectional range and B_w be the beam width. The optimal beam width, B_w^*, used is calculated

$$B_w^* = \arg\max_{B_w} \left\{ \left[r\sqrt{\frac{2\pi}{B_w}} \right] p_{nf} p_{nl} \right\}. \tag{A.1.5}$$

where p_{nf} and p_{nl} are the probabilities that at least one node is located in the first and last sectoral regions. Note that, we assume the path loss exponent to be equal to 2 and free space propagation. The choice of the two probability terms is dictated by the motivation of maintaining connectivity while maximizing beam length. Since an accurate estimate of maintaining bidirectional connectivity would require knowledge of the entire network, we use the term p_{nf} to estimate the probability that connectivity is maintained with the omnidirectional neighborhood. As the first region under the beam lies within the omnidirectional range, a higher number of nodes here increases the probability that omnidirectional neighbors can be reached. Increasing the beam length, however, is achieved by reduction in beam width. The narrow beam width reduces the probability of a node maintaining connectivity to its omnidirectional neighborhood. The second probability p_{nl} indicates

the probability that at least one node benefits from the increased beam length. If there are no nodes present in the last region, greater connectivity can be achieved by increasing the beam width while the improvements from the beam length stay the same. The nodes in the middle regions of the beam are not used in the expression since they do not represent the maximum benefits achievable by increased beam length nor are they the most affected in terms of connectivity as a result of reduced beam width. The values for p_{nf} and p_{nl} are obtained based on the node density in the network. Given that the number of nodes in the omnidirectional neighborhood of a node is n, the corresponding values are obtained as $p_{nf} = 1-(1-\frac{A_f}{\pi r^2})^n$ and $p_{nl} = 1-(1-\frac{A_l}{\pi r^2})^n$ where A_f and A_l are the area of the first and last regions respectively. Recall that the area under the beam is equal to that of the omnidirectional area when the same transmit power is used. The value of m, number of antenna elements, used is computed as $m = \lceil \left(\frac{r(B_w^*)}{r}\right)^2 \rceil$ where $r(B_w^*)$ is the transmission range after beamforming.

Further, nodes orientate their beams in a direction in which they record the maximum hop count based on earlier traffic flows, so as to minimize the network diameter. The simulation results shown here only consider the sector model. The results are compared for nodes distributed at two different node densities. A high node density of $d = 4.6$ nodes per unit area results from distributing 300 nodes over an 8x8 region, and a lower density of $d = 3$ is obtained by increasing the region size to 10x10 for the same number of nodes. Subsequently, the density is kept constant at either of these two values and only the the size of the region is varied.

We show the performance improvements available for increasing size of the network region, indicated using D as the maximum distance between any two nodes. Fig. A.2(b) shows the set of nodes chosen as beamforming nodes for the network setup shown in A.2(a). It can be seen that the majority of the nodes chosen (circled in red) are located towards the center of the network.

Fig. A.3 illustrates the impact on path length reduction. A reduction of more than 40% and 30% are shown to be achievable for $d = 4.6$ and $d = 3$ respectively in Fig. A.3(a). Here, it is notable that the improvements in both cases are higher than the corresponding value of p for randomized beamforming. As the network size increases, the beam length to diameter ratio reduces to $\frac{r(B_w^*)}{D} \approx 0.2$ for $d = 3$ and $\frac{r(B_w^*)}{D} \approx 0.25$ for $d = 4.6$. Using the results in [48] as a benchmark, we find that the path length reduction is greater for corresponding values of $\frac{r}{D}$. The results in Fig.

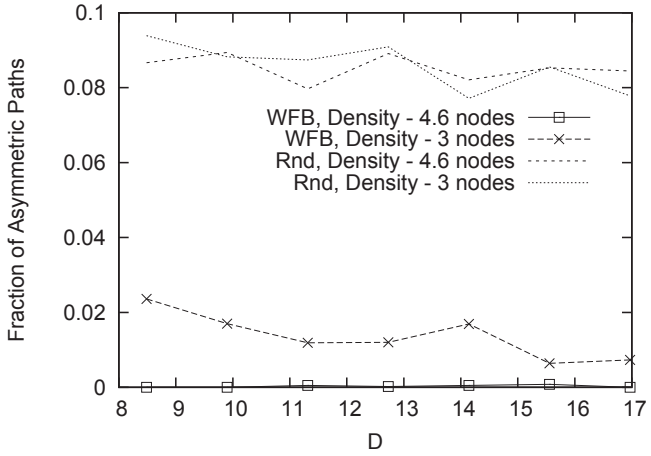
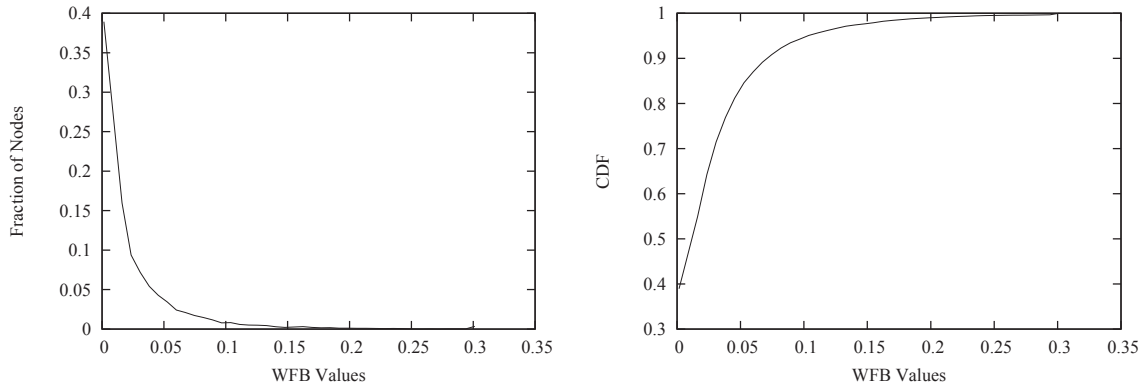


Figure A.4: Impact on unidirectional connectivity.

A.3(b) compare the growth of the average path length with the logarithm of the network size. The line corresponding to reduced path length using directional beams is shown to grow in parallel with the line corresponding to $\log D$, implying that the growth in path length is logarithmic to that of the network size.

Fig. A.4 shows the impact of directional beamforming using WFB on the fraction of node pairs that are unidirectionally connected. The effect on unidirectional connectivity is negligible for $d = 4.6$. For $d = 3$, a relatively higher fraction of node pairs are connected unidirectionally, though this is still lower than the value for randomized beamforming. The improvement in connectivity results from the fact that nodes with high value of centrality are better connected. For lower node density, the higher fraction of unidirectional connectivity results due to the choice of beamforming nodes. Nodes with high values of WFB are likely to be neighbors of each other as they are located towards the center of the network. As nodes with high values of WFB are chosen as the beamforming nodes, they are likely to be neighbors to each other. In the case of lower node densities, this results in a higher fraction of neighboring nodes which are beamforming thereby resulting in unidirectional connectivity. This is alleviated for higher node densities as, in spite of neighboring nodes creating beams, a higher number of neighbors continues to use omnidirectional beams thereby increasing the chances of bidirectional paths.



(a) Distribution of WFB values (b) Cumulative distribution of WFB values

Figure A.5: Distribution of WFB values in the network.

A.2 Distributed Small World Creation using WFB

We now focus on distributed algorithm design for small world creation using directional beamforming. The design centers around nodes determining their beamforming behavior based on their estimated importance in the network, thereby adapting small world creation to the network structure. Specifically, the beamforming behavior of a node is determined based on the relation between its WFB value and those of its neighbors. This is shown to result in significant reduction in average path length accompanied by negligible loss in connectivity, thereby realizing small world properties.

In the previous Section, we illustrated the benefits achievable by using top ranked nodes by WFB as beamforming nodes. However, as the set of nodes with highest values of WFB was identified using global network information, it is not suitable for distributed implementation. Further, as discussed before, only choosing nodes with high values of WFB can result in unidirectional connectivity at low node densities.

A.2.1 Distributed Beamforming Algorithm

The distribution of WFB values in the network is shown in Fig. A.5. As expected, the majority of nodes have low centrality values while very few nodes have very high values. A very small fraction of nodes has WFB values in the top 2%, and can therefore be identified easily. However, it was noted

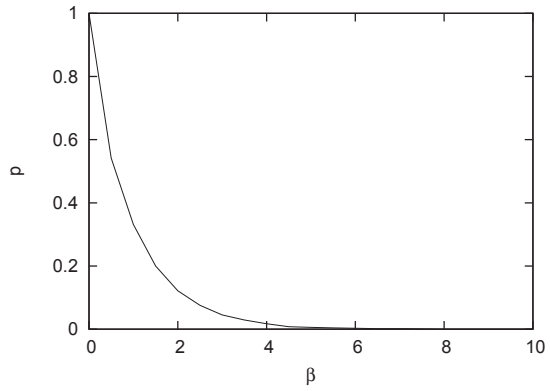


Figure A.6: Relation between β and p

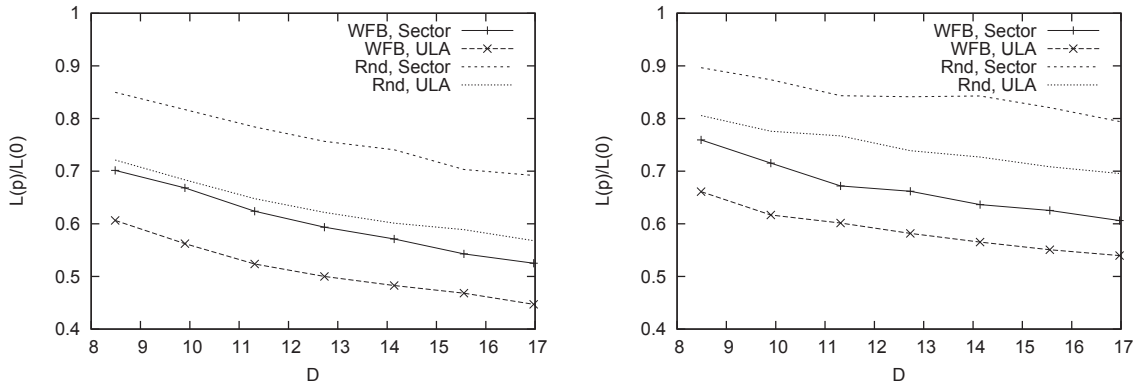
that a large fraction of nodes has WFB values lying in the 90-95 percentile. This implies a higher likelihood of these nodes being neighbors and therefore, are likely to result in loss of connectivity if they are all chosen as beamforming nodes. Based on this observation, an algorithm in which nodes decide on their behavior based on WFB values observed in the neighborhood is proposed.

As nodes broadcast their WFB values as part of packet transmissions, all nodes are aware of the values of their neighbors. As part of the algorithm, a node v decides on using a directional beam if its own value exceeds that of its neighbors by a *similarity factor* β. Thus, the beamforming condition for a node v can be expressed as

$$\frac{w(v)}{w_{avg}(\mathcal{N}(v))} > \beta \tag{A.2.1}$$

where $w_{avg}(\mathcal{N}(v))$ denotes the average WFB for v's neighborhood. The above condition ensures that only nodes with high values of WFB choose themselves for beamforming. The fraction of such nodes is determined by β. A higher value of β implies a stricter condition resulting in fewer beamforming nodes. Further, the condition also results in a lower chance of neighboring nodes creating beams. This is because, a node is chosen if its value exceeds those of its neighbors by the similarity factor β, the condition is unlikely to hold true for any of the neighbors themselves, and therefore excludes them from beamforming. The choice of beam width and beam direction is the same as for centralized choice of beamforming nodes.

The set of nodes that decide on using directional beams is illustrated in Fig. A.2(c) for the network setup in A.2(a). Note that, compared to the centralized choice of nodes in Fig. A.2(b),



(a) Path length reduction for $d = 4.6$ nodes per unit area (b) Path length reduction for $d = 3$ nodes per unit area

Figure A.7: Path length reduction by beamforming for distributed choice of nodes.

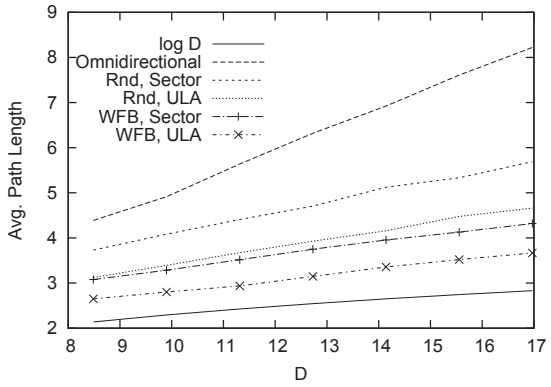
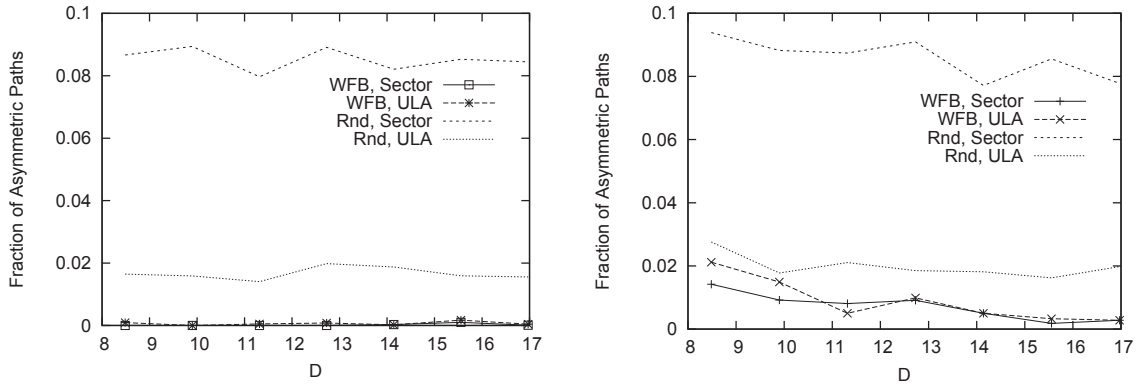


Figure A.8: Growth of average path length with the logarithm of the network size.

the distributed choice of nodes are relatively more spread out, thereby reducing the set of neighbors that simultaneously use beams resulting in better connectivity.

A.2.2 Simulation Results

We evaluate the proposed algorithm using simulations at different node densities as earlier. The value of β is chosen such that the fraction of beamforming nodes stays close to $p = 0.1$, which was used for the earlier set of simulations using centralized choice of nodes. Using simulations, we identify this value as $\beta = 2$ which results in values of p between $0.11 - 0.13$. The relation between β and p is shown in Fig. A.6. The results are obtained using the sector model as well as the realistic



(a) $d = 4.6$ (b) $d = 3$

Figure A.9: Impact on unidirectional connectivity for distributed choice of nodes.

ULA model and MATLAB. Each individual simulation is averaged over 40 different topologies and the results obtained as the mean over all possible node pairs in the network.

Fig. A.7 illustrates the path length reduction for node densities $d = 3$ and $d = 4.6$ nodes per unit area for both the sector and ULA models. When using the sector model, the path length reduction achieved is higher than in the case of centralized choice of nodes. This additional benefit is again due to the fact that fewer fraction of nodes that are omnidirectional neighbors beamform as a result of the proposed algorithm. Directional beams that are located close to each other tend to reduce the improvement in path length as only one of the two may get chosen most of the time, thereby rendering the other redundant. The performance for the ULA model is even more promising as greater than 40% reduction is achieved in both cases, with the improvement being greater than 50% for greater size of the network region with $d = 4.6$. Fig. A.8 shows the growth in path length for $d = 4.6$ with respect to the logarithmic of the network size. As the choice of beam width is the same as earlier, we obtain a similar values of $\frac{r(B_w^*)}{D} \approx 0.2$ for $d = 3$ and $\frac{r(B_w^*)}{D} \approx 0.25$ for $d = 4.6$ when the network size increases. The reduction in path length, is, therefore, much higher than corresponding values of $\frac{r}{D}$ and p shown in [48].

The impact on connectivity is illustrated in Fig. A.9. As in the case of centralized choice of nodes, the fraction of node pairs unidirectionally connected is negligible for higher node density, $d = 4.6$. For lower node density $d = 3$ as well, the impact on connectivity is lower than in the earlier case.

A.3 Conclusion

In this Appendix, we explored the use of directional beamforming for self-organization of a dense wireless *ad hoc* network as a small world and provided a simulation based analysis of the achievable performance benefits of randomized beamforming and identify the challenges involved. Subsequently, a detailed distributed algorithm for nodes to decide on their beamforming behavior was provided with a new measure of betweenness centrality, Wireless Flow Betweenness (WFB), which is used to identify beamforming nodes. Using simulation results it was shown that significant performance benefits can be achieved over randomized beamforming.

APPENDIX B

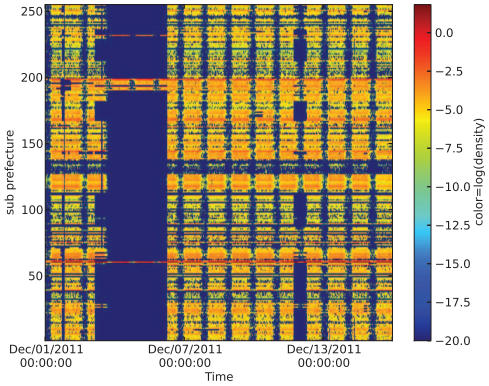
SET3 Details

The SET3 is a coarse grain dataset containing billion records which is sub divided into 10 small sub datasets each consisting of mobility patterns for two weeks. Table B.1 shows the start and end time for the sub datasets of SET3. The mobility patterns are generated after concatenating these sub datasets. Each entry in the sub dataset has been made when a user makes a call and consists of fields like associated subprefecture ID, user coordinates and time when the call was made. Such kind of entry makes SET3 a Call Detail Record (CDR) dataset. As stated in Chapter 5, due to the information pertaining to users, we are able to generate the density of the users over each sub dataset (cf. Fig. B.1 to B.3) and extract mobility information in terms of MP_c and MP_v as shown in Section 5.2.2. Both MP_c and MP_v are asymmetric matrices (cf. Fig. B.4(a) to B.4(b)). The asymmetrical nature of the movement rate matrices relates to the fact that first, users are not tracked continuously, second, the dataset is a CDR with a possibility of user making a call mainly from certain locations, third, the number of user going from subprefecture i to subprefecture j could be different from the number of those going from j to i.

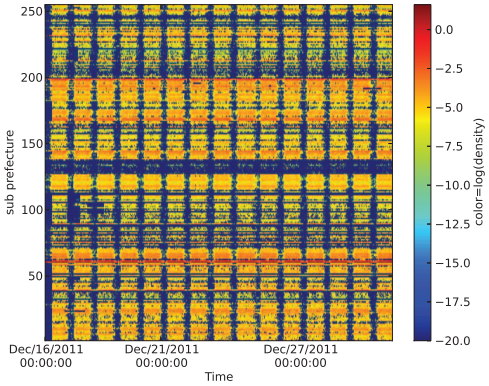
As SET3 is related to the mobility of the users it is essential to determine the mobility parameters associated to the data. We thus determine the distribution of jump length and radius of gyration for the user. Similar to the definitions in Section 5.2.2, consider a user u from the dataset. Let subprefecture entry for each entry for u form a set U such that $U = (u_1, u_2, \ldots, u_n)$ where u_i is the subprefecture entry. The jump length is calculated as the euclidian distance between two consecutive entries, i.e., jump length= $\sqrt{(u_i(x) - u_{i-1}(x))^2 + (u_i(y) - u_{i-1}(y))^2}$. The radius of gyration, r_g, is calculated as the root mean square distance of the jump length from the mean jump length. Consider X be the set of jump length samples. The mean jump length is defined by $\overline{x} = \frac{1}{|X|} \sum_{i \in X} i$. Radius of

Sub Dataset Number	Start	End
A	Dec/01/2011 00:00:00	Dec/15/2011 23:59:00
B	Dec/16/2011 00:00:00	Dec/30/2011 23:59:00
C	Dec/31/2011 00:00:00	Jan/14/2012 23:59:00
D	Jan/15/2012 00:00:00	Jan/29/2012 23:59:00
E	Jan/30/2012 00:00:00	Feb/13/2012 23:59:00
F	Feb/14/2012 00:00:00	Feb/28/2012 23:59:00
G	Feb/29/2012 00:00:00	Mar/14/2012 23:59:00
H	Mar/15/2012 00:00:00	Mar/29/2012 23:59:00
I	Mar/29/2012 23:00:00	Apr/13/2012 23:59:00
J	Apr/13/2012 23:00:00	Apr/28/2012 23:59:00

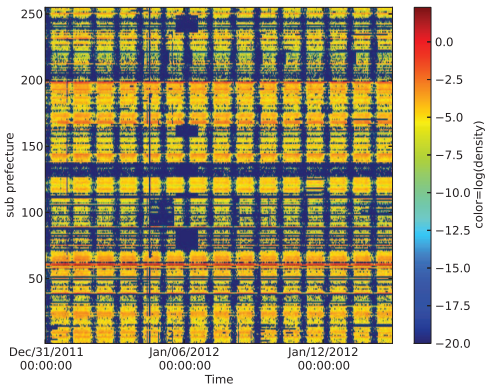
Table B.1: Start and End time for fragments of SET3.



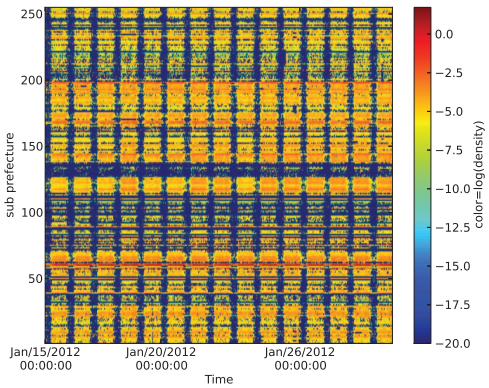
(a) User density for subset A.



(b) User density for subset B.

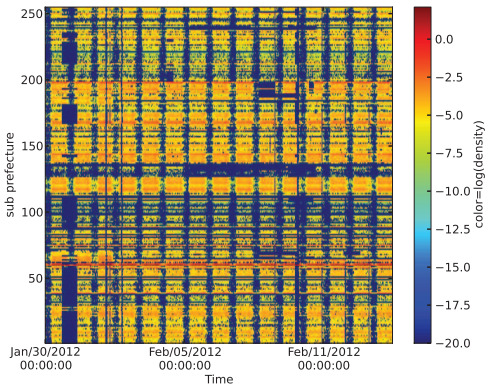


(c) User density for subset C.

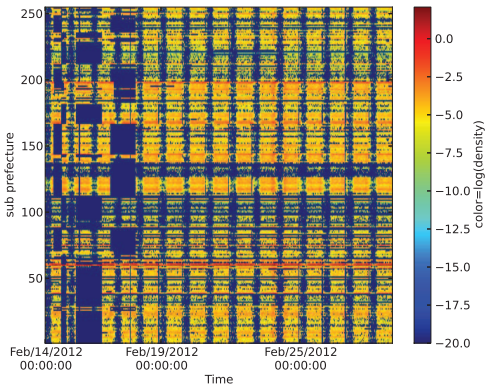


(d) User density for subset D.

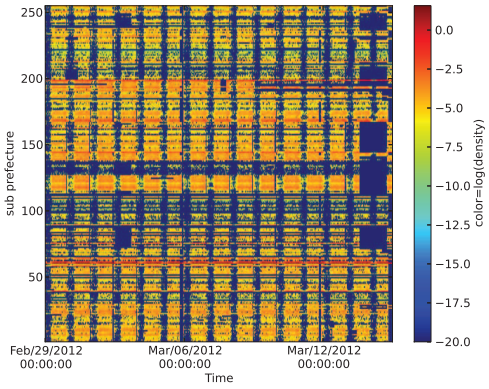
Figure B.1: User density for subsets from A to D.



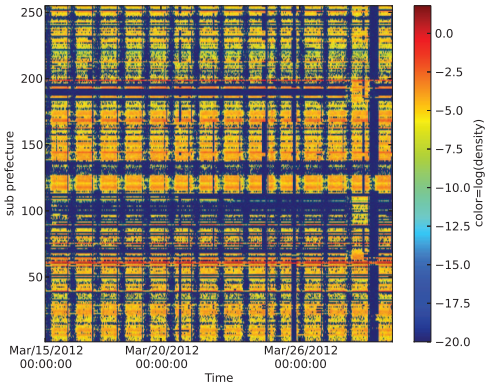
(a) User density for subset E.



(b) User density for subset F.

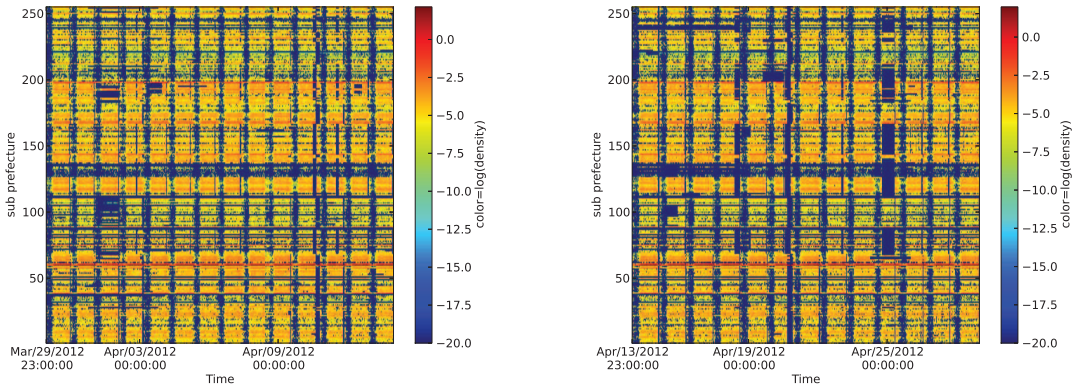


(c) User density for subset G.



(d) User density for subset H.

Figure B.2: User density for subsets from E to H.

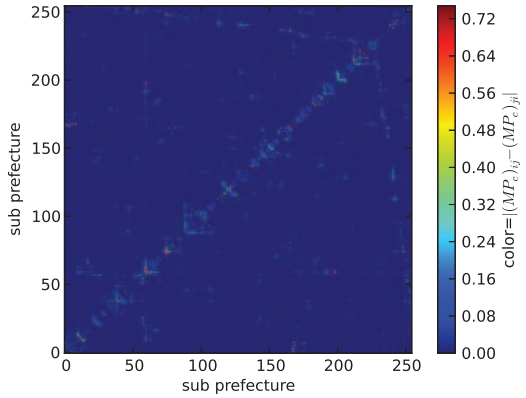


(a) User density for subset I.

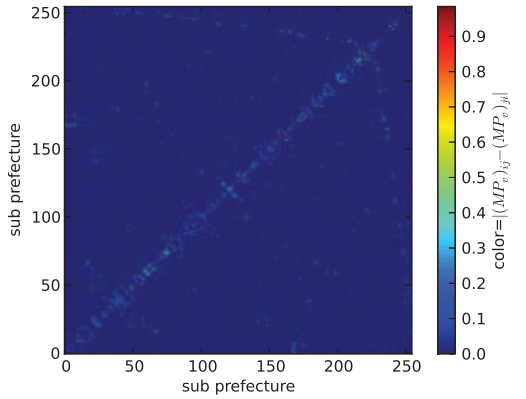
(b) User density for subset J.

Figure B.3: User density for subsets from I to J.

gyration is thus defined as $r_g = \sqrt{\frac{1}{|X|}\sum_{i \in X}(i - \overline{x})^2}$. Fig. B.5(a) and B.5(b) provide the $CCDF$ of the jump length and radius of gyration respectively on the loglog scale. It is observed that the jump length distribution in the Ivory coast had cuttoff distance as 122 ± 5km and power law exponent as 1.62 ± 0.03 [155].

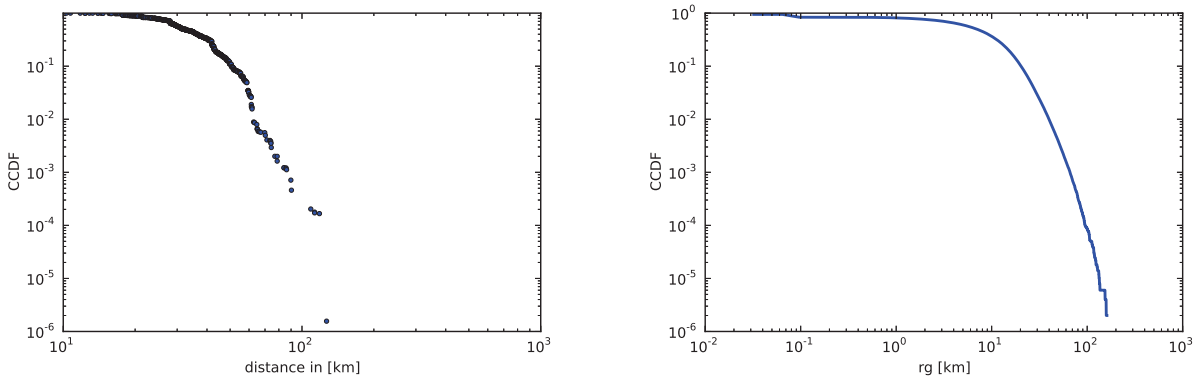


(a) $|MP_c(ij) - MP_c(ji)|$.



(b) $|MP_v(ij) - MP_v(ji)|$.

Figure B.4: Asymmetric property of conditional probability matrix for movement.



(a) Jump length.

(b) Radius of Gyration.

Figure B.5: Mobility Characteristics.

APPENDIX C

Résumé - Vers une amélioration de la diffusion des informations dans les réseaux sans-fils

Auteur: Rachit Agarwal

Abstract: Dans les systèmes d'alertes publiques, létude de la diffusion des informations dans le réseau est essentielle. Les systèmes de diffusion des messages d'alertes doivent atteindre beaucoup de nœuds en peu de temps. Dans les réseaux de communication basés sur les interactions "device to device", on s'est récemment beaucoup intéressé à la diffusion des informations et le besoin d'auto-organisation a été mis en évidence. L'auto-organisation conduit à des comportements locaux et des interactions qui ont un effet sur le réseau global et présentent un avantage de "scalabilité". Ces réseaux auto-organisés peuvent être autonomes et utiliser peu d'espace mémoire. On peut développer des caractères auto-organisés dans les réseaux de communication en utilisant des idées venant de phénomènes naturels. Il semble intéressant de chercher à obtenir les propriétés des "small world" pour améliorer la diffusion des informations dans le réseau. Dans les modèles de "small world" on réalise un recâblage des liens dans le réseau en changeant la taille et la direction des liens existants. Dans un environnement sans-fils autonome une organisation de ce type peut être créée en utilisant le flocking, l'inhibition latérale et le "beamforming".

Dans ce but, l'auteur utilise d'abord l'analogie avec l'inhibition latérale, le flocking et le "beam-

forming" pour montrer comment la diffusion des informations peut être améliorée. L'analogue de l'inhibition latérale est utilisé pour créer des régions virtuelles dans le réseau. Puis en utilisant l'analogie avec les règles du flocking, on caractérise les propriétés des faisceaux permettant aux nœuds de communiquer dans les régions. Nous prouvons que les propriétés des "small world" sont vérifiées en utilisant la mesure des moyennes des longueurs des chemins. Cependant l'algorithme proposé est valable pour les réseaux statiques alors que dans les cas introduisant de la mobilité, les concepts d'inhibition latérale et de flocking nécessiteraient beaucoup plus de temps.

Dans le cas d'un réseau mobile la structure du réseau change fréquemment. Certaines connexions intermittentes impactent fortement la diffusion des informations. L'auteur utilise le concept de stabilité avec le "beamforming" pour montrer comment on peut améliorer la diffusion des informations. Dans son algorithme il prévoit d'abord la stabilité du nœud en utilisant des informations locales et il utilise ce résultat pour identifier les nœuds qui réaliseront du beamforming. Dans l'algorithme, les nœuds de stabilité faible sont autorisés à faire du beamforming vers les nœuds de forte stabilité. La frontière entre forte et faible stabilité est fixée par un seuil. Cet algorithme ne nécessite pas une connaissance globale du réseau, mais utilise des données locales. Les résultats sont validés en étudiant le temps au bout duquel plus de nœuds reoivent l'information et en comparant avec d'autres algorithmes de la littérature.

Cependant, dans les réseaux réels, les changements de structure ne sont pas dus qu'à la mobilité, mais également à des changements de la densité des nœuds à un moment donné. Pour tenir compte de l'influence de tels événements sur la diffusion des informations concernant la sécurité publique, l'auteur utilise les concepts de modèle de métapopulation, épidémiologiques, "beamforming" et mobilité géographique obtenu à partir de données D4D. L'auteur propose la création de trois états latents qu'il ajoute au modèle épidémiologique connu: SIR. L'auteur étudie les états transitoires en analysant l'évolution du nombre de postes ayant reçu les informations et compare les résultats concernant ce nombre dans les différents cas. L'auteur démontre ainsi que le scenario qu'il propose permet d'améliorer le processus de diffusion des informations. Il montre aussi les effets de différents paramètres comme le nombre de sources, le nombre de paquets, les paramètres de mobilité et ceux qui caractérisent les antennes sur la diffusion des informations.

C.1 Introduction

Compte tenu des progrès technologiques récents et de la croissance des besoins humains, le déploiement des réseaux sans-fils a attiré récemment beaucoup d'attention. Beaucoup de propriétés qui concernent les performances des réseaux sans-fil ont été identifiées. [3] Puis, beaucoup d'études se sont concentrées sur les moyens de résoudre les problèmes de scalabilité, les problèmes de supervision et ceux qui sont introduits par les développements technologiques. D'immenses périls topologiques dans le déploiement des réseaux sans fils ont rendu difficiles pour les hommes de contrôler les besoins des réseaux sans-fils en termes d'énergie et de développement technologiques. [4]. Il faut prendre en compte les limites dues aux problèmes cités précédemment mais aussi de nombreux autres facteurs comme les contraintes réseaux, les interactions avec l'environnement, l'utilisation des ressources, le développement de différents types d'architectures hétérogènes. [4]. L'évolution des réseaux sans-fil *ad hoc* rend difficile pour les algorithmes existants de répondre au sujet de beaucoup de détails qui leur sont associés comme l'accroissement de la concurrence et des collision qui conduisent à rechercher des algorithmes qui donnent de bons résultats dans des auto * environnements[1]. La croissance des réseaux a amené les chercheurs à considérer des systèmes complexes, distribués et adaptatifs qui puissent répondre à la plupart de ces questions et croitre avec le réseau.

Pour qu'un système soit auto-organisé il faut qu'il réponde à quatre paradigmes synthétisés dans [5]. Il faut que les entités du système fonctionnent de manière autonome en utilisant des informations locales, soient très synchrones, aient une mémoire de faible durée de vie (historique court) et s'adaptent aux changements de l'environnement. Il est difficile d'avoir de bonnes performances globales en utilisant seulement des informations locales dans un environnement distribué. Si on considère la Nature, elle résout très efficacement la plupart des problèmes complexes en utilisant des règles très simples et en suivant les quatre paradigmes de [5]. L'application des phénomènes naturels en informatique remonte à l'époque de Sir Allan Turing, où il a modélisé les dessins sur la peau d'un animal en utilisant des règles simples. Sa recherche a largement ouvert un champ pour que les chercheurs étudient l'utilisation de phénomènes naturels dans leurs domaines de recherche respectifs.

[1] auto* se rapporte à auto-organisation, auto-similarité, auto-configuration, auto-gestion, auto-adaptation, auto-diagnostic, auto-réparation, auto-protection, auto-réparation

C.1.1 Les Défis

Dans les réseaux sans-fils, même si on utilise des informations locales dans un environnement autonome, le plus important est toujours de transmettre efficacement les informations d'un nœud à l'autre, en dépit des contraintes citées précédemment. Dans un environnement réel, dans les réseaux sans-fil, la transmission des informations est affectée par beaucoup de facteurs comme le déploiement non uniforme, la topologie dynamique, le manque d'infrastructure centralisée et les ressources limitées. De plus, dans différents types de réseaux mobiles, plusieurs propriétés associées affectent la diffusion des informations. Par exemple dans les réseaux mobiles sans fil, et en particulier dans les réseaux PSN (Pocket Switched Networks) et les réseaux de communication $D2D$ (Device to Device) les caractéristiques du comportement humain et les schémas de la mobilité humaine affectent la diffusion des informations.

Considérons un message d'alerte publique qui a à être diffusé dans la population. Dans les réseaux statiques (un réseau qui ne change pas pendant une durée fixée), le déploiement non uniforme de nœuds conduit à un réseau comportant plusieurs sous-réseaux déconnectés. La diffusion des informations entre deux sous réseaux ne sera pas possible avant qu'une connexion soit créée entre chacun des couples de sous-réseaux. Cela limite la diffusion des informations à une fraction de la population. Dans le cas d'un message d'alerte publique, il n'est pas suffisant qu'une fraction de la population reçoive le message. La totalité de la population doit pouvoir le recevoir. Considérons un réseau statique, où les nœuds ne bougent pas et n'ont aucune des caractéristiques des nœuds mobiles. Maintenant supposons qu'une perturbation est apparue dans le réseau et a causé quelques changements dans la topologie du réseau. Par exemple les facteurs qui peuvent causer une telle perturbation seraient les suivants: un nœud pont dans l'un des sous-réseaux peut avoir été enlevé du sous-réseau, à cause des limitations concernant l'énergie, ce qui amène le sous-réseau à être coupé en plusieurs sous-réseaux plus petits. Autres exemples les catastrophes naturelles comme les glissements de terrains, tremblements de terre. Dans ces cas un réseau fortement connecté peut être cassé en plusieurs sous-réseaux et on peut arriver à un réseau déconnecté. L'auteur appelle néanmoins le nouveau réseau un réseau statique car à moins d'un événement imprévu comme ceux qui ont été cités, la structure du réseau ne va pas changer pendant une période assez grande. Dans les catastrophes une diffusion rapide des messages d'alerte joue un rôle important. Plus les secours

arrivent vite, mieux c'est. Cependant à cause des changements dans la topologie du réseau, il se peut que le réseau soit coupé en plusieurs sous-réseaux déconnectés, et la diffusion des informations est fortement influencée.

Quand les nœuds d'un réseau sans-fils sont mobiles (Mobile *ad hoc* Network, Pocket Switched Networks, Delay Tolerant Networks, Device to Device communication network), la topologie du réseau change souvent et ces changements causent des changements dans le processus de diffusion. Les modèles de mobilité, principalement reliés aux modèles de mobilité pour les usagers qui transportent les appareils ou qui les supervisent peuvent amener le réseau à être déconnecté et entraver la diffusion des informations. Cependant certaines études révèlent que la diffusion des informations est améliorée par la mobilité. Dans les modèles de métapopulation, les modèles de communautés basés sur les interactions, il a été prouvé que la mobilité améliore le processus de diffusion des informations. Cependant dans un scenario où les messages d'alertes doivent être diffusés rapidement à la totalité du réseau, une tâche difficile est de concevoir un algorithme efficace et juste qui permette d'atteindre cet objectif le plus tôt possible Qu'il soit mobile ou statique, le réseau est très dynamique, compte tenu des besoins croissants et des changements de technologie. Dans les scenarii réels, il y a une population dynamique qui peut causer à n'importe quel moment des changements importants dans le réseau ; ces changements ont naturellement une influence forte sur la diffusion des informations. Quand le réseau est mobile, la diffusion des informations est aussi influencée par les caractéristiques des régions comme la densité et la distribution de la population dans la région. Les difficultés à résoudre ont amené l'auteur à étudier comment le processus de diffusion des informations peut être amélioré dans le réseau où les nœuds n'ont pas une connaissance globale du réseau et sont réduits à une connaissance à un saut. Si le réseau est statique et déconnecté, la question posée est: comment les propriétés d'un nœud peuvent-elles être utilisées pour diffuser les informations d'un sous-réseau à un autre? Si le réseau est mobile la question est: comment le processus de diffusion des informations peut-il être amélioré?

C.1.2 Motivation et Contributions

Intéressons-nous à un réseau sans-fils, dans un environnement qui change en permanence, il est important pour les chercheurs de concevoir et d'analyser des algorithmes utilisant des informations

locales qui peuvent minimiser les “overheads” et garantir les performances pour le fonctionnement du réseau sans-fils. Dans un réseau sans-fils où les nœuds ont une énergie limitée et une taille mémoire limitée, l’utilisation d’informations locales diminue les dépenses d’énergie et de mémoire. Les interactions des nœuds avec leur environnement jouent aussi un rôle très important pour définir les propriétés des nœuds pour obtenir des transferts de données efficaces et une bonne diffusion des informations. Les propriétés des “small world” [2] sont connues: on peut réduire la longueur moyenne des chemins en recâblant les liens, tout en maintenant la structure du réseau. Il y a une relation directe entre la vitesse de diffusion des informations dans le réseau et la longueur moyenne des chemins car le diamètre du réseau est réduit. Dans les réseaux sans-fil, cependant le recâblage peut être réalisé par la technologie $MIMO$ en adaptant le rayon de transmission au rayon de réception. Ceci peut être réalisé en utilisant du “beamforming”. Le “Beamforming” est une technique qui consiste à adapter les antennes d’un nœud pour que la zone de couverture contienne un rayon plus long dans une direction choisie. L’adaptation de la zone de transmission et de réception dans une direction donnée est un des points clés de ce travail. Il, s’agit de concevoir comment les nœuds vont auto-organiser leurs rayons de couverture pour faciliter la diffusion des informations. Dans l’ensemble l’auteur a cherché à déterminer quelles améliorations des performances peuvent être obtenues dans les réseaux sans-fils, comment optimiser la vitesse de diffusion des informations dans le réseau, en minimisant l’ “overhead”.

Dans ce but, dans la section C.2 l’auteur suggère comment les informations locales et les modèles provenant de la nature comme la migration en groupe et l’inhibition latérale peuvent être utilisés pour accroître la vitesse de la diffusion des informations tout en maintenant un faible “overhead” dans un réseau statique. Dans la section C.2, l’auteur s’intéresse surtout au déploiement non uniforme de nœuds statiques autonomes dans un réseau comprenant plusieurs sous-réseaux. L’autonomie amène l’auteur à proposer une approche centralisée où les nœuds communiquent directement comme dans une communication $D2D$. L’auteur utilise la création de faisceau, “beamforming”, afin de permettre des communications entre les sous-réseaux déconnectés. Comme le suggèrent les propriétés des “small world” un petit nombre de liens recâblés peuvent changer de façon radicale la longueur moyenne des chemins tout en conservant intacte la structure du réseau. Cependant le “beamforming” peut générer des dissymétries et réduire la taille des voisinages. Si tous les nœuds créaient des faisceaux,

cela pourrait générer des dissymétries et le réseau pourrait perdre sa structure. C'est pourquoi l'auteur veut limiter le nombre des nœuds qui créent des faisceaux. Le choix des nœuds qui créent un faisceau et les caractéristiques du faisceau sont les points essentiels de la section C.2. Ce choix, ainsi que le choix des nœuds vers lesquels le nœud crée un faisceau est réalisé en utilisant l'analogie avec les migrations en groupes et l'inhibition latérale. Comme dans la nature où on observe trois règles pour les migrations en groupe: alignement, cohésion et séparation, l'auteur crée des règles: il relie l'alignement à la décision des nœuds de créer ou non un faisceau, la cohésion au choix de la direction du faisceau vers le nœud le plus central, et la séparation au fait que deux nœuds ne doivent pas créer des faisceaux dans la même direction. Pour déterminer le nœud le plus central, l'auteur utilise l'inhibition latérale. C'est aussi cette inhibition qui permet de créer des régions logiques telles que seuls les nœuds dans la périphérie de la région peuvent créer un faisceau en direction des nœuds de la région. Cette règle limite le nombre de nœuds qui créent un faisceau. Les résultats montrent que l'algorithme améliore les performances du réseau en diminuant la longueur moyenne des chemins et le coefficient de clusterisation mais aussi améliore la connectivité.

Cependant dans les réseaux sans-fils mobiles, l'algorithme cité précédemment créerait un "overhead" important, à cause des changement fréquents de la topologie du réseau. Fasciné par les propriétés des PSN (pocket switch networks), l'auteur dans la section C.3 propose d'utiliser le concept de stabilité et le "beamforming" pour accélérer la diffusion des informations. Dans les PSns les propriétés du réseau dépendent beaucoup des modèles de mobilité, à chaque instant le réseau évolue et des communications $D2D$ existent. Récemment, il a été montré que la répartition des sauts de la mobilité de l'homme pouvait être estimée en utilisant une loi puissance tronquée [6] ce qui pourrait conduire à une répartition non uniforme de la population humaine dans la région. Ceci pourrait amener le réseau à être déconnecté et entraver la diffusion des informations. Par ailleurs la mobilité peut aussi améliorer la diffusion en créant de nouvelles connexions dans le réseau [7]. En utilisant les propriétés de la mobilité, l'auteur propose dans la section C.3 de calculer la stabilité d'un nœud en utilisant des informations locales et le modèle de mobilité et d'utiliser la valeur obtenue pour améliorer la diffusion des informations. Dans la section C.3 l'auteur définit la stabilité comme le nombre de connexions qui sont inchangées. La valeur de la stabilité est alors utilisée pour identifier les nœuds qui créent un faisceau et les nœuds vers lesquels ils le font.

Dans les cas réels, cependant, quelques chercheurs ont soutenu que la nature de la loi puissance tronquée est une estimation provenant de l'échantillonnage des données [8] et dépend de la région dans laquelle les échantillons sont prélevés [9]. Dans [6] les données sur la mobilité proviennent des Etats Unis alors que dans un travail récent [9] les échantillons provenaient du Portugal. [9] montra que la répartition de la mobilité humaine peut être modélisée par une loi lognormale. Les différences entre les modèles de mobilité ont conduit l'auteur à acquérir des données $D4D$ et à participer au challenge [10] organisé par Orange Labs et le MIT. Les données concernaient les traces des appels et les déplacements des usagers du réseau Orange en Côte d'Ivoire. L'analyse des données a montré que la distribution des déplacements humains de la Côte d'Ivoire est une loi Gamma.

Dans de grandes populations distribuées géographiquement, appelées souvent métapopulation, où les interactions sont basées sur les mouvements et sont des interactions de proximité, la diffusion des informations dans les métapopulations utilisant des modèles de mobilité humaine, ont été bien étudiées [11, 12] en utilisant des modèles épidémiologiques. Cependant les cas de réseaux sans-fils mobiles réels $D2D$ où la densité des nœuds du réseau change constamment ont été négligés jusqu'à présent. Les deux sections C.2 et C.3 de cette thèse ont utilisé des algorithmes de diffusion des informations traditionnels et considéré une densité des nœuds fixe au cours du temps. Dans la section C.4 l'auteur montre l'influence d'une densité dynamique sur la diffusion des informations et celle du "beamforming" et la mobilité peuvent être utilisés pour améliorer la diffusion dans un scenario de communication $D2D$. Dans la section C.4 l'auteur propose une mise à jour du modèle épidémiologique classique SIR, ajoutant 3 états, E_S, E_I et E_R, afin d'avoir une densité dynamique des nœuds. Chaque état du modèle SIR est ainsi associé à un état supplémentaire. L'auteur appelle ces états supplémentaires des états latents. Les nœuds en état latent ne participent pas à la diffusion des informations. Cependant si un nœud devient actif, il participe à nouveau à la diffusion. Les transitions entre les trois états actifs (S, I, R) et les états latents et vice versa, se produisent avec un taux donné.

C.1.3 Organisation

Le manuscrit est ensuite organisé de la façon suivante: il donne une vue d'ensemble des algorithmes développés par l'auteur dans la section C.2, la section C.3 et la section C.4 respectivement. Le

manuscrit conclut dans la section C.5 et présente des directions futures dans la section C.6.

C.2 Obtention des Propriétés des "Small World" dans le Réseau Sans-fils grâce à des Techniques Bio-inspirées

Dans cette section, l'auteur examine comment on peut obtenir une augmentation de la connectivité, une réduction de la longueur moyenne des chemins APL et une quasi-conservation du coefficient de clusterisation CC dans un réseau sans fil non uniformément distribué, en utilisant des informations locales uniquement et du "beamforming". L'auteur crée des grands faisceaux entre 2 nœuds qui ont une faible et une forte importance. La nécessité de calculs décentralisés et d'auto-organisation amènent l'auteur à s'inspirer des phénomènes naturels et à utiliser: l'inhibition latérale [73,75,76,74] et la migration en groupe [77]. Il utilise aussi le concept de centralité afin de construire la solution.

L'auteur utilise l'inhibition latérale pour créer de petites régions logiques dans un réseau statique. L'inhibition latérale est une technique où les nœuds empêchent les nœuds voisins de réaliser une tâche. Dans l'algorithme, l'inhibition latérale amène les nœuds non inhibés à inhiber les nœuds qui sont à moins de g hops d'eux. Les nœuds inhibés forment une région. L'utilisation de l'inhibition latérale ne réduit pas seulement la complexité du message, elle permet d'utiliser avec succès la migration par groupe, en effet elle permet d'identifier les nœuds qui vont créer des faisceaux et de définir les propriétés des faisceaux. Selon les règles de la migration: alignement, séparation et cohésion, il est important d'identifier les nœuds isolés, d'aligner les nœuds, de les séparer et de les déplacer en direction du centre de leur voisinage. De façon analogue, après avoir créé des régions en utilisant l'inhibition latérale dans un réseau sans-fil *ad hoc* non uniformément distribué, l'auteur utilise les règles de la migration en groupe pour identifier les nœuds qui créent un faisceau et diriger ce faisceau vers le centre de la région identifiée. Le nœud central est un nœud qui a une grande importance pour la structure du réseau. Les nœuds qui créent des faisceaux sont les nœuds les plus éloignés du nœud central de la région. Ce processus est analogue à l'indentification des groupes ou à la règle d'alignement. Créer un faisceau vers le nœud central d'une région contribue à diminuer la longueur moyenne des chemins car le nœud central est le nœud le plus connecté, a la plus grande mesure de centralité de voisinage et est le plus éloigné du nœud qui crée un faisceau.

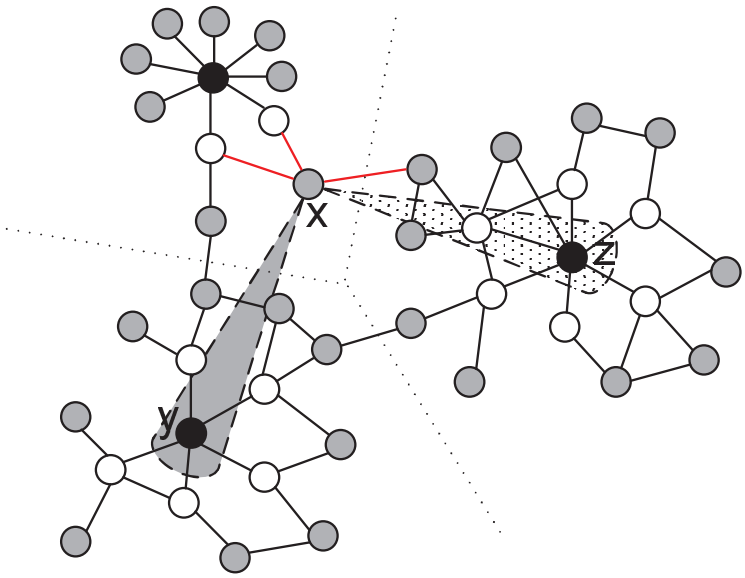


Figure C.1: Un sous réseau avec trois régions et $g = 4$. Ici le nœud x peut créer un faisceau vers y ou z, mais le *nombre de hops* vers y est plus que celui vers z, le nœud x crée un faisceau vers y. Les nœuds centraux sont noirs, les nœuds créant des faisceaux sont gris, les segments noirs représentent les connexions tandis que les segments rouges représentent les connections asymétriques.

Créer un faisceau en direction du nœud central est analogue à la règle de cohésion dans la migration par groupe. Ainsi créer un faisceau (beamforming) en direction du nœud central est analogue au comportement de *preferential attachment* pour un nœud qui crée un faisceau. L'auteur applique aussi une analogie avec la règle de séparation en imposant que deux nœuds qui créent un faisceau ne le dirigent pas vers le même centre. Il est donc nécessaire que l'algorithme crée d'abord les régions avant d'appliquer les règles de la migration. La Fig. C.1 montre trois régions et un nœud x qui crée un faisceau vers le centre le plus éloigné y.

Ainsi l'auteur a conçu un algorithme tel qu'il identifie d'abord les régions, en utilisant l'inhibition latérale, puis il identifie les nœuds centraux des régions et ensuite utilise les règles de la migration pour choisir les nœuds qui vont créer des faisceaux et les propriétés de ces faisceaux. Les résultats montrent qu'une réduction de la longueur moyenne des chemins APL peut être obtenue et la connectivité peut être améliorée en utilisant l'algorithme proposé.

C.3 Amélioration de la Diffusion des Informations dans les Réseaux Sans-fil Dynamiques en utilisant l'Evaluation de la Stabilité et le "Beamforming"

Dans [112, 113, 114] on voit que le temps moyen pour diffuser une informations peut être raccourci quand le "beamforming" est associé à la mobilité. Dans ces études les auteurs montrent aussi que seule une petite fraction des machines doivent être utilisées pour créer des faisceaux afin d'améliorer la diffusion des informations dans le réseau mobile. La qualité de la diffusion des informations dépend de la largeur et de la longueur des faisceaux. De meilleures performances sont atteintes sont obtenues quand on utilise des faisceaux plus longs et plus étroits qui utilisent la même puissance que celle d'un faisceau pluridimensionnel et des gains sont obtenus quand l'antenne tourne lorsque les nœuds se déplacent. Cependant ces études considéraient seulement une seule source envoyant un seul paquet d'informations. Un autre travail concernant le "beamforming" dans les communications mobiles est du à Li *et al.* [136]. L'objectif de [136] était d'étudier les effets du "beamforming" dans un réseau vehiculaire pour les hommes. Li *et al.* ont montré que le "beamforming" peut améliorer la diffusion des informations dans un réseau mobile en faisant des observations similaires à celles de [112, 113, 114]. Toutefois la seule différence entre [112, 113, 114] et [136] concernait la sélection des machines qui "beamform". Au contraire de [112, 113, 114], où la direction du faisceau et le choix des nœuds qui créent un faisceau était aléatoire, l'auteur pense que si le choix de ces nœuds est basé sur une mesure et si les faisceaux sont dirigés dans une direction où il y a plus de nœuds qui n'ont pas reçu les informations, le processus de diffusion peut être amélioré.

Dans ce contexte, la première question à laquelle il faut répondre est: "quel nœud doit créer un faisceau?" Après avoir fait ce choix, il faut choisir la direction du faisceau. Dans la section C.2, l'auteur a proposé l'utilisation de la mesure de la centralité avec d'autres concepts pour résoudre le problème dans le cas statique. Les concepts utilisés dans la section C.2 ne peuvent pas être appliqués au cas dynamique parce que le calcul de la mesure de la centralité deviendrait complexe, dépenserait beaucoup d'énergie et nécessiterait une informations sur l'historique [63, 54]. Sauvegarder une énorme quantité d'informations historiques dans un nœud dont la capacité mémoire est limitée poserait des

problèmes informatiques. Dans ce but, l'auteur propose de calculer une mesure de stabilité des nœuds à l'instant $t + \Delta t$ en utilisant leur degré (mesure locale) à l'instant t pour améliorer la diffusion des informations. Dans [156] il a été remarqué que la stabilité d'un nœud dans un graphe est étroitement reliée à la mobilité du nœud.

Si on considère le degré, si un nœud a une stabilité de $p\%$ cela signifie que le nœud a conservé $p\%$ de son voisinage initial. Si $p \to 1$ cela signifie que le voisinage du nœud n'a pas changé et qu'il ne communique qu'avec ses voisins. Cependant si $p \to 0$, cela signifie d'abord que le nœud a un voisinage complètement différent de son voisinage initial et d'autre part que le nœud envoyait des informations à ses voisins à l'instant t et sera capable de recevoir des informations en provenance de nouveaux voisins $t + \Delta t$. Par conséquent, il recevra les informations d'un plus grand ensemble de nœuds (voisinage à t et à $t + \Delta t$) et de plus le nœud était déconnecté à l'instant t.

Une méthode pour calculer la stabilité est proposée dans [102] où les auteurs utilisaient le degré du nœud observé au cours de successifs instantanés. Dans un scenario mobile, où le "beamforming" est appliqué aussi, le calcul de la stabilité proposé par [102] nécessiterait de connaître le nouveau voisinage à l'instant $t + \Delta t$. Pour déterminer le nouveau voisinage il faudrait immobiliser l'évolution jusqu'à ce que le nouveau voisinage soit connu. Ceci prendrait du temps de l'énergie et nécessiterait beaucoup d'espace mémoire. Ainsi, il vaut mieux estimer la stabilité d'un nœud avant de connaître son voisinage à $t + \Delta t$ en utilisant seulement son voisinage à t et les caractéristiques de sa mobilité. L'auteur a conçu un algorithme qui prédit la stabilité des nœuds. Ensuite les nœuds utilisent cette mesure pour décider s'ils doivent créer un faisceau et si oui, dans quelle direction. La Fig. C.2 montre comment la probabilité de connexion peut être calculée en évaluant la taille de l'arc qui est en gras. En utilisant cette probabilité, la mesure de la stabilité est calculée en faisant la moyenne des probabilités de conservation pour tous les liens du voisinage.

Afin de déterminer quel nœud doit créer un faisceau, l'auteur utilise la mesure de la stabilité et propose que les nœuds qui ont une stabilité faible créent un faisceau. Comme expliqué précédemment l'auteur propose qu'un nœud de faible stabilité crée un faisceau vers un nœud de forte stabilité. Ainsi, dans cette section, l'auteur utilise la discussion mentionnée pour construire l'algorithme qui permet d'étudier si la diffusion des informations peut être améliorée dans un réseau dynamique où les nœuds n'ont pas de connaissance globale du réseau. Les résultats suggèrent qu'on peut améliorer

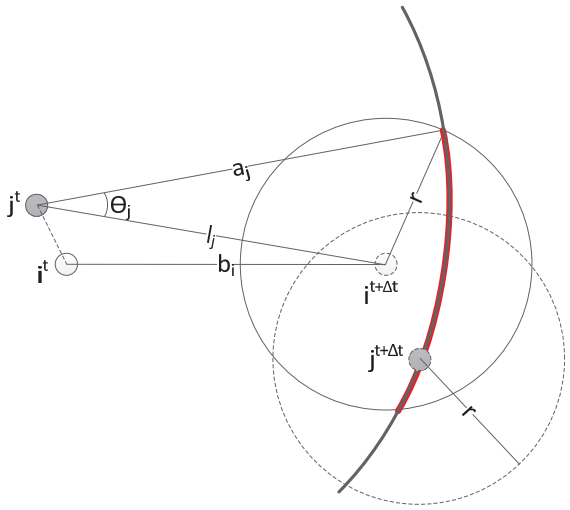


Figure C.2: La probabilité de connexion entre les nœuds i et $j \in N_i^t$ après un intervalle de temps Δt est donnée par la taille de l'arc en gras.

la diffusion.

C.4 Amélioration de la Diffusion des Informations dans une Métapopulation de "Device to Device" en utilisant des traces de la Mobilité Humaine et le "Beamforming"

Les communications $D2D$ sont définies comme des communications entre des nœuds physiquement proches sans aucune influence de l'infrastructure du réseau. Le $D2D$ présente beaucoup d'avantages comme le fait que les communications soient autonomes, aient des performances améliorées et puissent réutiliser les fréquences, dépensent peu d'énergie et réduisent la charge sur l'infrastructure. Comme nous nous intéressons à la diffusion des informations dans une région de grande échelle, l'approximation du champ moyen est le modèle qui convient le mieux pour modéliser la diffusion dans le réseau des contacts à travers le pays. Le processus de diffusion des informations a été considéré comme très proche d'un processus de diffusion d'une épidémie à travers une population pour lequel le modèle SIR existe. On peut trouver une synthèse sur les modèles épidémiologiques dans [139, 140].

Dans un modèle obtenu par une méthode de champ moyen, les états, S, I et R correspondent respectivement aux nœuds n'ayant pas de paquet, aux nœuds ayant un paquet et étant prêt à le transmettre, et aux nœuds qui ne participent plus à la diffusion des informations afin d'économiser l'énergie. Dans ce contexte il a été montré que la diffusion des informations est influencée par beaucoup de facteurs tels que: les communautés dans le réseau, les rafales [135], la force de la connexion [105], la source de l'infection, le nombre de nœuds infectés, les paramètres de la mobilité humaine (section C.3), les lieux les plus visités, [141], la structure du réseau [106], le type d'activité [107], les caractéristiques des machines [109] et dans la section C.3, l'altruisme [110], et les mesures sociales [142]. Cependant, comme l'a fait remarquer [7], les caractéristiques de la mobilité humaine sont les paramètres qui ont le plus d'importance, si on les compare aux autres.

Pour un modèle à l'échelle d'un pays, l'auteur a utilisé les données fournies par les organisateurs d'une étude de $D4D$ (Data for Development) et par Orange Labs [10]. L'analyse des données révèle que dans une grande population, il existe des communautés et que la densité de la population évolue en fonction du temps. Une telle structuration d'une grande population en communautés où les interactions entre les communautés sont dues à la mobilité, a déjà été étudiée en créant un modèle synthétique de population appelé métapopulation [12, 11] où il a été dit que l'existence des communautés dans une grande population est une évidence et qu'elles influent sur le processus de diffusion. Si une personne était réduite à rester dans une communauté, elle ne pourrait pas se déplacer librement. Par exemple, le mouvement d'une personne d'un pays à un autre est restreint par les douaniers. Ceci entrave la diffusion des épidémies. La Figure C.3 montre la différence entre le nombre de personnes infectées quand il y a c communautés et quand il n'y en a qu'une. Au début, la grande différence entre les courbes est due au fait que plus de personnes en ont rencontré d'autres quand elles ne sont pas contraintes à rester dans une communauté donnée et ainsi la diffusion est rapide. Récemment, il a été montré dans [143] que non seulement la structure de la communauté influe sur le processus de diffusion mais que la densité des communautés a aussi une grande influence.

Dans les réseaux de communication, cependant, on peut rencontrer des densités dynamiques dans deux cas: premier cas: processus de naissance et de mort, deuxième cas: processus de sommeil et de veille (cas qui a éte peu cité dans la littérature). Ainsi dans un réseau $D2D$ où la structure du réseau dépend aussi de la mobilité humaine et des caractéristiques des machines, l'auteur cherche

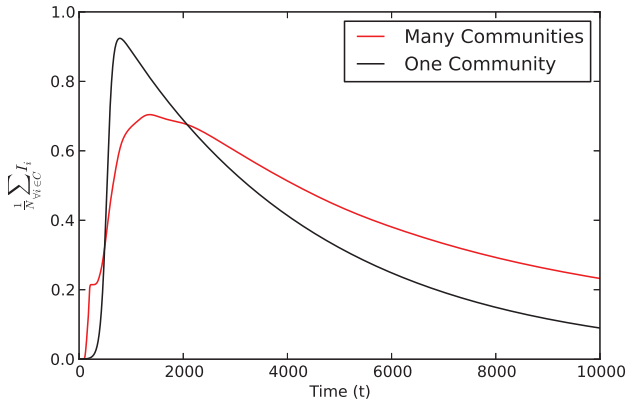


Figure C.3: L'évolution du nombre de machines ayant l'informations quand la région est divisée en beaucoup de petites communautés et quand la région n'a qu'une communauté. Les résultats ont été obtenus en utilisant des données du $D4D$ et le modèle épidémiologique de type SIR.

à savoir comment le processus de diffusion des informations à l'échelle de tout le pays est impacté et comment il peut être amélioré dans un cas de population à densité variable (sans considérer un processus de naissance et de mort) en comparant ces résultats au cas où la taille de la population était constante C.3 Pour faire varier la densité et analyser ses conséquences, l'auteur utilise un état Latent (E). Pour distinguer les états Latents mais Susceptible d'être infectés, les états Latents mais Infectés et les états Latents mais Recovered (guéris) l'auteur sépare E en trois états: E_S, E_I and E_R. Un nœud dans l'état E_S, dans l'état E_I et dans l'état E_R ne participe pas à la diffusion des informations. Seuls les nœuds dans les états S,IetR y participent.

Les données fournies par les organisateurs du $D4D$ contiennent des informations au sujet de la mobilité des utilisateurs et des relations entre lui et une communauté, (appelée sous-préfecture) dans une région de la Côte d'Ivoire. Afin d'étudier le modèle de la population dans tout le pays, l'auteur utilise les résultats du recensement [144]. En Côte d'Ivoire, un groupe de sous-préfectures forme une communauté plus grande: celle du département. Un ensemble de départements forme une région, et un ensemble de régions forme un pays. Comme les données concernent la structure des mobilités au niveau de la sous-préfecture, dans cette section, l'auteur est surtout intéressé par les sous-préfectures. On peut trouver des détails sur les taux de transition et la façon dont ils sont évalués dans la section 5.2.2. Dans les réseaux de communications comme les PSN, les réseaux tolérant aux délais (DTN)

et les réseaux de communication $D2D$, les machines peuvent transmettre des informations à des machines proches qui sont dans leur zone de couverture. Dans cette section, néanmoins, l'auteur considère surtout l'architecture du réseau $D2D$ avec de la mobilité, en effet le marché des applications reliées au contexte et des services reliés à la localisation, s'est beaucoup développé et les opérateurs commencent à considérer le déploiement des communications $D2D$ comme la couche sous-jacente des réseaux cellulaires.

En se basant sur le contexte et les informations sur le voisinage, un nœud peut découvrir d'autres nœuds dans son voisinage et établir des liens avec eux [145]. Un nœud peut parfois être considéré comme un relais multi-saut dans un réseau de communication $D2D$. Ceci peut augmenter la zone de couverture parce que les informations pourraient être transmises à des nœuds qui sont au-delà de la zone de couverture classique et même dans une zone où il y a peu de cellules. En se basant sur le contexte, les nœuds peuvent aussi former des groupes. Un nœud peut être le chef de groupe ou un membre du groupe. En utilisant les mécanismes de "collaborative forwarding" les nœuds peuvent partager efficacement les données pour optimiser l'utilisation des ressources et obtenir un meilleur débit [146].

Un nœud dans un réseau peut être équipé d'un vecteur d'antennes comportant beaucoup d'éléments répartis géométriquement. Il est ainsi possible que le nœud ait un rayon de transmission plus long dans une certaine direction et la largeur sera adaptable à cause des interférences entre les rayons provenant des éléments d'antennes. Cette technique est le "beamforming" qui a été étudiée et pour laquelle l'auteur a obtenu des résultats positifs, pour les réseaux statiques et les réseaux mobiles [112,136], section C.2 et section C.3. Ces études ont montré qu'on pouvait améliorer beaucoup la diffusion en appliquant le "beamforming" à un très petit nombre de nœuds. L'auteur utilise ce résultat et le concept de "beamforming" sur le modèle épidémiologique à six états et montre comment cette méthode peut être utilisée pour améliorer le processus de diffusion des informations.

Dans une population humaine réelle, un utilisateur est associé à un domicile. Chaque fois que l'utilisateur quitte son domicile, il a une certaine probabilité d'y retourner depuis son nouvel emplacement. L'auteur met ainsi à jour le modèle de mobilité pour tenir compte des retours des utilisateurs vers leur domicile. (Cf. Fig. C.4). Cette probabilité de retour est déduite des données que l'auteur a reçu. En utilisant deux modèles différents de mobilité, l'auteur montre l'effet des

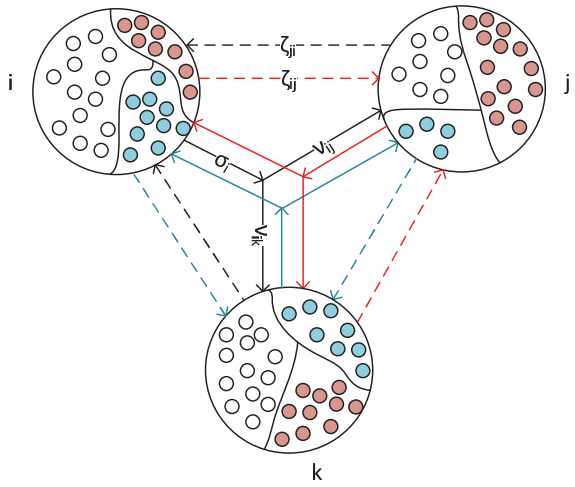


Figure C.4: Taux de retour pour le modèle de mobilité. La ligne pointillée montre le taux de retour depuis une autre communauté et est noté $\zeta_{cc'}$ et la ligne pointillée correspondant aux mouvements dans la communauté est notée σ_c, tandis que la ligne en trait plein correspondant au départ vers une autre communauté est notée $\sigma_c \nu_{cc'}$. Chaque communauté correspond à un nombre de nœuds. Pour les différentes communautés, les nœuds ont des couleurs différentes (blanche pour i, rouge pour j, bleue pour k). La couleur des lignes représente aussi à quelle communauté elles sont associées.

modèles de mobilité sur le processus de diffusion.

C.5 Conclusion

L'évolution des réseaux mobiles et la nécessité de procurer aux nœuds des possibilités d'auto-organisation afin qu'ils puissent mieux interagir avec l'environnement et avoir de meilleures performances ont conduit l'auteur à proposer des algorithmes dans la section C.2, la section C.3 et la section C.4. Les performances sont meilleures car la longueur moyenne des chemins est plus petite, la connectivité est augmentée et les informations peuvent être diffusées plus vite dans tout le réseau de manière décentralisée en n'utilisant que des données locales. La méthode consiste à appliquer aux systèmes complexes des concepts tels que l'auto-organisation, l'inhibition latérale, les migrations en groupes et en réalisant des modèles de mobilité, de stabilité, du "beamforming" et en mesurant les densités dynamiques. Dans la section C.2 l'auteur a essayé d'étudier comment les propriétés des "small world" pouvaient être appliquées efficacement aux réseaux sans-fil statiques en utilisant des concepts classiques de systèmes complexes comme des modèles bio-inspirées comme l'inhibition

latérale et l'analogie avec les migrations en groupe. Pour atteindre son but l'auteur utilise aussi le concept de "beamforming". L'algorithme fonctionne en utilisant des informations locales et les nœuds n'ont pas besoin de connaître tout le réseau. L'inhibition latérale est utilisée pour former des communautés virtuelles dans le réseau afin de réduire la complexité des messages tandis que l'analogie avec la migration en groupe est utilisée pour déterminer les propriétés des faisceaux. Les résultats montrent que la longueur moyenne des chemins APL est diminuée, et aussi que la connectivité est améliorée. Une plus petite APL signifie moins de sauts d'un nœud à un autre et par conséquent le temps de transmission d'un nœud à un autre demande moins de temps.

Un exemple d'interaction de nœuds avec l'environnement est celui qui correspond à des machines mobiles. La mobilité crée de la diversité, des clusters et cause des changements dans l'environnement des interactions. Quand il y a mobilité, à chaque instant les interactions entre les machines changent et la structure du réseau change. Cependant il y a certain pourcentage d'invariance dans les interactions, que l'auteur appelle stabilité. Utilisant la stabilité l'auteur définit un algorithme dans la section C.3. Cet algorithme obtient des résultats analogues à ceux des algorithmes existants déjà, mais d'une façon plus efficace (en dépensant moins d'énergie et en utilisant moins de mémoire). Dans la section C.3 l'auteur utilise une approche probabiliste pour calculer la stabilité d'un nœud en environnement mobile. L'algorithme utilise aussi du "beamforming" dans le but d'accélérer la diffusion des informations dans le réseau. Il fonctionne en n'utilisant que des informations locales. Le "beamforming" est utilisé par les nœuds qui ont une faible stabilité en direction des nœuds qui ont une forte stabilité. Dans cette section l'accélération de la diffusion signifie que plus de nœuds sont capables de recevoir les informations circulant dans le réseau, plus vite. Il est aussi montré dans cette section que l'algorithme proposé donne aussi de meilleurs résultats pour différents cas et les effets sur la diffusion de l'informations des variations des paramètres du réseau sont analysés.

Toutefois l'algorithme de la section C.3, ne considère pas le cas où la densité dans les réseaux mobiles évolue au cours du temps. L'auteur, afin de tenir compte des cas de réseaux à densité dynamique, introduit trois états latents,. Les nœuds dans l'un des états latent ne participent pas à la diffusion dans la section C.4. L'auteur considère un modèle de diffusion de type SIR et relie chaque état dans le modèle SIR à un état latent. Seuls les nœuds qui sont dans les états S, I ou R participent au processus de diffusion. Les transitions entre les états actifs et les états latents

se produisent avec un taux donné. Pour accélérer la diffusion la mobilité est combinée avec le "beamforming" au dessus du processus de diffusion décrit par le modèle enrichi des états SIR. Les résultats montrent que le processus de diffusion des informations dans le cas d'un réseau à densité dynamique peut aussi être amélioré: plus de nœuds reçoivent les informations en moins de temps.

Dans chacune des trois contributions l'auteur propose que les algorithmes limitent le nombre de nœuds qui créent des faisceaux. Dans la section C.2, l' inhibition latérale est utilisée pour limiter le nombre de nœuds qui créent un faisceau alors que dans la section C.3 seuls les nœuds de faible stabilité sont autorisés à le faire. Dans la section C.4 le nombre de nœuds qui créent un faisceau est fixé à une valeur faible. Le fait de limiter ce nombre de nœuds permet au phénomène des "small world" de se produire comme cela a été analysé dans [2].

C.6 Directions Futures

Les algorithmes proposés par l'auteur fournissent une méthode pour améliorer la diffusion des informations dans un environnement autonome. Elle pourrait aider à diffuser efficacement des messages d'alertes dans la population dans n'importe quelle sorte de réseau (statique ou mobile). Un cas d'application de communication $D2D$ concerne la nécessité de diffuser efficacement les messages d'alertes. Utiliser les trois algorithmes proposés avec les environnements et règles du $D2D$ serait évidemment utile. Les algorithmes proposés peuvent aussi être utiles dans différents cas de problèmes de diffusion.

Cependant, des extensions des algorithmes proposés peuvent être prévues. La décision pour un nœud de créer un faisceau est une décision autonome où en utilisant les interactions locales on peut vouloir obtenir des améliorations globales. En théorie des jeux, la décision pour les nœuds de créer un faisceau est un jeu joué par les nœuds. Ici voici quelques questions dont les nœuds aimeraient connaître la réponse: les questions comme celle de savoir si pour le nœud une coopération avec d'autres nœuds ou celle de savoir si un changement des propriétés du faisceau améliorerait les résultats globaux. Dans cette direction, l'application de la théorie des jeux évolutionnaires pourrait être étudiée.

De plus, dans la section C.4, l'auteur n'a pas utilisé un modèle markovien de processus de naissance et de mort avec des additions de nouvelles machines et la suppression de vieilles machines.

Ces évènements peuvent être dus à plusieurs causes: pannes de matériel, changement de technologies, besoins croissants des machines à des prix raisonnables. Ajouter de tels concepts aux modèles le rendrait plus réaliste. Afin de modéliser un tel processus, il est nécessaire de connaître les taux d'addition ou de suppression des machines du réseau. En obtenant de telles informations depuis les traces des appels: CDRs (Call Data Record) on améliorerait le modèle et l'étude de la diffusion de l'informations serait plus réaliste. Aussi pour étudier des cas plus proches de la réalité, l'auteur suggère d'utiliser la probabilité de connexion dans une direction donnée pour le calcul de $\langle k \rangle$ [138]. Enfin comme perspective, un cas de communication $D2D$ qui pourrait être étudié serait la diffusion de messages d'alertes. On pourrait étudier le modèle décrit dans la section C.4.

Les caractéristiques de la mobilité utilisées dans la thèse sont déduites de l'analyse d'échantillons de données de mobilité humaine. Dans la section C.3, les caractéristiques de la mobilité utilisées étaient basées sur les schémas observés dans les données collectées dans une région des USA, alors que dans la section C.4, les caractéristiques de la mobilité utilisées étaient dérivées d'échantillons obtenus en Côte d'Ivoire. Dans ces deux sections, à cause de l'utilisation des CDR, la mobilité peut seulement être approximée. L'utilisation des échantillons de données pour la modélisation de la mobilité devrait être extensive et exacte. Ceci exige un grand nombre d'échantillons obtenus à grande fréquence et nécessite un mécanisme efficace d'enregistrement des données. Si on considère des individus autonomes, les données collectées jusqu'à présent ne réussissent pas à rendre compte des aspects comportementaux attachés à la mobilité humaine. Les chercheurs citent des problèmes de sécurité concernant le fait d'avoir des informations de ce type dans les données. Cependant si des informations de ce type étaient disponibles, les modèles de mobilité pourraient être plus subtils. Utiliser ces modèles améliorerait sûrement les algorithmes dans la C.3 et dans C.4. Le manque de données de ce type et d'enregistrements de CDR amènent l'auteur à proposer une autre direction de recherche future grâce à laquelle les CDRs pourraient être rendus plus complets et gérables.

APPENDIX D

Publications

- Journals
  - **R. Agarwal**, V. Gauthier, M. Becker, T. Toukabri, H. Afifi, "Enhancing Information Dissemination in Device to Device Metapopulation using Human Mobility Trace and Beamforming", *(Submitted)* IEEE JSAC on D2D Communications in Cellular Networks.
  - **R. Agarwal**, V. Gauthier, M. Becker, "Enhancing Information Dissemination in Dynamic Wireless Network using Stability and Beamforming", *(Submitted)* AdHoc Network.
  - **R. Agarwal**, A. Banerjee, V. Gauthier, M. Becker, C. K. Yeo, and B. S. Lee, "Achieving Small World Properties using Bio-Inspired Techniques in Wireless Networks", The Computer Journal, vol. 55, no. 8, pp. 909-931, Aug-2012.
  - A. Banerjee, **R. Agarwal**, V. Gauthier, C. K. Yeo, H. Afifi and B. S. Lee, "A Self-Organization Framework for Wireless Ad Hoc Networks as Small Worlds", IEEE Transaction on Vehicular Technology, vol. 61, no. 6, pp. 2659-2673, Jul-2012.

- Workshop/Conferences
  - **R. Agarwal**, V. Gauthier, M. Becker, "Information Dissemination using Human Mobility in Realistic Environment (E-Inspire)", in Proc. of the D4D special session (NetMob 2013), Boston, 1-3 May 2013.
  - **R. Agarwal**, A. Banerjee, V. Gauthier, M. Becker, C. K. Yeo, and B. S. Lee, "Self organization of Nodes Using Bio-Inspired Techniques for Achieving Small World Properties", in Proc. of the workshop on complex networks and pervasive group communication (IEEE GLOBECOM 2011), pp. 89-94, Houston, 5-9 Dec 2011.
  - A. Banerjee, **R. Agarwal**, V. Gauthier, C. K. Yeo, H. Afifi and B. S. Lee, "Self-Organizing of Wireless Ad Hoc Network as a Small World using long range Directional Beams", in Proc. of the workshop on complex networks and pervasive group communication (IEEE GLOBECOM 2011), pp. 120–124, Houston, 5-9 Dec 2011.

Bibliography

[1] A. Banerjee, R. Agarwal, V. Gauthier, C. K. Yeo, H. Afifi, and B. S. Lee, "A Self-Organization Framework for Wireless Ad Hoc Networks as Small Worlds," *IEEE Transactions on Vehicular Technology*, vol. 61, pp. 2659–2673, July 2012.

[2] D. J. Watts and S. Strogatz, "Collective Dynamics of Small World Networks," *Nature*, vol. 393, pp. 440–442, June 1998.

[3] J. Andrews, R. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A Primer on Spatial Modeling and Analysis in Wireless Networks," *IEEE Communications Magazine*, vol. 48, pp. 156–163, November 2010.

[4] F. Dressler and O. B. Akan, "A survey on Bio-Inspired Networking," *Computer Networks*, vol. 54, pp. 881–900, April 2010.

[5] C. Prehofer and C. Bettstetter, "Self-Organization in Communication Networks: Principles and Design Paradigms," *Communications Magazine, IEEE*, vol. 43, pp. 78–85, July 2005.

[6] M. González, C. Hidalgo, and A. L. Barabási, "Understanding Individual Human Mobility Patterns," *Nature*, vol. 453, pp. 779–782, June 2008.

[7] M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, J. Saramäki, and M. Karsai, "Multiscale Analysis of Spreading in a Large Communication Network," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, pp. P03005(1–32), March 2012.

[8] S. Mossa, M. Barthélémy, E. Stanley, and L. N. Amaral, "Truncation of Power Law Behavior in Scale-Free Network Models due to Information Filtering," *Physical Review Letters*, vol. 88, pp. 138701(1–4), March 2002.

[9] B. C. Csáji, A. Browet, V. Traag, J.-C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, and V. D. Blondel, "Exploring the Mobility of Mobile Phone Users," *Physica A: Statistical Mechanics and its Applications*, vol. 392, pp. 1459–1473, March 2013.

[10] V. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, E. Huens, F. Morlot, Z. Smoreda, and C. Ziemlicki, "Data for Development: The D4D Challenge on Mobile Phone Data." http://arxiv.org/pdf/1210.0137v1.pdf, 2012.

[11] J. Arino and P. Van den Driessche, "Disease Spread in Metapopulation," *Fields Institute Communications*, vol. 48, pp. 1–13, 2006.

[12] D. Watts, R. Muhamad, D. Medina, and P. Dodds, "Multiscale, Resurgent Epidemics in a Hierarchical Metapopulation Model," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 11157–11162, August 2005.

[13] W. Ashby, "Principles of the Self-Organizing Dynamic System," *The Journal of general psychology*, vol. 37, pp. 125(1–8), October 1947.

[14] H. Sayama, "Complex System Organization Map." http://en.wikipedia.org/wiki/File:Complex_systems_organizational_map.jpg, 2010.

[15] A. Lotka, "Contribution to the Theory of Periodic Reaction," *Journal Physical Chemistry*, vol. 14, no. 3, pp. 271–274, 1910.

[16] P. Bak, C. Tang, and K. Wiesenfeld, "Self-Organized Criticality," *Physical Review A*, vol. 38, pp. 364–374, July 1988.

[17] D. L. Turcotte, "Self-Organized Criticality," *Reports on Progress in Physics*, vol. 62, pp. 1377–1429, October 1999.

[18] R. Albert, H. Jeong, and A. L. Barabási, "The Diameter of the World Wide Web," *Nature*, vol. 401, pp. 130–131, 1999.

[19] S. Dorogovtsev and J. Mendes, "Scaling Behaviour of Developing and Decaying Networks," *Europhysics Letters (EPL)*, vol. 52, pp. 33–39, October 2000.

[20] R. Albert, H. Jeong, and A. L. Barabási, "Error and Attack Tolerance of Complex Networks," *Nature*, vol. 406, pp. 378–382, July 2000.

[21] L. N. Amaral, A. Scala, M. Barthélémy, and E. Stanley, "Classes of Small-World Networks.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, pp. 11149–11152, October 2000.

[22] R. Albert and A. L. Barabási, "Topology of Evolving Networks: Local Events and Universality," *Physical Review Letters*, vol. 85, pp. 5234–5237, December 2000.

[23] P. Krapivsky, G. Rodgers, and S. Redner, "Degree Distributions of Growing Networks," *Physical Review Letters*, vol. 86, pp. 5401–5404, June 2001.

[24] D. Pennock, G. Flake, S. Lawrence, E. Glover, and L. Giles, "Winners don't take all: Characterizing the Competition for Links on the Web," *Proceedings of the National Academy of Sciences*, vol. 99, pp. 5207–5211, April 2002.

[25] K. Lerman, R. Ghosh, and J. H. Kang, "Centrality Metric for Dynamic Networks," in *Proceedings of the 8th Workshop on Mining and Learning with Graphs - MLG '10*, (Washington DC), pp. 70–77, ACM New York, NY, 25-28 July 2010.

[26] A. Scherrer, P. Borgnat, E. Fleury, J. L. Guillaume, and C. Robardet, "Description and Simulation of Dynamic Mobility Networks," *Computer Networks*, vol. 52, pp. 2842–2858, October 2008.

[27] F. Kuhn and R. Oshman, "Dynamic networks: Models and Algorithms," *ACM SIGACT News*, vol. 42, pp. 82–96, March 2011.

[28] P. Borgnat, E. Fleury, J. L. Guillaume, C. Magnien, C. Robardet, and A. Scherrer, "Evolving Networks," in *NATO Advanced Study Institute on Mining Massive Data Sets for Security*, (Villa Cagnola, Gazzada, Italy), pp. 198–204, IOS Press, 10-21 September 2007.

[29] P. Grindrod and D. Higham, "Evolving Graphs: Dynamical Models, Inverse Problems and Propagation," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 466, pp. 753–770, November 2009.

[30] P. Grindrod, M. Parsons, D. Higham, and E. Estrada, "Communicability across Evolving Networks," *Physical Review E*, vol. 83, pp. 046120(1–10), April 2011.

[31] P. Grindrod and M. Parsons, "Social Networks: Evolving Graphs with Memory Dependent Edges," *Physica A: Statistical Mechanics and its Applications*, vol. 390, pp. 3970–3981, October 2011.

[32] N. Santoro, W. Quattrociocchi, P. Flocchini, A. Casteigts, and F. Amblard, "Time-Varying Graphs and Social Network Analysis: Temporal Indicators and Metrics," in *Social Networks and Multiagent Systems Symposium*, (York, UK), pp. 32–38, 4-5 April 2011.

[33] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro, "Time-Varying Graphs and Dynamic Networks," *International Journal of Parallel, Emergent and Distributed Systems*, pp. 1–22, April 2012.

[34] P. Holme and J. Saramäki, "Temporal Networks," *Physics Reports*, vol. 519, pp. 97–125, October 2012.

[35] R. K. Pan and J. Saramäki, "Path Lengths, Correlations, and Centrality in Temporal Networks," *Physical Review E*, vol. 84, pp. 016105(1–10), July 2011.

[36] D. Braha and Y. Bar-Yam, "From Centrality to Temporary Fame: Dynamic Centrality in Complex Networks," *Complexity*, vol. 12, pp. 59–63, November 2006.

[37] D. Braha and Y. Bar-Yam, "Time-Dependent Complex Networks: Dynamic Centrality, Dynamic Motifs, and Cycles of Social Interaction," in *Adaptive Networks* (T. Gross and H. Sayama, eds.), Understanding Complex Systems, ch. 3, pp. 39–50, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[38] J. Travers and S. Milgram, "An Experimental Study of the Small World Problem," *Sociometry*, vol. 32, pp. 425–443, Dec 1969.

[39] M. Newman and D. Watts, "Scaling and Percolation in the Small-World Network Model," *Physical Review E*, vol. 60, pp. 7332–7342, December 1999.

[40] M. Newman, "Models of the Small World," *Journal of Statistical Physics*, vol. 101, pp. 819–841, 2000. 10.1023/A:1026485807148.

[41] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[42] A. L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, pp. 509–512, Oct 1999.

[43] R. Monasson, "Diffusion, Localization and Dispersion Relations on Small-World Lattices," *The European Physical Journal B*, vol. 12, pp. 555–567, December 1999.

[44] R. Kulkarni, E. Almaas, and D. Stroud, "Evolutionary Dynamics in the Bak-Sneppen Model on Small-World Networks," 1999.

[45] A. Barrat and M. Weigt, "On the Properties of Small-World Network Models," *The European Physical Journal B*, vol. 13, pp. 547–560, January 2000.

[46] J. Kleinberg, "Navigation in a Small World," *Nature*, vol. 406, p. 845, August 2000.

[47] J. Kleinberg, "The Small World Phenomenon: An Algorithm Perspective," in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, (Portland, OR), pp. 163–170, ACM, New York, 21-23 May 2000.

[48] A. Helmy, "Small Worlds in Wireless Networks," *IEEE Communications Letters*, vol. 7, no. 10, pp. 490–492, 2003.

[49] A. L. Barabási and E. Bonabeau, "Scale-Free Networks," *Scientific American*, vol. 288, pp. 60–69, May 2003.

[50] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[51] L. C. Freeman, "Centrality in Social Networks Conceptual Clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.

[52] M. Everett and S. Borgatti, "Ego Network Betweenness," *Social Networks*, vol. 27, pp. 31–38, Jan 2005.

[53] E. Daly and M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs," in *Proceedings of 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, (Montreal, Canada), pp. 32–40, ACM, New York, 9-14 September 2007.

[54] H. Kim, J. Tang, R. Anderson, and C. Mascolo, "Centrality Prediction in Dynamic Human Contact Networks," *Computer Networks*, vol. 56, pp. 983–996, February 2012.

[55] L. C. Freeman, S. P. Borgatti, and D. R. White, "Centrality in Valued Graphs: A Measure of Betweenness Based on Network Flow," *Social Networks*, vol. 13, pp. 141–154, June 1991.

[56] S. Dolev, Y. Elovici, and R. Puzis, "Routing Betweenness Centrality," *Journal of the ACM*, vol. 57, pp. 1–27, April 2010.

[57] A. Banerjee, R. Agarwal, V. Gauthier, C. K. Yeo, H. Afifi, and B. S. Lee, "Self-Organization of Wireless Ad Hoc Networks as Small Worlds using Long Range Directional Beams," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, (Houston), pp. 120–124, IEEE, New York, 5-9 December 2011.

[58] P. Marsden, "Egocentric and Sociocentric Measures of Network Centrality," *Social Networks*, vol. 24, pp. 407–422, Oct 2002.

[59] M. Newman, "The Mathematics of Networks," 2008.

[60] L. Katz, "A New Status Index Derived from Sociometric Analysis," *Psychometrika*, vol. 18, pp. 39–43, March 1953.

[61] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths," *Social Networks*, vol. 32, pp. 245–251, July 2010.

[62] D. Katsaros, N. Dimokas, and L. Tassiulas, "Social Network Analysis Concepts in the Design of Wireless Ad Hoc Network Protocols," *Network*, vol. 24, pp. 23–29, November 2010.

[63] H. Kim and R. Anderson, "Temporal Node Centrality in Complex Networks," *Physical Review E*, vol. 85, pp. 026107(1–8), February 2012.

[64] C. Balanis, *Antenna Theory-Analysis and Theory*. New York: Wiley, 1997.

[65] C. Bettstetter, C. Hartmann, and C. Moser, "How does Randomized Beamforming Improve the Connectivity of Ad Hoc Networks?," in *Proceedings of the IEEE International Conference on Communications (ICC)*, vol. 05, (Seoul, Korea), pp. 3380–3385, IEEE, New York, 16-20 May 2005.

[66] R. Vilzmann, C. Bettstetter, D. Medina, and C. Hartmann, "Hop Distances and Flooding in Wireless Multihop Networks with Randomized Beamforming," in *Proceedings of the 8th International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile systems*, (Montreal, Canada), pp. 20–27, ACM, New York, 10-13 October 2005.

[67] R. Vilzmann, J. Widmer, I. Aad, and C. Hartmann, "Low Complexity Beamforming Techniques for Wireless Multihop Networks," in *Proceedings of 3rd Annual IEEE Communication Society on Sensor and Ad Hoc Communication and Networks*, (Reston, VA), pp. 489–497, IEEE, New York, 25-28 September 2006.

[68] M. Kiese, C. Hartmann, J. Lamberty, and R. Vilzmann, "On Connectivity Limits in Ad Hoc Networks with Beamforming Antennas," *EURASIP Journal on Wireless Communication and Networks*, vol. 2009, pp. 1–15, 2009.

[69] Z. Yu, J. Teng, X. Bai, D. Xuan, and W. Jia, "Connected Coverage in Wireless Networks with Directional Antennas," in *Proceedings of IEEE International Conference on Computer Communications*, (Shanghai, China), pp. 2264–2272, IEEE, New York, 10-15 April 2011.

[70] P. Li, C. Zhang, and Y. Fang, "Asymptotic Connectivity in Wireless Ad Hoc Networks Using Directional Antennas," *IEEE/ACM Transaction on Networking*, vol. 17, pp. 1106–1117, Aug 2009.

[71] S. Durrani, X. Zhou, and H. M. Jones, "Connectivity of Wireless Ad Hoc Networks with Random Beamforming: An Analytical Approach," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, (Cannes), pp. 1–5, IEEE, 15-18 September 2008.

[72] X. Zhou, S. Durrani, and H. Jones, "Connectivity Analysis of Wireless Ad Hoc Networks With Beamforming," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 5247–5257, November 2009.

[73] P. Lawrence, "The Making of a Fly: The Genetics of Animal Design," *Molecular Reproduction and Development*, vol. 37, pp. 120–120, Jan 1994.

[74] Y. Afek, N. Alon, O. Barad, E. Hornstein, N. Barkai, and Z. Bar-Joseph, "A Biological Solution to a Fundamental Distributed Computing Problem," *Science*, vol. 331, pp. 183–185, Jan 2011.

[75] R. Nagpal and D. Coore, "An Algorithm for Group Formation in an Amorphous Computer," in *Proceedings of International Conference of Parallel and Distributed Systems*, (Las Vegas, NV), pp. 1–4, IASTED Press, Calgary, Canada, 1-4 October 1998.

[76] R. Nagpal and M. Mamei, "Engineering Amorphous Computing Systems," in *Methodologies and Software Engineering for Agent Systems* (F. Bergenti, M. P. Gleizes, F. Zambonelli, and G. Weiss, eds.), vol. 11 of *Multiagent Systems, Artificial Societies, and Simulated Organizations*, pp. 303–320, Boston: Springer, 2004.

[77] C. Reynolds, "Flocks, Herds and Schools: A Distributed Behavioral Model," *ACM SIGGRAPH Computer Graphics*, vol. 21, pp. 25–34, July 1987.

[78] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, "Effective Leadership and Decision-Making in Animal Groups on the Move," *Nature*, vol. 433, pp. 513–516, Feb 2005.

[79] B. A. Kadrovach and G. B. Lamont, "A Particle Swarm Model for Swarm-Based Networked Sensor Systems," in *Proceedings of ACM Symposium on Applied Computing*, (Madrid, Spain), pp. 918–924, ACM, New York, 11-14 March 2002.

[80] P. Antoniou, A. Pitsillides, A. Engelbrecht, T. Blackwell, and L. Michael, "Congestion Control in Wireless Sensor Networks Based on the Bird Flocking Behavior," in *Proceedings of the 4th IFIP TC 6 International Workshop on Self-Organizing Systems*, (Zurich, Switzerland), pp. 220–225, Springer-Verlag, Berlin, 9-11 December 2009.

[81] C. Bettstetter, M. Gyarmati, and U. Schilcher, "An Inhomogeneous Spatial Node Distribution and its Stochastic Properties," in *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, (Chania, Greece), pp. 400–404, ACM, New York, 22-26 October 2007.

[82] U. Schilcher, M. Gyarmati, C. Bettstetter, Y. W. Chung, and Y. H. Kim, "Measuring Inhomogeneity in Spatial Distributions," in *Proceedings of Vehicular Technology Conference*, (Singapore), pp. 2690–2694, IEEE, New York, 11-14 May 2008.

[83] J. Y. Le Boudec and M. Vojnovic, "The Random Trip Model: Stability, Stationary Regime, and Perfect Simulation," *IEEE/ACM Transactions on Networking*, vol. 14, pp. 1153–1166, Dec 2006.

[84] W. J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling Time-Variant User Mobility in Wireless Mobile Networks," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, (Anchorage), pp. 758–766, IEEE, New York, 6-12 May 2007.

[85] J. Riihijarvi, M. Petrova, and P. Mahonen, "Influence of Node Location Distributions on the Structure of Ad Hoc and Mesh Networks," in *Proceedings of Global Telecommunications Conference*, (New Orleans, LO), pp. 1–5, IEEE, New York, 28 Nov-30 Dec 2008.

[86] L. Hu and L. Dittmann, "Heterogeneous Community-Based Mobility Model for Human Opportunistic Network," in *Proceedings of the 2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, (Marrakech, Morocco), pp. 465–470, IEEE, New York, 12-14 October 2009.

[87] N. Aitsaadi, N. Achir, K. Boussetta, and G. Pujolle, "Artificial Potential Field Approach in WSN Deployment: Cost, QoM, Connectivity and Lifetime Constraints," *Computer Networks*, vol. 55, pp. 84–105, Jan 2011.

[88] A. Clauset, C. R. Shalizi, and M. Newman, "Power-Law Distributions in Empirical Data," *SIAM Review*, vol. 51, pp. 661–703, November 2009.

[89] C. Song, T. Koren, P. Wang, and A. L. Barabási, "Modelling the Scaling Properties of Human Mobility," *Nature Physics*, vol. 6, pp. 818–823, September 2010.

[90] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Human Mobility Models for Opportunistic Networks," *IEEE Communications Magazine*, vol. 49, pp. 157–165, December 2011.

[91] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: A New Mobility Model for Human Walks," in *IEEE INFOCOM 2009 - The 28th International Conference on Computer Communications*, (Rio de Janeiro), pp. 855–863, IEEE, New York, 19-25 April 2009.

[92] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: Self-Similar Least-Action Human Walk," *IEEE/ACM Transactions on Networking*, vol. 20, pp. 515–529, April 2012.

[93] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the Levy-Walk Nature of Human Mobility," *IEEE/ACM Transactions on Networking*, vol. 19, pp. 630–643, June 2011.

[94] B. Pincombe, "Detecting Changes in Time Series of Network Graphs using Minimum Mean Squared Error and Cumulative Summation," in *Proceedings of the 13th Biennial Computational Techniques and Applications Conference, CTAC-2006* (W. Read and A. J. Roberts, eds.), vol. 48 of *ANZIAM Journal*, pp. 450–473, October 2007.

[95] G. Jurman, R. Visintainer, and C. Furlanello, "An Introduction to Spectral Distances in Networks," in *Proceedings of the 2011 conference on Neural Nets WIRN10*, (Vietri sul Mare, Salerno, Italy), pp. 227–234, IOS Press Amsterdam, 27-29 May 2010.

[96] S. Hill and D. Braha, "Dynamic Model of Time-Dependent Complex Networks," *Physical Review E*, vol. 82, pp. 046105(1–7), October 2010.

[97] S. Hanneke, W. Fu, and E. Xing, "Discrete Temporal Models of Social Networks," *Electronic Journal of Statistics*, vol. 4, no. 2010, pp. 585–605, 2010.

[98] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, "Small-World Behavior in Time-Varying Graphs," *Physical Review E*, vol. 81, pp. 055101(1–4), May 2010.

[99] M. H. Zayani, V. Gauthier, and D. Zeghlache, "Improving Link Prediction in Intermittently Connected Wireless Networks by Considering Link and Proximity Stabilities," in *13th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, (San Francisco, CA), pp. 1–9, IEEE New York, 25-28 June 2012.

[100] L. Lu and T. Zhou, "Link Prediction in Complex Networks: A Survey," *Physica A: Statistical Mechanics and its Applications*, vol. 390, pp. 1150–1170, March 2011.

[101] C. Song, Z. Qu, N. Blumm, and A. L. Barabási, "Limits of Predictability in Human Mobility," *Science*, vol. 327, pp. 1018–1021, February 2010.

[102] M. R. Brust, C. H. C. Ribeiro, and S. Rothkugel, "Heuristics on Link Stability in Ad Hoc Networks," in *IEEE Network Operations and Management Symposium*, (Salvador, Bahia), pp. 738–741, IEEE New York, 7-11 April 2008.

[103] M. Brust, S. Rothkugel, and A. Andronache, "Node Stability in Dynamic Communication Networks," in *First Asia International Conference on Modelling & Simulation (AMS'07)*, (Phuket, Thailand), pp. 206–211, IEEE New York, 27-30 March 2007.

[104] L. Sattenspiel and K. Dietz, "A Structured Epidemic Model Incorporating Geographic Mobility among Regions," *Mathematical Biosciences*, vol. 128, pp. 71–91, July 1995.

[105] G. Miritello, E. Moro, and R. Lara, "Dynamical Strength of Social Ties in Information Spreading," *Physical Review E*, vol. 83, pp. 045102(1–4), April 2011.

[106] V. Nicosia, F. Bagnoli, and V. Latora, "Impact of Network Structure on a Model of Diffusion and Competitive Interaction," *EPL (Europhysics Letters)*, vol. 94, pp. 68009(1–6), June 2011.

[107] A. Vazquez, B. Rácz, A. Lukács, and A. L. Barabási, "Impact of Non-Poissonian Activity Patterns on Spreading Processes," *Physical Review Letters*, vol. 98, pp. 158702(1–4), April 2007.

[108] P. Wang, M. González, C. Hidalgo, and A. L. Barabási, "Understanding the Spreading Patterns of Mobile Phone Viruses," *Science*, vol. 324, pp. 1071–1076, May 2009.

[109] D. Wang, Z. Wen, H. Tong, C. Y. Lin, C. Song, and A. L. Barabási, "Information Spreading in Context," in *Proceedings of the 20th international conference on World wide web - WWW '11*, (Hyderabad, India), pp. 735–744, ACM Press, New York, 28 March-1 April 2011.

[110] P. Hui, K. Xu, V. Li, J. Crowcroft, V. Latora, and P. Lio, "Selfishness, Altruism and Message Spreading in Mobile Social Networks," in *IEEE INFOCOM Workshops 2009*, (Rio de Janeiro), pp. 1–6, IEEE, New York, 19-25 April 2009.

[111] N. Valler, B. A. Prakash, H. Tong, M. Faloutsos, and C. Faloutsos, "Epidemic Spread in Mobile Ad Hoc Networks: Determining the Tipping Point," in *Networking 2011* (J. Domingo-Pascual, P. Manzoni, S. Palazzo, A. Pont, and C. Scoglio, eds.), vol. 6640 of *Lecture Notes in Computer Science*, pp. 266–280, Valencia, Spain: Springer-Verlag, Berlin, 9-13 May 2011.

[112] F. Peruani, A. Maiti, S. Sadhu, H. Chate, R. Choudhury, and N. Ganguly, "Modeling Broadcasting using Omnidirectional and Directional Antenna in Delay Tolerant Networks as an

Epidemic Dynamics," *IEEE Journal on Selected Areas in Communications*, vol. 28, pp. 524–531, May 2010.

[113] R. Maiti, N. Ganguly, and A. Gupta, "Epidemic Broadcasting in DTNs using Directional Antenna," in *Proceedings of Fourth International Conference on Communication Systems and Networks*, (Bangalore), pp. 1–2, IEEE, New York, 3-7 January 2012.

[114] R. Maiti, N. Ganguly, and A. Gupta, "Analyzing the Performance of Epidemic Broadcasting in DTNs using Directional Antenna," in *Proceedings of Fifth International Conference on Communication Systems and Networks*, (Bangalore), pp. 1–10, IEEE, New York, 7-10 January 2013.

[115] D. Shah, "Gossip Algorithms," *Foundations and Trends in Networking*, vol. 3, no. 1, pp. 1–125, 2007.

[116] Y. M. Ko and N. Gautam, "Epidemic-Based Information Dissemination in Wireless Mobile Sensor Networks," *IEEE/ACM Transactions on Networking*, vol. 18, pp. 1738–1751, December 2010.

[117] I. Akyildiz and I. Kasimoglu, "Wireless Sensor and Actor Networks: Research Challenges," *Ad Hoc Networks*, vol. 2, pp. 351–367, Oct 2004.

[118] F. Dressler, *Self-Organization in Sensor and Actor Networks*. Chichester, UK: John Wiley & Sons, Nov 2007.

[119] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*. New Jersey: Princeton University Press, 2003.

[120] H. A. Simon, "On a Class of Skew Distribution Functions," *Biometrika*, vol. 42, no. 3-4, pp. 425–440, 1955.

[121] S. Manna and P. Sen, "Modulated Scale-Free Network in Euclidean Space," *Physical Review E*, vol. 66, pp. 1–4, dec 2002.

[122] M. Barthélemy, "Crossover from Scale-Free to Spatial Networks," *Europhysics Letters*, vol. 63, pp. 915–921, Sept 2003.

[123] G. Sharma and R. Mazumdar, "A Case for Hybrid Sensor Networks," *IEEE/ACM Transaction on Networking*, vol. 16, no. 5, pp. 1121–1132, 2008.

[124] D. Guidoni, R. Mini, and A. Loureiro, "On the Design of Resilient Heterogeneous Wireless Sensor Networks Based on Small World Concepts," *Computer Networks*, vol. 54, pp. 1266–1281, June 2010.

[125] C. K. Verma, B. R. Tamma, B. S. Manoj, and R. Rao, "A Realistic Small World Model for Wireless Mesh Networks," *Communications Letters, IEEE*, vol. 15, pp. 455–457, April 2011.

[126] M. R. Brust, C. H. C. Ribeiro, D. Turgut, and S. Rothkugel, "LSWTC: A Local Small World Topology Control Algorithm for Backbone-Assisted Mobile Ad Hoc Networks," in *Proceedings of 35th Conference on Local Computer Network*, (Denver, CO), pp. 144–151, IEEE, New York, 10-14 October 2010.

[127] M. Brust, A. Andronache, and S. Rothkugel, "WACA: A Hierarchical Weighted Clustering Algorithm Optimized for Mobile Hybrid Networks," in *Proceedings of 3rd International Conference on Wireless and Mobile Communication*, (Guadeloupe, French Caribbean), pp. 23–30, IEEE, Washington DC, 4-9 March 2007.

[128] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Micro-Sensor Networks," *IEEE Transaction on Wireless Communication*, vol. 1, pp. 660–670, Oct 2002.

[129] O. Younis and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 3, pp. 366–379, Oct 2004.

[130] S. Dorogovtsev, J. Mendes, and A. Samukhin, "Giant Strongly Connected Component of Directed Networks," *Physical Review E*, vol. 64, pp. 025101–025104, Jul 2001.

[131] J. Kleinberg, "Computing: The Wireless Epidemic," *Nature*, vol. 449, pp. 287–288, ?September 2007.

[132] Y. Schwarzkopf, A. Rákos, and D. Mukamel, "Epidemic Spreading in Evolving Networks," *Physical Review E*, vol. 82, pp. 036112(1–8), September 2010.

[133] P. Bajardi, C. Poletto, J. J. Ramasco, M. Tizzoni, V. Colizza, and A. Vespignani, "Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic," *PloS one*, vol. 6, pp. e16591(1–8), January 2011.

[134] B. Birand, M. Zafer, G. Zussman, and K. W. Lee, "Dynamic Graph Properties of Mobile Networks under Levy Walk Mobility," in *2011 IEEE Eighth International Conference on Mobile*

Ad-Hoc and Sensor Systems, (Valencia, Spain), pp. 292–301, IEEE New York, 17-22 October 2011.

[135] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A. L. Barabási, and J. Saramäki, "Small but Slow World: How Network Topology and Burstiness Slow Down Spreading," *Physical Review E*, vol. 83, pp. 025102(1–4), February 2011.

[136] Y. Li, Z. Wang, D. Jin, L. Zeng, and S. Chen, "Collaborative Vehicular Content Dissemination with Directional Antennas," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 1301–1306, April 2012.

[137] G. Németh and G. Vattay, "Giant Clusters in Random Ad Hoc Networks," *Physical Review E*, vol. 67, pp. 036110(1–6), March 2003.

[138] C. Bettstetter, "On the Connectivity of Ad Hoc Networks," *The Computer Journal*, vol. 47, pp. 432–447, April 2004.

[139] H. W. Hethcote, "The Mathematics of Infectious Diseases," *SIAM Review*, vol. 42, pp. 599–653, Jan. 2000.

[140] T. Britton, "Stochastic Epidemic Models: A Survey," *Mathematical biosciences*, vol. 225, pp. 24–35, May 2010.

[141] B. Wang, L. Cao, H. Suzuki, and K. Aihara, "Safety-Information-Driven Human Mobility Patterns with Metapopulation Epidemic Dynamics," *Scientific reports*, vol. 2, pp. 1–8, January 2012.

[142] M. J. Keeling and K. T. D. Eames, "Networks and Epidemic Models," *Interface, Journal of the Royal Society*, vol. 2, pp. 295–307, September 2005.

[143] H. Lund, L. Lizana, and I. Simonsen, "Effects of City-Size Heterogeneity on Epidemic Spreading in a Metapopulation: A Reaction-Diffusion Approach," *Journal of Statistical Physics*, January 2013.

[144] http://www.statoids.com/yci.html.

[145] M. Corson, R. Laroia, J. Li, V. Park, T. Richardson, and G. Tsirtsis, "Toward Proximity-Aware Internetworking," *IEEE Wireless Communications*, vol. 17, pp. 26–33, December 2010.

[146] N. Golrezaei, A. Molisch, and A. Dimakis, "Base-Station Assisted Device-to-Device Communications for High-Throughput Wireless Video Networks," in *Proceedings of the 2012 IEEE*

International Conference on Communications, (Ottawa), pp. 7077–7081, IEEE, New York, 10-15 June 2012.

[147] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, and Z. Turanyi, "Design Aspects of Network Assisted Device-to-Device Communications," *IEEE Communications Magazine*, vol. 50, pp. 170–177, March 2012.

[148] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator Controlled Device-to-Device Communications in LTE-Advanced Networks," *IEEE Wireless Communications*, vol. 19, pp. 96–104, June 2012.

[149] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-Device Communication as an Underlay to LTE-Advanced Networks," *IEEE Communications Magazine*, vol. 47, pp. 42–49, December 2009.

[150] S.-Y. Liu, A. Baronchelli, and N. Perra, "Contagion Dynamics in Time-Varying Metapopulation Networks," *Physical Review E*, vol. 87, pp. 032805(1–9), March 2013.

[151] V. Belik, T. Geisel, and D. Brockmann, "Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases," *Physical Review X*, vol. 1, pp. 011001(1–5), August 2011.

[152] R. Louf, P. Jensen, and M. Barthelemy, "Emergence of Hierarchy in Cost-Driven Growth of Spatial Networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 8824–8829, May 2013.

[153] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Energy-Efficient Broadcast and Multicast Trees in Wireless Networks," *Mobile Network Application*, vol. 7, pp. 481–492, December 2002.

[154] L. A. Maglaras and D. Katsaros, "New Measures for Characterizing the Significance of Nnodes in Wireless Ad hoc Networks via Localized Path-based Neighborhood Analysis," *Social Network Analysis and Mining*, vol. 2, pp. 97–106, June 2011.

[155] http://perso.uclouvain.be/vincent.blondel/netmob/2013/D4D-book.pdf.

[156] S. Hanson, "Perspectives on the Geographic Stability and Mobility of People in Cities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 15301–15306, October 2005.